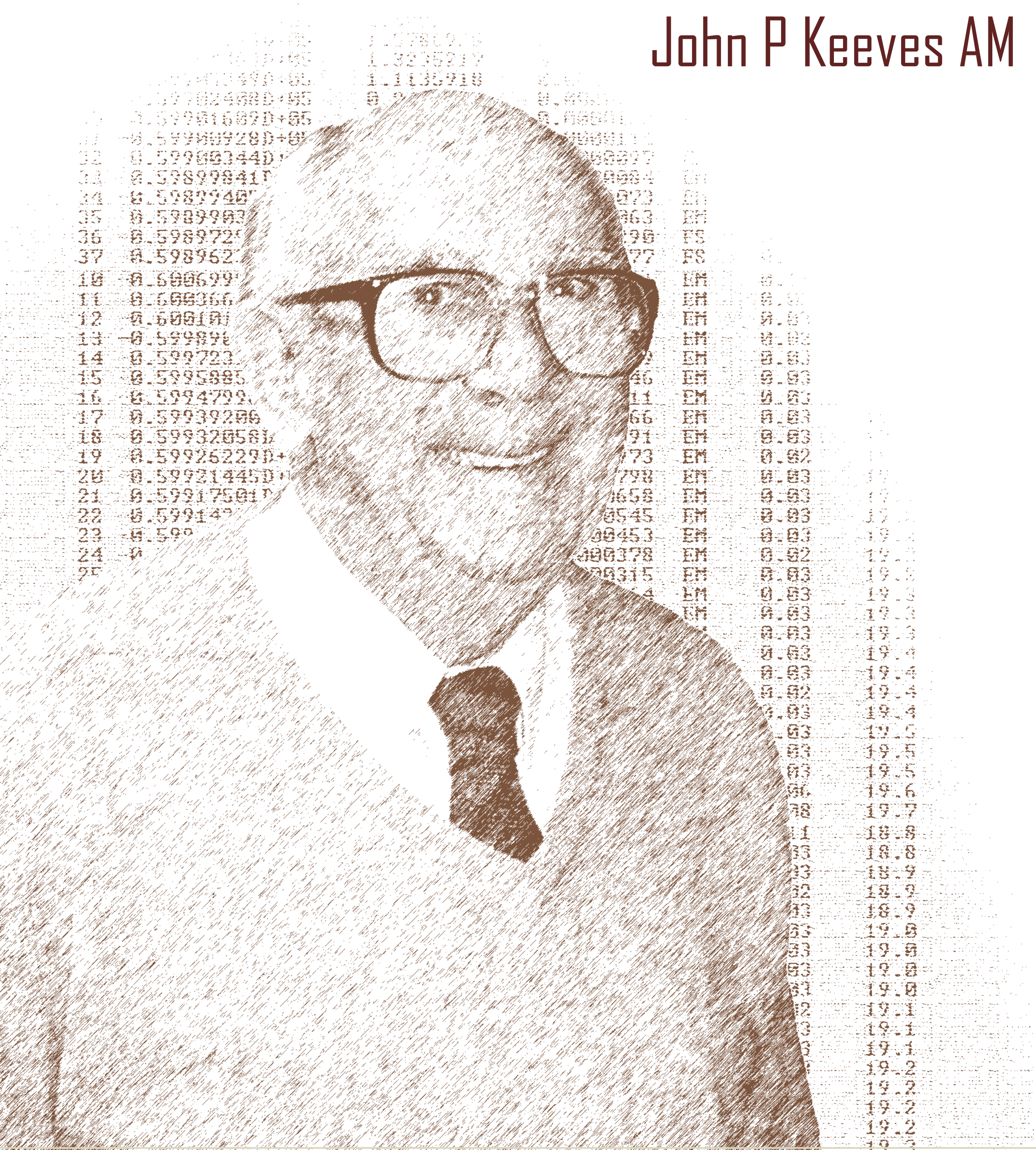
The Process of Research in Education

A festschrift in honour of John P Keeves AM



The Process of Research in Education

A festschrift in honour of John P Keeves AM

EDITED BY

BOBBIE MATTHEWS AND TONY GIBBONS

Produced by the School of Education, Flinders University
Sturt Road, Bedford Park, Adelaide SA 5000

Cover and design by Katherine L. Dix: Stylised photo of John Keeves (2004)
Published by Shannon Research Press, South Australia
ISBN: 978-1-920736-39-2

Acknowledgements

The authors of this volume gratefully acknowledge the financial support provided by the School of Education, Flinders University. The funds have assisted with the publication and printing of this book.

The editors would like to extend their gratitude and thanks for their assistance and expertise in the production of this special volume of work.

- Jeni Thomas for her time in the extensive English language editing, formatting, and preparation of the volume,
- Jonathan Anderson for 'casting his eye' over the final draft,
- Lawrence Saha for his encouragement and editorial assistance,
- Petra Leitz for her sensitive introduction to the article written by the late Neville Postlethwaite,
- Carol Aldous for enlisting the support of School of Education, Flinders University,
- Katherine Dix for final book preparation and cover design.

Contents

Preface

John Keeves is acknowledged internationally by colleagues, students and friends as a superb teacher and researcher. Now in his eighty-fifth year, a number of these same people have come together to write papers that, in their content, express the foci of his teaching and research activities and, that pay tribute to the considerable effect he has had on the world of educational research and, on the authors themselves. This book is the result. It is no surprise that the authors are international and diverse.

Petra Lietz has written an introduction to the paper by the late Neville Postlethwaite. She writes with insight and humanity of both John and Neville and provides a unique picture of two eminent teachers and researchers who argued, wrote and collaborated over many years. This introduction sets the scene for the papers that follow. In his paper, Neville recalls the beginnings of the IEA, the six-subject study, the second science study, and the first mathematics study, which was the first large-scale international achievement study. John Keeves was the Australian national research coordinator in the first mathematics study and the prime mover in both the first and second science studies.

Ted Sandercock of the South Australian Institute for Educational Research (SAIER) worked with John Keeves when he joined the SAIER Executive Committee in 1989. The contacts that John Keeves had with the international educational community and his involvement in the Flinders Institute for International Educational greatly assisted the SAIER Executive to rethink its direction and modify its annual programme of seminars and workshops.

Jan Lokan, the former Assistant Director of the Australian Council for Educational Research (ACER) notes John Keeves' role in the early studies following the establishment of the International Association for the Evaluation of Educational Achievement (IEA). John's own work benefited from his knowledge of IEA models, procedures and analyses, as he himself freely admits, but IEA studies benefited in their turn from the contributions he made to them. His experiences in the IEA also provided the bases from which he was able to produce many papers, books and encyclopaedias for the benefit of research endeavours that go beyond the IEA studies.

Penny Van Deur looks at models from the viewpoint of an educator. There is no branch of science which does not make use of models. While it may be agreed that they represent something, to leave it at that conceals a number of issues. Does the model represent some range or selection of phenomena? Is this different from a model representing a range or selection of data? What of a model that seems to be an exemplification of a theory? She describes the process of educational research as taught by Professor Keeves to higher degree students from Australia and the Asia-Pacific region. Her work considers John's emphasis on building models and testing them to see if they are consistent with the real world, his knowledge of qualitative and quantitative methods of analysis, his stress on academic rigour, his attention to detail, and his concern that writing should clearly communicate results and insights gained from the research.

Nordin Abdul Razak reaffirms what Penny Van Deur has said and discusses what he considers to be the qualities and capabilities of a good mentor. Nordin, who came from Penang in Malaysia to do a doctorate considers that John manifests (a) a desire and a willingness to give up time to help others, (b) a readiness to share skills, knowledge, and expertise with his students, (c) demonstrates a positive attitude and acts as a role model, (d) provides guidance and constructive feedback to his students, (e) sets and meets ongoing personal and professional goals, (f) values the opinions and initiatives of others, and (g) motivates others by setting a personal example of a high standard of research

John's interests have always extended beyond the confines of the Anglo-Saxon world. Yasmeen Faruqi's paper discusses the basic tenets of Islam and the Islamic view of nature and its role in the development of scientific thinking during the so-called 'Golden Age of Islam'. She argues that the Muslims contributed original and highly significant works to the different fields of science. The paper also examines the views of present day Muslim scholars, who have emphasised the importance of Islamic science and the need for greater dialogue between Islamic science and modern or Western science. Faruqi explores ways in which Islamic values and the adoption of Western sciences can be reconciled and the need for discourse between modern or Western science and Islamic scientific points of view.

Mochtar Marhum's argument for the cultural and linguistic retention of the local language dialect in parts of Indonesia involves positive assertions on the issues surrounding pluralism and multiculturalism that foster retention of difference. Thus many local people are still proud of the use of *Bahasa Indonesia* as their national language and a means of communication among people of different ethnic and linguistic backgrounds that also retain and utilise their local language.

Albert Tuijnman describes the gradual development of the European Union (EU) co-operation in education and training with reference to the original Treaty of Rome as well as to subsequent Council decisions that have shaped the development of action programmes in the education sector. The policies on the structure, organisation and content of education are a national rather than a European responsibility. As a result the sole official sources of broad guidance as to possible EU policy directions are the decisions and related reports from the European Parliament and the European Commission.

Helen Askell-Williams and Michael Lawson investigated childhood mental health issues. They reported the assessment of students' mental health based upon data from the evaluation of the KidsMatter mental health promotion, prevention and early intervention pilot program in 100 primary schools across Australia in 2007 and 2008. Goodman's (2005) Strength and Difficulties Questionnaire (SDQ) was completed by parents or caregivers and teachers of 4980 primary school students. A second measure, specifically developed for the KidsMatter evaluation, canvassed the five core groups of indicators of students' social and emotional competencies identified by the Collaborative for Academic, Social and Emotional Learning (CASEL, 2006), namely, self-awareness, self-management, social awareness, relationship skills, and responsible decision making. This measure was also completed by the students' teachers and parents or caregivers. A third measure was based on a non-clinical assessment by school staff, who identified students considered to be 'at risk' of social, emotional or behavioural problems. Correspondence analysis, which overcomes limitations of the collected data in that it works with frequency data that does not impose restrictive distributional assumptions, was used to examine relationships in these measures. It was able to illustrate clear patterns of relative distances between alternative measures of positive mental health.

Katherine Dix has also investigated teacher and parent assessments of students' mental health based upon data from the evaluation of KidsMatter, the Australian mental health promotion, prevention and early intervention initiative. She notes that the prevalence of child and adolescent mental health disorders has increased over the last 50 years according to investigators to the extent that it is now of major concern in Australia. In order to counter this trend, national initiatives, such as KidsMatter, have been trialled with the intention of improving the mental health outcomes for Australian primary school children. She discusses the use of multiple instruments from multiple informants and the challenge of classifying each child's mental health status by bringing these various assessments together. A technique called Latent

Class Analysis was considered to be the best approach and is the focus of the chapter.

Lawrence Saha and Gary Dworkin point to research suggesting that work conditions can create alienation and powerlessness in the worker which lead to stress, and that stress leads to burnout. A national sample survey of Australian adults, has examined the roles of educational attainment, centrality of education to self-identity, and the disjuncture between educational attainment and occupational status in bringing about burnout. The multi-block model suggests that when occupational status is at parity or greater than educational attainment, job stressors and work conditions are less likely to produce a sense of work alienation (burnout) than when occupational status lags behind educational attainment. The perceived relative centrality of the educational attainment to self identity is implicated in the burnout process.

Ian Blackman and Margaret Hall examine the complexity of applied English language skills to the perceived ability of non-English speaking nursing students using Rasch analysis. The application of the Rasch scaling (the Partial Credit method) to derive and analyse data from a series of Likert-based rating scales that measure student perceptions of the degree of difficulty in undertaking and completing communication tasks is considered and the preparedness of non-English speaking background nursing students to engage in language tasks associated with communication, related to their clinical nursing work is examined. Emphasis is given to how the principles of Rasch scaling can be applied to the calibration and analysis of four different rating scales that estimate participants' English reading, writing, listening, and speaking skills in a nursing context.

Ratna Rintaningrum points out that children who (a) do not learn to read early enough, (b) do not learn to read with comprehension and fluency, or (c) do not learn to read for pleasure, are likely to encounter difficulties in their lives. Moreover, their chances of succeeding in an academic environment, obtaining satisfying jobs that are rewarding financially, are, in practice, slight. However, if these children are equipped with an adequate reading literacy, they are able to participate fully in their communities or in society. Her results have been analysed using PLSPath Analysis (Sellin, 1990) as has the work of Tilahun Afrassa who has examined factors influencing mathematics achievement in Ethiopia. Previous research studies in Ethiopia indicate that there are substantial differences between students in their achievement levels in school mathematics. Researchers have identified some of the student level factors that influence mathematics achievement. They argue that students' attitudes towards mathematics are representative of the student level factors influencing achievement in mathematics in Ethiopian schools.

Kwok-cheung Cheung notes that in 2003, under the auspices of OECD, PISA (Program for International Student Assessment) collected mathematical literacy data in 41 countries and regions to compose indicators of educational effectiveness for guiding policy directions (OECD, 2004a). Macao, a special administrative region of China, participated for the first time as a non-OECD member. Preliminary analyses of the sampled data revealed that Macao, in 2004, ranked ninth in terms of mean mathematical literacy in the League Tables, and sixth in terms of students' attainment of minimum mathematical literacy (Lo & Cheung, 2005). In addition, the impact of the social, economic and cultural status of the home on mathematical literacy was been found to be the lowest amongst the participating countries and regions – the percentage of explained variance in student performance was only 1.9 per cent (OECD, 2004a, p.397; OECD, 2004b, p. 21). Although Macao was able to combine equity with relatively high literacy performance, the affective aspects of mathematics learning were far from satisfactory. The findings revealed that students' sense of belonging to schools were amongst the lowest in all the participating countries and regions (OECD, 2004a, p.368). Another indicator was that teachers' morale and commitment as perceived by their principals needed more enhancement (OECD, 2004a, p.412).

Petra Lietz and Dieter Kotte have examined League Tables of countries' average student achievement that seem to have become the main media-effective focus of internationally comparative studies in education. However, the original aim of such studies by the founders of the International Association for the Evaluation of Educational Achievement (IEA) was to improve the conditions and outcomes of learning through the provision of (a) comparable information through which educational systems could learn from each other and (b) reliable and valid data from national representative samples of certain age groups in order to provide sound evidence as a basis for policy discussions of topical issues in participating countries. This chapter examines this issue in two ways. First, a two-level hierarchical linear model is examined with a particular view to the way in which school type operates together with other student and school-level factors to influence mathematics achievement. Second, given the apparent achievement differences in mathematics, a logistic regression analysis explores which variables discriminate between United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) and government schools in the Gaza Strip and the West Bank

Shirley Yates has completed a four-year study that focuses on the inherent multilevel effects of schooling. For the majority of their time at school, students are grouped within classes where they share learning experiences and are taught by particular teachers. Explanations for change within educational settings must take these clustering effects into account (Raudenbush & Bryk, 1988) as well as the effects of growth over time. The longitudinal study presented here addresses these deficiencies as it was conducted over four years following the introduction of coeducation into a single sex school, involved students across elementary, middle and secondary levels and employed appropriate statistical techniques to allow for the assessment of particular situational and environmental variables that may have been operating across cohort groupings and grade levels over time.

I Gusti Ngurah Darmawan and Firenius Sitinjak have examined the hierarchical effects of student-related and school-related variables on science performance in Indonesia in 2006. They have used the Programme for International Student Assessment (PISA) to assess science performance in the selected sample that consisted of 10,647 15-year-old students who attended 352 schools. Four student level variables, grade, gender, socio-economic status, and attitudes toward science are seen to have an impact on science performance. Two group composition variables, average socio-economic status and average grade and two other school level variables, school capacity and ability grouping within classes are also seen to have an impact on science performance.

Bobbie Matthews, using the same statistical procedure, investigated and sought to comprehend whether Confucian heritage culture (CHC) students' approaches to learning have been retained or modified in a Western social, cultural, and educational environment. A bilingual version of the *Study Process Questionnaire* (Biggs, 1987) (SPQ) was used to measure sojourner students' approaches to learning on five occasions over a period of two years in Australia. A two-level analysis was undertaken at the intra and inter-student levels to see if there has been change in the students' approaches to learning over time. These results showed that five of six approaches to learning changed during the period, but not for all groups of students.

Carol Aldous examines the role that feeling, intuition and imagination have, together with cognitive and non-cognitive approaches to reasoning, in the solving of novel mathematics problems. This outcome is intriguing because the identification, description and measurement of such non-cognitive activities are notoriously elusive. Nevertheless this paper describes the journey of one such undertaking. The first path entails the use of protocol analysis, and the development of a conceptual model of creative problem solving. The second involves testing this model using multilevel structural equation modelling procedures. The conclusion that no model of creative problem solving was found that did not involve a feeling approach to reasoning is profound.

Pauline Goh reports on an investigation conducted to examine relationships among English language competency, perceptions of the learning environment, approaches to learning, and learning related outcomes of students from Malaysian Private Higher Educational Institutions (PHEIs) undertaking twinning programs with overseas partner universities in Malaysia. The investigation has been able to establish specific issues within students' teaching and learning environments that are able to encourage or inhibit students' full adoption of a deep approach to learning. In addition, the unexpected existence of a personality trait better known as *kiasu* that has been defined variously as being 'afraid to lose out' has been reflected in students' (a) not questioning; (b) using passive memorisation; (c) exhibiting overly careful behaviour; and (d) conformity within students' learning environment. The study has shown that the perceptions of being *kiasu* are prevalent in the three Malaysian ethnic groups (Malay, Chinese, and Indian) and may inhibit student's adoption of deep approaches to learning.

Arief Liem's study has investigated how students' values and perceptions of the cultural climate of the classroom learning environment were related to their approaches to learning and perceived academic performance in cross-cultural contexts. To this end, three measures, the *Portrait Values Questionnaire* (Schwartz et al., 2001), the *Cultural Learning Environment Questionnaire* (Waldrip & Fisher, 2002), and the *Learning Process Questionnaire* (Biggs, 1987), were administered to senior secondary school students in Australia, Singapore, the Philippines, and Indonesia. Partial least square (PLS) path modelling analysis, allowed the simultaneous analysis of the many variables at both the individual cultural groups and the pan-cultural group level.

Tony Gibbons provides the last, but not final word on the process of research. He takes a philosophical look at the concept of reflection. While reason and reasoning attracts much analysis and research, reflection does not. Yet the word and its synonyms appear frequently in research documents, books and pronouncements. Reflection is linked to reasoning yet is not simply reasoning. Reflection is linked to the core beliefs that a person holds and it is to these beliefs that a person turns when reflecting. This raises the question of how we develop core beliefs and the question of what should be researched.

Many people supervise research, some even edit their students' work, others present challenges, but very few inspire the excellence that exemplifies these processes with grace, humility and dignity in the manner that John Keeves does. It is the conclusive unity in the diversity of the work in this volume that has made it a challenge and ultimately satisfying mission to edit.

Bobbie Matthews and Tony Gibbons, September 2009

1

Memories of IEA achievement surveys

T. Neville Postlethwaite

Foreword by Petra Lietz

The following article was written by Neville Postlethwaite who, as was characteristic of him, was one of the first people to submit his contribution to this book. As many readers would be aware, Neville passed away on 12th April 2009 after a brief but very intensive battle with a brain tumour. The following is written in recognition of the important role Neville played in John's life.

To me, the stories of John and Neville are inextricably linked. According to John, he first noticed a young student rushing past him on a pushbike on her way to Hamburg University in 1988 where Neville had just recruited me as an undergraduate student to assist with the production of the report for the Second International Science Study (SISS) undertaken by the International Association for the Evaluation of Educational Achievement (IEA).

At this time, John travelled back and forth between Stockholm, where IEA's headquarters were then located, and Hamburg, where Neville was in charge of the International Coordinating Centre for SISS. Neville had managed to obtain funding for the analyses and report writing for SISS and had asked John to do this work. John accepted the challenge and moved from his Directorship of the Australian Council for Educational Research (ACER) in Melbourne to Stockholm to undertake that task.

One of the first interactions between them which I witnessed was a difference of opinion concerning the priorities for the analyses.

Whereas Neville's focus was on all 23 countries that participated in SISS, John was keen to explore the analytical possibilities stemming from the fact that ten of these countries had also participated in the first international science study in 1970/71. This resulted in two books being written: One edited by Neville together with David Wiley focusing on science achievement in 23 countries and one edited by John addressing changes in science education and achievement: 1970 to 1984.

In addition to having occasional differences in opinion, Neville and John also presented as very different personalities. Neville was a man of action with rolled-up sleeves and a "let's-do-it" attitude. He, for example, took on the challenge of setting up the International Coordinating Centre for the Reading Literacy Study to show the world that an international study of reading involving 40 countries – the largest study of its kind until then – could be undertaken effectively. This occurred in 1990/91 at a time before e-mail and Internet existed. Neville came across as a very colourful and vibrant person: You knew when Neville was around as his voice could be heard from far away even if doors were closed. He loved to tell the occasional joke, enjoyed a glass or two of wine and smoked until a few years ago. John, on the other hand, is of a quieter nature, more analytical and humble. In brief, the archetype of a "gentleman", although a couple of stories circulate which portray John as showing less restraint.

Differences between the two can also be noted regarding the teaching of students at university. For Neville, this was a necessary part of his job as a professor which is why he would often get one of his many visitors to regale his classes with stories of their respective home countries, which, although interesting, were not necessarily related to the topic of the course. John, on the other hand, regards teaching as a way to enlighten people. He is a dedicated teacher who will put in many hours into lesson preparation as well as into guiding and marking students' work. It is not without reason that John was awarded the Member of the Order of Australia (AM) in recognition of his assistance to research students among other things. Also, I often thought that one could write textbooks just by recording what John says in class as his sentences are so well formed and logically build upon each other.

But John and Neville also had much in common. First and foremost, they shared an enthusiasm for educational research that pervaded everything they did and which was contagious. You couldn't be around either John or Neville without thinking that whichever problem or topic regarding educational research they were discussing was of utmost importance. One also could not help but be intrigued by the breadth of experience and knowledge on all matters regarding educational research and educational systems that both demonstrated. Their research projects were their passion and they didn't do anything half-heartedly. They worked extremely hard and expected the same of everyone around them (did I hear the word 'slave-driver' somewhere?). Also, each in his own way could be said to be very forceful and resourceful in order to get things done the way they considered to represent the best course of action.

Both have been extremely productive as illustrated by the countless articles, reports and books that they have written. They also shared a passion for accuracy in expression which stemmed from a long history as editors. Indeed, Neville kept reminding me that tables or figures in texts were inert objects and could never be active. Therefore, sentences starting with 'Table 8.1 shows…' or 'As Figure 13.7 illustrates…' should be rephrased to 'As can be seen in Table 8.1…' or 'From Figure 13.7 it can be inferred…' Similarly, I learnt from John not to split infinitives (e.g. 'To boldly go…') and to avoid hanging participles ('Keeping in mind…'). I hope that I have managed to do the right thing in these respects by both masters in this piece of writing.

Both were very careful with their money. I remember Neville running through the corridors at Hamburg University when he learnt about some damage to the roof of his house in St Albans and bemoaning the expense its fixing would incur. We – his student assistants – felt so sorry that we were contemplating setting up a collection to assist him. A similar illustration can be given for John: He very kindly invited me to his summer house at Port Willunga on my first Christmas as an international graduate student here in Australia; only the bread that was bought had to be yesterday's bread at reduced prices.

However, both were – and in John's case continue to be – extremely generous with their time. Only last year, Neville took the time to work carefully through a chapter for an international report with which I was struggling while John keeps taking the time to think through and answer meticulously any of my questions. Without doubt many people can cite numerous ways in which Neville and John have helped them with their work.

This position of expert, however, also means that the number of people to whom one can relate about the highly specific problems and arguments one is considering is rather limited. This was the reason for John and Neville having regular chats over the telephone on Sundays. Again only last year, when my family and I visited Neville and his wife, Trudy, on their beautiful estate in Southern France, Neville was quite concerned as John was experiencing a stretch of ill health at the time. Neville said that if John wasn't around he wouldn't quite know with whom to discuss a number of topics as only John would truly understand him. This respect and deep appreciation of the other appeared to be genuinely reciprocal.

Only recently, I learnt that it was Neville who, in the early 1960s, introduced John to the world of international comparative research in education and many of the people who have been instrumental in shaping and advancing this area of research. So, what he said when we learnt of Neville's passing also applies to John himself: "Many people in the world wouldn't be doing what they're doing now if it wasn't for Neville!"

Petra Lietz, Adelaide, June 2009

The beginnings of IEA

In the report (Hotyat, 1958) from a meeting held at the UNESCO Institute for Education in Hamburg in March 1958 nearly all of the participants were also those involved in the pilot study of IEA. Interestingly, the meeting was mostly concerned with how test results could be used to supplement pupil records in school on the general development of the child. They were also interested in the use of tests at critical stages in school life (e.g. transfer from kindergarten to primary school, during primary education, transfer from primary to secondary school, at entry to higher education, and at the transfer from technical school to industry). Georg Rasch[1] was at the IEA meeting in 1958 but it took another 30 years before Rasch scaling was used in IEA studies (in the IEA Second Science Study in 1989 and the IEA Reading Literacy Study in 1990). The k-factor (spatial) in intelligence testing was identified by El-Koussy in 1935 when he was working for Spearman. El-Koussy also attended the original IEA meeting.

Encouraged by Wall, the then director of the National Foundation for Educational Research in England and Wales (NFER) who had helped organise the first Hamburg meeting, many of the participants from the March 1958 meeting, met again in Hamburg in June 1959 and the week after at Eltham Palace, London in order to explore what might be learned by conducting an international study. First they had to see if it

[1] On page 11, there is the following paragraph: The attempt has also been made to grade tests irrespective of the age of the pupils. Thus Dr. Rasch of Denmark has designed attainment tests which are graded according to the intrinsic progressive complexity of subject matter; the marks, in this case, aim to assess the absolute level which the child has reached in a particular subject.

was possible to undertake such a study. Thus, they planned and executed the pilot study in twelve countries. It is of interest that there were four Americans at the Eltham Palace meeting who had not been at the Hamburg meeting. At Hamburg, the American had been Willard Olson (University of Michigan) but following the Hamburg meeting Husén told Arnold Anderson about the new initiative. Anderson told Bloom (both University of Chicago). Foshay, Thorndike, and Passow (all from Teachers College, Columbia University, New York) were also at the Eltham meeting. All either paid themselves to go to meetings or had their institution pay. On one occasion Thorndike, who taught his classes on a Wednesday in New York, left the Eltham meeting on Tuesday evening, taught in New York on the Wednesday and flew back to London on Wednesday night – all paid for out of his own pocket. Indeed, over the time that John Keeves and I were associated with IEA, none of the key people were ever paid.

In the publication of the pilot study results, Foshay wrote ‘If custom and law define what is educationally allowable within a nation, the educational systems beyond one’s national boundaries suggest what is educationally possible. The field of comparative education exists to examine these possibilities’. This has often be cited as the *raison d’être* of IEA-type studies. The pilot study involved several tests: geography, science, mathematics, reading, and a non-verbal reasoning test. The tests were translated into seven different languages. Test scores were examined in relation to gender, home background, size of community, grade repetition, and nationality of parents. Specific hypotheses were formed and tested. Tests were examined for their item factor loadings in each country. It was also of interest to note that the between-country variance was larger for the so-called culture fair test of non-verbal reasoning than for reading comprehension!! There were technical deficiencies in the work but the overall conclusion was that it was possible to conduct work of this kind. The group decided to embark on a technically ‘better’ study of one subject area only – mathematics.

The early players and the vision that drove them

When IEA obtained money from the United States Office of Education for the study they decided they needed a full time person working on the study, which, in the course of events, became known as the first international mathematics study (FIMS). They advertised. I applied and was interviewed and obtained the job. I had been working as a test constructor at the National Foundation for Educational Research in London. I started working for IEA on 2 December, 1962, and on 14 December I attended a meeting of the Standing Committee of IEA in New York. At that meeting were Bob Thorndike, Ben Bloom, Arnold

Anderson, Gaston Mialaret, Douglas Pidgeon, Wells Foshay, Harry Passow and Torsten Husén[2]. Lee Cronbach popped in. I sat in awe. Some of these people were demi-gods for me. I knew them through their writings and by reputation. I was to work with them over the next several years and I learned a lot. Dick Wolf[2] joined the study in late 1963 and he was the one who had to analyse all of the data on a mainframe computer at the University of Chicago. In 1964 Gilbert Peaker[3] joined IEA and he, too, was very impressive. His work on sampling has had a great effect on IEA, PISA and SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality). He was also instrumental in directing the analysis away from covariance and towards multiple regression. It was Peaker who used path analysis for the first time in education in the Plowden report. It was during the late 60s and early 70s that LISREL was being developed by Jöreskog in Sweden and PLS was being developed by Herman Wold, also in Sweden and in this way the IEA was linked to them.

Although, at the first meeting, some time was devoted to administrative matters (about half a day) the rest of the week was spent on trying to decide what were the major questions that should be answered by the forthcoming mathematics study (later – but only when IEA decided to repeat the mathematics study – this study was re-named as the first international mathematics study (FIMS). Others such as Jack (John B) Carroll, John Tukey, James Coleman, and Ralph Tyler were to have an association with IEA. It is worth emphasising that the researchers involved were not so much interested in mathematics *per se* but it was regarded as an outcome against which all sorts of tests of the effect of many independent variables could be made.

It was Arnold Anderson who was keen to have good measures of outcomes. Until that time the economists had used the percentage of an age group graduating at certain points in the system as a measure of quality. It was, however, glaringly obvious, that quality should be measured by what the students knew rather than on the number graduating. Ben Bloom saw the world as an educational laboratory and was keen on exploiting the variance both within and between systems to shed light on features of systems that were malleable. So, he was interested – for example - in the class size problem, in grade repetition and ability grouping, age of entry to school, selective versus comprehensive schooling at the junior secondary level (a popular

---

[2] Dick Wolf was a graduate student of Bloom's.

[3] Gilbert Peaker was an inspector of schools from the Ministry of Education in England but was the key statistician and sampling expert used in all Ministry surveys.

theme not only at that time but also in 2008 when one can see in several States of Germany the *Orientierungsstufe* (Grades 5 and 6, which students previously already undertook in the respective secondary school) being integrated into a 6 year primary school to be attended by all children) and in an optimum size of school at primary level and at junior secondary level. There was also the general problem of the extent whether the regression line for socio-economic-status and achievement was the same in all countries or not and, if not, why not. In the first mathematics study book a research question was stated, the literature reviewed, and then the data analysed to answer the question. In short the study was driven by openly stated research questions that were of interest to the participating countries. Some examples of the kinds of questions stated and answered are presented in Table 1.

The first science study as part of the six-subject survey

The number of participating countries increased from twelve to nineteen between FIMS and the Six Subject Survey. Most national centres were institutional (research centres or university departments) and some were units within Ministries of Education. At that time, the overall concern was that the centres participating should have the necessary technical competence to participate. At the same time it was felt that the balance between researchers and ministries about the types of questions to be asked in the research studies was not enough in terms of Ministries.

One major question was if the factors related to achievement were the same for other subjects or were they different? Thus began the six-subject study. The subjects studied were Reading Comprehension (Thorndike), Literature (Purves), Science (Comber and Keeves), English as a foreign language (Glyn Lewis), French as a foreign language (John B. Carroll), and Civic Education (Torney, Oppenheim and Farnen). A system case-study report was prepared by Passow, Noah, Eckstein and Mallea and a technical report was written by Peaker. The results of whether the same factors affected all subject areas were written up by David Walker. The actual data collection for the six-subject study was in 1970 and 1971. The analysis and write-up took until 1973-5. I consider it to have been a weakness of the six-subject that it did not have explicit research questions or hypotheses as in the mathematics study. Rather they were implicit. But, the multivariate analyses conducted were mostly replicated national analyses. It was very important that the analyses were conducted centrally but in a way that national reports could be written from them. This was done to help the national centres in their work.

Table 1. Examples of hypotheses (research questions) from the IEA first mathematics study

- The level of mathematics achievement at age 13 is not related to the age at which compulsory schooling begins
- The mathematical achievement of each country is related to the mean age of the students forming the population within that country; (this was calculated for each of the populations)
- The mean level of achievement in a school will be related to the total enrolment at the school; (this was calculated for each of the populations)
- The level of mathematics achievement is not related to the size of class; (this was calculated for each of the populations)
- In Population 3a (the last grade of secondary school's specialisation in maths) the level of mathematics achievement will be higher where the number of subjects studied is smaller
- The level of mathematics achievement for students will be higher and the variability lower in specialised schools than in comprehensive schools
- Students at the 13 year-old level will have more favourable interests in mathematics in schools and countries which emphasise comprehensive and non-selective secondary education
- There will be a systematic difference in socio-economic status between students in Populations 1 (still in primary school) and those in Populations 3 (last grade of secondary school)
- The average level of mathematics achievement in both terminal groups will be lower in countries with larger percentages of the age group still in school
- When equal proportions of the total age group are compared, countries will not differ in the terminal level of mathematics achievement
- In countries retaining larger proportions of an age group in school, higher levels of mathematics achievement will be attained by a smaller proportion of those still in school but by a larger proportion of the total age group
- When level of mathematics instruction is held constant, inquiry-centred approaches to learning will produce higher and less variable scores in mathematics than will more traditional approaches
- The profile of test performance in each country will be related to the national emphasis on each topic (as reported by teacher ratings) in each school program
- Performance on the mathematics test will be related to per student financial expenditure (1) as a whole, and (2) specifically for teacher salaries

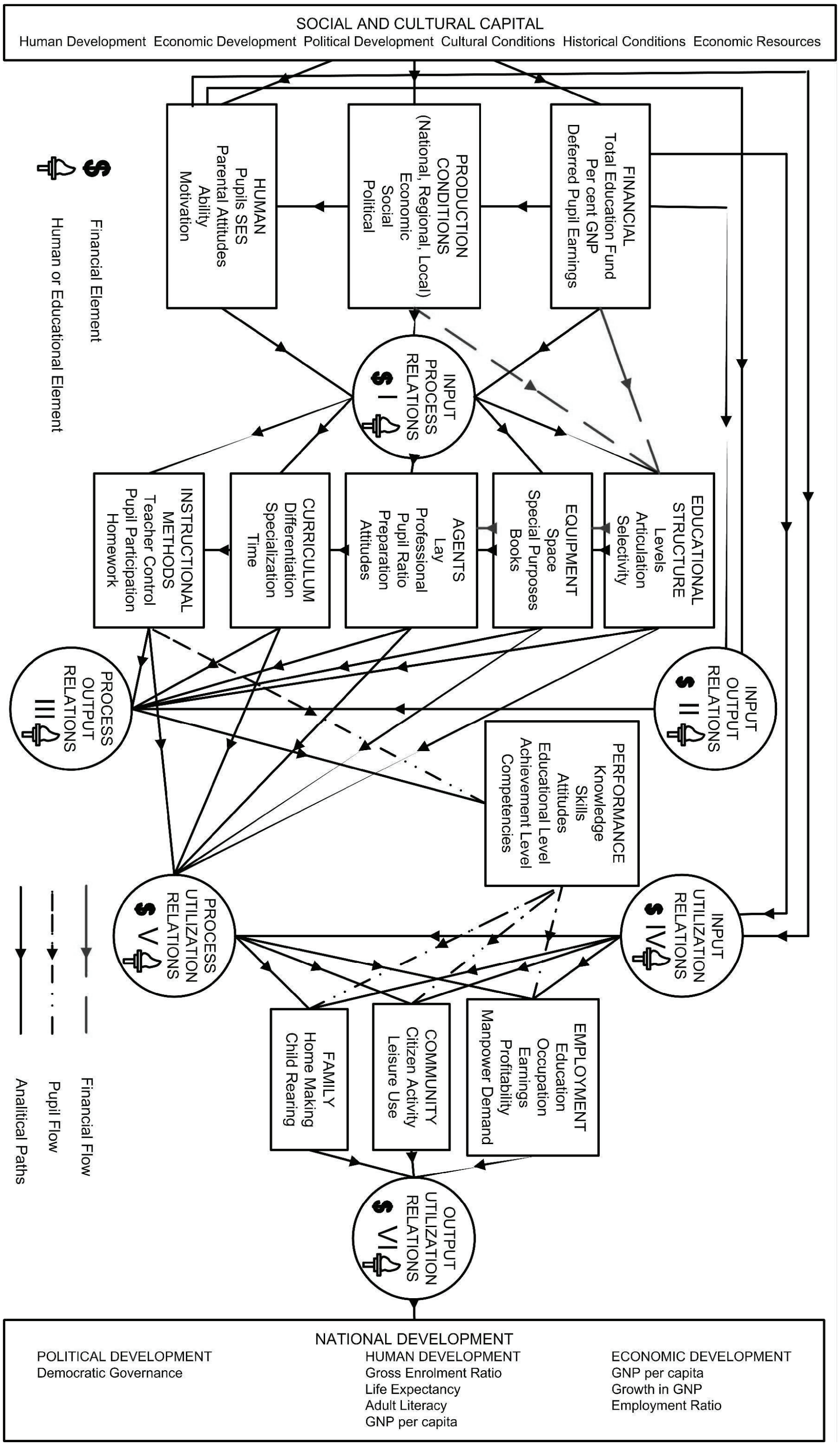


Figure 1. The Lake Mohonk model (amended by Keeves)

What else do I remember from those days? In brief, some of the highlights were visiting all of the participating centres in the first mathematics study together with Dick Wolf to ensure that no grave errors were being committed (we did discover some); incorporating IEA legally (1966). This was done in Belgium (on the legal advice of UNESCO) because Belgium was the only country at the time whose laws allowed an incorporated NGO to have its hypothetical seat in that country but its actual seat elsewhere. I wrote a proposal to that effect in 1965 with Ben Bloom in a hotel in Berlin. It was for one million dollars (very big money at that time) and we were successful. The grant-giving agency was the Ford Foundation; at the same time in Berlin both Bloom and I were refused entry (on the grounds that we were not members) to a Comparative Education Society of Europe (CESE) meeting taking place in Berlin. Husén was attending the meeting and wanted us to attend one speech by a friend of ours. Bloom was President of AERA at the time. After some haggling they allowed Bloom to go but not me. Bloom said that if I could not go, then neither would he. Ben Bloom was using the MESA (Measurement, Evaluation and Statistical Analysis program) graduate students to work with the IEA members in the running of analyses, interpreting them and writing them up at a two week meeting in Chicago in February, 1965. Some of those graduate students later became famous researchers; going to Hungary in 1967 to discuss with the researchers in a quasi-governmental research institute called OPI (Országos Pedagógiai Intézet) to bring them into IEA; going to New Zealand in 1968 to discuss possible New Zealand membership of IEA. The head of the New Zealand Council for Educational Research and the head of the relevant unit in the Ministry of Education were both at the airport to greet me. I knew that both wanted to be the New Zealand member of IEA Both stuck out their hands to greet me at the same time. Which hand to shake first, I wondered; visiting the centres in Chile, India, Iran, Italy, Israel, Poland, and Thailand to check on their progress with the data collection; moving the IEA office from Hamburg to Stockholm in 1969 – closing the Hamburg office on 11 July and opening the Stockholm office on 14 July. A very smooth transition! The results written up were of the analyses in enormous tables on a very long blackboard in Stockholm together with John Keeves. This was for the first three subjects (Science, Reading comprehension, and Literature) in the six-subject survey; organising (with Bloom) the Gränna workshop on 'Curriculum Development and evaluation' for teams from 23 countries. Some of the work was based on results from IEA and many of the faculty were from IEA. This was done at the same time as writing the books for the first three subjects; and, finally, helping to organise and run three workshops (one in English, one in French, and one in German) for recent doctoral students on 'Learning and the Educational Process'. Many of the faculty involved were from IEA.

John Keeves had been the Australian National Technical Officer for the first mathematics study. He only came to his first meeting in 1964, but he conducted the Australian part of the IEA study. He also wrote the Australian national report entitled 'Variation in Mathematics Education in Australia: some interstate differences in the organisation, courses of instruction, provision for and outcomes of mathematics education in Australia' (I suspect that John would not write such a long subsidiary title some 45 years later). No data of publication has been given because the Australian Council for Educational Research (ACER) shredded the report at the request of one of the State Director Generals of Education because one finding in it did not sit well with the educational policy of his state. But John had also been a Science teacher and was also the head of the science teachers association of Australia. IEA appointed the Chief Inspector of Science from England (L.C. Comber) to be the head of the IEA science committee and asked John to assist him. This proved to be a very good combination but it was John who did the analyses for the book (Comber and Keeves, 1973) and who wrote most of it.

The second science study

In 1978 at the Tokyo IEA General Assembly meeting, I became chair of IEA. By that time there was virtually no money in the bank, only one project (the second mathematics study) and only 20 countries involved. It was decided that it would be desirable to start some new projects, encourage more countries to join and to acquire some more money into IEA's bank account!! We did start some new projects. I had thought that if we started four new projects, then at least two of them should prove to be successful. I learned my lesson about it being virtually impossible to stop a project once it had been started. More countries joined and we did acquire more money. But money was very difficult to acquire in the 1970s and even more so in the 1980s. The general policy at that time was that countries should not pay fees (partly they did not want to and secondly it was difficult to devise a policy that would generate enough money but allow the poor countries to pay much less or even nothing). We did not solve the money problem but now it seems to have been resolved (thanks, as far as I can see, to the efforts of people like Tjeerd Plomp and Hans Wagemaker). I was chair and executive director and had to run IEA out of my university office with a part-time university secretary. That part of the operation did not cost IEA one cent. It is perhaps worth mentioning that all communication was by snail mail and telephone (very expensive during those years). With e-mail, Skype, and fax, cheap telephone communications are much easier these days.

One of the projects we started was a second science study. The first one had been done in 1970 and it was time for a repeat. I asked John if he would become director. John was director of the ACER at that time and he hummed and hawed quite a bit. Then he said 'yes' on condition that he could have Malcolm Rosier as the coordinator of the study. I agreed. The work was not difficult but getting money for the study was extremely difficult. We had to get the money in bits and pieces. This we managed – but the study was run on a shoestring. John paid all of his own costs to attend the science meetings. At one point we had to submit an interim report in order to get some more money for data analysis. For some reason the coordinator was unable to write the report so that John and I had to sit down and wrote it. We did and we got the money. That was probably the only IEA publication without an author. We put IEA (IEA, 1988) as the author. In 1986 on my last day as IEA chairman I was at a formal dinner in Sweden and sitting next to the director of the Bank of Sweden Tercentenary fund and the chairman of its Board. By the end of the meal I had convinced them to give the science study one million kronor but the condition was that it had to be spent in Sweden. John finished being director at ACER and he went to Stockholm for more than a year in order to finish the analyses and write two books of results.

At the same time there was the Classroom Environment Study (Anderson, Ryan and Shapiro, 1989) in which many partial Least Squares analyses were run and reported. The key data processor was Norbert Sellin from my office in Hamburg but he was very greatly helped by 'Uncle John'. IEA owes John Keeves a lot.

Some thoughts about current studies

IEA has continued with a series of other projects. Two of its better known assessment programs are TIMSS and PIRLS. Many people are unaware that these are IEA studies. But three other major assessment programs have appeared on the world scene. These are PISA and SACMEQ (run by Andreas Schleicher and Ken Ross respectively, both of whom had much of their training within IEA) and the Laboratorio study in Latin America. PISA is the Program of International Student Achievement, focuses on 15 year-olds in school and runs in three-year cycles for reading comprehension, mathematics, and science. Nearly 70 countries participate. SACMEQ is the Southern and Eastern Africa Consortium for Monitoring Educational Quality, has 15 ministries of education as members and the governing board consists of the ministers of education of the participating countries. It focuses on students in Grade 6 and the subjects are reading, mathematics, and knowledge and awareness of HIV/AIDS. The Laboratorio study focuses on Grades 4 and 6 in about 20 Latin American countries and focuses on reading and mathematics.

I have not been involved with IEA's work since 1994 but I have had a lot to do with SACMEQ and some dealings with PISA. It seems to me that an enormous amount of money is spent on international studies and that it is time to ask a few questions about what is going on.

Who should be compared?

Is it better to have a large international study or regional studies?

PISA and IEA have large international studies. SACMEQ has a regional study. Regional studies tend to have fewer countries. But the countries often have large differences. For example, the difference in the GDP per capita in the SACMEQ countries is four times greater than the difference in the OECD countries. On the other hand, there is a greater feeling of relevance and coherence than in very large studies. Should IEA consider having a series of overlapping regional studies? And what would these regions be? Would they be geographically proximate countries or economically competitive nations or what? In my experience, many Ministries would like to have some comparative information about achievement in neighbouring states – and America! In the IEA Civic Education project several extra questions have been asked for certain regions of the world. Thus the European countries will have special extra questions about Europe, Asia about Asia and Latin America about Latin America. It is not clear if IEA will scale the regional countries just for each region or if it will only leave the international scores and then predict these using some of the special extra questions for each region. But, if the countries within a region are very close on the international scale this will not help much for examining factors explaining between-country differences within the region.

Is there any point in a country participating in a large international study if its mean score is very low and with little variance? Can IEA create an item bank with items and their logits in it?

The position that a country has among many countries in terms of the mean score is a matter of which countries are in the list. If a country has a low score and little variance it cannot exploit its variance for analysis purposes. Furthermore, it is not necessary for a country to be in TIMSS to compare itself with other TIMSS countries on condition that it has the logit values for the TIMSS items it may use in its own national tests. Is there not a case for a country to develop its own tests but then join TIMSS (at a much cheaper rate than now) for having the anchor items classified by described skill level and the international logits for the items?

There is also the possibility of having a filter test and then deciding which set of tests a country should have. This would allow a lot of variance but still allow a country to know where it is on the grand international score.

Given that teacher subject-matter knowledge is a major predictor of pupil performance in the same subject (see SACMEQ and other national results) in many countries, can IEA not test teachers?

SACMEQ succeeded in having the Grade 6 teachers tested despite strong union opposition in some countries. It was astounding how many primary school teachers teaching mathematics were not very good in mathematics and in some cases having a lower performance than their pupils. But, at the beginning Mauritius and South Africa had strong trade union opposition and the teachers were not tested. As soon as these two countries saw the results for other countries, it was agreed that the teachers would be tested.

How can the international studies be made more relevant?

It seems to me that there are two ways in which this can be done. The first is to have general questions (rather like in the first math study) and answer them. These educational questions should be answered in the first publication to come out. (I have heard it said by countries trying to decide between joining PISA or IEA that PISA has analyses presented in the first publication to appear and that IEA's analyses - at least in TIMSS - only come later). The second is to make the research useful to national policy-makers. This implies working with the national research coordinators (NRCs) in the production of national reports. IEA does run workshops for national coordinators and also regional training seminars.

One reason for having replicated national analyses in the first mathematics study and in the six-subject study was that several NRCs were conducting this kind of survey for the first time and it was important that each and every country should profit from the research. It is an approach that was appreciated by all countries since some of them did not have the technical skills to conduct their own analyses.

In the national and international sample surveys with which I have been associated since leaving IEA it has been my experience that the research questions agreed upon by the participating countries form a very good basis for organising a study and the use of its results. In general I have worked with research units from Ministries of education. They have been able to work with the heads of divisions within their Ministries and to know the kinds of research questions they have. Furthermore, when the analyses have been completed then the national researchers have communicated with the heads of divisions about the type of action might be taken by a unit in the Ministry or provincial office of education. These suggestions for action have normally formed the final chapter of a national report with an indication of whether the suggested action requires very little funding to a lot of funding and whether it will take only a short time, medium time or long time to implement And, incidentally, this makes it very easy to demonstrate the impact that the research has had. The questions may be similar to those I cited from the first mathematics study but of course the analyses required to answer them may be very complicated.

It is recognised that where the federal Ministry does not have power over sub-national systems, then it is the sub-national systems that are important.

It has sometimes been said that it is not good research to base policy suggestions on one study. Replications are always useful as are time series data but all I can say is that the Ministries of education have welcomed the kinds of policy suggestions that they have formed based on data rather than anecdote.

How can the sampling be made more appropriate? And, how important are dummy tables?

In my view, all countries (systems of education) want information at the sub-national level, usually for each major administrative unit – whether it be a province or state level. All are interested in the intra-class correlation between provinces and between schools within provinces. It is important for the sampling statisticians to be aware of the analyses that should be conducted later. To this end dummy tables are useful. It is desirable for the researchers to think of the research questions posed and then develop the dummy tables. These tables show the variables (and derived variables) required and the kinds of analyses required. I have seen examples of where the sampling statisticians did not see the kinds of analyses to be carried out and therefore did not plan the sampling such that regions were strata (explicit or implicit). There are many examples where two or three-stage samples have been designed and employed but completely ignored in the analyses of the data.

What should be compared?

Which outcomes should IEA be measuring and comparing?

I assume that IEA undertakes a regular review of what its members want measured. I have two points here. I have noticed that the higher the officials with whom one discusses in the Ministry (head of division and above) the more the officials know what they want. The lower in the hierarchy one goes the less idea the official has. Secondly, it was of interest to me that in SACMEQ where the Governing Board consists of the Ministers of Education of the participating countries that a test of awareness and knowledge of HIV and AIDS was suddenly requested – and subsequently delivered. I am also surprised that there has not been a demand for English as a foreign language. There is a delicate balance to be struck between ideas from researchers and ideas from Ministry officials. Both are needed but bearing in mind that questions about how school systems function are most important.

Can and should systems be compared in other ways than just mean score?

The mean score is only one way of comparing countries. Ken Ross developed one way of comparing systems for mean score as well as social gradient, as well as how large the variance was, and how a system would have fared on scores if it had had the average socio-economic status of all countries in the study. An example is presented in Figure 2. The beauty of this graph is that all of the above information is present in one graph. IEA and PISA have both compared the social gradients but not the other aspects that Ross used. I wonder why not. I once asked this and was told that over 60 countries are too many to get into one graph. But there are many ways of grouping countries. And what is nice about these comparisons is that there is no country that is best at everything. A detailed description of the graph and how to make the calculations has been presented in Ross and Zuze (2004).

There are many other ways in which systems can be compared. For example, how well has a system equalised its sub-systems like provinces or regions? This is relatively easy to do but has not been done. What are the other types of comparisons that systems would find useful?

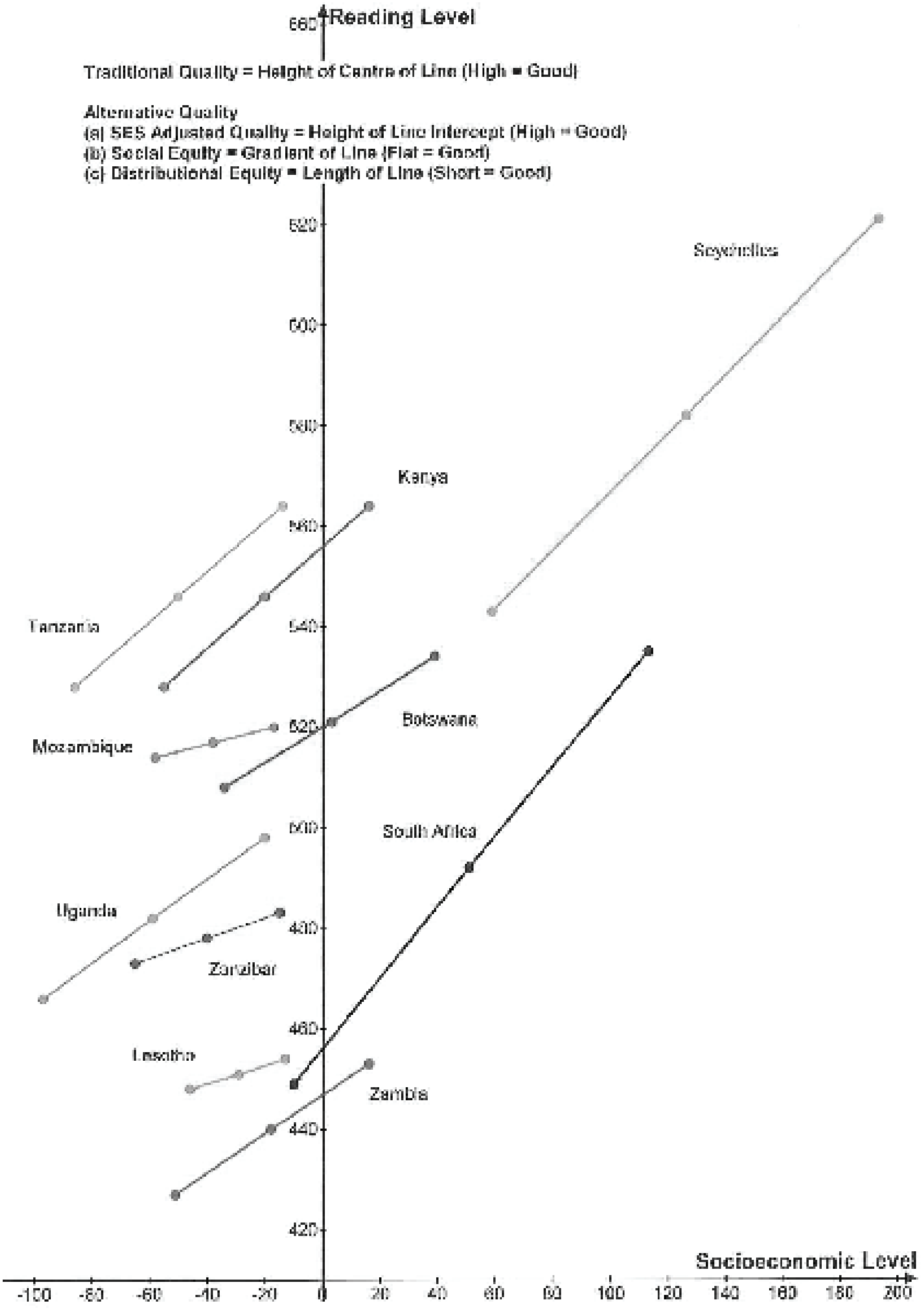


Figure 2. Alternative views of system performance - SACMEQ II (Reading)

How can school systems be helped to know if they are wasting money?

School and classroom resources cost money. Are some schools or classrooms over-resourced? Which are the kinds of resources that are essential if pupils are to learn, for example, to read? Which are not essential? Above which resources (and which kinds of resources) does the score distribution begin to plateau?

True variance?

Whereas from a measurement point of view I can see that conditioning to arrive at plausible values is important, I still have problems understanding how then it is possible to conduct multivariate analysis to estimate the relationship between the independent variables and the dependent score variable – the same independent variables that were used to condition the scores. It is the apparent circularity that bothers me. All of my statistician friends tell me that this is perfectly all right to do despite the fact that it appears to be counter intuitive. But, I am still thinking about it!

Good independent variables

A great deal of effort has been put into improving the measures of the dependent variables but what about some of the independent variables – even SES. In the famous Heyneman and Loxley article it was said that SES had a flatter slope in developing countries than in developed countries. As soon as SACMEQ developed a 'better' measure of SES for the developing countries in Africa, this was no longer the case. Indeed, the slopes were steeper in the developing countries. Is it possible to estimate the effect of constructs that measure the same thing but that might be somewhat differently measured in different countries? How many new kinds of independent variables have been constructed and tried out?

Should IEA be extending its coverage?

IEA started to have good measures of outcomes in education and to test specific educational hypotheses. This is what drove them. They operated in primary, and lower and upper secondary school levels. They trained many researchers in educational research techniques and also in curriculum development and evaluation. Some of the world's best researchers had their training with IEA.

Evaluating vocational education and university education

I do not know *all* of what IEA does now. I have seen that IEA has entered the field of pre-service teacher training and that it has successfully developed a mathematics test for the graduates of such programs. This is excellent and could pave the way for the evaluation of the outcomes of university education especially those in the Bologna scheme. The between country and between institution within country differences in outcomes seem to be enormous. Are they? And what are some of the determinants? Can 12^{th} grade results on national examinations predict university achievement at all? There are many, many such questions.

There are also enormous challenges in identifying the most successful forms of vocational education. The planning and acceptability of different forms of vocational education are complex. The IEA-DPC is in Germany, the home of very successful vocational education. For them to take a lead in this kind of research is a great challenge. There is also the problem of linking these two forms of education to economic productivity and social cohesion and harmony.

Different types of training courses

There is also the problem of training researchers. Hands-on training is not sufficient and needs to be complemented by university training. Take a simple example. It is rare for researchers to understand the many aspects of, say, ConQuest simply by being associated with a project that uses ConQuest. They also need – in my view – a three or four months' university type course in order to understand it fully. Although some universities offer such programs, many do not. What role can IEA play in all of this?

Training of policy analysts

As more data bases become available and as Ministries want more data-based answers to questions, how can IEA help train the new breed of policy analysts in Ministries who must be able to know the different data sets and to know how to analyse them, in order to answer questions quickly?

A final remark

In the different eras of research since its inception, IEA has always been at the forefront. In the beginning there was the challenge of changing the definition of quality of a school system from the proportion of an age group reaching a certain grade level to the measurement of achievement in various subject areas – not just what grade they were in but how much they knew. In a phase that followed quickly on from the first there was the challenge of the identification of the important malleable variables affecting achievement. IEA entered this and with the best analyses available at the time dealt with this problem. IEA refined this by using descriptive skill levels of achievement in its studies. There was then a breakthrough in statistical analyses in multilevel modelling and IEA used HLM and will soon use M plus. The work of the data processing centre is impressive. But, there seems to be a lack of good research questions – the raison d'être of IEA's existence.

IEA's success has never depended on money, but on good ideas, academic excellence, and the commitment of the key people in the study. IEA was capable of 'inventing' international assessments because its strength was research and innovation. If it wants to invent the next generation of international studies, it will have to continue to invest in research and innovation. (It has started to do by setting up the Research and Analysis Unit (RANDA) at the DPC in Hamburg). If it wants to make a living by bringing only repeating studies but in more countries, then it will have to compete with other players and some other group will take over the research and innovation.

References

Anderson, C. A. (1961). Methodology of Comparative Education. In *International Review of Education* Vol. VII, No. 1.

Comber, L. C. & Keeves, J. P. ((1973). *Science Achievement in Nineteen Countries*. Oxford: Pergamon Press.

Hotyat, F. (Ed.) (1958). *Evaluation in Education*. Report of an international meeting of experts held at the UNESCO Institute for Education, Hamburg 17-22 March, 1958.

IEA (1988). *Science Achievement in Seventeen Countries: A Preliminary Report*. Oxford: Pergamon Press.

Ross, K. N.. and Zuze, L. (2004). Traditional and alternative views of school system performance. IIEP Newsletter, October-December.

2

Professor John Keeves and SAIER

Ted Sandercock
President, SAIER

Introduction

This Chapter overviews the contribution John Keeves has made to the South Australian Institute for Educational Research (SAIER) since he joined its Executive Committee in 1999. His contacts with the international educational community and his involvement in the Flinders Institute for International Educational greatly assisted the SAIER Executive rethink its direction and modify its annual program of seminars and workshops.

SA Institute for Educational Research (SAIER)

The Institutes for Educational Research were formed in the late 1920s as supports for and promotion of the Australian Council for Educational Research (ACER). The preliminary meeting of ACER was held in Melbourne on 16 August 1929. All State Institutes were present except for Queensland. The ACER constitution was adopted and sent to the Carnegie Corporation for approval. With the support of the Carnegie Corporation ACER commenced operations in 1930.

The first recorded meeting of SAIER was held on 6 November 1929 at Adelaide University, with Dr A. J. Schultz in the chair.

The Aims of SAIER are:

1. To provide a forum for promotion of research and development activities at all levels of education in South Australia.

2. To provide support and advice to individuals and groups conducting educational research and development.
3. To maintain an information network for national research and development work in education.
4. To provide links between research findings and general practice, in order to assist teachers to benefit from the results of research.
5. To provide liaison, and interact cooperatively, with the Australian Council for Educational Research (ACER).

Membership is open to all persons interested in educational research, especially teachers, lecturers and policy makers as well as educational researchers.

John Keeves and SAIER

During the tenure of John Keeves as Director of ACER, 1977 to 1984, SAIER provided an Institute for an Educational Research member on the ACER Council. Thus members of the SAIER Executive had direct link with John Keeves and ACER. During this period ACER provided speakers at no cost to SAIER. This enabled the SA educational research community to meet with ACER researchers and hear about current research projects undertaken by ACER. The SAIER representative on ACER Council also participated in ACER research forums that were held before ACER Council meetings.

After his retirement from ACER and his appointment as Emeritus Professor at Flinders University, John Keeves attended SAIER Seminars. He was regularly consulted by the SAIER Executive during the planning for the annual seminar program. In 1999 the SAIER Executive of SAIER invited John Keeves to join the SAIER Executive. Fortunately for SAIER, he agreed and his influence created two important opportunities for SAIER. The first was an invitation to SAIER to be represented on the Flinders Institute for Teaching and later the Flinders Institute for International Education. This relationship led to the initiation of the Spring Seminars, a joint undertaking with the Flinders Institute for International Education. The other opening was an invitation for SAIER to support and promote the Flinders Education School November Research Conference.

Springs seminars

John Keeves paved the way for SAIER to hold a series of three Seminars in a major educational topic during August or September. This event became known as the Spring Seminars. The Seminars have involved national and local speakers and attracted audiences from schools, universities, TAFE and the community.

Until 2008, this event had been held at Flinders University, jointly with Flinders Institute for International Education for several years. John's encyclopaedic knowledge of education and his perceptive knowledge and awareness of current issues in education proved to be valuable in setting challenging topics and inviting informative experts for these seminars. Some examples are:

2000 Seminar Series –	School Governance and South Australian Education
2003 Seminar Series –	Retention Rates
2004 Seminar Series –	Australia's Future: Australia's Teachers
2005 Seminar Series –	Education for Sustainable Development
2006 Seminar Series –	Research Issues on the Future of Post-Compulsory Secondary Schooling in South Australia
2007 Spring Seminars –	The Challenge of Change
2008 Spring Seminars –	Towards a National Curriculum

John Keeves was always involved in the initial discussions for a theme and on the planning group designing the seminar topics and finding speakers. He has on occasion been an invited speaker.

South Australian Research in Education Forum

During John Keeves' membership of the SAIER Executive, he has been concerned to improve the communication between the School Systems and the university Schools of Education so that researchers might be informed by current and perceived education trends and needs. The SAIER Executive had maintained contact with the CEOs of the Department of Education and Children's Services (DECS) and normally had a DECS representation on the Executive. However, SAIER noted that there was no formal communication that would assist educational researchers and post-graduate students to easily gain information on current educational directions that might inform their choice of research topics. Through John Keeves' contacts the Deans of Education and other key researchers were often invited to participate in workshops designed to improve such sharing.

In 2008 SAIER was invited to join the South Australian Research in Education (SARIE) Forum, a group initiated by DECS in 2007 to investigate ways of sharing research information across educational and tertiary sectors to achieve greater synergies in research processes. The SARIE Forum members come together with a commitment to strengthen communication between the major educational institutions about educational research needs, projects and outcomes in and for South Australian children's services, preschools and schools.

The SARIE Forum currently has members from

- Flinders University School of Education
- University of Adelaide School Education
- University of South Australia School of Education
- Department of Education and Children's Services (DECS)
- Department of Further Education, Employment, Science and Technology (DFEEST)
- Tabor Adelaide
- South Australian Institute for Education and Research (SAIER)

It appears that John Keeves' concern about better interaction across the educational research community has become a reality.

SAIER Life Membership

At the SAIER Annual General meeting held in March 2005, the SAIER awarded John Keeves, Honorary Life Membership for his outstanding contribution to educational research and to the work of SAIER. John Keeves with his networks and experience in educational research has enabled the SAIER to present current and relevant seminars and workshops for the SA education community. The SAIER Executive appreciates his assistance and guidance in its deliberations and planning.

The current SAIER Executive Committee is

- Dr Ted Sandercock – President
- Mr Ian Graham – Treasurer and ACER Liaison
- Professor John Keeves – Flinders University
- Dr Marietta Rossetto – Flinders University
- Dr Tilahun Alfrassa – DECS
- Dr Carol Aldous – Flinders University
- Dr Kelvin Gregory – Flinders University
- Dr Margaret Scrimgeour – University of SA
- Dr Petra Lietz – ACER

3

John Keeves and international studies – Some personal perspectives

Jan Lokan
Former Assistant Director, ACER

At the outset, I feel I should confess that I was what could be called an 'IEA sceptic' for many years (more on this later). I first found out about the IEA (International Association for the Evaluation of Educational Achievement) quite early on, when I was an enthusiastic young mathematics lecturer at Western Teachers College in Adelaide in 1965 and was interested in finding out how much mathematics the incoming students knew at the beginning of the year. At about that time, John Keeves, who was by that time on the staff of the ACER, gave a talk at the Raywood Inservice Centre in the Adelaide Hills, which I attended and afterwards chatted to him about what sort of test I might use. This was not long after the First Mathematics Study had happened and John helpfully offered me copies of the test from that to use with my students. (Needless to say, from the results of that testing I was disillusioned about how little my students knew! After all, they were training to be primary, not secondary, teachers and were mostly not very interested, nor accomplished, in mathematics.)

Three years later I went to live in Canada, where I worked for several years in the Research Branch of the Ottawa Board of Education. In this role I had the opportunity to go to several American Educational Research Association annual conferences and found that there was usually at least one symposium on the program, sometimes more, about the IEA's work. I was curious and went to listen to a few of these. While I was awed by the line-ups of speakers, 'demi-gods' (as

Postlethwaite says in his article in this book) such as Bloom, Carroll, Coleman, Cronbach, Tyler and Thorndike, I came away with the feeling that the IEA organisation was rather a ‘club for the old boys’ and I wasn’t, at that stage, very interested in the research they were doing. It seemed to me that the problems associated with undertaking such large comparative studies were just too difficult to surmount and that the findings of the studies would therefore not be very useful.

Wheels turn and when I came back to Australia from Canada in 1978 I was offered a staff position at ACER, at the time the Second International Science Study (SISS) was first being planned. I worked for a short while on further analyses of attitude scale data from the earlier mathematics and science studies, but soon began working on other projects and had few further thoughts about the IEA until the early 1990s (more of this later, too).

I don’t need to say much about the IEA itself, or its purposes, as Neville Postlethwaite was much more closely involved and has written about these in his article in this book. He has also described the evolution of the first few studies undertaken – the ones in which John Keeves had most involvement – and touched on the methodological problems experienced. However, I should like to quote from two sources that have provided excellent ‘nutshell’ descriptions. The first, which is from current IEA Chair Tjeerd Plomp’s (1995) Foreword to one of John Keeves’ own publications, touches on the impetus for IEA studies and some basic aspects:

> In the late 1950s (the) group of leading educational research workers (who) met … recognised the need for a comparative research program that was empirically oriented and that investigated problems which were common to many national systems of education. They saw the world of education as a natural laboratory in which different countries were experimenting with different strategies of teaching and learning. By examining the naturally occurring differences between countries in both the conditions of learning and educational outcomes, they thought it might be possible to identify the significant factors that influenced educational achievement. The program of research would be both comparative and cooperative. Decisions were to be made through scholarly debate and not political pressure. Members of the organisations would be research centers and scholars with the competence to undertake survey research and not necessarily governmental instrumentalities.

The second, which is from an IEA Guidebook (1998), referring to a 1995 article by John Keeves himself, provides a good overview of general features of IEA studies as well as some of their procedural aspects, under the heading of ‘What Characterizes an IEA Study?’:

An IEA study:

- is a collaborative effort, all participants have an equal say and all decisions are made consensually in a democratic manner
- involves comparisons across countries but can also identify a relationship that is unique to a particular system of education
- is multidisciplinary, it draws on ideas from the area of education under examination and from disciplines such as psychology, sociology, demography, economics, history, anthropology and statistics
- addresses issues that are important to many audiences with many different interests ─ scholars, research workers, policy-makers and educational practitioners
- is not limited to the descriptions of the effects, but looks for their causes and explains the relationships
- involves a concern for both the process and the product of education, so as to contribute to a better understanding of how things work
- approaches educational reality with all its complexity
- collects information on many important factors and on the many important levels at which they operate
- adopts a broad scope of measures of students' performance to increase data accuracy and consistency
- employs large-sample survey procedures to ensure the comparability of data
- provides opportunities for the professional development of researchers, with each study providing training for its national research coordinators.

IEA studies are indeed complex beasts, but have lofty ideals, which John Keeves has had almost continuous connections with for 45 years, having attended his first international meeting in 1964. The National Foundation for Educational Research (NFER) in England and ACER are sister organisations, set up in similar ways and with similar purposes; Dr W D (Bill) Wall, then Director of NFER, was a key member in the original group whose discussions led to the formation of the IEA and so it is not surprising that ACER's Director at the time, Dr W C (also Bill) Radford, was persuaded to attend the first planning meeting for the inaugural mathematics study. The decision was taken for Australia to join the 11 other countries participating in this study and John, with his background in mathematics and science, was the logical person to implement the national component as the 'Australian Technical Officer'. Thus he became Australia's representative at further international meetings on the study.

The experiences he had in this project, collaborating with survey researchers, data analysts and measurement experts internationally, proved invaluable to him as he pursued further studies of his own. In

1966 he completed his Masters degree in education at the University of Melbourne, on students' attitudes to mathematics teaching, making use of and doing further work with an instrument from the 1964 IEA mathematics study. Beginning in 1967, he took leave of absence from ACER and embarked on doctoral studies at the Australian National University (ANU) in Canberra.

Despite being energetically committed to his PhD research, John simultaneously contributed to the first IEA study of science, which began with an extensive instrument development, trial and revision phase from 1966 to 1969 as part of what was then known as the 'Six Subject Survey'.[1] Given his strong background in science and science teaching, plus his work with the Australian Science Teachers Association, John was invited to be a member of the International Science Committee for this study. The committee's nine members, whose role was to develop specifications for and guide the development of the instruments, came from Australia, England (two members), Germany, Israel, Japan, Scotland, Sweden and the United States – elite company indeed. As for the earlier mathematics study, ACER was the national centre responsible for providing input to the instruments, carrying out the field trial and then the main study in Australia, this time with Malcolm Rosier as the National Technical Officer. Main data collection for science took place in 19 countries in 1970.

Meanwhile, John was making good progress with his ambitious doctoral work. He had the courage to attempt to collect good and sufficient data with which to test a model, almost as complex in appearance as the Lake Mohonk one shown in Postlethwaite's article in this book, of how environmental factors influence student achievement. His study was simplified a little because he could ignore big picture community factors and school and classroom resource factors, which were reasonably uniform in Canberra where the data were collected. The majority of both school buildings and people's houses were 'of recent construction' and 'of high standard', as John acknowledged in the first chapter of his thesis. That said, his study was more complex than almost all IEA investigations in that it incorporated a longitudinal design – his sample of about 240 students were tested at the beginning and end of their first year of secondary school – and also included classroom observations of teaching and interviews with the sampled students' parents as well as many other measures of educational environment.

[1] Australia did not take part in the other five components of this survey, which pertained to literacy, languages and civic education.

For such a project to be undertaken for an individual's PhD is remarkable and hardly anyone would have been in a position at the time to complete it successfully. But 'pull it off' John did and his study, variations of which were published in 1972 in both Australia and Sweden, still provides a relevant example for and a strong contribution to current research. John recognised that his work could be criticised for being set in a relatively homogeneous community, but argued that 'the origins of this inquiry into the educational environment in Australia must be appreciated and its significance seen in a wider context'. He went on to acknowledge the work of the IEA up to that time and justified his own study as complementing that work

> in being an intensive study carried out in a restricted setting, rather than one of the extensive surveys that are being undertaken … It seeks to explore in depth and to fill in some of the detail necessarily lost in a large-scale project. It owes a great deal to help and guidance received from individual members of the IEA body and draws extensively on research undertaken by IEA personnel and their co-workers in their respective countries.

Not only was his PhD study comprehensive in the range of constructs addressed and types of data collected, it was unusual in that a random sample of students was selected and sophisticated, very new methods of data analysis were used. John always had a strong interest in methodology, but readily acknowledged his indebtedness to Gilbert Peaker, a sampling and analysis expert associated with the IEA, who had himself carried out a longitudinal follow-up to the UK's Plowden Report four years after the initial inquiry and published the results in 1971. John knew Peaker through the involvement of both of them in the IEA and sought his advice, admitting that 'the procedures of analysis that he has suggested from his own investigations have had a profound influence on the way in which the analyses reported in this monograph have been carried out'. Help was also offered by Torsten Husén, IEA Chairman, who allowed John 'to draw on the materials and wisdom gathered together by IEA from many corners of the world'. It seems eminently sensible that John should make use of what he had learnt from his own involvement in an IEA project to enhance his own doctoral work, which in turn very much enhanced the profile of Australian educational research.

Having presented his thesis at the ANU, John turned his attentions back to the IEA Science Study. He spent a year as a Visiting Scholar at the IEA International Office in Stockholm, in 1971−72, where a data analysis unit had been set up. He spent his time primarily conducting analyses, but also wrote many components of the first report of the study, published in 1973. He returned to ACER in 1972 and became its

Associate Director. ACER was a member of the IEA and its Director, or representative, was expected to attend the annual General Assembly meetings, which provided a way for Dr Radford and John to keep in touch with developments. After the Six Subject Survey, though, some years passed before anyone had energy or funds to think about the next phase of studies. Then, from the late 1970s, several proposals were made for studies to be done over the next decade. These entailed a second mathematics study (SIMS), a second science study (SISS) and new projects to investigate classroom environment, writing and reading literacy.

It is useful at this point to make some comments on how the IEA studies were funded. Initially the international aspects were supported by grants from entities such as the United States Office of Education, the Spencer Foundation and the Ford Foundation in the United States, and later from sources in England, Japan, Germany and Sweden as well. Participating countries were always expected to cover their own national data collection, analysis and reporting costs. As the studies increased in scope the international costs became too high for the IEA to continue to function in this way, particularly after extensive training programs for national research personnel were begun, and 'participation fees' were introduced.

The situation regarding national costs in Australia was interesting. For more than its first 60 years, the ACER received a 'core grant' from the Federal, State and Territory governments. The Director and senior staff each year prepared a suggested program of research, which was discussed, and sometimes amended, by the governing Council, members of which were primarily academics or bureaucrats. By the late 1970s, when the proposals for further mathematics and science studies were on the international agenda, John Keeves had become ACER's Director and was keen for Australia to be able to participate in them. He obtained the Council's endorsement for both and also, later, for the Classroom Environment study. A budget was prepared for data collection in SIMS to be undertaken in 1978, when new components of the tests were supposed to be ready to add to selected 'link' items from the 1964 tests. The test developers were running late, however, leaving John with the dilemma of allocated budget money and undelivered instruments. His solution was to go ahead and collect data with the instruments that were on hand, namely the ones that had been used in 1964. This enabled change over time to be validly assessed within Australia, a main objective of SIMS, but Australia found itself excluded from the international reports because a different way of measuring change had been adopted.

No such problem occurred with the operational aspects of SISS. John was considered to be the best qualified person to lead the project internationally and the Study Centre was duly established at ACER, with Malcolm Rosier again leading the Australian national component. Data were collected in 24 countries (counting English and French Canada as two) in 1983-84. Problems were encountered at the reporting stage however, when international funds were short, resulting in delays in the release of reports. As Postlethwaite recounts in this book, money was finally obtained to finish the reports, but had to be spent in Sweden. John had by this time moved from ACER to Melbourne University and was recruited from there to go to Stockholm in 1986 to undertake data analyses and report writing.

After John left ACER, the new Director, Professor Barry McGaw, decided that costly IEA studies were not a priority in the allocation of core funds. This meant that Australia did not participate in the written composition study, the larger 30-country reading literacy study, the computers in education study or the pre-primary study that took place during the second half of the 1980s and early 1990s. A civic education study also occurred during this period; Australia participated but with the University of Canberra, not ACER, as the national study centre (although some ACER staff members provided input to the project).

Once the Third International Mathematics and Science Study (TIMSS) was proposed in the early 1990s, the Australian governments, through the Australian Education Council (from mid-1993, the Ministerial Council on Education, Employment and Youth Affairs (MCEETYA)) agreed that Australia should take part, given that this would enable achievement trends to be measured from the first, second and third studies. ACER re-entered the IEA study fray, this time with designated funds for TIMSS from the various Australian governments. That is when I became involved with the IEA again, in 1993, as the TIMSS National Research Coordinator. It is also when I lost my IEA scepticism, over the course of the first year or so, when I could see at first hand the thoroughness of the instrument development stages, translation, pilot testing, field trials, reviewing and so on, that were undertaken. A lot of effort was put into development of standardised procedures for sampling, test administration and scoring of free-response items, which were presented in manuals for use in national centres and supported by comprehensive training sessions over several days. In short, many stones were turned to ensure that tests, data collection and scoring would be as comparable and appropriate across countries as humanly possible.

To be fair, the demi-gods who were influential in setting up the movement for international comparative studies via the IEA recognised that there were many problems to overcome before educators would accept the studies as making important contributions to knowledge about teaching and student learning. Sometimes these problems loomed large. Writing in the Foreword to the First Science Study, Husén (1993) commented:

> Administratively, financially, and—not least—technically, a study of this magnitude was a gigantic task larger than any previous study in the social sciences. We certainly at the outset greatly underestimated the problems we would be faced with later.

Every study undertaken to this day has had at least one associated Technical Report produced – and who better to write these than the gamut of experts in their fields who contributed much effort to attempting to overcome the problems. In the 50 years since the IEA began, huge strides in methodology have been made and described in reports and other publications so that others may benefit.

Other large-scale comparative studies now exist, the most wide-ranging being the OECD's Program for International Student Assessment (PISA), which first tested 15-year-old students in reading, mathematics and science literacy in 32 countries in 2000, with its main focus on reading. Testing occurs every three years, with one of the three areas as main focus on a cyclical basis. PISA differs from TIMSS in that it emphasises critical and reasoning skills more like the skills that students will need in their life beyond school, rather than solely curriculum-based learning. It also differs in that it has had more involvement of education ministries rather than academic research institutions and hence policy implications are tending to be better understood. So far, MCEETYA has supported Australia's participation in both PISA and the continuing cycles of TIMSS (the 'T' now stands for 'Trends'), believing that both programs provide valuable information on our students' performance. PISA has added even more stringent procedures, including more translation steps and requiring that trained test administrators go to the schools rather than allowing teachers to administer the tests, but there is much commonality between its methods and those developed over the years by the IEA. Given that several of the prime movers for PISA had worked on IEA studies, this is not surprising. Once again ACER took on the role of International Study Centre in 1998, this time for PISA, having, with some European and American collaborators, won the contract from the OECD.

Before finishing, I should like to react to Postlethwaite's comment in this book that 'IEA's success has never depended on money'. I understand what he is getting at, but I want to put forward the view that money aspects have led to significant problems in many instances. At least two studies have had to be aborted internationally because funding could not be obtained to continue them, and the long delays in release of reports of some of the studies were one of the reasons the OECD decided to introduce PISA. Substantial IEA participation fees are now charged, which must preclude countries from signing up for studies – for example, the participation fee for the current Teacher Education and Development Study in Mathematics (TEDS-M) is $US120 000, in four annual instalments. The number of countries involved in TIMSS has grown, but many African and Middle-Eastern countries' participation is supported by the World Bank. In Australia MCEETYA has paid the participation fees for TIMSS and PISA as well as the national component costs, but decided not to support our involvement in the first two Progress in Reading Literacy Studies (PIRLS) at middle primary level in 2001 and 2006. Our participation in PIRLS 2011 is still under consideration, but we will have missed out on being able to assess achievement trends through not having the earlier data.

An overview of John Keeves' involvement in IEA work would not be complete without mentioning some of his publications beyond the reports of results he produced in his role as international or national coordinator of some of the studies. In 1991 he and Neville Postlethwaite contributed chapters on students' achievement to a book on science education edited by Torsten Husén. In 1992 and 1995 the IEA published two summary books written by John, the first drawing on results from the 1970 and 1984 science studies and the second being an overview of the first 15 IEA studies and their findings. John was uniquely positioned to prepare these books, given his thorough knowledge of the science studies and overall interest in the IEA. That he did so is a tribute to his commitment to that organisation. He and Postlethwaite went much further, in that both have edited large encyclopaedic volumes on educational research as well as contributing to other books and many journals.

As an example, John edited a large handbook on educational research, methodology and measurement which was published in 1988. This book, of more than 800 pages, is divided into four main sections and 11 subsections within these. John's versatility and breadth of knowledge shows in that he not only wrote the Introduction but also contributed a dozen or so chapters, which appear in eight of the 11 subsections spread across all the sections. Again, he was in a good position to produce such a volume as he had developed much technical expertise over the years, partly from his involvement in IEA studies, and had met many other professionals in related fields whom he could ask for contributions. He also had the skills and stamina to produce a second edition of this comprehensive work, published in 1997. These are but a few examples of his prodigious contribution to educational research more generally, over and above publications that are more directly related to IEA projects and procedures.

Now that the original instigators of the IEA are no longer with us, I would like to conclude by saying that Keeves and Postlethwaite, perhaps together with others such as Albert Beaton, clearly became the IEA demi-gods themselves.

References

Quoted text comes from the following (listed in order as quoted):

Plomp, T (1992), Preface. In J P Keeves, *Learning Science in a Changing World: Cross-national Studies of Science Achievement: 1970 to 1984*. The Hague: IEA, p. v.

IEA (1998), *IEA Guidebook: Activities, Institutions and People*. Amsterdam: IEA, p. 17, citing Keeves, J P (1995), The case for international comparisons. In J L Lane (ed.) *Ferment in Education: A Look Abroad*. Chicago: NSSE.

Keeves, J P (1972), *Educational Environment and Student Achievement. ACER Research Series No. 89*. Melbourne: ACER, pp. 13, 14 & 17.

Husén, T (1973), Foreword. In L C Comber & J P Keeves, *Science Education in Nineteen Countries*. Stockholm: Almqvist & Wiksell and New York: Wiley (Halsted Press), p. 10.

Other sources consulted

Burstein, L (ed.) (1992), *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. Oxford, UK: Pergamon.

Degenhart, R E C (1990), *Thirty Years of International Research: An Annotated Bibliography of IEA Publications (1960-1990)*. The Hague: IEA.

Husén, T & Keeves, J P (eds) (1991), *Issues in Science Education: Science Competence in a Social and Ecological Context*. Oxford, UK: Pergamon.

Husén, T & Postlethwaite, T N (1996), A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, 3(2), 129−160.

Keeves, J P (ed.) (1988), *Educational Research, Methodology and Measurement: An International Handbook*. Sydney, NSW: Pergamon.

Keeves, J P (ed.) (1992), *The IEA Study in Science III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon.

Keeves, J P (1995), *The World of School Learning: Selected Key Findings from 35 Years of IEA Research*. The Hague: IEA.

Keeves, J P (ed.) (1997), *Educational Research, Methodology and Measurement: An International Handbook*, 2nd edition. Oxford: Pergamon.

Rosier, M J (1980), *Changes in Secondary School Mathematics in Australia, 1964−1978*. Hawthorn: ACER.

Tuijnman, A C (1994), A tribute to J P Keeves. In A C Tuijnman & T N Postlethwaite, *Monitoring the Standards of Education: Papers in Honor of John P Keeves*. Oxford: Pergamon and Elsevier Science.

4

Guiding the development of educational researchers

Penny Van Deur
School of Education, Flinders University

When I began my PhD I had a strong interest in research and some experience investigating educational problems having recently completed a Master of Arts (MA) by research. Even so I, like many research higher degree students, did not feel confident about undertaking the level of research required to complete a PhD. Fortunately, I discovered that I would be able to sit in on classes about research design and evaluation run by Professor John Keeves. When I went to the class I found local and international higher degree students who had come to Flinders University enthusiastic about research but, like me, needing instruction in the process of carrying it out. In the first three-hour seminar I heard that we would learn how to design a research study, develop a proposal, select appropriate techniques for collecting and analysing data, and report on our findings in a scholarly way. I was pleased that I would have the opportunity to learn about these topics from an expert in the field of research design who showed an obvious passion to share with his students his knowledge about research. This Chapter enables me to acknowledge and share the knowledge that I gained from Professor Keeves.

Research Design and Evaluation course structure

The course 'Research Design and Evaluation' was to be conducted over 14 weeks and had a strong theoretical basis that was linked with practical assessment tasks. Listed below is the structure of the course:

Week 1: Research in Education: nature, impact, needs and priorities
Characteristics of Educational Research

Needs and priorities for Research in Australia and Asia-Pacific countries

Week 2: Evaluation and accountability in Education

Selection and definition of a problem for Research

Discussion of proposed problems

Week 3: Formulation of models and hypotheses.

Problems of sampling, measurement and significance

Preparation of a research plan and research proposal

Week 4: Review of related research and scholarly work

Locating educational information

Computer searches for information

Evaluating research reports

Week 5: Types of research – Descriptive studies

Introduction to evaluation-A taxonomy for learning, teaching and assessment

Week 6-Types of research – Correlational research studies

Introduction to evaluation-The classification of educational goals: Attitudes and values

Week 7: Types of research – Causal-comparative research

Modelling and the testing of models

Introduction to evaluation- Measuring instruments

Week 8: Types of research – Experiments and interventions and Design-Based Research

Introduction to evaluation and monitoring of educational outcomes

Week 9: Types of research – Policy research

Problems of sampling and size of sample

Sampling schools and classrooms, students, occasions

Sampling for interviews and sampling tasks

Week 10: Types of research – Case study and systemic research

Ethical considerations in research

Week 11: Types of research – Historical and Hermeneutic research

Interviewing

Week 12: Types of research – Action research

Preparing a research report

Week 13: Presentation of a proposal to staff and students for assessment.

Each class member to present a 30 minutes research or evaluation proposal that is to be summarised in a paper of no more than 3,000 words together with a PowerPoint display of the key ideas associated with the proposal.

Week 14: Assessment of learning and performance

Evaluating topics and courses

Assessing effectiveness of instruction

Assignment and thesis writing

Completion of Assignments which could take the form of one of the following:

1. Review and evaluation of a published research report in a field of interest; or
2. Preparation of a review of research in a field of interest; or

3. Preparation of a research proposal for a dissertation or thesis; or
4. Preparation and presentation under supervision of a PowerPoint presentation on an assigned topic

Evaluation and accountability in education

In our introductory class, Professor Keeves stressed that evaluation is an important aspect of research and is particularly important in educational research because researchers need to evaluate the quality of teaching and learning. He also introduced the concept of modelling as important in educational research because models can be derived from the body of knowledge and then tested through construction of hypotheses. Importantly, he stressed models allow researchers to observe the real world of education and ask questions of it. He also discussed assessment and evaluation as important ways to examine the way schools operate with new approaches to education. Assessment was described in terms of the performance of people, while evaluation involves judging the worth of curriculum or institutions. It is through evaluation that researchers can examine what the school system is trying to do in its curriculum. Formative and summative evaluations track progress in student learning. Importantly, evaluation has to involve research because information is collected and interpreted in relation to the context of the school or perspective of the educational system. Professor Keeves stressed that evaluation cannot be overlooked when planning and carrying out research and added that accountability is important also because it considers value for money of institutions.

Characteristics of educational research

Professor Keeves described the process of educational research in terms of three worlds. World one is the real world which can be changed by action, while world two comprises the human mind which is influenced by teaching and instruction (meta-cognition examines what it happening in the human mind) and collective views. Finally, world three is the body of educational knowledge, which now includes information and communication technology that is used to learn in world two.

Researchers carry out educational inquiry and research by collecting data, and use measurement and experimentation to construct models, hypotheses and theories. Then they test to see whether the constructed model is consistent with the real world. Professor Keeves described some research paradigms such as: hermeneutic and historical research through which researchers take ideas from the real world and look to see how they exist in the body of knowledge; critical theory, which uses collective views to plan to change the real world, and action research which uses the process of collecting information about the real world as a way of changing it.

Professor Keeves discussed how the impact of educational research goes beyond the production and dissemination of knowledge to user-oriented action research. Research is important because it can respond to critical issues and problems and utilise new ideas from other disciplines such as Information Communication Technology (ICT). ICT can be applied to the investigation of problems through the use of simulation studies and virtual inquiry where researchers can build models and create random effects to test them. He stressed that educational research is not just applied psychological research but is multi-disciplinary, multivariate and multi-level.

When priorities are set in educational research there are a number of criteria to be considered including social and economic needs; equity in terms of social and racial justice; needs for professional development; existing research strengths, recent advances and important gaps in knowledge. Current priorities in Australian research include the teaching of thinking skills; learning in pre-school and adult years; assessment of student learning; learning to learn; problem-based learning, self-directed learning and valuing the way children learn out of school. I was pleased to hear this because many of these aspects were issues I was interested to investigate in my own PhD.

Types of educational research

Throughout the course, Professor Keeves described many types of educational research which could be carried out depending on the research questions being investigated. Some of these types of research are described below.

Policy research examines educational policy changes which could result from changes in economic conditions, developments in science and technology, political changes, social and cultural evolution, and advances in education. When researchers act as policy analysts they use indicators and statistics and seek public opinion. Policy research can examine the influence and effect of persuasion and economic issues, and consider what other educational systems are doing in policy development.

Case study research focuses on how something operates as a system so that the researcher looks for patterns of operation within the system. It often involves a single, in-depth study which uses participant observation and interview. This type of in-depth investigation can take a holistic or systematic approach where patterns and networks are sought and a deep understanding of a phenomenon is developed. Researchers carrying out this research select cases, negotiate access, undertake field work, keep organised records, analyse these records and write a report on the findings of the case study.

Historical research is carried out to provide understanding through an interpretative account of educational phenomena, which leads to the introduction of change through social action. Here the emphasis is on looking at the past to develop the future so that knowledge of the past is used to initiate change. Primary and secondary sources are used to collect information. These include newspapers, legislative documents, photographs, oral histories, committee meeting records, relics and statistical data. A completed historical study has propositions and questions, uses primary and secondary sources of data which are carefully and thoroughly documented, employs external and internal criticism, shows a logical analysis of relationships and has sound conclusions.

Descriptive research is valuable for identifying or describing the way things are or the way sub-groups view particular topics or issues. It involves collecting data in order to answer questions about the opinions of people about a topic or issue. To carry out this research the researcher needs to identify a problem, select an appropriate sample of participants, collect valid and reliable data by questionnaire, interview, observation, telephone or online survey, and then analyse the data and report any conclusions.

Correlational research studies are carried out to determine whether or to what degree a relationship exists between two or more variables with the degree of relationship expressed as a correlation coefficient. If a relationship, or correlation, is established it is possible to use one variable to predict the other. A high correlation between two variables leads to more accurate prediction than a low correlation. Types of analysis used in correlational research include discriminant analysis, factor analysis, cluster analysis, multi-dimensional scaling and canonical correlation.

Causal and comparative research employs models to explain or account for variability and uses arrows to show causal relationships. In this type of research a model is derived from and also contributes to theory. An analogue model describes a relationship, such as the factors influencing intake and dropout at school. A semantic model uses words to describe a relationship such as Carroll's description of school learning. A schematic model uses a flow diagram, picture or map to explain what is happening, while mathematical models, such as Rasch models, use formulae to state relationships that can be tested statistically.

Two other useful strategies for conducting educational research are the experimental design of pre-test/post-test and the longitudinal study where data are collected at a number of time points.

Process of educational research

Professor Keeves explained that there are many factors that influence educational outcomes and identified control as a potential problem in educational research. When a researcher uses the control/treatment approach stronger causal inferences can be made. However, control is usually thought of in psychological terms where a key factor is identified and the researcher assigns a causal relationship due to a treatment being assigned to one group and withheld from another group. This can be difficult to carry out in a school or classroom setting. Procedures such as regression analysis can be used for statistical control. In this case, least squares requirements, or structural equation modelling using maximum likelihood procedures are valuable forms of analyses.

Generally, educational research studies need the characteristics of realism, representation and randomisation. Due to the nature of schools and classrooms, there can be problems setting up experiments which result in realism being destroyed. Resulting studies can also lack representation and randomisation. Randomisation can be used to balance the effects of many factors when they are matched. Professor Keeves explained that alternative procedures to experiments in school settings are quasi-experimental studies and interventions. These types of studies investigate explanatory variables and controlled, disturbance, randomised, outcome or criterion variables. It is important that statistical control is applied to both quasi-experimental studies and interventions. In these types of studies, correlations can be investigated within treatment and control groups as well as between treatment and control groups. Two levels of data can be collected: within students and between occasions, and between students.

When monitoring educational outcomes a valuable approach is the pre-test/post-test design because the researcher can examine how well a student does in a post-test after controlling for prior performance such as prior achievement. The pre-test/post-test approach can be used also to examine the effects of the classroom on attentiveness and questioning while keeping in mind that prior achievement has an effect on later achievement.

Model building

Professor Keeves placed a strong emphasis on building models because they are ‘collected views that have to be tested against the real world’. He explained a model as a collection of inter-related hypotheses that is built from accumulated evidence or theory. A model has a structure, or system, and allows for deductive fertility, which leads to further inquiry. It allows the researcher to predict what is happening, aids imagination, contains a causal mechanism, and provides explanation. It can be assumed to be true, but should be tested through measurement where data is collected so that the researcher has confidence in its usefulness. A model should be causal, extend inquiry and have structural relationships for the way the data are generated.

The type of model developed by a researcher is important because it influences the nature of the study carried out. A model can guide research but can never be accepted as the whole truth because evidence could arise leading to its rejection. However, repeated replication of the research design increases acceptance of the findings.

The first step to building a model is the development of propositions which are combined to become a hypothesis. Each proposition can be investigated and tested as an acceptable statement about the real world. Importantly, propositions can be confirmed or rejected but never proven.

Educational researchers usually want to know why things are happening rather than just describing what is happening. Professor Keeves explained the rule that things that are different are not the same but researchers want to know why they differ. In other words, researchers want the best explanation for why things happen. Models or hypotheses can be used to explain a problem situation through the use of causal statements. To attribute cause, there must be three conditions: X must precede Y (earlier events influence later events), X and Y must co-vary together, and no alternative explanation accounts as well for why X and Y co-vary together. Once it is developed, a useful model allows the researcher to predict what is happening, aids imagination, contains a causal mechanism, and provides explanation. Any model should be able to be tested so researchers can have confidence in its usefulness.

The ten characteristics of a model are that no model is true but it can be adequate; it is based on theoretical knowledge; it explains how data are generated; it contains data that has a systematic and a random part because researchers can make errors in recording or calculation (statistical analysis is used to examine the size of the systematic part); a model can be accepted as adequate or rejected as unlikely to be correct; it can help show why things happen; the size and practical significance of an effect; unusual events or outlier data which may provide useful information; data obtained from a sample (of the population) and inferential statistics are used to test models.

Professor Keeves described the model building process as beginning with an understanding of relevant theory in the relevant literature. Importantly, carrying out a literature review puts the study into the context of what is known already, points out research strategies and instruments found to be useful and not useful, and later facilitates the interpretation of results. The model constructed from the literature has propositions and hypotheses and must be based on sound reasoning, provide a reasonable explanation for the predicted outcome, state the expected relationship of the variables and be testable in a reasonable time frame. Models can be inductive and based on observed relationships or deductive because they are derived from theory. Some models are declarative and are non-directional or non-causal, while some are directional and show causal relationships.

Evidence in the form of information or observations is used to test models although the use of logical argument is important. Qualitative data analysis is used for counts and quantitative data analysis is used for measurement. Computers and Information Communication Technology bring together the procedures of logic, mathematics and statistics.

The process of developing a research topic

The selection and definition of a problem to be researched is an important issue for any researcher. In educational research a topic should be identified that is researchable and can be managed in terms of skills, cost and time. It should be significant in that it will contribute to knowledge and practice and be interesting to the researcher. Researchers could select a research topic in which aspects of a theory are confirmed or rejected. A topic may be selected based on personal experience, a desire to replicate previous studies, or the researcher may know of a practical problem that needs investigation. Before carrying out a study, the researcher needs to consider whether the study will be generalisable and contribute to knowledge and if the results are testable, coherent or consistent with what is known. The researcher should consider whether the conclusions provide a plausible explanation of the real world. As a rule a study should be simple so that the findings will be simple and clear and it should also be fruitful in that it will lead to further studies.

Once an area or topic has been selected the researcher needs to talk with his/her supervisor, read the literature to be aware of the research that has been done on this area, and then narrow the topic. If doing humanistic research, the researcher generally narrows the topic before beginning the investigation. If doing scientific research, it is more common to start with a general idea, summarise relevant theory, identify a general framework, identify a specific framework relating to the topic and then summarise specific theory relating to the topic.

An important part of the research process is the development of research questions. General research questions (big picture) can be identified first, followed by more specific ones (fine grain). Research questions should include the variables of interest and there should be a relationship between the variables and participants involved. When the researcher is developing the background to the topic consideration should be given to justifying the study and showing any practical applications of this work.

Preparing a research plan and research proposal

Professor Keeves described the components of a research plan as including an introduction, literature review, methods section, data analysis outline, time schedule and budget. In the introduction the researcher needs to state the topic, describe the background to the research, state the research questions, the model/hypothesis/proposition as well as the significance of the study and any delimitations of it.

The literature review section or chapter should include what is known about the topic and give any literature supporting the model/ hypothesis/proposition and research questions. A literature review for a research proposal can be developed through a ten-step process. The researcher first defines the problem, then consults experts, locates secondary sources, selects then reads these, refines the problem statement, reads primary sources, evaluates the material covered, develops research questions then writes a review of the relevant literature. This section concludes with a brief summary of the literature and its implications for the proposed study to be undertaken.

The methods section should describe the participants by defining the source of the population and describing the sample. Next, there should be a description of any instruments to be used and a statement made about how each one will measure the variables. There should be a discussion of the process of administering and scoring any instruments. If the instrument to be used is a published one, it is important to say why it is suitable for the study and give an indication of its measurement properties in terms of reliability and validity. If the researcher has developed a research instrument, the methods section should include a description of how the instrument was developed and how it will be evaluated. There should also be a statement to show how any instruments used are related to the model being used to guide the investigation or to the sample used.

Professor Keeves described many measuring instruments including observation, field notes, observation schedules, interviews, and questionnaires. Observation can be carried out by participant or non-participant observers or a combination of the two. Field notes are a form of data that is used later on and any notes should be clear, extensive, and descriptive. Summary words or statements should be avoided and only activities seen should be described.

If the researcher wants to include a descriptive aspect, he/she could use an observation schedule to record immediately any ideas and thoughts on the topic along with the date, time and setting. This observation schedule becomes a list of issues that guide observation and provides the researcher with a focus during the observation and provides a common framework for field notes. A simple observation schedule includes participants and their roles; a map to show interactions; activities or unusual events; what the observer was doing during the session (eg walking around), and memo writing ('as you observe you are analysing the situation you are observing').It is important to realise that observer bias is a threat to the quality of the observation because the observer may note invalid information and the observer effect could impact on behaviour or perception of those being observed. Some ways to enhance validity and reduce observer bias are to look for consistency by staying in the field for long periods, broadening the database to include other additional participants, and working to ensure the trust or comfort of the participants. The observer can try to recognise his/her own bias by comparing data with another researcher as well as allowing participants to review and critique notes and recordings made during an observation. As well as personal observation and noting, data can be collected on audio-tape or by camera. This can assist the researcher to examine unusual or contradictory results. Triangulation, or the use of different data sources, will assist validity. Professor Keeves advised that it is important for researchers to collect data until they are getting the same information over and over again. He recommended 160 minutes observation of a normal situation.

Interviews are a valuable source of data in educational research and allow the researcher to obtain information that cannot be obtained from observation because respondents supply reasons for why things happened. Interviews may be totally unstructured, free-flowing and open-ended or structured where the interviewer has a protocol or list of questions to ask. Often interviews combine both structured and open-ended sections so that respondents have the opportunity to talk about aspects of the topic they are especially interested in or the interviewer can follow up issues raised by the respondent.

The research proposal will detail the design of the study that is selected by the researcher and is based on the model/propositions/hypotheses. It will describe the procedure to be used, the steps to be taken, give details about the participants and make clear any assumptions as well as detail any limitations of the investigation. It will define any new terms early in the text. The section on data analysis will have a description of any techniques to be used. A time schedule will be calculated and extra time should be allowed because there can always be delays when educational research is being carried out. An expenses budget will include such considerations as travel, computer programs, printing, and mailing costs. A research proposal should be evaluated by the student and his/her supervisor and then it should be presented to an audience so that feedback can be received. It is advisable to do a pilot study to enable the researcher to trial data collection and analysis techniques and make adjustments before the main study is carried out.

Attention to detail

Professor Keeves stressed frequently that all researchers should pay attention to detail at all stages of the research and write-up. His discussion of the use of primary and secondary sources is an example of this. He advised that secondary documents are useful in the initial stages of reading about the topic. However, he discussed the value of primary sources, which should be used wherever possible because the researcher is able to check what has been said rather than trust the interpretation of another writer. He described the value of internal criticism for building up a story and authenticating it so that it is possible to check the consistency of two people reporting on an event. If using primary sources, he suggested the steps of finding people and talking to them to find out what they know about an issue, talking to particular people closely involved with the issue and then getting ideas from other relevant people. Then secondary sources, such as books, could be used to establish the context of the research. Following this, magazines and journals could be examined for articles and studies closely related to the topic being investigated.

Reporting on findings in a scholarly way

Writing-up is an important aspect of any research investigation. A clear, simple and straightforward style should be used but at the same time the report should reflect scholarship. Professor Keeves advised that a research report should be prepared by making an outline. The main chapters should be broken down into section headings and then sub-headings. The headings and sub-headings should be checked to see that they show logical steps which make a story. Next, a table of contents helps the writer to see that the structure of the report is logical because a thesis is really an argument and the writer needs to trace or think through it. It should accurately describe what was done and what was found. When writing a report, care should be taken to ensure the use of correct spelling, grammar and punctuation. An important tip for thesis writers is that the body of the document should be written in the past tense except when referring to what is in the thesis when the present tense should be used. Professor Keeves discussed abbreviations and advised that they should be avoided unless they are common abbreviations. They must be defined if used in tables, figures, footnotes and references. Other tips included the following: authors cited in reports should be cited by surname only; numbers are expressed as Arabic numerals unless the number is less than nine or the first word of a sentence when it is written in full.

The report should have a title, an abstract that describes the most important aspect of the study, the participants, instruments used, data collection procedures, major results and the conclusions of the study. In the introduction there is a statement of the problem and some background to it, the purpose and aims of the study, specific research questions, a short discussion of the significance and limitations of the study, and definitions of any uncommon terms.

The literature review describes what is known about the topic and concludes with a brief summary of literature and its implications. The methods section should accurately describe what was done. It should include a description of the participants and the selection process used, sample size and the major characteristics such as age and grade. It details any measurement instruments (such as questionnaires, tests, observation schedules, interview schedules) and indicates the validity and reliability of each one. This section also details the design of the study and justifies why the design was chosen. It should discuss any problems associated with the design and how these were handled. The final part of the methods section should describe the procedure or steps used to conduct the study including details about how participants were assigned to groups, any treatments undertaken, how data were collected and any conditions under which the data were collected.

The results section describes the techniques used to analyse the data and presents and interprets the results. Often tables and figures are used to do this. Professor Keeves stressed that each table or figure should be introduced and placed following its introduction and that each one should be clear enough to be interpreted without related text. The conclusions section is used to discuss the results in relation to the topic and research questions. In this section there should be a statement to show whether the original hypothesis, model or propositions were or were not supported by the data. It should also show how the findings are related to what has been found in other studies. The section on recommendations and implications should discuss the theoretical implications of the findings, any possible revisions or additions to existing theory, implications of the findings for educational research in particular, and make recommendations for future research or action. This would be followed by the references section, which should list all sources directly used in the text of the thesis and follow the style manual of either the American Psychological Association or the Chicago system. The Appendices follow this section and provide information regarding the study that is too lengthy or detailed to include in the main text. They may include instruments such as tests or questionnaires, covering letters and raw data.

Clear communication of results and insights from research

Professor Keeves stressed that research reports should be written clearly so that any results and insights could be communicated to an audience of interested researchers or practitioners. There is unity between the processes of creating knowledge through research, its diffusion in an incidental and undirected way and dissemination in a purposeful and directed way, and ultimately its utilisation. Information is disseminated through publication in books, research monographs, journal articles, summary reports and newsletters set up on the internet. This volume of information available via Information Communication Technology means that researchers and practitioners need training in how to access and select information which is high quality and most useful for their needs.

Even though Information Communication Technology furthers communication, it is still the case that interpersonal contact is a powerful means of communication. Presenting papers at research conferences can assist researchers to make this initial contact with other researchers working in the field. For this reason the course on research design and evaluation gave students the opportunity to make a 30 minute presentation to staff and students of a research or evaluation proposal. Each student was to be assisted to make a PowerPoint display of the key ideas associated with the proposal which was to be summarised in a paper of no more than 3,000 words. In addition, students could choose to review and evaluate a published research report in their field of interest, prepare a review of research in a field of interest, prepare a research proposal for a dissertation or thesis, or prepare and present, under supervision, a PowerPoint presentation on an assigned topic.

Design-Based Research methods of inquiry

I was beginning my PhD when I attended Professor Keeves' classes on research design and evaluation. My topic was an investigation of the concept of effective Self-Directed Learning (SDL) in primary students. I was interested in the role and function of knowledge of SDL in developing self-directed learners. I was keen to develop a model of SDL in primary students, scaffold their knowledge of it and assess its development over time at the school and student level. In order to do this I needed to design an innovation to assess knowledge of SDL in primary schools. At this time, I was grappling with the difficulty of research design that would enable me to do the tasks of collecting primary students' views about Self-Directed Learning (SDL) and describing SDL in primary school students.

In week eight of our course on research design and evaluation Professor Keeves described a new approach to educational research called Design-Based Research (DBR). As I listened, I realised that this method could be what I was looking for as a way of organising my PhD investigation. The way DBR was described it seemed that these methods of inquiry could enable me to achieve my aims of developing an assessment of knowledge of self-directed learning in primary schools through the processes of discovery, exploration, confirmation, and finally dissemination of the assessment instrument developed and the results of teaching and assessing students' knowledge of SDL. As it was described, DBR involves continuous cycles of design, enactment, evaluation and re-design. Importantly for my PhD investigation, DBR was described as suitable for developing an innovation or intervention that is informed by knowledge from outside as well as inside the school setting. This approach also allows the use of both qualitative and quantitative methods.

The more I read about DBR method of inquiry the more it seemed that this was an innovative way of carrying out research in schools. A compelling aspect for me of DBR was that it would allow the use of mixed-methods that are guided by the questions and issues surrounding the investigation. This emphasis on the use of quantitative and qualitative methods of inquiry meant that I could use it in my investigation to clarify what SDL means for students in primary school, and evaluate the effectiveness of teaching students about SDL and assessing their knowledge of it over time. It seemed that this approach would accommodate movement along a data analysis continuum from coarse-grained analysis to fine-grained highly refined statistical techniques needed to analyse multi-variate and multi-level effects. It also included the concept of failure that is used in DBR to identify ways in which an innovation could fail in the context of its implementation in school settings.

I could see that it would be possible to use DBR methods of inquiry to devise the innovation of an assessment of SDL then trial and revise it as required, based on experience with students in different school contexts. Using this process it would be possible to refine both theory about SDL itself and students' knowledge of it. An advantage of DBR for my PhD investigation was the emphasis on the production of useable knowledge that can be transferred to curricula and fitted into the real work of teachers and learners. This appealed to me because I wanted to develop an innovation that would have credibility with classroom teachers and be able to be used in schools to assist students to be self-directed learners who could carry out inquiry.

My research investigation involved assessment of knowledge of SDL in primary schools but there were no currently available instruments that could be used for this assessment. For this reason I developed an assessment of school support for inquiry, teachers' ratings of students as self-directed learners at school, and students' knowledge of SDL. I found that there was no guide available for a reporting format for DBR so I would also need to develop one for my thesis write-up.

Design-Based Research seemed to be a worthwhile approach for my investigation which aimed to initiate and support change in educational practice. DBR allowed me to employ descriptive research and analysis at the school and student level to describe and assess school support for inquiry and primary students' knowledge of SDL. I could use statistical techniques in design-based ways so that the analyses could go from coarse grained (descriptive) to fine grained (highly refined multi-variate and multi-level) analysis. By using rigorous statistical analyses I was be able to test whether there were interaction effects of school and student factors on knowledge of SDL and its development over time.

Carrying out educational research

Professor John Keeves was not my PhD supervisor but became a very helpful 'critical friend' during my candidature. I benefitted greatly from his classes on research design and evaluation and went on to complete my PhD, graduating in 2007. I used Design-Based Research methods of inquiry to guide my PhD research. The final thesis detailed the research from its initial conception through cycles of describing SDL, modelling SDL and school inquiry environments, then designing, trialling, refining and administering assessment instruments. Following this I analysed (with the help of my supervisor, and Professor Keeves and his colleagues) results of school and student assessments; revised the model of effective SDL in primary students and the assessment of knowledge of SDL and teaching about it. I considered the potential pitfalls for implementing the assessment of SDL in primary schools. I also developed a format to report my DBR research in stages of Development, Implementation and Review.

Disseminating knowledge of the process of educational research

The students in Professor Keeves' classes on research design and evaluation all completed their higher degrees at the masters or doctoral level. Many of the students were from countries in the Asia-Pacific region and have returned home now with a deep knowledge of the process of educational research. We were taught by a generous teacher who stressed that knowledge of educational processes is important because it influences how and what is learnt, as well as how and what is taught. He emphasised that as educational researchers we should be able to analyse and interpret research information so that we could translate it into policies and practices that would improve teaching and learning in schools.

Professor Keeves shared his passion, skills and knowledge because of his desire to develop educational investigators who would understand the 'design of research and evaluation studies in education from the selection of a problem for investigation, through to reviewing relevant theory and previous research reports, procedures for data collection and the writing of research reports.' To this end, he was highly successful.

5

John Philip Keeves: A transformational mentor

Nordin Abd Razak
Universiti Sains Malaysia, Penang, Malaysia

"Thy sword of nobility shalt spruce green fields"

Introduction

Not everyone can be a good mentor. They must have special and unique qualities as well as capabilities which may be different from those of others. Guest (2000) suggested the key qualities of a mentor, namely: (a) is a wise and trusted counsellor; (b) is suitably experienced; (c) has travelled the mentee's path; (d) acts as a confidential adviser and guide; and (e) stimulates professional development. Furthermore, Guests also included some aspects of mentoring, such as (a) a long term process; (b) sharing experiences; (c) offering encouragement; (d) insight through reflection; and (e) mutual learning – it is a two-way relationship. Stanley and Clinton (Clinton, 1992) identified some essential characteristics of successful mentors. According to Stanley and Clinton, good mentors "take a passionate interest in seeing young people grow", has "good intuitive judge of potential", and "knows the importance to young people of crucial formative experiences." (p. 116). Rick and Van Gyn (1997, p 88) identify the characteristics of an effective or ideal mentor as "wise and experienced with regard to the needs of the protégé, accepting of alternate views, flexible in their behaviour, patient and unbiased".

However, it is difficult to have an opportunity to be intellectually guided by an effective and empathetic mentor. You can be (considered as) assumed to be a lucky individual or student if you have one, especially at a time that you really need somebody to guide you – to share your problems and concern as well as to get ideas about what you are doing. Throughout my quest to accomplish my doctoral study, the experience that I have had under the supervision of Professor John Philip Keeves (after this I write as Professor Keeves) qualifies him as a good mentor. He is the one of those individuals who has the qualities of a good and empathetic mentor and has made a difference to students' learning. As a good mentor and leader, Professor Keeves (a) shows a desire and a willingness to give up time to help others, (b) willingness to share skills, knowledge, and expertise with his students, (c) demonstrates a positive attitude and acts as a role model, (d) takes a personal interest in the mentoring relationship, (e) exhibits enthusiasm in the field of study, (f) values ongoing learning and growth in that field, (g) provides guidance and constructive feedback to his students, (h) is respected by colleagues and employees at all levels of the organisation in which he works, (i) sets and meets ongoing personal and professional goals, (j) values the opinions and initiatives of others, and (k) motivates others by setting a personal example of a high standard of research.

In this chapter, I would like to link the mentor qualities and capabilities possessed by Professor Keeves and his leadership in general, and transformational leadership in particular. The discussions focus on the associations of these qualities with the "Four Is" (idealised influence, inspirational motivation, intellectual stimulation, individualised consideration) of transformational leadership (Avolio, Waldman, & Yammarino, 1991). In relation to academic advising, transformational leadership can influence major changes in the attitudes and assumptions of students and help build a commitment for academic goals and objectives (Slack, 1997). According to Burns (1978), transformational leadership involves leaders and followers (in this case, advisers and advisees) who raise one another to higher levels of motivation, as the leader appeals to the ideals and values of subordinates. One of the best descriptions of transformational leadership has been penned by Bass (1985, p.171), a renowned exponent of the idea. He says succinctly:

> Leaders are truly transformational when they increase awareness of what is right, good, important and beautiful, when they help to elevate followers' needs for achievement and self-actualisation, when they foster in followers high moral maturity, and when they move followers to go beyond their self-interests for the good of their group, organisation or society.

Linking mentoring and transformational leadership

There are a variety of ways the word mentoring can be defined. According to Wright (2004, p.55), mentoring is an "intentional, exclusive, intensive, voluntary relationship between two persons—a teaching or learning connection…in which both persons work to nurture the relationship and contribute to the connection". The CNA (2004, p.24) states, "Mentoring involves a voluntary, mutually beneficial and usually long-term professional relationship. In this relationship, one person is an experienced and knowledgeable leader (mentor) who supports the maturation of a less experienced person with leadership potential (mentee)". Studies of this relationship show that mentoring provides two distinct functions for the mentee: one, a psychological function; and two, a career-facilitation function (Claire & Deluga, 1998; Kram, 1985; Levinson et al., 1978). Furthermore, Poon (2006) has suggested that these definitions underscore a number of key facets about mentoring, namely: (a) mentoring involves a relationship; (b) mentoring entails learning; and (c) mentoring is a mechanism to achieve significant leadership development.

The relationship between the mentor and the mentee seems to be one of determining factors for the success of the mentoring processes. This relationship is believed to be dependent on the effective accomplishment of the roles and responsibilities between the mentor and the mentee (Bally, 2007). Moreover, Bally (2007) suggested that the mentoring responsibilities such as coaching, counselling, advising, confirmation, accepting, and friendship influencing the quality and the outcome of the career path of both the mentee and the mentor which are clearly consistent with leadership behaviours. The qualities and practices adapted from transformational leadership provide an appropriate framework for the implementation and positive support of mentorship.

According to Burns (1978, p. 4), transformational leadership results in "mutual stimulation and elevation that converts followers into leaders and may convert leaders into moral agents." The effectiveness of transformational leaders has been linked to their ability to clearly articulate goals for their followers, convey images of how those goals can be reached, demonstrate confidence in their ability to achieve those goals, and motivate followers to work to make those goals a reality (Kuhnert & Lewis, 1987, p. 650). Transformational leaders or mentors are responsible for performance beyond expectations of their mentees as they transmit a sense of mission, stimulate learning experiences, and arouse new ways of thinking (Hater & Bass, 1988). Bass (1994) has proposed an approach that is particularly relevant as it focuses on aligning internal structures to reinforce values, morals, and ethics specific to the environment in the quest of knowledge culture. In relation to academic advising, transformational leadership can influence major changes in the attitudes and assumptions of students and help build a commitment for academic goals and objectives (Slack, 1997). By utilising Bass' (1994) perspective of four leadership initiatives, if understood and implemented, they may enhance the quality of the mentoring processes between the mentor and the mentee as well as produce high quality mentoring results. The long-term goal then is to facilitate the development of a set of transformational mentoring processes and techniques that can be used to effectively move the students through the highest stages of learning. Scandura and Schriesheim (1994) have argued that the characteristics of transformational leadership are consistent with requirements for effective mentoring. Therefore, transformational mentor behaviour may be more congruent with protégé receipt of mentoring functions compared to transactional or laissez-faire mentoring behaviours. The four dimensions of transformational leadership proposed by Bass (1994) are *inspirational motivation, individualised consideration, idealised influence,* and *intellectual stimulation,* also known as the *'Four Is'.*

Mentoring experiences with Professor Keeves from Four I's perspective

Enrolled as a mature-aged student at the Flinders University, my intention was not only to obtain the doctoral degree but also to learn and acquire skills to conduct research, learn good supervision skills, statistical concepts and recent data analysis techniques using different statistical software. I was in dire need for a mentoring environment that was able to provide good models and guides, with the provision of support, challenge and vision. I believed that support enabled the development of a constructive relationships environment, and encouragement to meet new challenges. The challenges offer an opportunity for me to develop a new learning experience and the vision provided me with a view of the future and my work as a lecturer. I strongly believe that Professor Keeves created these productive mentoring environments which are applicable to the Four I's transformational leadership qualities. As demonstrated below, logical connections can be made between each of the four factors and mentoring activities and possessed by Professor Keeves and the Transformational leadership Four I's and discussed below. The descriptions of the "Four I's" below are drawn from Bass and Avolio (1994).

Inspirational motivation

According to Geijsel, Sleegers, Leithwood, & Jantzi (2003), motivation and inspiration are the aids for overcoming barriers to the successful for implementation of mentoring, and for the mentee/students to make a commitment to accomplish their task. An effective mentor, as part of the mentoring processes, should be able to generate enthusiasm and optimism to convince mentees that they can meet the challenge that has been set forth for them. The ability to motivate and inspire their mentee is critical to the success of the transformational mentor that shows the inspirational motivation quality of a mentor. Professor Keeves demonstrates inspirational motivation by not just simply encouraging his students to complete a task, but rather combines meaning with a challenge to achieve increasingly higher levels of performance while providing a vision in which he encourages his students to share. He provides a clearly articulated vision of what a mentee or student can achieve, and not only provides guidance, but may also inspire the student, encourage a sense of purpose, and foster the attachment of that purpose to his or her work.

Professor Keeves inspired and motivated students and never made his students frustrated even when we (or I) encountered problems. This type of help made the works or tasks meaningful for me as a student and motivated me to accomplish more than what I planned to accomplish.

Individualised consideration

Individualised consideration refers to the emphasis by the mentor as a leader's attention to each individual's need for achievement and growth. Individualised consideration has been specifically identified with mentoring individuals (Clair & Deluga, 1998). A mentor's or leader's use of individual consideration was a crucial element in mentee or student achievement of their full potential through a close consideration of their developmental needs (Yammarino, Spangler & Bass, 1993). In providing individual consideration, the mentor was not only cognisant of and sensitive to the current needs of followers, but was also aiming to lift those needs to a higher level. This was done by Professor Keeves, for example, by coaching and mentoring, as well as by setting examples and tasks, that were developmentally consistent with the needs of each of his students. Professor Keeves provided empathic attention to his mentee and helped them to perceive and understand their unique contributions, gifts and talents. He monitored the work of the students, not in a controlling way, but from a position to provide additional feedback and guidance when needed. He recognised and valued diversity while providing each student with specifically tailored opportunities for learning and development in a supportive manner.

Idealised influence

Idealised influence is identified in mentors or leaders who behave in a manner that causes mentees or followers to want to emulate them. The idealised influence results when transformational leaders behave in ways that cause them to be role models for their followers which is consistent with the concept of role modelling in the mentoring literature (Clair & Deluga, 1998). According to Clair and Deluga (1998), to be effective as a role model, a mentor must show respect and trust and this may be earned by giving precedence to the needs of others rather than to their own personal needs.

Professor Keeves productively creates an environment of trust and friendliness, open relationships, emphasising credibility and attentiveness. Being authentic and treating people with respect and dignity are all the characteristics that he demonstrates to achieve a sense of trust. Trust relates to open communication channels, requesting, receiving, and providing feedback which are integral to the success of his mentoring quality and essential for providing adequate support and a challenging environment. He promotes the utilisation of interpersonal communication skills and active listening for each individual student or mentee. He shows commitment, modelled and share experiences, best practices, hard work and knowledge excellence. No wonder most of his students emulate his transformation mentoring qualities.

Intellectual stimulation

The leader or mentor is also expected to provide intellectual stimulation to his mentee or student. In providing intellectual stimulation, the mentor or leader is said to bring into the light of the mentee “an awareness of problems, to their own thoughts and imagination, and to the recognition of their beliefs and values” (Yammarino & Bass, 1990, p.153). Furthermore, by providing an intellectually stimulating environment, transformational mentors are able to foster the development of creative solutions to problems, learning through mistakes, being honest and be willing to engage in difficult conversations. Professor Keeves encourages his students to come up with new ideas and they are never publicly criticised for their mistakes. As a transformational mentor, Professor Keeves emphasises on the intellectual stimulation which is consistent with the career development activities of coaching and providing challenging assignments to his mentees. He encourages creativity and novel approaches that result in mentees diverse ability of conceptualising and understanding of problems.

The coherent connections between the mentoring qualities and transformational leadership behaviour identified above strongly suggest that Professor Keeves is the transformational mentor that shows an effective and empathetic mentoring from the standpoint of providing career development and psychosocial support to his mentees or students. Prof Keeves effectively unconsciously adapts approaches and corresponding targeted behaviours from the transformational model as proposed by Bass and Avolio (1994, pp. 3-4) such as: (a) questioning assumptions; (b) reframing problems; (c) suggesting approaches to old situations in new ways; (d) encouraging creativity; (e) soliciting new ideas and creative problem solutions from students; (f) encouraging students to try new approaches; (g) encouraging students to develop their own ideas, even when they do not agree with the mentor's ideas; (h) encouraging students to revisit problems; (i) creating a "readiness" for changes in thinking; (j) creating a "holistic" picture that incorporates different views of the problems; and (k) listening to seemingly foolish ideas.

As mentoring activities should work from the premise that all students are individuals and, as such, the mentoring process needs to be personalised for each student. Some students are very capable in their research abilities and can take a project, go with it and make significant progress on the project. Other students need to have more intentional guidance with their projects. Taking a personalised approach Professor Keeves has successfully empowered his student's to:

Become self aware - evaluating themselves, developing an awareness of their personal strengths and weaknesses, their skills, the valuable contributions they make, their interpersonal relationships with others.

Become self directing - following their own path, going where their instinct and intuition take them, choosing the directions they want to take.

Develop a sense of their own purpose - understanding their personal needs, what interests them, what they want, where they are going in life.

Experience their own success - the reward from feedback from accomplishments in which they have shared. Learning from their mistakes. Gaining deeper insights into themselves and others and the world in general through "doing".

Research conducted by Ricks and Van Gyn (1997) has found three characteristics of a successful mentoring relationship which are congruent with the characteristics possessed by Professor Keeves. First, he promotes the mentee as a whole person, responding to their intellectual, physical, social, emotional and moral development. Secondly, he places the mentee in the centre of the mentoring relationship. He facilitates in a manner, which allows the mentee to explore choices, and make decisions. Thirdly, he acknowledges the uniqueness of the mentee. All these arguments provide strong contention to accept Professor Keeves as a transformational mentor.

Conclusions

Successful transformational mentoring relationships established by Professor Keeves have long been recognised as central to student learning and have a significant influence on the activities, relationships, learning, development and growth of students' professional careers. He has provided the opportunity to develop a creative and supportive environment for me as a student to achieve my peak performance. He provides an important role model in creating an environment for fostering significant learning experiences for his students. Thank you to Professor John P. Keeves for being my mentor and for your unending encouragement and support throughout my pursuit of a Ph.D. I have been given invaluable guidance by Professor Keeves. This precious experience has made a great impact on my academic life. I am very grateful to him.

DEDICATION TO
PROFESSOR JOHN P. KEEVES

This chapter is dedicated to Professor John P. Keeves, who patiently guided me in the realisation of my work. John, you have encouraged me, you believed in me, you held me up when I was down. You gave me faith; you gave me hope; and most of all you have been my inspiration because you made this journey possible. I will miss your smile; your words of encouragement and your enthusiasm for my work. This article is for you John, my source of motivation and my mentor.

References

Avolio, B. J., Waldman, D. A., & Yammarino, F. J. 1991. Leading in the 1990s: The Four I's of transformational leadership. *Journal of European Industrial Training,* 15(4): 9-16.

Bally, J.M.G. (2005). The Role of nursing leadership in creating a mentoring culture in acute care environments. *Nursing Economics*. 25(3), 123-148.

Bass, B. M. (1985). Leadership and performance beyond expectations. New York: Free Press.

Bass, B.M. (1994). *Improving Organisational Effectiveness Through Transformational Leadership*. Thousand Oaks, CA: Sage Publications.

Bass, B.M. & Avolio, B.J. (eds.) (1994). *Improving Organisational Effectiveness: Through Transformational Leadership*. Sage.

Burns, J. M. (1978). *Leadership.* New York: Harper & Row.

Canadian Nurses Association (CNA). (2004). Achieving excellence in professional practice a guide to preceptorship and mentoring. Ottawa, Ontario.

Clair, L.S. & Deluga, R.J. (1998). *Transformational Mentoring*. Paper presented at the Eastern Academy of Management annual meeting, organisational behavior track, May 15, 1998, Springfield, MA.

Deranger, S. (1997). The relationship between leadership identity formation and motivation for further educational pursuits as seen through a first nations summer leadership institute. Unpublished Master of Education thesis. University of Regina.

Hater, J. J., & Bass, B. M. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied Psychology, 73*, 695-702.

Geijsel, F., Sleegers, P., Leithwood, K., & Jantzi, D. (2003). Transformational leadership effects on teachers' commitment and effort toward school reform. *Journal of Education Administration, 41*(3), 228-256.

Guest, G. (2000) Coaching and Mentoring in Learning Organisations. Unpublished: Presented at 5th International ASHRM Conference in Bahrain in Mentoring and Coaching: Definitions and Differences. [Online 25/03/02] http://wwwreobertsoncooper.com/coaching_mentoring/defintions_differences.html.

Kuhnert, K. W., & Lewis, P. (1987). Transactional and transformational leadership: A constructive/developmental analysis. *Academy of Management Review, 12*, 648-657.

Kram, K. E. (1985.) *Mentoring at work: Developmental relationships in organisational life*. Glensview, IL: Scott, Foresman.

Levinson, D. J., Darrow, C. N., Klein, E. B., Levinson, M. H., & McKee, B. (1978). *The Seasons of a Man's Life.* New York: Knopf.

Poon, R. (2006). A Model for servant leadership, self-efficacy and mentorship. [Online 3/03/09] http://www.regent.edu/acad/global/publications/sl_proceedings/2006/poon.pdf

Ricks, F. V. G., G. H. (1997). Mentoring relationships as learning opportunities. *The Journal of Cooperative Education, 22 (3)*(Spring).

Scandura, T. A., & Schriesheim, C. A. (1994). Leader-member exchange and supervisor career mentoring as complementary constructs in leadership research. *Academy of Management Journal*, 37(6): 1588-1602.

Slack, T. (1997). Understanding sport organisations: The application of organisation theory. Champaign, IL: Human Kinetics

Stanley, P. D., & Clinton, J. R. (1992). *Connecting: The mentoring relationships you need to succeed in life*. Colorado Springs: Navpress.

Wright, W. C. (2004). *Mentoring: The Promise of Relational Leadership.* Carlisle, UK: Paternoster Press.

Yammarino, F.J. & Bass, B.M. (1990). Long-term forecasting of transformational leadership and its effects among naval officers: some preliminary findings. In K.E. Clark & M.B. Clark (Eds), *Measures of Leadership*, (pp. 151-170). West Orange, NJ: Leadership Library of America.

Yammarino, F.J., Spangler, W.D. & Bass, B.M. (1993). Transformational leadership and performance: A longitudinal investigation. *Leadership Quarterly*, 4(1), 81-102.

6

Islamic Science: The golden age of Islam and the need for discourse between Islam and modern science in contemporary societies

Yasmeen M. Faruqi
School of Education, Flinders University

Introduction

In contemporary societies the production of scientific knowledge has become "Eurocentric through the dominance of western modernity", and has led to a westernisation of knowledge at a global level (Ur-Rehman, 2002, p. 65). Therefore this has resulted in a concerted effort by contemporary Islamic scholars to call for an 'Islamic science' or the 'Islamization of knowledge' (Golshani, 2000; Davies, 1991; Ravetz, 1991). Islamic scholars have also been calling for an investigation to reflect back to the so-called 'Golden Age of Islam' when there seemed to have been major developments in science; and an acceptance of harmony between religion and science. This chapter examines how the tenets of Islam and the Islamic view of nature facilitated the development of science in the so-called 'Golden Age of Islam'; and how the Western view of nature conflicts with the Islamic view. Another aspect that is explored in this chapter is whether common or universal values can help build bridges between modern[1] or Western science and Islamic science.

[1] The terms Western science and modern science have been used interchangeable in this chapter.

Islam is not just a way of life but a civilisation, in which the way of life may vary "from one Muslim country to another, but is animated by a common spirit far more humane than most Westerners realize" (Mazrui, 1997, p. 118). The religious life of Muslims all over the world and in Islamic countries is governed by the tenets of Islam. Very briefly these tenets are considered, in the following section, along with the values that they exemplify.

Tenets of Islam

Islam is not so much about believing in the faith but how Muslims respond to the faith, that is, it "not so much a noun but a verb, an action" (Ball and Haque, 2003, p. 317). Ball and Haque (2003, p. 315) argue that "Islam is tantamount to accepting a way of life – spiritually, politically, and socially, about how to behave in family and public life", and that Islamic values impact on all areas of society. In Arabia in the early seventh century, God or *Allah*[2] through the Prophet Muhammad (PBUM), revealed the religion of Islam, which means 'to surrender' or 'submission to the will of God' (Cornell, 1999).

At the core of the belief system of Islam are the Five Pillars of Belief, along with core values, laws, and behaviours outlined in both the *Quran*; and the practice (*Sunna*[3]) and teaching (*Hadith*[4]) of the Prophet Muhammad (PBUM[5]) (Moore, 2006). Monotheistic in nature, Islam claims that human beings believe in the same God, have similar needs, wants, and experiences, and can relate to a set of universal moral principles (Cornell, 1999). Thus the Islamic set of moral principles are good for all mankind, regardless of race, ethnicity, gender, or origin (Moore, 2006). Islam's rapid expansion in the seventh and eighth centuries as well its vigour today lies in its claim of universal validity across time and space (Cornell, 1999).

Individuals can convert to Islam by practising the Five Pillars of Belief and adhering to the core values, laws, and behaviours outlined in the *Quran* and the *Hadith*, teachings of Prophet Muhammad (PBUM). The first pillar of belief in Islam is called *Shahadah* (the act of bearing witness). The *Shahadah* requires that a Muslim declare his or her devotion to Allah or God by declaring 'there is no God but Allah and

[2] Allah: in this paper Allah will be used to represent God as stated in the Islamic belief system.
[3] Sunna: the habits and the religious practice of the Prophet Muhammad (PBUM).
[4] Hadith: documented traditions of the teaching, actions, and sayings of the Prophet Muhammad (PBUM).
[5] PBUM: means peace be upon the Prophet Muhammad and is written after his name as a mark of respect.

Muhammad is the messenger of Allah.' Therefore by this profession of faith Muslims assert that Allah is the only God, and not part of the pantheon (Moore, 2006). Thus clearly Islam rejects the concept of 'Trinity' and presents a direct theological challenge to Christianity. For 14 centuries this has been at the heart of the tension between Christianity and Islam (Smith, 1999) with intractable views held by both religions that prevent any compromise.

The second pillar of belief in Islam requires Muslims to pray at five specified times during the day. The establishment of prayers is stated in the *Quran* and were demonstrated by the Prophet Muhammad (PBUM) during his lifetime (*Sunna*). While performing prayers Muslims must face the Great Mosque (*Holy Kaaba*[6]) in Mecca, the holiest city of Islam (Cornell, 1999). Prayers or *al-salat* are viewed as a vital component of submission to the will of Allah, and "involve a variety of important rituals, each signifies the centrality of prayer in Islamic life". By performing the prayers five times a day, "Muslims acknowledge humanity's total dependence on the will of Allah" (Moore, 2006, p.141).

The third pillar of belief is *Zakat* or charity, giving of one's wealth for the benefit of the poor. The *Quran* mandates this as 2.5 per cent of the individual's saving, to be used to help the less fortunate in the society. The *Zakat* reflects the importance of charity and emphasises the *Quranic* view of social justice and compassion.

The fourth pillar of belief involves fasting or *al-saum* from sunrise to sunset in the month of Ramadan[7]. During Ramadan, Muslims abstain from food, drink, and sex during the day. By fasting, Muslims develop a deep sense of devotion to Allah, and that helps them participate in, and be responsible to, a larger moral community (Cornell, 1999). By fasting Muslims develop a deep sense of devotion to Allah, and find recognition as members of the Islamic community (Cornell, 1999).

The fifth pillar is the *hajj* or annual pilgrimage to Mecca, Saudi Arabia. The *hajj* must be preformed by every healthy and financially able Muslim once in their lifetime. It is undertaken in *Dhul-Hijjah*, the twelfth month of the Islamic lunar calendar. The pilgrims perform *hajj* repeating the rituals that were performed by the Prophet during his last pilgrimage. The *hajj* symbolises the believer's entry into the earthly House of God in Mecca, a replica of the cosmic House of God in the Seventh Heaven (Cornell, 1999).

[6] Holy Kaaba: the cube-shaped shrine in the holy city of Mecca is considered the earthly house of God and the most sacred in the Islamic world.
[7] Ramadan: is the ninth month of the Islamic lunar calendar.

Furthermore, Muslims have a belief in Allah's angels; a belief in Allah's revealed texts, including the *Quran*; a belief in Allah's messengers; a belief in the day of judgement (the world has been created for a fixed period of time); and a belief of Allah's complete control over worldly affairs (Cornell, 1999). Moreover, in Islamic societies religion and politics are closely linked. The Islamic society is supposed to be governed by the *Shariah* and *Fiqh*, two complex sources of *Quranic* law. These *Quranic* Islamic laws provide justifications for formation and implementation of laws that govern religious practices and obligations, social life, marriage and divorce, commerce and business, taxation, government, criminal justice, economics, and other areas (Kamali, 1999).

Like Christianity and Judaism, Islam "is a highly complex religion that has scholars debating interpretations of Islamic values, history, laws and practices, the above mentioned core tenets and beliefs are not debatable" (Faruqi, 2007, p.463). Based on these tenets Man (human being) has a designated role to play on earth.

Role of man as stated in the Quran

Another important aspect that must be considered is the role of man as described in the *Quran*. Man's role, as stated in the *Quran*, is that human beings have been placed on the earth as God's representative or '*Khalifah*'. The *Quranic* doctrine of vice-regent or '*Khālifah*' placed Man in the role of *Amāna* or trustee and custodian of the earth, and Man is responsible for building the earth and utilising its resources with a sense of justice to oneself and to fellow mankind (Kamali, 2003). While nature can be said to be man's testing ground, Man is instructed to read it's 'signs' (Manzoor, 1984, p.156) in order to understand God. Thus Muslims have contributed to the development of natural science in order to understand God and fulfil their role as God's representatives.

The role assigned to 'Man' by the *Quran* includes accountability of the numerous resources given by God or *Allah*. Moreover, the Islamic concept of knowledge includes both the transcendental knowledge as well as the knowledge based on sense perception and observation. Consequently, all knowledge gained through scientific activities aims to result in human welfare; and seeks to utilize the resources of the universe for beneficial purposes; (Kamali, 2003), that is, there is both social justice and compassion. Hence all scientific endeavours by Muslim scientists need to be scrutinized by the values, ethics and theological standards as encompassed in Islam.

Islamic view of nature versus Western view of nature

The Islamic view of nature sets man the task of 'studying nature in order to discover God and to use nature for the benefit of mankind'. Nature can be used to provide food for mankind and its bounty is to be equally distributed among all peoples. All activities that cause harm to mankind and in turn destroy nature are forbidden. Destruction of the natural balance is discouraged, for example, unnecessary killing of animals or removal of vegetation may in turn lead to starvation due to lack of food. This view is an extension of the idea that Man has been placed on earth as God's representative (Faruqi, 2006a; Said, 1989; Zaidi, 1991). Contemporary Muslims scholars advocate that scientists and scholars are best motivated by these underlying values when undertaking scientific endeavours.

The Islamic view of nature has its roots in the *Quran*, the very word of God and the basis of Islam. The following passages from the *Quran* illustrate the relationship between Nature and Man and how this relationship inspires Muslim scholars to study natural phenomenon, in order to understand God (Wersal, 1995). The following verses also show the way the *Quran* presents the whole universe:

> We created not the heavens, the earth, and all between them, merely in (idle) sport; we created them not except for just ends. But most of them do not understand (Surah Al-Dukhān 44: 38-39[8], [Ali, 1989, p. 1289]).
>
> Behold! In the creation of the heavens and the earth; In the alternation of the night and the day; In the sailing of the ships through the ocean for the profit of mankind; In the rain which Allah sends down from the skies; And the life which He gives therewith to an earth that is dead; In the beasts of all kinds that He scatters through the earth; In the change of the winds, and the clouds which they trail like their slaves between the sky and the earth - (here) indeed are Signs for a people that are wise (Surah Ad-Baqarah 2: 164 [Ali, 1989, p. 64-65]).

The point to note is the general empirical attitude of the *Quran* that engendered in its followers a feeling of reverence and thus made them founders of an enlightened society (Iqbal, 1986). This view of nature influenced the scholars of the so-called 'Golden Age of Islam' to undertake scientific activities that resulted in the vast corpus of scientific works of that period.

[8] In the Quran, the name of the Surah is followed by its individual number and then numbers for the individual verses.

The Western view of nature that emerged after the Scientific Revolution was that “no footprints of the divine can be discerned in the sands of the natural world” (Peters, 2003, p.33). Furthermore, any commonality that existed between the sciences that emerged in Europe and those that had developed in the Islamic civilization “was rent asunder by the rise of modern science” (Nasr, 1996, p.129). Seyyed Hossein Nasr, notes in his work *Religion and the Order of Nature*, (1996, p.133)

> From the idea of cosmic order and laws created by God through His Will and applicable to both men and nature to the idea of ‘laws of nature’ discoverable completely by human reason and usually identified with mathematical laws, divorced from ethical and spiritual laws, there is a major transformation that played a central role in the rise of modern science. This new idea of laws of nature also eclipsed the earlier Christian understanding of the subject, although later theologians tried to ‘Christianize’ the seventeenth-century scientific concept of laws of nature. Interestingly enough, such an event did not take place in other civilizations with a long scientific tradition such as the Chinese, Indian, and Islamic, and this is of great significance in the parting of ways between the modern West and other civilizations as far as the understanding of the order of nature and its religious significance are concerned.

In the final analysis, it seems that Europe decided to transform the medieval science that had been influenced by the Islamic scientific traditions. Plato replaced Aristotle, and mathematics was the new tool of science. With contributions from Nicholas Copernicus (1473-1543), to Galileo Galilei (1564-1642), and Johannes Kepler (1571-1630) it climaxed with Charles Darwin’s work ‘*The Origins of Life*’ in the biological sciences and had philosophical implications that has continued till today.

Koyré (1892-1964), a respected French historian of science, (cited in Iqbal, 2002, p.29) stated that

> What the founders of modern science did was neither refinement, nor improvement of what they had inherited; they had to actually destroy one world and to replace it with another. They had to reshape the framework of our intellect itself, to restate and to reform its concepts, to evolve a new approach to Being, a new concept of knowledge, a new concept of science.

Some historical lessons: Science in the Islamic period

The Islamic Empire consisted of a society that was multicultural in terms of languages, customs, traditions and religion. It encompassed vast lands with peoples of different faiths and cultures. Thus the population of the Islamic Empire consisted of Muslims from three continents, Arabs, Persians, Turks, Africans, Indians and other Asians, but also Jews, Christians and other faiths. Therefore scholars from all faiths worked under the umbrella of Islam to produce a unique culture of knowledge and learning. Considered in this section are just some of the Islamic scholars and their work from the so-called 'Golden Age of Islam'. In this period inspiration for the development of sciences was found in the *Quran*. Moreover, scientific activities were undertaken for the betterment of mankind; therefore the sciences that initially attracted the attention of Islamic scholars were medicine, mathematics, pharmacy, and pharmacology (Faruqi, 2006b).

Medicine

In medicine, Al-Razi, known in Latin as Rhazes, excelled in the powers of observations and wrote some 184 works on topics that he studied as a practising doctor. One of Al-Razi's books, *Treatise on Smallpox and Measles*, was translated into Latin, then English and other European languages, and "went through forty editions between the fifteenth and nineteenth century" (Turner, 1995, p.135). Another scholar, Ibn Sina (980-1037) was renowned throughout medieval Europe as Avicenna, his famous book *The Canon of Medicine* was translated into Latin in the twelfth century and it was used in medical schools throughout Europe until the advent of Western science (Beshore, 1998; Meyers, 1964).

Ibn An-Nafīs contradicted the theories of blood circulation as put forward by Galen. He advanced a theory of blood circulation between the compartments of the heart and the lungs, and of pulmonary circulation or lesser circulation. In 1553, three centuries later, a Spaniard Miguel Serveto (Michael Servetus) forwarded a similar theory (Meyerhof, 1935). Ibn An-Nafīs's theory from the thirteenth century was largely ignored. But he was among the initial precursors to Harvey's scholarly work that revealed the circulation of blood in the human body. Furthermore, Muslims established hospitals that were far superior to any that existed in ancient times or in lands beyond the Islamic Empire. These hospitals were to become models for hospitals as we know them today (Turner, 1995).

Chemistry, pharmacology and pharmacy

In chemistry, the works of Jaber ibn Haiyan and Al-Razi formed the basis of Western science. Jaber, know as Geber in Latin, described in his works the preparation of many chemical substances: the sulphide of mercury, oxides and arsenic compounds. Al-Razi in his book *Secret of Secrets* know as *Liber secretorum bubacaris*, described the chemical processes and experiments he conducted. Hill (1993, p.83) has stated that Al-Razi's book *Secret of Secrets* 'foreshadows a laboratory manual' it deals with substances, equipment and procedures. The discovery of inorganic acids during chemical experiments had valuable industrial applications for the centuries that followed.

In the fields of pharmacology and pharmacy Muslims carried out scientific investigation into the composition, dosages, uses and therapeutic effects of drugs. Having translations of Dioscorides' *De Materis Medica*, along with knowledge from Syria, Persia, India and the Far East, Muslim scholars and physicians showed great innovative skills. They developed the procedures for the manufacture of syrups and juleps, and established apothecary shops (Turner, 1995).

Mathematical sciences, optics and astronomy

The mathematical sciences as practised in the Islamic world during this period consisted of mathematics, algebra, and geometry as well as mathematical geography, astronomy and optics. Muslims derived their theory of numbers (*'ilm al-a'dad*) in arithmetic from translations of the Greeks sources such as Books VII through to IX of Euclid's *Elements* and the *Introduction to the Science of Numbers* by Nicomachus of Gerasa (Berggren, 1997). Furthermore, they acquired numerals from India (Hindu) and possibly China and made their use widespread. Mohammad Bin Ahmed in the tenth century advanced the concept of zero or *sifr* (Badawi, 2002). Thus replacing the cumbersome Roman numerals and creating a revolution in mathematics. This led to advances in the prediction of the movement of the planets and advances in the fields of astronomy and geography.

Muslim mathematics had inherited both the Babylonian sexagecimal system and the Indian (Hindu) decimal system, and this provided the basis for numerical techniques in mathematic (Folkerts, 2001; Rajagopal, 1993). This enabled to build mathematical models using the decimal system, expressing all numbers by means of ten symbols, and each symbol accorded the value of position as well as absolute value (Kettani, 1976). Many creative methods of doing multiplications were developed by Muslims; methods of checking by casting out nines, and decimal fractions (Anawati, 1976). Thus Muslim scholars contributed and laid the foundations of modern mathematics and the use of mathematics in the fields of science and engineering (Høyrup, 1987).

Al-Khwarizmi wrote the first book of algebra, the word 'algebra' transliterates into the term '*al-jabr*'. Al-jabr represents the two basic operations used by al-Khwarizmi in solving quadratic equations. In the latter half of the twelfth century, the first part of al-Khwarizmi's *Kitab* *al-Jabr wa al-Muqabalah* was translated and made available in Europe (Kettani, 1976; Sarton, 1927). Al-Battani (d.929) systematically developed trigonometry and extended it to spherical trigonometry (Kettani, 1976; Sarton, 1927), with important consequences for astronomy, geography and exploration beyond the known world, thus making the construction of better maps and the reconceptualisation of the structure of the planet Earth.

Ibn al-Haytham was acclaimed in both the East and West as the founder of modern optics on account of his seminal work *Kitab al-Manazir*. Ibn al-Haytham's empirical optical discoveries and formulation of the scientific research method involved the close and critical engagement with various ancient Greek optical theories, in both their philosophical and empirical aspects. Ibn al-Haytham systematically set out to construct counter-theories that would stand up to rigorous, objective logic-mathematical analyses and refined, innovative observational testing. Through translation of al-Haytham's work Europeans became aware of Ibn al-Haytham's remarkable achievements in the field of Optics (Meyers, 1964, p.32). John Peckham in the late-thirteenth century used Ibn al-Haytham's *Kitab al-Manazir*, as well as Witelo's *Optics* that has echoes of *Kitab al-Manazir*. Witelo's work was used by Johannes Kepler. Furthermore by promoting the use of experiments in scientific research, Ibn al-Haytham played an important role in setting the scene in Western science (Rashed, 2002). Roger Bacon, the founder of experimental science, probably used the original Arabic works of Ibn al-Haytham as well as Latin translations, since he had contacts with the Islamic centres in Spain (Meyers, 1964).

In the thirteenth century, Al-Tusi, a Persian astronomer put forward his concept known as the "Tusi Couple", that is, a hypothetical model of "epicyclic motion that involves a combination of motions each of which was uniform with respect to its own center"(Turner, 1995, p.68). This model was applied by Ibn al-Shatir to the motions of the heavenly bodies in the fourteenth century. Ibn al-Shatir's formulations were the beginnings of verifying theoretical astronomy through systematic observations (Turner, 1995).

Philosophy

In Islam it must be understood that there is no 'philosophy' as recognised by Western standards. For traditional Muslims, answers to questions pertaining to God, the creation of the universe, and the destiny of mankind, could be sought in the *Quran*, (Faruqi, 2006b). Some orthodox Muslims subjected *Quranic* verses to *Kalam*, that is, a theological discipline involving rational dialectical examination (a form of Muslim scholastic theology). Philosophers like Al-Kindi (800-870) and Al-Farabi (d. 950) were inspired by the translations of the works of Aristotle. They attempted to reconcile Aristotelian and Platonic ideas with revelation. Thus trying to build a bridge between belief and reason Taton, 1963). Al-Farabi's works illustrated that Aristotelian logic had scriptural support in the *Quran* and the prophetic *hadith* (Bakar, 1999). Al-Farabi and Ibn Sina (Avicenna) (980-1037) tried to develop the use of logic within the framework of religious consciousness of the Transcendent. Al-Farabi and Ibn Sina wrote works which sought to demonstrate that logic when used correctly, could in relation to religious truths help explain their rationality and clarify overall consistency (Armstrong, 2000; Sarton, 1927). However, there emerged opposition to Aristotelian logic from within both religious and intellectual quarters.

Abu Bakr al-Razi (Rhazes) (d. 925) was probably the first to write a critic on Aristotle's logic. Al-Ghazzali (1058-1111) wrote his famous critic of the earlier philosophers such as Al-Farabi and Ibn-Sina, titled '*Tahâfut al-falâsifa*' or 'The Incoherence of the Philosophers' who had been inspired by Aristotle. Al-Ghazzali accentuated the unacceptability of the three metaphysical claims: (a) the denial of bodily resurrection; (b) the limitation of divine knowledge to universal, eternal truths; and (c) the doctrine that the world is eternal (King 2004, p.58). Al-Ghazzali denounced these claims and all who held these beliefs were disbelievers. But Al-Ghazzali also wrote works which encouraged the use of logic for enhancing religious understanding, but reason always was subservient to revelation. According to Al-Ghazzali, "the *Quranic* term *al-mĩzān* usually translated as the balance, refers among other things to logic. Logic is the balance with which man weighs ideas and opinions to arrive at the correct measurement or judgement" (Bakar, 1991, p.4). Al-Ghazzali being a scientist and religious scholar was able to combine religious beliefs with the scientific ideas of the time (Faruqi, 2006b).

However, Ibn Rushd (Averroes) (1126-98) wrote '*Tahâfut al-Tahâfut*' or the 'Incoherence of the Incoherence' as a rebuttal to the arguments presented against philosophers such as Al-Ghazzali (Taylor, 2000). Ibn Rushd sought to prove that there was nothing either philosophically or religiously objectionable in Aristotelian doctrine of the eternity of the world (Faruqi, 2006b). Ibn Rushd's works demonstrated the relationships that existed between religious thinking and the scientific developments of this period. These scholarly works demonstrated the intellectual struggle between Islamic science based on the Quranic worldview and Greek thought (Faruqi, 2007).

Muhammad Iqbal (d. 1939) undertook an incisive analysis of the Greek philosophy and its comparison with the worldview of the *Quran* (Kamali, 2003). Muhammad Iqbal refuted some of the hitherto parallels that had been drawn between the two. Iqbal acknowledged that "Greek philosophy had been a great cultural force in the history of Islam" (Iqbal, 1986, p.3), but the worldview of the *Quran* which inspired the Muslim scholars was different from the Greek thought. For example, Aristotle wrote extensively on physics without undertaking a single experiment; and on natural history without determining the most easily verifiable facts (Kamali, 2003). Socrates postulated that the study of man alone, was sufficient in the study of the human world, whereas the *Quran* encompasses that all of nature must be studied, the 'humble bee a recipient of Divine inspiration' and 'to observe the perpetual changes of the wind, the alternation of day and night, the clouds and the planets swimming through infinite space' (Kamali, 2003).

Furthermore, the *Quran* deems 'hearing' and 'sight' as valuable instruments in the process of learning. Thus Islamic science developed in scientific inquiry the method of observation and experimentation. Kamali argued that the experimental method that developed in 'Islam was not due to a compromise with Greek thought but to a prolonged intellectual warfare with it' (Kamali, 2003). Consequently, this resulted in the magnificent developments in science during the period from the twelfth to the fifteenth centuries in the various territories of the Islamic Empire, Baghdad, Andalusia, and Sicily (Faruqi, 2006b). Therefore one can infer that Islamic scholars contributed original works to all fields of science. These ideas, theories and works once transferred to Europe proved fertile matter on which the development of science could occur.

Summary

The Islamic science that developed in the Islamic lands made its way to Europe through North Africa, Sicily and Spain. Knowledge also began to filter back to Europe through the translations of Arabic versions of the Greek knowledge, the original Greek treatises (Burnett, 2001) and the seminal contributions of scholars of the Islamic world. Western science as we know it today works with theories and models that must be tested empirically. This was standard practice in the fields of mathematics, astronomy and medicine in the Islamic world of 1000 years ago. The Muslims developed the procedures for testing knowledge both empirically and logically (Faruqi, 2006b). More over, an important characteristic of Islamic science was its experimental character. Islamic scientists were interested especially in the applied sciences, in the construction of apparatus, in testing theories by undertaking observations, and analysis of results through mathematics

(Bammate, 1959). These ideas and procedures were all available before the times of Galileo and Newton to whom they have been largely attributed in Western Europe (Faruqi, 2006a).

Need for discourse between Islamic science and Western science

There are ample historical resources to reconstruct a picture as to how Islamic scientific knowledge came to Europe. But it must be noted that modern science is not simply based on Islamic science or the scientific thinking that was present in Europe between the twelfth to fifteenth centuries (Faruqi, 2006b). Modern science is based on a concept of nature essentially totally different from that on which Islamic science was based. “The process of *transmission* of Islamic scientific tradition to European thought was followed by a process of fundamental *transformation* of this tradition” (Iqbal: 2003, p.144).

The development of Western science and technology led to the separation of facts from values and this resulted in destructive consequences for humanity that arose from some scientific discoveries (Golshani, 2003). Scientific progress raised serious ethical issues in terms of human or animal subjects and public safety (Golshani, 2003). Consequently, in the Islamic world and in the West, Muslim scholars need to tackle these issues using all the tools available including religious knowledge.

Moreover, in Islam, the acquisition of knowledge, be it scientific or non scientific was argued not to be an end in itself but “one method to comprehend the glory of God” (Al-Hayani, 2005, p.565). Muslim scholars displayed their “diversities in theme and orientation that demonstrated the dynamic nature of Islam, far from its image that has been portrayed in certain media as a monolithic and stagnated system of ideas” (Moaddel and Talattof, 2000, p.1). Throughout history there have been clashes between religion and science, with all religions trying to find solutions to problems and ramifications of new scientific discoveries. With the advent of Western science the central theological problem that Islamic scholars were confronted with was “the question of the validity of the knowledge derived from sources external to Islam and the methodological adequacy of the four traditional sources of jurisprudence: the *Quran*, the dicta attributed to the Prophet (*hadith*), the consensus of theologians (*ijma*), and juristic reasoning by analogy (*qiyas*)” (Moaddel and Talattof, 2000, p.1).

Some intellectuals have tried to formulate reform in Islamic sources of knowledge, in line with the prevailing standards of scientific rationality and modern social theory. These have included intellectuals and theologians such as Sayyid Jamal al-Din al-Afghani, Sayyid Ahmad Khan, Chiragh Ali, Muhammad Abduh, Amir Ali and Shibli Nu'mani (Moaddel and Talattof, 2000, p.2). These Islamic scholars were influenced by the West, especially by its achievements, "ranging from scientific and technological progress, the Newtonian conception of the universe, Spencer's sociology, and Darwinian evolutionism, to the Western style of living" (Moaddel and Talattof, 2000, p.2). Clearly there is debate and disagreement between the scholars who have tried to interpret development of Western scientific thought within the context of Islamic historical and religious perspectives.

Conclusion

The author is of the view that in order to enhance understanding there is a need for further input and active participation of scholars engaged in research in Western scientific development in the West with Islamic thought. Therefore there is a need for discourse between Islam and Western science, rooted in the *Quran* (Iqbal, 2002) on the one hand, and between Western science and religious thought on a global setting on the other hand. In both cases, the origins of Western science over the past 1000 years needs to be better understood since it appears to be different from the views that have been promulgated in the United Kingdom and the United States during the nineteenth and twentieth centuries.

For Muslims to tackle the contemporary modern world with self-assurance and confidence there is the need to rediscover and restore the "mainsprings of Islamic civilization. They need to re-build an idea of Islam which includes justice, integrity, tolerance and the quest of knowledge-like the classic Islamic civilization" (Ahmed, 2002, p.44). There is a need for Muslim scientists and religious scholars to examine the relationship between modern science and Islam in the context of present day Islamic science and the *Quranic* worldview. Furthermore other peoples need to learn from the holistic Islamic approach and benefit from studying the medieval Islamisation of Western science. This aspect needs to be fully researched, accepted and incorporated in specialised works and in the teaching materials of schools and colleges around the world. Thus by acknowledging the Muslim contributions to scientific knowledge bridges can be built both within Islamic communities and others approaches to science, including Western or modern science.

I wish to acknowledge the support given by Professor John Keeves for his assistance with my research which clearly required a broad and

deep understanding of the diverse subject matter that unites Islamic scientific history and the development of Western scientific thought.

References

Ahmed, A. (2002) Ibn Khaldun's understanding of civilizations and the dilemmas of Islam and the West today. *Middle East Journal*, 56(1), 20-45.

Ali, A. Y. (1989) *The Meaning of the Glorious Qur'an*: (New edition with revised translation and commentary). Brentwood, Maryland, USA: Amana Corporation.

Al-Hayani, F. A. (2005) Islam and Science: Contradiction or Concordance. *Zygon*, 40 (3), 565-576.

Anawati, C. G. (1976) The significance of Islam's scientific heritage for the Moslem world today. *Impact of Science on Society*, 26(3), 161-167.

Armstrong, K. (2000) *Islam: A Short History*. New York: Modern Library.

Badawi, J.A. (2002) Islamic worldview: prime motive for development. *Humanomics*, 18(3/4), 3-25.

Bakar, O. (1999) *The History and Philosophy of Islamic Sciences*. Cambridge: Islamic Texts Society.

Bakar, O. (1991) *Tawhid and Science: Essays on the History and Philosophy of Islamic Science*. Kuala Lumpur, Penang: Secretariat for Islamic Philosophy and Science.

Bammate, N. (Apr/Jul 1959) The Status of Science and Technique in Islamic Civilization. *Philosophy East and West, Preliminary Report on the Third East-West Philosophers' conference*, 9(1/2), 23-25.

Ball, C., and Haque, A. (2003) Diversity in Religious Practice: Implications of Islamic Values in the Public Workplace. *Public Personnel Management*, 32 (3), 315-330.

Berggren, J. L. (1997) Mathematics and Her Sisters in Medieval Islam: A Selection Review of Work Done from 1985- 1995. *Historia Mathematica*, 24, 407-440.

Beshore, G. (1998) *Science in Early Islamic Culture*. New York, NY: F Watts.

Burnett, C. (2001) The Coherence of the Arabic-Latin Translation Program in Toledo in the Twelfth Century. *Science in Context*, 14 (1/2), 249-288.

Cornell, V. J. (1999) Fruit of the tree of knowledge: The relationship between faith and practice in Islam. In J. L. Esposito (ed.) *The Oxford History of Islam*. New York: Oxford University Press.

Davies, M. W. (1991) Rethinking Knowledge: 'Islamization' and the Future. *Futures*, 231-247.

Faruqi, Y. M. (2006a) Contributions of Islamic scholars to the scientific enterprise. *International Education Journal*, 7 (4), 391-399.

Faruqi, Y. M. (2006b) The Development of Scientific Thinking in Islamic Countries during the Period Between the Seventh and the Fifteenth Centuries (Common Era [C.E.] till the Fall of Constantinople (1453). Unpublished Doctorial Thesis, The Flinders University of South Australia.

Faruqi, Y. M. (2007) Islamic view of nature and values: Could these be the answer to building bridges between modern science and Islamic science. *International Education Journal*, 8 (2), 461-469.

Folkerts, M. (2001) Early Texts on Hindu-Arabic Calculation. *Science in Context*, 14(1/2), 13-38.

Golshani, M. (2003) Values and Ethical Issues in Science and Technology: A Muslim Perspective. *Islamic Studies*, 42(2), 317-330.

Golshani, M. (2000) How to Make Sense of 'Islamic Science'? *American Journal of Islamic Social Sciences*, 17 (3), 1-19.

Hill, D.R. (1993) *Islamic Science and Engineering*. Edinburgh: Edinburgh University Press.

Høyrup, J. (1987) The Formation of "Islamic Mathematics" Sources and Conditions. *Science in Context*, 1(2), 281-329.

Iqbal, M. (2003) Traditional Islam and Modern Science. In T. Peters, & G. Bennet (eds.) *Bridging Science and Religion*. Minneapolis: First Fortess.

Iqbal, M. (2002) Islam and Modern Science: Questions at the Interface. In T. Peters, M. Iqbal, & S. N. Haq (eds.) *God, Life, and the Cosmos: Christian and Islamic Perspectives*. Aldershot, UK: Ashgate.

Iqbal, M. (1986) *The Reconstruction of Religious Thought in Islam*. Iqbal Academy Pakistan: Institute of Islamic Culture.

Kamali, M. H. (1999) Law and society: The interplay of revelation and reason in the Shariah. In J.L. Esposito (ed.) *The Oxford History of Islam*. New York: Oxford University Press.

Kamali, Mohammad Hashim. (June 2003) Islam, rationality and science. In *Islam & Science*, 1, 115. Expanded Academic ASAP via Thomson Gale. Flinders University Library. [10 November 2006].

Kettani, M. A. (1976) Moslem contributions to the natural sciences. *Impact of science on society*, 26 (3), 135-147.

King, D.A. (2004) Reflections on some New Studies on Applied Science in Islamic Societies (8th-19th Centuries). *Islam & Science*, 2 (1), 43-56.

Manzoor, S. P. (1984) Environment and values: the Islamic perspective. In Z. Sardar (ed.) *The Touch of Midas: science, values and environment in Islam and the West*. Manchester, UK: Manchester University Press.

Mazrui, A. A. (1997) Islamic and Western Values. *Foreign Affairs*, 76 (5), 118-132.

Meyers, E. A. (1964) *Arabic Thought and the Western World in the Golden Age of Islam*. New York: Frederick Ungar Publishing Co.

Meyerhof, M. (Jun 1935) Ibn An-Nafis (XIIIth Cent.) and His Theory of Lesser Circulation. *Isis*, 23(1), 100-120.

Moaddel, M. and Talattof, K. (eds). (2000) Contemporary debates in Islam: an anthology of modernist and fundamentalist thought. New York: St. Martin's Press.

Moore, J. R. (2006) Islam in Social Studies Education: What We Should Teach Secondary Students and Why It Matters. *The Social Studies*, 97 (4), 139-144.

Nasr, S. (1996) *Religion and the Order of Nature*. New York: Oxford Press.

Peters, T. (2003) *Science, Theology, and Ethics*. Aldershot, Hants, England; Burlington, VT: Ashgate.

Rashed, R. (2 Aug 2002) A Polymath in the 10th Century. *Science*, 297 (5582), 773.

Rajagopal, P. (1993) Indian Mathematics and the West. In R. Hayhoe (ed.) *Knowledge across Cultures: Universities East and West*. Columbia: Hubei Education Press & OISE Press.

Ravetz, J. R. (April 1991) Prospects for an Islamic science. *Futures*, 262-272.

Sabra, A. I. (1996) Situating Arabic Science: Locality versus Essence. *Isis*, 87, 654-670.

Said, A. A. (1989) The Paradox of Development in the Middle East. *Futures*, 619-627.

Sarton, G. (1927) *Introduction to the History of Science, Volume 1*. Washington: Carnegie Institution of Washington.

Smith, J. I. (1999) Islam and Christendom: Historical, Cultural, and Religious interactions from the Seventh to the Fifteenth centuries. In J. L. Esposito (ed.) *The Oxford History of Islam*. New York: Oxford University Press.

Taton, R. (ed). (1963) *Ancient and Medieval Science from Prehistory to AD 1450*. Translated by A. J. Pomerans. London: Thames and Hudson.

Taylor, C. R. (2000) “Truth Does Not Contradict Truth”: Averroes and the Unity of Truth. *Topoi*, 19, 3-16.

Turner, R. H. (1995) *Science in Medieval Islam: An Illustrated Introduction*. Austin: University of Texas Press.

Ur-Rehman, S. (2002) Secular Knowledge versus Islamic Knowledge and Uncritical Intellectuality: The Work of Ziauddin Sardar. *Cultural Dynamics*, 14 (1), 65-80.

Wersal, L. (1995) Islam and Environmental Ethics: Traditional Responds to Contemporary Challenges. *Zygon*, 30 (3), 451-459.

Zaidi, I. (1991) On the Ethics of Man’s Interaction with the Environment: An Islamic Approach. *Environmental Ethics*, 3 (1), 35-47.

7

The linguistic awakening in the eastern part of Indonesia

Mochtar Marhum
School of Education, Flinders University

Introduction

The Eastern Part of the Indonesian Region is identified as one of the most linguistically diverse regions in the world. It has also a multicultural and pluralistic background of ethnic groups. As reported in several Linguistic and Cultural Studies there are more than four hundred local languages found in the Eastern Part of Indonesia. However, many of them have been declared moribund and a few of them are even extinct following the emergence of top-down language policy and planning of the New Order Regime. The Soeharto Regime urged the promotion of *Bahasa Indonesia* as a unifying language. On the one hand, it can be argued that the Central Government's top-down policy of promoting national language can strengthen nationalism and patriotism. Furthermore, it can promote the philosophy of unity within diversity which is based on the Pancasila Ideology of Indonesia to the people of different linguistic and cultural backgrounds. On the other hand, as revealed in several studies many school children were not allowed to use their local language during school hours. Thus children are discouraged and reluctant to use their local language. This situation prompted the young generation to pay less attention to their local languages.

The introduction of Regional Autonomy post political reform brought about an attitude change in the language of the speakers. Many speakers of local languages have become more proud to use their local language in public. In addition, some Local Governments have encouraged the local people to get involved in the promotion of their local languages either through schools such as local language lessons and bilingual instructional assistance or the promotion of local languages outside through the use of local languages in such events as the wedding ceremony, ritual ceremony and local cultural show. In addition, there has been wider use of local language as the medium of communication among local people. The local language awakening seems to have brought several consequences that might influence multicultural life if it is not carefully anticipated and well managed by the Local Government.

The emergence of globalisation has brought both positive and negative consequences to the linguistic ecology in Central Sulawesi. The local children have become much more proud to learn and practise their English. The teaching and learning of English has become the first priority of the language learning program. Furthermore, the number of private English Schools has mushroomed over the last ten years. In the formal education context, English has become the compulsory instructional language in International Rating Schools. So far many parents have proudly sent their children to the International Rating School of SMP N. 2. Palu. The events of English Speech and Story telling competitions in the English language are more frequently held. The events are sponsored by the Department of Education and involve students of public and private schools.

It is interesting to study the linguistic ecology in Central Sulawesi where people of different generations, both in cities, towns and rural villages, are able to speak *Bahasa Indonesia* fluently. It is different from the situation in some other parts of Indonesia in which many elderly people are still not able to speak *Bahasa Indonesia* fluently. Though, in the era of Globalisation and Decentralisation *Bahasa Indonesia* remains the effective medium of communication among people of different ethnic background of Central Sulawesi. This chapter will focus on the linguistic awakening issues and include more discussion on the local language situation and the promotion of English in the educational context. In addition, discussion on *Bahasa Indonesia* as a unifying language used by people of different ethnic and linguistic backgrounds in Central Sulaewesi is included. Finally, this chapter will discuss Multilingualism and Language Policy, Issues in Multilingualism and Educational Policy, and the Discussion of Languages Situation versus Language Awakening.

Multilingualism and language policy

The situation of local, national and global languages in Indonesia has become a reality over many years alongside the history of the Republic of Indonesia which consists of bilingual and multilingual communities (Alisjabana, 1976; Abas, 1987). Furthermore, the language awareness among Indonesian communities has grown significantly following the development of Indonesia. It is both essential and urgent to define and develop the functions of local, national and global languages.

The status and function of English in Indonesia in a globalisation era

English is the first foreign language now being taught in Indonesia. It is based on the *Regulation of Ministry of Education and Culture No. 096/1967* (Kartono,1976: Alwasilah, 1997) regulating the status and function of English. It has been the first foreign language in Indonesia since 1955. Halim (1976, p.146) argued that English has some official functions in Indonesia: (1) a means of communication among Nations, (2) a means of development as a supporter of Indonesian language to become a modern language, and (3) a means of science and technology transfer for national development.

English has several functions in the globalisation era. First, many nations all over the world have a growing rate of interdependence (Crystal 1994; Pennycook, 1994; Alwasilah, 1997). International relations are not limited to the economic and political arena, but also operate in many other aspects of life. Their functions include the establishment of close relations with other nations and the implementation of foreign policy based on the *Regulation of the Ministry of Education, No 096/1967*. Thus, in general, foreign languages function as a means of global communication in all aspects of life.

Globalisation brings about an increase in international competition. The mastery of science and technology becomes the key to winning the competition. For that reason, English should function as an instrument of applying and advancing science and technology to accelerate the developmental process (Crystal, 1997; Ingram 1993). This function includes that of acquisition, use and development in a general sense. This function also covers the use of English as an instrument of development that supports the use of Indonesian as a modern language (Alwasilah, 1997; Huda, 2000).

According to historical records, following the independence of the Republic of Indonesia, the Dutch language was replaced by English as the first foreign language, and has been recognised as such in Indonesia since 1955 (Alisjabana, 1976; de Han, 2003). Since the 1980s, English has been considered to be the most important foreign language in Indonesia. The government's and community's interest in English has been growing since the early 1990s (Alwasilah, 1997; Supriadi, 1999). This position of English can be traced from government documents on the results of Parliamentary meetings. In the *GBHN* (The Guidelines of the State Policy) 1983 and 1988, foreign language policy was not incorporated. However, in the GBHN 1993, the policy on foreign languages, particularly English, was clearly stipulated. The policy related to the use and mastery of English. In 1988, *Government Regulation No. 55, 56 and 57/1988* changing *Government Regulation No. 28, 29/990* was introduced. It confirmed the use of English in schools. Moreover, *Government Regulation of No 57/1957/1988* confirmed the use of English as a foreign language and as a means of communication in universities. Subsequently, it was incorporated into *Government Regulation No 60/1999* on the use of English in all higher education. Alwasilah (1997, p.89) suggested that:

> The need for mastery of English in the globalisation era was absolutely necessary. In addition, it would be ideal if the mastery of English became the mastery of second language. Yet, there were several obstacles that would necessarily be encountered.

Abas (1987), argued that it had to conform to the national interest which gave high priority to the development of the Indonesian language as a national language of unity and unification.

The role and function of Bahasa Indonesia in the globalisation era

In order to prepare for the globalisation era, it was necessary to maximise the role and function of *Bahasa Indonesia* (Alwi et. al. 2000; Moeliono, 2000). *Bahasa Indonesia* should be studied and evaluated. In addition, the problems of learning *Bahasa Indonesia* must be identified and solved. Gunarwan (2000) suggested that establishing positive attitudes toward language learning could become the target of education for language maintenance.

Moelinono (2000, p.26) argued that there were several factors which were not conducive to the development of *Bahasa Indonesia*. First, many Indonesian people in particular areas are not yet able to use *Bahasa Indonesia*. Furthermore, some are not motivated to learn *Bahasa Indonesia* because they think it is not necessary. According to the linguists, many Indonesian people have not considered the importance of learning the skills of reading and writing of *Bahasa*

Indonesia. Second, many people still use local languages in certain local situations. This leads to problems with *Bahasa Indonesia* development. In other words, lack of support for *Bahasa Indonesia* might contribute to the inadequate promotion of *Bahasa Indonesia* (Alwasilah et al., 2000). However, the use of local languages might cause suspicion among those people who do not understand the local languages in spoken form. Third, the use of the foreign language and English, as a prestigious language, would be beyond the role, status and function of language by individuals and communities (Alwasilah, 1997; Crystal, 2000). Such habits and attitudes might not show significant respect of *Bahasa Indonesia* as well as the people. Fourth, casualness of *Bahasa Indonesia* teaching or *Bahasa Indonesia* development have resulted in several errors and mistakes in the use of *Bahasa Indonesia* (Moeliono, 2000). The linguistic factors that have needed to be corrected have included word choice, euphemism and language structure.

The status and function of local languages in the globalisation era

Taha (2000, p.34) suggested that the status and function of local languages in Indonesia in the current situation were based on several factors.

The explanation of *Article 36 of the 1945 Constitution* stated that "in the areas where local languages are actively spoken, the local people maintain and develop their local languages".

At a national language seminar in February 1975 on the status and function of local languages in relation to the national language, the languages mentioned previously were said to function as local languages. There were several functions of local languages: a symbol of local pride, local identity, and as a *lingua franca* of local family and community. Furthermore, in relation to national language, local languages had several functions such as: (a) supporter of national language, (b) instructional language at primary schools in particular areas to support the teaching of *Bahasa Indonesia* and other subjects, and (c) as supporting instrument of local cultural development (Halim, 1980). Listing and intensifying the use of local languages were also formulated in the seminar of national language politics in 1975.

The 1993 *Guidelines of State Policy (GBHN), Number 3 f,* stipulated that "the maintenance of local language should be sustained in order to develop and enrich *Bahasa Indonesia* vocabulary and Indonesian cultural diversity" (Alwi et al., 2000). Consequently, it is necessary to conduct research, as well as study and develop the local language and literature and publish them in the mass media and other publications.

One of the agreements of the fourth Indonesian Language Congress in 1993, particularly in the area of local language and literature, was that the local languages in certain areas could be taught to the student speakers without hampering the teaching of Indonesian language and literature (See Mahsun 2000; Mu'adz, 2000). For that reason, curricula, textbooks of teaching and learning methodology and other infrastructure to support local language education should be developed.

Issues in multilingualism and education policy

The incorporation of multilingual teaching into the school curriculum needs to have positive implications for the development and promotion of local, national and global language education, particularly in an era of globalisation and the expansion of democracy world wide (Nagai, 1997). The promotion of multilingualism must cover linguistic diversity, such as vernacular language, national language and English as a global language. The following sections discuss the relevant issues.

A vernacular language

According to Walton and Eggington (1990, p. 54), a vernacular language means an indigenous local language that is used by people in a certain district or place. They further state that the term 'vernacular language' is used in a more political sense by UNESCO to mean a language dominated or oppressed by another and with the implication that vernacular languages are generally spoken by relatively small groups of people, who typically have little or no tradition of writing and the language is unlikely to be standardised. Holmes (1992, p.80-86) argues that the term 'vernacular' is used in a number of ways. It generally refers to a language which has not been standardised and which does not have official status.

Minority language speakers are motivated by religious and political factors. Anthropological linguists and Ministries of Culture recommend preserving and maintaining the identity of minority language (Crystal, 2000). However, the central government and the border controls perhaps want rapid assimilation into the main stream of national culture, economy, religion and politics. Mugler and Lynch (1996) argued that when it came to the policy making stage, including educational policy, the two goals were necessarily in conflict. Consequently, it was really important to consider both top down policy and bottom up policy making in order to manage the conflict of interest that arose.

The South East-Asian governments have much in common in their attitudes toward minority groups. They agree that the territories of minority language speakers need to be protected. The minority groups need to be allowed to assimilate, migrate, disappear, or stay intact where they are, but the tendency for separatism needs to be prevented from developing (Walton and Eggington 1990). Even though they may be isolated from national educational planning, the language they speak needs to be considered.

The strengthening of regional and national pride may lead to further research on minority languages. A strong national language is an important issue, in what Gonzales (1979 in Noss, 1984) calls 'language welfare'. Then a language firmly rooted in natural or regional antiquity is considered to be even more important.

A national language

Indigenous languages which are given official priority in particular countries by the government concerned can be defined as national languages (Nos, 1984; Grimes, 1992; Digest 1999). For example, in South East Asian countries the national languages are *Bahasa Indonesia* in Indonesia, *Bahasa Malaysia* in Malaysia, and Thai in Thailand. *Bahasa Indonesia*, *Bahasa Malaysia*, Malay and the Tagalog language of Philippines belong to the same family.

In terms of national languages, two categories need to be considered. One is the nature of the national language which refers to what the national language is supposed to be (Nos, 1984). The other is the role of the national language which is associated with what the national language is supposed to do. In ASEAN countries, every country is in a different situation regarding the nature and the role of its national language.

English as a global language and its impact

Crystal (1997) argued that a language can be defined as a global language when it is recognised in many countries around the globe. In addition, it gains a genuinely global status and plays a key role. Crystal (1997, p.2) obviously agrees that English is a global language because it is spoken as a first language, a second language or third language by people around the world. He further explains that in order to gain a global status, two things should be considered. First, the language must be recognised as an official language and it is widely used. Second, although the language does not have any official status, it is widely and popularly taught as a foreign subject.

English meets these criteria as it is currently widely spoken around the world and popularly taught in the largest number of schools, including the schools of the non-English speaking countries (Pennycook, 1994; Crystal, 1997).

What makes a global language? According to the history of English, there is a close link between the dominance of language and power. No language can be recognised as a global means of communication without a strong political, military or economic power base (Pennycook 1994; Crystal 1997).

However, becoming a global language has nothing to do with the number of language speakers. It is associated more with who speaks the language (Cook, 1994; Crystal, 1997). For example, Latin used to be an international language. This had nothing to do with the number of Latin speakers. It had a link with the power of the Roman Empire. Furthermore, Chinese has the biggest number of speakers in the world but it is not considered to be a global language like English.

An international language can result from a militarily powerful nation. Moreover, a militarily powerful nation can contribute to the maintenance and expansion of an international language. The development of international business and information technology, for instance, requires the use of a global language (Lo Bianco, 1987b; Ingram, 1993; Crystal 1997). English is currently playing a key role in this area which is supported by the wide use of English as the first and foreign language in many developed countries such as United Kingdom, United States, Australia, Canada, New Zealand, Singapore and Hongkong.

The use of English around the globe is appreciated by millions of people. To communicate over the Internet with people in Australia, Germany and Singapore, for instance, needs a single *lingua franca* or a common language (Grabe, 1988; Crystal 1997; Graddol 1997; Bruthiaux, 2002). Furthermore, it is more complicated to use a three-way electronic translation in international business meetings involving three nations than the use of a single global language.

Skutnabb-Kangas (2000) suggests that language policy, including the promotion of English, must be inspired by an equitable vision of how all languages can be permitted to flourish. If English is to be a force for democracy and human rights, much needs to change, in Northern countries as much as in the South, and in North-South relations.

Language policy needs to and can play an important role in such a transition.

Linguistic power. The existence of a global language may result in linguistic power. Currently people who have English as their first language, second language or those who have a good common grounding of English are assumed to have power and access to develop their career in the international arena (Pennycook1997; Bruthiaux, 2002). On the other hand, those who do not have English might have some problems, for example, scientists who do not have a good command of English do not get access to international publications in journals. Business people cannot run international trade if they are not able to communicate in English.

Linguistic Death. According to history, thousands of languages have died since humans were first able to speak. A dominant language may contribute to the language loss if a particular ethnic group adopts a dominant language and ignores its own language (Day, 1980; 1985; Pennycook, 1994; Nagai, 1997).

Crystal (1997, p.17) states that:

> A lot of indigenous languages in North America, Brazil, Indonesia and parts of Africa have been lost. Other estimates are that within the next century around 80% of the world's 6,000 or so living languages will disappear.

When languages disappear, there will be an intellectual and social tragedy. Many languages have not yet been written down, or have only recently been written down (Crystal, 2000; Skutnab-Kangas 2000). Language is a medium of serving the history of people. A language can never be recaptured when it is lost. It is similar to the loss of an endangered species and environment degradation.

Crystal (2000) reported that the early history of English contact with minority language speakers in North America, Australia and in the Celtic parts of the British Isles was indeed one of conquest and assimilation. But currently, the existence of English as a global language has a positive effect which supports the local languages.

Discussion of language situation versus language awakening

Issues in local language endangerment

Ethnologue reported that eastern Indonesia is considered one of the regions that has hundreds of local languages and most are found in the Island of Papua. The second largest number of local languages was found in Sulawesi Island. However, many of the local languages were endangered or threatened to die if the Government and the speakers do not have the political will to keep maintaining the local languages and passing them on to the young generation (Crystal, 2000; Skutnab-Kangas, 2000; Himmelman, 2001; Marhum,2005).

Crystal in his book on Language Death (2000) predicted that around 90 per cent of languages will die in the next 50 years, particularly those found in the developing countries. In Indonesia itself, many local languages which have less than 1000 speakers have been moribund or threatened to die. In central Sulawesi many weak local languages have been endangered or died. There are several reasons why many local languages in Central Sulawesi were endangered or threatened to die. First, the incidence of intermarriage from year to year has significantly increased. The incidence of intermarriage brought about growing numbers of young *Bahasa Indonesia* speakers. Thus, this intermarriage contributed to the decrease of local language speakers. Second, a language shift has occurred over the last three decades in central Sulawesi. This case was mostly found in the Districts of Tolitoli and Donggala.

The language shift happened in Central Sulawesi when the local native people shifted their local languages to Buginese as a strong and dominant language in those two particular areas mentioned above (Marhum, 2003; 2005; 2008). The native people who were involved in the language shift were such groups as Dampal, Dondo, Tolitoli and Dampelas. The language shift happened because of economic, cultural and political reasons. Bugis ethnic group are migrant people who were originally from South Sulawesi. They were well-known as industrious or hard working people. Many of them are successful traders, fishermen, farmers, bureaucrats and politicians. Since they dominated many sectors of development, their language and culture became very dominant and is a powerful language in some particular areas of Central Sulawesi particularly in the Districts of Tolitoli and Donggala.

The language attitude of the young generation also contributed to the incidence of language death in central Sulawesi. For example, Marhum (2005, p.150) in his studies reported that in the District of Toltioli many young people were shy about using their local languages in the public arena. Furthermore, some parents did not pass on their local languages to the young generation and they did not encourage their children to use their local languages either at home or outside. Furthermore, teachers in some schools in the District of Tolitoli reported that in the past during *Orde Baru* Regime, children were strongly prohibited from using their local language at schools. Instead, many teachers were encouraged to promote *Bahasa Indonesia* as a unifying language at schools. Thus, it can be argued that this language policy on the one hand was good in which the Government has good political will to maintain *Bahasa Indonesia* as a unifying language and promote harmonious multicultural life as well as trying to prevent ethnic fanaticism. On the other hand, the language policy had discouraged the children from getting involved in their native language preservation.

Children had the wrong perception on their local native language. Some thought that their local languages were not important. The local language was the language of uneducated people. Such wrong perception had led to the decreasing numbers of local language speakers particularly among the young generation. However, this case applied only to the particular weak local languages or the language with a lower significant number of speakers such as Dondo, Toltioli and Dampal. In short, the local Government had a key role in either maintaining the local languages or contributing to the language endangerment.

Local language awakening

In the Regional Autonomy Era, the local language situation in some particular districts of Central Sulawesi has changed quite significantly. The local native elites who have influential and powerful positions in the political arena have contributed to the language awakening in the District of Tolitoli (Marhum, 2005; Marhum, 2007). For example, the native Tolitoli elites have encouraged their people to use and promote Tolitoli language as a native language of Tolitoli. Thus, Native Tolitoli people who work for the Local Government have more frequently used their local language of Tolitoli in their workplace. They also become more proud of the native language of Tolitoli. In the market place and other public places the use of *bahasa* Tolitoli by the native people has become more frequent compared to the past before the Regional Autonomy Era.

Furthermore, local languages have been promoted throughout formal education and the annual cultural and arts festival and media. In the formal education context some local languages have been incorporated into local curriculum content. In 2006, Kailinese language has been taught at primary schools in the district of Donggala, Parimo and city of Palu. In the Cultural and Arts event, the promotion of local language and culture were held in the form of a local language song festival, local language poet festival and local language speech festival. Two famous annual cultural festivals are Central Sulawesi Week Festival and Lake Poso Festival. In addition, in each district of Central Sulawesi, a number of local language and cultural festivals were held every year. Those festivals take place every year and aim at promoting the local language and culture. The local Government and the local people fully support those festivals. They are also aimed at encouraging the young generation to be proud to use their local languages. Local languages have also been used by Radio announcers in their particular programs such as news and other related cultural programs of FM Radio.

Post Regional Autonomy, language awakening of Tolitoli emerged and so did Kaili language of Donggal District and Palu City. The local native people tend to show their nativity and cultural identity through the use of their local language in daily communication. The local language jargons associated with ethnic fanaticism and native political identity is often heard. For example in the District of Buol, people often say '*Anak itai to*' this sentence is implicitly associated with racism and discrimination meaning. The same local language expression and meaning is often used by Kailiniese people of Palu City, District of Parimo and District of Donggala. For example, they often say, '*Seima iko*'. These two expressions also often refer to racial and prejudicial meaning. Furthermore, such negative expressions could also be associated with racial prejudice and stereotype towards people from other different ethnic groups identified as migrant groups. Ironically, such racial expressions were hardly ever heard before the Regional Autonomy era.

In short, the Regional Autonomy Era has contributed to the local language and ethnic awakening. Many people and experts of Regional Autonomy are worried about the detrimental impact of regional autonomy towards multicultural life. So it is recommended that the Government pay close attention to both the positive sides and negative sides of Regional Autonomy in relation to the local language and culture awakening (Muladi, 2007; Marhum,2007; Lampe 2008). It can be speculated that the implementation of regional autonomy has likely resulted in the detrimental political identity and ethnic fanaticism. At last, some new linguistic jargon which is associated with ethnic fanaticism and political identity has emerged. People in the Districts of Donggala and Palu City use local language jargon in Kailinise such as '*Seima Iko*?' and people in the District of Buol often use the same meaning of prejudicial expressions such as, *anak itai to*?'.

Bahasa Indonesia and the emergence of local dialect influence

Bahasa Indonesia is identified as an effective unifying language of Indonesia in general and in particular regions of Central Sulawesi. There is no linguistic barrier for the people of different ethnic and cultural background to communicate since their effective communication was supported by *Bahasa Indonesia* as their *lingua franca* and as a unifying language of Indonesian people from different ethnic group (Abas, 1987; Sumampow, 2003; Marhum, 2005). Almost all people of Central Sulawesi are able to speak *Bahasa Indonesia* fluently. People of different age and different geographical origin always speak *Bahasa Indonesia* in addition to their local languages. *Bahasa Indonesia* has always been an effective and popular unifying language of people in Central Sulawesi.

People of Central Sulawesi speak *Bahasa Indonesia* with local language accent and specific local language styles. Their *Bahasa Indonesia* is also influenced by their local dialects. Very often people from outside Central Sulawesi did not understand some particular sentences of *Bahasa Indonesia* of Central Sulawesi people since it is influenced by local dialect and local accent. For example, one day a migrant child asked Ali who was carrying a football heading to the football field. The boy asked Ali, '*ngana mo pi mana*?'. Ali answered, '*Sa pi main bole di sana'* or in *Bahasa Indonesia* it means '*anda mau ke mana*? and the answer, '*saya pergi main bola di sana*'. In English translation, it means, 'Where are you going?', 'I am going to play football over there' (answer). In local Indonesia Malay language, the word, '*Sa pi*' is the short form of local colloquial dialect meaning '*saya pergi*'. In fact, the meaning of the sentence above could sound confusing since *sa pi main bola di sana* sounds like 'a cow (sapi) plays a football over there';. Other *Bahasa Indonesia* sentences that confuse people from outside Central Sulawesi are related to some particular expression. For example, if someone makes a telephone call and asks to talk to someone else and those who are wanted on the phone are not around. The answer will be '*oh ada keluar*' This sentence means that the one who is wanted on the phone is not around. because *ada keluar* also means someone is somewhere just in the front yard or in the backyard of their house. In fact, *ada keluar* means someone went out at the moment. Thus, those sentences and phrases given previously are confusing.

Another anomaly meaning of *Bahasa Indonesia* sentence made by people of Palu and the surrounding districts follows. For example, when a policeman walks on the street wearing plain clothes someone who happens to recognise him or her might say, '*Polisi nanti dia itu*' which means 'he will become a policeman later'. The word *nanti* should mean later or in the near future according to standardised language. Those expressions given above often confuse many migrant people who just got settled in Central Sulawesi. Finally, though almost all people of Central Sulawesi are able to speak *Bahasa Indonesia* or Malay, some particular words or sentences are confusing to the people from outside Central Sulawesi because the words or sentences are influenced by the local dialects.

The promotion of English in schools

English is one of the first foreign languages taught in Indonesian schools since the beginning of political independence. Now English is the most popular foreign language taught in both private and public schools (Groeneboer,1998; Alwasilah, 1997; Ali, 2000; Jazadi, 2003). Many people of Indonesia have a positive perception on the promotion of English in Indonesia. The teaching of English has been promoted through all levels of education. Following the Introduction of Educational Law, the teaching of English has started from the primary school level to university levels.

In Central Sulawesi, the emergence of the Globalisation Era has been simultaneously responded to by the significant growing number of private English Schools. Furthermore, many people have realised the importance of English in the Globalisation Era since it is considered an important global language (Crystal, 1997; Ali, 2000; Marhum, 2003; Marhum, 2005; Marhum, 2008). For those reasons, many parents sent their children to the private English schools. Parents thought that the learning and teaching English received by their children through formal education was not enough. Many parents have encouraged their children to learn English at private schools because they also thought that the private English schools are the right institutions for their children to learn the English language.

There are two types of private English schools in Central Sulawesi. They are big private English schools and small private English Schools. The big ones have links with foreign education agencies such as English Language centre with IDP Australia network. The small English courses were mostly sponsored and managed by fresh graduates from an English Education Department in collaboration with local private training and educational stakeholders.

However, none of the private English schools in Central Sulawesi currently employ fulltime native speaker teachers. All English schools are looking forward to having native speakers as full-time English teachers. Since some English speaking countries have introduced travel warnings, many English schools have no access to recruit native speakers of English teachers. It can be argued that the recruitment of native speakers as teachers of English can upgrade the popularity of their English schools.

The majority of English schools in Central Sulawesi offer only general English to their learners whereas many public servants and private employees need to learn English for Specific purposes (ESP) which are applicable and relevant to their career.

In the formal education context, the use of English as an instructional language at International Rating Schools (Sekolah Bertaraf Internasional) is a must. The International Rating Schools (SBI) have become much more popular since English is a compulsory instructional language used at those schools. Many parents revealed that they are very proud to send their children to International Rating Schools. Yet, many parents have to be patient to wait for an available opening in the SBI schools. Thus, some parents also had to accept the reality that their children were put on the waiting list if there are not enough places.

There are only two International Rating Schools in Central Sulawesi. One is found in the city of Palu and the other one is found in the District of Luwuk and in addition there are five recommended International Rating Schools. Though International Rating Schools have become the most popular education institution in Central Sulawesi, many people are still worried about the insufficient human resources and insufficient supporting facilities of teaching and learning. Yet, a lot of parents in Central Sulawesi are very proud of the presence of International Rating Schools where their children have enough access to learn and practise their English continuously.

Conclusion

In summary, the status of the multilingual country of Indonesia, particularly in the Era of Decentralisation and Globalisation, has changed compared to the past. The language situation and the language awakening has shown a clearer situation on the status of local language, national language and English as a foreign language and global language.

In the Eastern Part of Indonesia where the region is identified as the most diverse linguistic region, many local languages have been endangered. In Central Sulawesi, some local languages have been moribund or eventually disappeared and some others were ignored. However, following the introduction of Regional Autonomy, the situation has changed since a lot of speakers of local languages have become more proud of their local language status as cultural identity and native political identity. Furthermore, *Bahasa Indonesia* has a key role as the means of communication among people of different ethnic and linguistic backgrounds. It can be argued that *Bahasa Indonesia* is an effective means of unifying language.

English has become a popular foreign language taught and used as the instructional language at International Rating Schools (Sekolah Bertaraf Internasional) in Central Sulawesi, particularly in the Era of Globalisation. The presence of International Rating Schools and the implementation of English Language Teaching and Education Policy have indicated how important English is in the Era of Globalisation. Finally, it can be argued that the status of English used and taught at schools of Indonesia should be upgraded to the second language status or third language status instead of foreign language status.

John has been a very inspiring Professor. I was very proud to be his student. He encouraged his students to work hard and make great progress in their studies.

References

Abas, H. (1987) *Indonesian as A Unifying Language of Wider Communication:* A Historical and Sociolinguistic Prospective; Pacific Linguistics, the Australian National University.

Ali, L. (2000), Lengser Ke Prabon: Kumpulan Kolom tentang Pemakaian Bahasa Indonesia. Pustaka Firdaus, Jakarta.

Alip, F.B. (1993), The Development of Bahasa Indonesia: A Historical Review. Unpublished Masters Thesis, SUNY at Stony Brook.

Alisjabana, S.T. (1976) *Language Planning for Modernisation,* the Case of Indonesia and Malaysia, Mouton, The Hague.

Alwasilah, A.C.(1997) *Politik Bahasa dan Pendidikan (Language Politics and Education)*. PT Remaja Rosda Karya, Bandung, pp. 31, 61-89.

Annamalai, E., 2003, Reflections on a language policy for multilingualism. *Language Policy Electronic Journal, Volume 2.* Kluwer Academic Publishers, pp.113-132.

Alwi, H. dan Sugono, D. (2000) *Politik Bahasa: Risalah Seminar Politik Bahasa.* Pusat Bahasa. Departemen Pendidikan Nasional. Jakarta, pp.1-25.

Basri, H. (1996), Languages Situation in Indonesia: Anticipating the Future of Minority Languages. Unpublished Report, SUNY, Stony Brook, p.20-27.

Bruthiaux, P. (2002) Predicting challenges to English as a global language in the 21st century. In Dasgupta & Tonkin (eds), *Language Problems & Language Planning Journal,* 26, 129-157.

Crystal D. (1997), *English as a Global Language.* Cambridge University Press.

Crystal, D. (2000) *Language Death*, pp.2-6. Cambridge University Press.

Day, R. (1980), 'ESL: a factor in linguistic genocide?' In J.C. Fisher, M.A. Clarke & J. Schacter (eds), *On TESOL '80, Building Bridges: Research and Practice in Teaching English as a Second Language.* Washington, DC: TESOL.

Djité,G.D.(1994), *From Language Policy to Language Planning: An Overview of Languages Other than English*, pp. 63-67. National Languages and Literacy Institute of Australia.

Eggington, W. and Wren, H. (1997), *Language Policy: Dominant English, Pluralist Challenges.* John Benjamins Company, Australia.

Grabe, W. (1988), English information access, and technology transfer: A rationale for English as an international language. *World English* 7, 63-72.

Grimes, B.F. (1988), *Ethnologue: Languages of the World*, 11th Edn. Summer Institute of Linguistics, Dallas.

Grimes, B.F. (1992), *Ethnologue: Languages of the World*. Dallas, TX: Summer Institute of Linguistics.

Groeneboer, K. (1998), *Gateway to the West: The Dutch Language in Colonial Indonesia, 1600 – 1950: A History of Language Policy*, pp. 249-290. Amsterdam University Press. Amsterdam.

Gunarwan, A (2000) Kedudukan dan Fungsi Bahasa Asing di Indonesia dalam Era Globalisasi. In Alwi at. Al. (eds) *Bahasa Indonesia dalam Era Globalisasi.* Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan Nasional, Jakarta.

Holmes, M. (1992), *Educational Policy for the Pluralist Democracy: The Common School, Choice and Diversity*, pp. 112-137. The Falmer Press. London.

Huda, N. (2000), Kedudukan dan Fungsi Bahasa Asing (the Status and Function of Foreign Language). In Alwih and Sugono, D (Eds) *Politik Bahasa: Risalah Seminar Politik Bahasa*, pp. 60-62*,* Pusat Bahasa, Jakarta.

Ingram, D.E, (1993) *Language Policy in Australia 1990s.* Paper to the Pre-conference International Workshop on Foreign Language Planning, pp. 11-12, National Foreign Language Center, the John Hopkins University, Washington DC.

Jakarta Post (2001), *Education Expert Call from Greater School Autonomy.* The Jakarta Post, 7/02/2001. Jakarta.

Jazadi, I. (2003a), Mandated English Materials and their Implications to Teaching and Learning: the Case of Indonesia. In W.A. Renandya (Ed), *Methodology and Materials Design in Language Teaching*, pp.2-7. SEAMEO Regional Language Centre, Singapore.

Keeves, J.P. and Sowden, S. (1997), Descriptive data analysis. In J.P. Keeves, (ed) *Educational Research, Methodology and Measurement: An International Hand Book*: Third Edition, pp. 296-305. Pergamon: New York.

Keeves, J. P. (1990), Legitimatory research. In J.P. Keeves, (ed) *Educational Research, Methodology and Measurement: An International Hand Book.* Second Edition, pp.146-150. Pergamon: New York.

Lampe, I. (2008), Ethnicity and Political Communication by General Election 2009. Koran Umum Harian Radar Sulteng (23 October 2008), Palu.

Lo Bianco, J. (1987a) *National Policy on Languages.* Canberra: Australian Government Publishing Service.

Lo Bianco, J (1987b) the National Policy on Languages. *Australian Review of Applied Linguistics.* 10 (2), 23-32.

Lucas, T.R. at.al. (1990) "Promoting the Success of Latino Language Minority, Students: An Exploratory Study of 6 High Schools". *Harvard Education Review* LX. 3, 315-340.

Marhum, M. (2003) Language Maintenance and Education Decentralisation in Tolitoli Central Sulawesi, a paper presented in the Seminar at the Department of Languages and Arts Education, University of Tadulako, 2003, Palu.

Marhum, M (2005) Language Policy and Educational Decentralisation in Indonesia. Unpublished Ph.D Thesis. Flinders University: South Australia.

Marhum, M (2006), *Language Maintenance in the Era of Decentralisation in Kabupaten Tolitoli.* A paper presented in the International Conference on Language, Gender and Sustainability in Collaboration with Frankfurt University of Germany. Universitas Tadulako, Palu.

Marhum, M. (2007), *Local Language and Ethnic Awakening and the Potential of Ethnic Conflict in District of Tolitoli in the Era of Decentralisation.* A fundamental Research sponsored by Directorate General of Higher Education, Ministry of Education, Jakarta.

Marhum, M. (2008), Language Policy in the Era of Globalisation; The Use of English as a Compulsory Instructional Language at International Rating Schools (SBI) in Central Sulawesi, A fundamental Research sponsored by Directorate General of Higher Education, Ministry of Education, Jakarta.

Moeliono, A.M, (2000), Kedudukan dan Fungsi Bahasa Indonesia dalam Era Globalisasi. in Alwih H. at al (eds) *Bahasa Indonesia dalam Era Globalisas (Indonesan Language in the Globalisation era)i: Risalah Kongres Bahasa Indonesia VIII.* Pusat Pembinaan dan Pengembangan Bahasa. Jakarta, pp. 26-29.

Mugler, F. and Lynch, J. (1996), *Pacific Languages in Educatio.* Institute of Pacific Studies. University of South Pacific.Suva.

Muladi, (2007), *Penerapan Otonomi Daerah Sarat Konflik.* Koran Harian Kompas (30 November 2007), Jakarta.

Nagai, Y. (1997), Educational Change from Within: Developing a Vernacular Elementary School in Papua New Guinea. PhD. Unpublished Thesis. Flinders University: South Australia.

Noss, R.B. (1984), *Overview of Language Issues in South-East Asia, 1950-1980*. Oxford University Press, Singapore.

Pennycook, A. (1994), *The Cultural Politics of English as an International Language*. Pearson Education: UK.

Phillipson, R. (2002) Globalising English and Local Language Policies. In A. Kirkpatrick, (Ed), *English in Asia: Communication, Identity, Power and Education. Language Australia*, pp. 7-14.

Pusat Pembinaan & Pengembangan Bahasa (1984), *Politik Bahasa Nasional (The National language politics)*. PN Balai Pustaka, Jakarta, 23-51.

Skutnab-Kangas (2000), *Linguistic Genocide in Education – or Worldwide Diversity and Human Rights?* Mahwah, New Jersey & London: Lawrence Erlbaum Associates.

Sneddon, J.N. (2003), *The Indonesian Language: Its History and Role in Modern Society*, pp. 196-217. University of New South Wales Press: Australia.

Somare, M , (1974), The butine oration – Education for self reliance. In *Educational Perspective in Papua New Guinea,* edited by the Australian College of Education. Carlton, VIC: The Australian College of Education, pp. 9-14.

Supriadi, D (1999), Restructuring the schoolbook provision system in Indonesia: some recent initiatives. *Educational Policy Analysis Archives,* 7 (7), http://epaa.asu.edu/epaa/v7n7.html

8

European Union co-operation in the education sector

Albert Tuijnman
Human Capital Division, European Investment Bank

Early developments in European education

Education is at the heart of national policies promoting the Knowledge Economy in all 27 Member States of the European Union (EU) but the Union does not have any explicit common education policy as such. Rather, under the principle of 'subsidiarity', the role of the Union is to promote co-operation between the Member States while preserving their sovereign right to determine the structure, organisation and content of education and training systems.

According to Article 149 of the Treaty of Rome, the Community's role is to

> … contribute to the development of *quality education* by encouraging co-operation between Member States and, if necessary, by supporting and supplementing their action, while fully respecting the responsibility of the Member States for the content of teaching and the organisation of education systems and their cultural and linguistic diversity. (emphasis in italics added to original) (European Commission, 2006)

Thus, according to the Treaty, the pursuit of *quality education* is the central objective of co-operation between the Members States and Community action in the education sector. As will be explained below the emphasis on 'quality' is significant.

In keeping with the spirit of this Article the European Union does not devise or implement any common education policies. Yet it nevertheless has developed ways of promoting co-operation through various action programs set up at the European level. These programs have grown incrementally over decades. In 1976, education ministers first decided to set up a joint information network, *Eurydice*, as a means for collecting statistics and improving understanding of education systems in the then nine-nation European Communities. The *Erasmus Mundus* program was initiated in 1988. Another popular EU program, *Socrates,* became operational in 1995 to strengthen relationships between schools and universities through staff and student exchanges and joint research programs. Education became formally recognised as an area of EU responsibility in the Treaty on European Union signed in Maastricht in 1992.

Support for European co-operation in vocational training was secured much earlier than in education. Article 150 of the Treaty states: "The Community shall implement a vocational training policy, which shall support and supplement the action of the Member States while fully respecting the responsibility of the Member States for the content and organisation of vocational training" (European Commission, 2006).

The Council of Ministers established the European Centre for the Development of Vocational Training (CEDEFOP), based first in Berlin and subsequently in Thessaloniki, in 1975. Several Community initiatives launched in the 1980s were precursors to what eventually became the *Leonardo da Vinci* program, adopted in 1994 as a means to contribute to EU vocational training policy. The European Training Foundation (ETF) was established in 1995 as the EU agency assisting over 40 neighbourhood and partner countries in modernising their vocational training systems.

The European Investment Bank and education

The EIB was established in 1958 under the Treaty of Rome, at the same time as the other institutions of what was then known as the European Economic Community (EEC), such as the European Commission and the European Court of Justice. Today it is the long-term policy-driven lending bank of the European Union (EU). It is part of the EIB Group, which also includes the European Investment Fund, and enjoys its own legal personality and financial autonomy within the EU. Its main task is to contribute towards the integration, balanced development and economic and social cohesion of EU Member States. The EIB is owned by the 27 Member States of the EU. The Bank reports to its Board of Governors, made up of a Government Minister from each Member State, usually the Minister of Finance.

The EIB is a large International Financial Institution in terms of its annual lending volume. Such institutions, which include the World

Bank and the International Finance Corporation, have significance in the global financial infrastructure because of the magnitude of the capital flows they mediate and the sectors they opt to support. The EIB finances itself by borrowing on the international capital markets rather than drawing on the budgetary resources of Member States. The Bank raises substantial volumes of funds on the capital markets – about EUR 55 billion annually in recent years – which it lends on favourable and long-term conditions to large-scale capital projects furthering EU policy objectives. It operates in keeping with strict banking practice and in close collaboration with the wider banking community, both when borrowing on the capital markets and when financing projects.

The early experience of the European Investment Bank (EIB) in lending to the education and training sector reflects the gradual development of European co-operation described above. With the exception of the European School in Luxembourg and one vocational school in Italy the Bank did not fund any projects specifically targeting education during the 1950s, 1960s and 1970s. In line with the clearer mandate given to the Community (cf. Article 150) to support vocational training in the 1970s, the Bank did start to finance the construction and rehabilitation of vocational and technical schools in the early 1980s, albeit on a rather modest scale. Only 15 such projects were financed until 1997, mainly in Ireland and Italy. Support for the general education sector and particularly universities only commenced after the EIB had received a new mandate at the Amsterdam Summit in 1997.

The Amsterdam European Council

By the mid-1990s the EU had built up a varied legacy of education and training initiatives, while steering clear of actions that might impact on the design of education services delivery as such. Although these actions were not designed to set specific EU policies or to involve the Union in national policy development, they reflected a shared desire to strengthen the role of education in the development of the common European space. Recognition of the need for the EU to become more directly involved in education and training in order to promote Europe's economic and social future was given voice in a White Paper prepared by the Commission in 1995: "Teaching and Learning: Towards the Learning Society". This identified five general, common objectives for European education systems: (a) Encourage the acquisition of new knowledge; (b) Bring schools and the business sector closer together; (c) Combat exclusion; (d) Develop proficiency in three Community languages; and (e) Treat capital investment and investment in education and training on an equal basis (European Commission, 1995).

This White Paper, and subsequent discussion of it at the Madrid European Council meeting in 1996, led to the Amsterdam Summit in 1997, which stressed the importance of education to economic growth and employment. At the Amsterdam meeting the European Council requested the European Investment Bank "… to examine its scope of intervention in the areas of education, health, urban environment and environmental protection" (European Council, 1997). This led, in July 1997, to EIB Board approval of the Amsterdam Special Action Program designed, *inter alia,* to encourage economic growth and employment in Europe by increasing investment in education and health (EIB, 1997). Thus the EU Council, while recognising the sovereign role of national governments in determining education policy and its implementation, clearly mandated the Bank to contribute to the funding of education services delivery. The EIB's mandate for education sector lending was initially limited in time, extending from 1997 to 2000. As will be seen below the Bank reacted promptly to the conclusions of the Amsterdam Council, with several general education projects, two of which were in higher education in Portugal and Spain and approved in 1997.

The Lisbon European Council

By 1999 the Bank had funded education projects in most of the then EU-15 Member States. Experience gained, and analytical work conducted, led to the recognition by the Bank's services that it should regroup its education and health sector lending activities under the general heading of 'human capital', defined as the productivity relevant knowledge, skills, competencies and other attributes of the population, including health. As part of the Bank's first Corporate Operational Plan for 1999-2000, four issues were identified as being particularly important when selecting education projects for funding: (a) Priority to be given to regional areas with the most pronounced human capital deficiencies; (b) All levels of education and training to be supported; (c) Both private and public sector education operations to be supported; and (d) Encouragement for the financing of research, development and innovation (RDI) as a complement to investment in human capital (EIB, 1999a-c).

Meanwhile, the centrality of education to European economic growth and jobs was emphasised at the Lisbon European Council in 2000, and reinforced at subsequent Council meetings held in Stockholm (2001) and Barcelona (2002). The Lisbon Council adopted ambitious targets to be achieved by 2010. It also positioned education, RDI and Information and Communication Technologies (ICTs), seen as the principal building blocks of the larger strategy to make Europe the most competitive economy in the world, at the centre stage of policy. The EIB has since placed these Lisbon targets at the core of its lending strategy for the so-called “Innovation 2010 Initiative” and the “Knowledge Economy” objectives.

Common EU objectives for education

Subsequent to the Lisbon Summit, on the basis of a Commission proposal and contributions from the Member States, the Council adopted the "Report on the Concrete Future Objectives of European Education Systems" in February 2001 (European Commission, 2001a; European Council, 2001). This was the first ever European Council document to set out a coherent approach to national education policies in the context of the EU. The Council also asked that a detailed work program for education be drawn up by the Commission.

In response, the Commission prepared the “Detailed Work Program on the Future Objectives of Education and Training Systems” (European Commission, 2001b). This was to be implemented using the "open method of co-ordination" between Member States. This instrument would be used to gradually build coherent policies in education, an area where a “common policy” still was not feasible but where there existed a recognised political need for a stronger European dimension (European Commission, 2002).

The “open method of co-ordination” (OMC) was a typically European ‘invention’ at the time. It is a method designed to help the Member States progress jointly in the reforms they need to undertake in order to reach the ambitious goals adopted by the Lisbon Council in the year 2000. The method includes the following elements (European Commission, 2009):

- Fixing **guidelines and timetables** for achieving short, medium and long-term goals;
- Establishing quantitative and qualitative **indicators and benchmarks,** tailored to the needs of Member States and sectors involved, as a means of comparing best practices;
- Translating European guidelines into **national and regional policies**, by setting specific measures and targets; and
- Periodic **monitoring** of the progress achieved in order to put in place **mutual learning** processes between Member States.

Initially the OMC was only applied to employment and economic policy. When the European Council set the three per cent of GDP objective for RDI investment by the year 2010, the Commission suggested that the OMC should be applied for this objective as well. Subsequently the OMC also was to be applied to the common education and training objectives of the European Union, agreed on in 2002. The OMC process is expected to contribute to enhancement of **mutual learning** and peer review; identification of **good practices** and of their conditions for transferability among Member States; development of **joint policy initiatives** among several Member States and regions; and identification of areas where **Community initiatives** can reinforce actions taken at Member State level.

The Barcelona Council (March 2002) underlined these ambitions by concluding that education was one of the foundations of the European social model and that Europe's education systems should become a *world quality reference* by 2010 (European Council, 2002). Since 2002 there have been three, formally recognised, EU objectives for the education sector:

- Improving the quality and effectiveness of education and training systems;
- Facilitating the access of all to the education and training systems; and
- Opening up the education and training systems to the wider world.

For each of these three objectives a sub-set of concrete goals has been formulated, and for each goal one or more international education indicators has been defined, many of them derived from OECD sources, particularly the *Education at a Glance: OECD Indicators* reports and the OECD Program for International Student Assessment (PISA). These indicators allow the European community to measure, through the open method of co-ordination, the actual progress being made towards achieving the common goals by the year 2010. At the time of writing the most recent benchmarking report for education was published in 2007 (European Commission, 2007). The objectives, goals and associated indicators are listed in the Appendix.

Five years after the launch of the EU Lisbon strategy, a High-Level Group advising the Commission acknowledged that the results achieved had been mixed at best. Although the reform and modernisation of Europe's education systems would remain the prerogative of the Member States, the Council and Parliament agreed that certain steps would have to be taken at the European level to contribute to the renewal of the Lisbon agenda. Among these are measures taken to promote a common European Higher Education Area, notably through the Bologna process, and the creation of the common European Research Area, to which the 7th Framework Program and the newly established European Research Council are expected to contribute. ICTs are another priority area, supported through the *e-Learning* program.

EIB contribution to EU education policies

In 2002 the Board of Governors of the EIB requested the Bank's services to devise a system that would allow it to more explicitly assess the value added of its projects in three broad domains: (a) contribution to EU policies and strategies; (b) Technical viability and economic soundness; and (c) financial profitability. In response the Bank catalogued its entire range of lending operations by principal sectors, and sought to demonstrate how each contributed to EU policies and strategies.

In the absence of a common education policy across the Community, the Bank based its value added criteria for the education sector on the three principal objectives and the associated goals and indicators adopted by the Barcelona Council in 2002. Accordingly, the Bank's mandate in the education sector is not defined on the basis of an explicit EU policy but derived from the Council's and Parliament's endorsement of the education objectives associated with the Lisbon strategy.

Since 2000 education lending has become a permanent feature of the Bank's portfolio, largely unrestricted geographically, and treated as a top priority, especially in relation to the connections made between the development of human and social capital, on the one hand, and desirable outcomes pursued as part of the Lisbon strategy, such as employment, cohesion and economic growth, on the other.

The Bank's contribution to the funding of the education sector has gone from little to substantial in the course of the decade. Since 2005 loans for education and training have amounted to about five per cent of the EIB's annual turnover, or some EUR 2.5 billion a year. Single signature education loans, or the education related components of structural program loans, have ranged from as little as EUR 1.8 million up to EUR 475 million. Over the decade since the launch of the Amsterdam Special Action Program for human capital, education lending has exceeded EUR 20 billion, the bulk of which has been for capital investments in education infrastructure.

In recent years the EIB has increased its lending for projects that are not strictly classified under an ‘education‘ heading, including the spill over into research and development activities. Accordingly, the sector portfolio has evolved from an exclusive reliance on tangible infrastructure projects to also include projects dealing with the intangible assets and processes of education and knowledge production, such as distance learning technologies, academic research in universities, and loans for higher education students. Such projects are justified because they focus on the ’enabling resources‘ that are needed to make schools and universities function well. Clearly, as one diversifies from tangibles such as school buildings and other hardware to intangibles, such as sector-wide restructuring or university-based academic research, the Bank moves closer to policy areas that are politically more visible, hence sensitive and there is potentially also more scope for reputation impact.

Lifelong learning and EIB lending priorities

The European Council of March 2005 adopted the policy framework of lifelong learning as the *sine qua non* for achieving all of the Lisbon objectives, taking into account the desirability of *high quality at all levels*. Lifelong learning policy concepts define the scope and means of Community action programs in education and training from January 2007.

As part of its updated Lisbon strategy, the European Council and Parliament have endorsed a new program for Community lifelong learning policies, replacing the *Socrates* and *Leonardo da Vinci* programs that expired at the end of 2006. This program will run throughout the period 2007-2013 and comprises four sectoral programs on school education (*Comenius*), tertiary education (*Erasmus*), vocational training (*Leonardo da Vinci*) and adult education (*Grundtvig*), complemented by a transversal program focusing on policy co-operation, languages, ICTs, and dissemination of best practices. The aim of the new program is to contribute, through its support for lifelong learning, to the development of the Community as an advanced knowledge society, with sustainable economic development, more and better jobs, greater social cohesion, active citizenship and personal development. It aims to foster interaction, co-operation and mobility between the education and training systems of the Member States, so that they will become a world quality reference.

Because the EIB’s mandate, eligibility and value added criteria for the education sector are all based on the three common objectives for education and training associated with the Lisbon strategy, the adoption of lifelong learning as *the* reference framework for the structure, organisation and content of Europe's education systems will have ramifications for how the Bank classifies and values its own operational activities in the sector. The adoption of the lifelong learning framework has had real operational implications. For example, it has provided the justification for Bank lending to the Kindergarten sector, which previously had not been considered a priority.

Since the late 1960s much has been written about the concept of lifelong learning (Tuijnman, 2002; Tuijnman and Boström, 2002). In the report of the UNESCO Commission chaired by Edgar Faure (Faure et al., 1972), a former French minister of education, lifelong education was considered not as a system of education but, rather, as a principle with respect to the organisation of education. Generally, lifelong learning is based on the idea that the organising principle for education cannot be based on the traditional, 'front-loaded' approach, according to which learning is mainly confined to a sequence of compulsory schooling and formal education at a young age. Instead, it is suggested that learning should be seen as a fundamental and lifelong process of human development.

In the mid-1990s the theory of lifelong learning made a strong come back in the international policy arena (Delors, 1996; OECD, 1996). The latter report adopted the following definition: "Lifelong learning is best understood as a process of individual learning and development across the life-span, from cradle to grave – from learning in early childhood to learning in retirement. It is an inclusive concept that refers not only to education in formal settings, such as schools, universities and adult education institutions, but also to 'life-wide' learning in informal settings, at home, at work and in the wider community" (OECD, 1996, p.15). This view of learning embraces social and individual development of all kinds. The approach is system-wide; it focuses on the standards of knowledge and skills needed by all, regardless of occupation or age. It emphasises the need to prepare and motivate all children at an early age for learning over a life-time, and directs effort to ensure that all adults, employed or unemployed, who need to retrain or upgrade their skills, are offered opportunities for doing so.

Despite its all-embracing nature, the concept of lifelong learning has several features that give it an operational significance for education policy in distinction from other approaches:

- The centrality of the learner and learner needs: that is, an orientation towards the 'demand side' of education rather than just the supply of places;

- An emphasis on self-directed learning and the associated requirement of ‘learning to learn’;
- Recognition that learning takes place in many settings, both formal and informal;
- An emphasis on acquiring not only knowledge but also useful skills and competencies;
- A long-term view, that takes the whole course of an individual’s life into consideration; and
- Recognition of the principle that all learning is cumulative and that early success in learning contributes to later achievement.

Ramifications for EIB education lending

As previously noted, the European Council and the Parliament have endorsed a new program for lifelong learning policies that will run throughout the period 2007-2013. This aims to foster interaction, co-operation and mobility between the education and training systems of the Member States.

The adoption of this lifelong learning framework as an umbrella for all Community action programs in the education sector, in essence, extends an invitation to the Bank to reinterpret its mandate, eligibility criteria and lending priorities for the sector on the basis of similar principles. So far, the Bank’s mandate has evolved in parallel with the gradual development of Community interests in the education and training sector, from support for technical and vocational education in the 1970s and 1980s, and funding for primary and secondary schools and universities in the late 1990s, to the first demand-oriented and intangible projects in student lending and public science and academic RDI in recent years. A recasting of the Bank’s eligibility guidelines in accordance with the lifelong learning framework will not, in reality, lead to a major shift in priorities because the Bank has already interpreted its mandate for education lending operations rather inclusively.

From the above a few straightforward principles for interpreting the Bank’s eligibility guidelines for the education sector can be derived.

The **lifelong dimension** stresses the importance of all learning interventions taking place from infancy through to life in retirement. Hence the Bank could support pre-school interventions, kindergartens, primary and secondary schools, technical and vocational education, higher education and universities, distance education and adult education for senior citizens.

The **life-wide dimension** implies that not only the learning going on in formal institutions is eligible for Bank support, but also organised learning interventions that occur in other life settings, notably the workplace and the wider community. Examples are youth apprenticeship training and active labour market programs for at-risk adults, human resource development programs in business enterprises, Internet-based learning platforms, and media and broadcasting programming with an educative orientation.

The above are components of the lifelong learning system that are, in principle, eligible for Bank consideration. In reality, however, the Bank is unlikely to pursue all these components at the same time. Project selection and approval in practice still depend on the findings of technical, economic and financial analysis, and on the priorities that flow from this.

The education quality imperative

The lifelong learning framework accentuates 'learning' rather than 'education' per se. This is significant because it somewhat reduces the traditional preoccupation with the formal structures, institutions and processes of education, and instead focuses attention on the individual learners and the knowledge, skills and competencies acquired by them. Placing the individual at the heart of the lifelong learning 'system' will mean paying more attention to the quality of learning outcomes since the realisation of lifelong learning policies will depend to a large degree on the capacity and motivation of individuals to handle their own learning pathways over time.

The OECD PISA results have come as a shock for politicians and the general public in a number of European countries. This has contributed to increased awareness of the urgent need to improve quality and learning outcomes in European education systems. As a result many countries have embarked on new school effectiveness programs and other reforms intended to improve the quality of education though increased accountability, school autonomy and better teacher training. It is also important to note that the PISA country averages mask significant regional disparities in student achievement in reading literacy, mathematics and problem solving within countries. Finally, the PISA results also have draw attention to structural aspects of secondary education, notably the divisive issue of school differentiation at an early age.

In Europe, following the huge and unprecedented *quantitative* expansion of enrolments in secondary and tertiary education during the 1980s and 1990s, which was accompanied by policies such as 'automatic' promotions and the opening up of universities to a broader clientele – which incidentally created a demand for more tangible investments in infrastructure to alleviate problems of overcrowding –- more policy emphasis is now being placed on efforts to improve the *quality* of education.

Politically the stakes have been raised in a number of countries as public awareness of the results of comparative studies such as PISA and surveys of mathematics and reading literacy undertaken by the International Association for the Evaluation of Educational Achievement (IEA) have become more widely known. Given these raised stakes, decision makers now more keenly demand information about the policy malleable factors that influence student achievement in primary and secondary schools, and the factors that can promote excellence in tertiary education. The knock-on effect is that the common European education benchmarks adopted under Lisbon strategy and the Open Method of Co-ordination have become even more prominent politically. Consequently the OECD and EU international education indicators and statistics have gained additional prominence and significance.

In tertiary education too there is a general thrust in Europe to improve quality and effectiveness and to pursue 'excellence'. Rather than spreading the available resources thinly by treating all universities similarly, with the mediocrity and duplication this might bring, countries are more keenly pursuing strategies to introduce more competition for funding to create critical masses in specific disciplines or areas of applied science and innovation. The EIB could potentially play a significant role in the modernisation of the common European Higher Education Area and the common European Research Area. This role need not be limited to the financing of university buildings and laboratories but could also include the funding of academic R&D projects directly.

Conclusions

The Lisbon strategy has brought an impetus for education reform in Europe, oriented towards the development of lifelong learning systems and quality in education. In 2000, in part in response to the challenges brought by the Lisbon agenda, the mandate of the EIB for lending to the education sector was extended indefinitely. Education has since become a permanent feature of the Bank's operations, applicable in both EU Member States and partner countries.

Even though policies on the structure, organisation and content of education continue to be the responsibility of the Member States, the

EU, in accordance with Article 149 of the Amsterdam Treaty, has adopted lifelong learning as the guiding principle for the development of education and training systems in Europe **along both the life-long and life-wide axes.** This development has been accompanied by a marked and still on-going shift in policy emphasis from the quantitative expansion of enrolment capacity to the pursuit of quality and excellence in European education.

Ideas about lifelong learning are not static but evolving; they vary according to context and various political, economic, social and cultural factors. For this reason, the emphasis placed at any given time on any specific elements that collectively constitute the policy framework of lifelong learning may differ over time and between Member States. This is deemed appropriate, considering that the education systems of European countries continue to differ in both structure and content. Policy priorities in the education sector will therefore vary too.

The EIB should be attentive to these differences and ensure that the lifelong learning framework is used as a flexible instrument for supporting the coherent development of education systems in Europe. Depending on a country's requirements and willingness to borrow, this will necessarily mean supporting early childhood interventions in one country, tertiary education in a second, or active labour market policies and skill training programs for the unemployed in a third.

In a lifetime perspective learning is cumulative, and early achievement builds later success. For this reason the Bank should be attentive, in addition to the adequacy of the physical infrastructure of schools and other educational institutions, also to the processes that enhance the attainment of high-quality learning outcomes.

In the absence of a common education policy across the EU, the Bank has based its value added criteria for the education sector on the three common EU objectives for education and the associated indicators and benchmarks. Accordingly, the Bank's role in the education sector is not defined on the basis of an explicit EU policy but derived from the Council's and Parliament's endorsement of the education objectives associated with the Lisbon strategy.

Postscript

This chapter has traced the gradual, at times torturous, evolution of European Union competence in the education and training sectors, from initially encompassing only limited actions in vocational and adult education, to eventually addressing the entire lifelong learning system, from early childhood education to learning in retirement. In the process the EU also adopted common education objectives, ambitious targets to be achieved by 2010 and beyond, and an elaborate system of international education indicators and benchmarks to measure progress in EU education reform.

The author has made a modest contribution to several of the underlying processes. Examples are his work on international education indicators for the *Education at a Glance* reports; the seminal publication, *Lifelong Learning for All: Meeting of OECD Education Ministers at Ministerial Level – January 1996*; and the recent EIB *Education Lending Policy: Focus on Improving Skills and Enhancing Employment*, the official policy statement on education lending priorities and targets adopted by the Board of Governors of the European Investment Bank in May 2008 that has, for the first time, provided the policy justification for EU-wide lending to Kindergarten projects and other investments in recurrent costs deemed of exceptional relevance to the pursuit of quality education in Europe.

At the basis of this work lies the rich experience the author gained while writing a Ph.D. thesis supervised by Professors Torsten Husén, Ingemar Fägerlind and John P. Keeves at the Institute of International Education, Stockholm University, in the mid to late 1980s. The knowledge and interests acquired at that time about educational measurement, comparative education and lifelong learning, but also improved writing skills, and access to international networks of scholars and policy officials, laid the foundation for all work undertaken subsequently. I will forever be indebted and grateful to John P. Keeves for his hugely important contribution to my development.

References

Delors, J. (1996). Education: The Necessary Utopia. In *Learning: The Treasure Within*. Report from the International Commission on Education for the 21st Century. Paris: UNESCO.

European Commission (1995). *White Paper on Education and Training. Teaching and Learning: Towards the Learning Society* [COM (95) 590 final, 29 November 1995]. Luxembourg: Office for Official Publications of the European Communities.

European Commission (2001a). *The Concrete Future Objectives of Education and Training Systems.* Communication from the Commission, COM (2001) 59, 31 January 2001. Brussels: European Commission.

European Commission (2001b). Retrieved14 February 2009 from http://europa.eu/scadplus/leg/en/cha/c11049.htm

European Commission (2002). European Benchmarks in Education and Training: Follow-up to the Lisbon European Council. Communication from the Commission, COM (2002)629, 20 November 2002. Brussels: European Commission.

European Commission (2006). Retrieved14 February 2009 from http://ec.europa.eu/education/policies/pol/policy_en.html#traité.

European Commission (2007). Progress Towards the Lisbon Objectives in Education and Training: Indicators and Benchmarks 2007. Staff Working Document, SEC(2007)1284. Brussels: European Commission.

European Commission (2009). Retrieved14 February 2009 from http://ec.europa.eu/invest-in-research/coordination/coordination01_en.htm

European Council (1997). Resolution of the European Council on Growth and Employment, Amsterdam, 16 June 1997. *Official Journal of the European Communities*, C 236, 2.8.1997. Luxembourg: Office for Official Publications of the European Communities.

European Council (2001). Presidency Conclusions on the Follow-up of the Report on the Concrete Future Objectives of Education and Training Systems. *Official Journal of the European Communities*, C 204 of 20.07.2001. Luxembourg: Office for Official Publications of the European Communities.

European Council (2002).Presidency Conclusions on the Follow-up to the Report on the Concrete Future Objectives of Education and Training Systems in View of the Joint Council/Commission Report Adopted by the Barcelona European Council on 15-16 March 2002 Official Journal of the European Communities, C 58 of 05.03.2002. Luxembourg: Office for Official Publications of the European Communities.

EIB (1997). Retrieved14 February 2009 from http://www.eib.org/about/press/1997/1997-124-eibs-growth-and-jobs-action-program-results-up-to-dec.-12.htm

EIB (1999a). *Corporate Operational Plan 1999-2000* (Doc. CA/325/99). Luxembourg: European Investment Bank.

EIB (1999b). *Background Paper No. 1: The Bank's Policy in the Health and Education Sectors* (Doc. CA/99/03). Luxembourg: European Investment Bank.

EIB (1999c). *Background Paper No. 3: The Bank's Strategy in the Education Sector* (Doc. CA/99/05). Luxembourg: European Investment Bank).

Faure, E., Herrera, F., Kaddoura, A.-R., Lopes, H. Petrovsky, A.V., Rahnema, M., and Ward, F.C. (1972). *Learning to Be: The World of Education Today and Tomorrow*. Paris: UNESCO.

OECD (1996). *Lifelong Learning for All*. Meeting of the Education Committee at Ministerial Level, 16-17 January 1996. Paris: OECD.

Tuijnman, A.C. (2002). Lifelong learning: Evolution of a Conceptual Map and some Implications. In J. Cribbin and P. Kennedy (eds.), *Lifelong Learning in Action* (pp.3-14). Hong Kong: Hong Kong University Press.

Tuijnman, A.C. and Boström, A.-K. (2002). Changing Notions of Lifelong Education and Lifelong Learning. *International Review of Education*, 48(1/2), 93-110.

Appendix: Objectives, key issues and indicators for European education and training systems

Source: European Commission,
http://europa.eu/scadplus/leg/en/cha/c11086.htm

Objective 1: Improving the quality and effectiveness of education and training systems

Strategic Objective	Key Issues	Indicators for Measuring Progress
1. Improving education & training for teachers & trainers Starting period: during 2002	1. Identifying the skills that teachers & trainers should have, given their changing roles in the knowledge society; 2. Providing the conditions which adequately support teachers & trainers as they respond to the challenges of the knowledge society, including through initial & in-service training in the perspective of lifelong learning; 3. Securing a sufficient level of entry to the teaching profession, across all subjects & levels, as well as providing for the long-term needs of the profession by making teaching & training even more attractive; 4. Attracting recruits to teaching & training who have professional experience in other fields.	- Shortage/surplus of qualified teachers & trainers on the labour market, - Progression in number of applicants for training programs (teachers & trainers), - Percentage of teachers & trainers who follow continuous professional training.
2. Developing skills for the knowledge society Starting period: 2nd half of 2001	1. Identifying new basic skills, & how these skills together with the traditional basic skills can be better integrated in the curricula, learned & maintained through life; 2. Making attainment of basic skills genuinely available to everyone, including those less advantaged, those with special needs, school drop-outs & to adult learners; 3. Promoting official validation of basic skills, in order to facilitate ongoing education & training & employability.	- People completing secondary education, - Continuous training of teachers in areas of emerging skills needs, - Literacy Attainment Levels (PISA), - Numeracy/ Mathematics Attainment Levels (PISA), - Learning to learn Attainment Levels, - Percentage of adults with less than upper secondary education who have participated in any form of adult education or training, by age group.
3. Ensuring access to ICT for everyone Starting period: 2nd half of 2001	1. Providing adequate equipment & educational software so that ICT & e-Learning processes can be best applied in teaching & training practices;	- Percentage of teachers that have been trained in ICT use in schools, - Percentage of pupils & students using ICT in

Strategic Objective	Key Issues	Indicators for Measuring Progress
	2. Encouraging the best use of innovative teaching & learning techniques based on ICT.	their studies, - Percentage of learning sessions in teaching & training institutions in which ICT is used.
4. Increasing recruitment to scientific & technical studies Starting period: 2nd half of 2001	1. Increasing the interest in mathematics, science & technology from an early age; 2. Motivating more young people to choose studies & careers in the fields of mathematics, science & technology in particular research careers & scientific disciplines where there are shortages of qualified personnel, in a short & medium term perspective, in particular through the design of strategies for educational & vocational guidance & counselling; 3. Improving gender balance among people learning mathematics, science & technology; 4. Securing a sufficient numbers of qualified teachers in mathematics & scientific & technical subjects.	- Increase in number of entries into mathematics, science & technology courses (upper secondary advanced levels & tertiary levels, by gender), - Increase in number of graduates in mathematics, science & technology, by gender, - Increase in number of scientists & engineers in society, by gender, - Increase in number of qualified teachers in MST (secondary level).
5. Making the best use of resources Starting period: During 2002	1. Increasing investment in human resources while ensuring an equitable & effective distribution of available means in order to facilitate general access to & enhance the quality of education & training; 2. Supporting the development of compatible quality assurance systems respecting diversity across Europe; 3. Developing the potential of public-private partnerships.	- Increase in per capita investment in human resources (Structural indicator).

Objective 2: Facilitating the access of all to the education and training systems

Strategic Objective	Key Issues	Indicators for Measuring Progress
1. Promoting open learning environments Starting period: between 2nd half of 2002 & end of 2003	1. Broadening access to lifelong learning by providing information, advice & guidance, on the full range of learning opportunities available; 2. Delivering education & training so that adults can effectively participate & combine their participation in learning with other responsibilities & activities; 3. Ensuring that learning is accessible for all, in order to better respond to the challenges of the knowledge society; 4. Promoting flexible learning paths for all; 5. Promoting networks of education & training institutions at various levels in the context of	- Percentage of the population between 25 & 64 participating in education & training (structural indicator).

Strategic Objective	Key Issues	Indicators for Measuring Progress
	lifelong learning.	
2. Making learning more attractive Starting period: between 2nd half of 2002 & end of 2003	1. Encouraging young people to remain in education or training after the end of compulsory education & motivating & enabling adults to participate in learning through later life; 2. Developing ways for the official validation of non-formal learning experiences; 3. Finding ways of making learning more attractive, both within the formal education & training systems & outside them; 4. Fostering a culture of learning for all & raising the awareness of potential learners of the social & economic benefits of learning.	- Percentage of working time spent by employees on training per age groups, - Participation in tertiary education, - Proportion of the population aged 18-24 with only lower secondary education & not in education or training (Structural indicator).
3. Supporting active citizenship, equal opportunities & social cohesion Starting period: During 2002	1. Ensuring that the learning of democratic values & democratic participation by all school partners is effectively promoted in order to prepare people for active citizenship; 2. Integrating fully equal opportunity considerations in the objectives & functioning of education & training; 3. Ensuring fair access to acquisition of skills for the less privileged or those currently less well served & motivating them to participate in learning.	- Proportion of the population aged 18-24 with only lower secondary education & not in education or training (structural indicator).

Objective 3: Opening up the education and training systems to the wider world

Strategic Objective	Key Issues	Indicators for Measuring Progress
1. Strengthening the links with working life & research, & society at large Starting period: between 2nd half of 2002 & end of 2003	1. Promoting close co-operation between education & training systems & society at large; 2. Establishing partnerships between all types of education & training institutions, firms & research facilities for their mutual benefit (1); 3. Promoting the role of relevant stakeholders in developing training, including initial training, & learning at the work place.	- Percentage of students & trainees in initial training benefiting from placement arrangements (*éducation en alternance*).
2. Developing the spirit of enterprise Starting period: between 2nd half of 2002 & end of 2003	1. Promoting the sense of initiative & creativity throughout the education & training system in order to develop the spirit of enterprise ('entrepreneurship'); 2. Facilitating the acquisition of skills needed to set up & run a business.	- Proportion of self-employed in various sectors of the knowledge economy (particularly age group 25-35), - Percentage of education & training institutions providing counselling & guidance for setting up business.

Strategic Objective	**Key Issues**	**Indicators for Measuring Progress**
3. Improving foreign language learning Starting period: between 2nd half of 2002 & end of 2003	1. Encouraging everyone to learn two, or where appropriate, more languages in addition to their mother tongues, & increasing awareness of the importance of foreign language learning at all ages; 2. Encouraging schools & training institutions in using efficient teaching & training methods & motivating continuation of language learning at a later stage of life.	- percentage of pupils & students who reach a level of proficiency in two foreign languages, - percentage of language teachers having participated in initial training or in-service training courses involving mobility providing direct contact with the language/culture they teach.
4. Increasing mobility & exchange Starting period: During 2002	1. Providing the widest access to mobility to individuals & to education & training organisations, including those serving a less privileged public & reducing the remaining obstacles to mobility; 2. Monitoring the volume, directions, participation rates as well as qualitative aspects of mobility flows across Europe; 3. Facilitating validation & recognition of competencies acquired during mobility; 4. Promoting the presence & recognition of European education & training in the world as well as their attractiveness to students, academics & researchers from other world regions.	- Proportion of national students & trainees carrying out part of their studies in another EU or third country, - Proportion of teachers, researchers & academics from other EU countries employed at different educational levels, - Number & distribution of EU & non-EU students & trainees in education & training.
5. Strengthening European co-operation Starting period: During 2002	1. Enhancing the effectiveness & timeliness of recognition processes for the purpose of further study, training & employment throughout Europe; 2. Promoting co-operation between responsible organisations & authorities in view of more compatibility in quality assurance & accreditation; 3. Promoting transparency of information on education & training opportunities & structures in view of the creation of an open European area for education; 4. Promotion of the European dimension of teaching & training.	- Proportion of undergraduate & postgraduate students & researchers continuing their studies in another EU or third country, - Percentage of graduates obtaining joint degrees in Europe, - Percentage of students & trainees within ECTS or Europass and/or obtaining Diploma or Certificate Supplement.

9

Correspondence between measures of children's mental health

Helen Askell-Williams and *Michael J. Lawson*
Centre for the Analysis of Educational Futures, Flinders University

Overview

This chapter reports the use of correspondence analysis to investigate teacher and parent/caregiver assessments of students' mental health based upon data from the evaluation[1] of the KidsMatter mental health promotion, prevention and early intervention pilot program in 100 primary schools across Australia in 2007 and 2008. Goodman's (2005) Strength and Difficulties Questionnaire (SDQ) was completed by parents/caregivers and teachers of 4980 primary school students. The SDQ was developed as a brief mental health screening instrument and is widely used in many nations, including Australia (Levitt, Saka, Romanelli, & Hoagwood, 2007). A second measure, specifically developed for the KidsMatter evaluation, canvassed the five core groups of indicators of students' social and emotional competencies identified by the Collaborative for Academic, Social and Emotional

[1] The data for this paper were gathered as part of the evaluation of the KidsMatter Stage 1 (2007-2009). We wish to acknowledge the contribution of members of the evaluation consortium based in the Centre for Analysis of Educational Futures at Flinders University in designing and undertaking this evaluation. The evaluation team consisted of Phillip Slee, Mike Lawson, Alan Russell, Helen Askell-Williams, Katherine Dix, Larry Owens, Barbara Spears, and Grace Skrzypiec. We wish to thank the other members of the evaluation team and *beyondblue: the national depression initiative* for permission to use the data for the preparation of this paper.

Learning (CASEL, 2006), namely, self-awareness, self-management, social awareness, relationship skills, and responsible decision making. This second measure was also completed by the students' teachers and parents/caregivers. A third measure was based on a non-clinical assessment by school staff, who identified students considered to be 'at risk' of social, emotional or behavioural problems. Correspondence analysis, which overcomes limitations of the collected data in that it works with frequency data that does not impose restrictive distributional assumptions, was used to examine relationships in these measures. The Correspondence Analysis illustrates clear patterns of relative distances between alternative measures of positive mental health.

Background

The recent Australian Research Alliance for Children and Youth Report card ranks Australians' mental health status at 13th out of 23 OECD countries, with the ranking for Indigenous Australians being even more concerning, at 23rd (ARACY, 2008). With respect to Australian children and adolescents, Sawyer and colleagues (Sawyer, Arney, Baghurst, & al., 2001; Sawyer, Miller-Lewis, & Clark, 2007; Sawyer, Sarris, Baghurst, Cornish, & Kalucy, 1990) record the prevalence of mental health disorders at approximately 13 to 21 per cent, according to self-report or parent/caregiver information.

One system-wide response to concern about population mental health status has been to implement mental health promotion and early intervention programs in schools, with a view to specifically targeting the mental health of children. Compelling reasons are advanced to support schools as settings for health promotion initiatives. For example, Spieldenner (WHO Europe, 2006) argued that school settings provide the entry gate to "get to" young people; allow long-term interventions to be planned, devised and implemented; and provide health promotion role models such as teachers and other students. One such initiative in the field of mental health in Australia is the KidsMatter Initiative.

The KidsMatter Initiative

The KidsMatter pilot Initiative (KMI) was developed in collaboration with the Australian Psychological Society, the Australian Principals Association Professional Development Council (APAPDC), *beyondblue: the national depression initiative* and the Australian Government Department of Health and Ageing, and has been supported by the Australian Rotary Health Research Fund. The overall purpose of the KMI was to develop and trial a multi-faceted intervention to improve mental health outcomes for primary school students. The pilot program involved school-based interventions in four areas, namely: 1) A positive school community, 2) Social and emotional learning programs for all students, 3) Parenting education and support, and 4) Early intervention for students at risk of experiencing mental health difficulties. Intervention in these four areas is predicted to strengthen protective factors, and weaken risk factors that reside in (a) the school context, (b) the family context and (c) the psychological world of the child.

The KMI was implemented in 100 primary schools in Australia. It began with preliminary work with teachers and leadership staff in the latter half of 2006, was implemented in 50, Round 1 schools in 2007, and extended to the remaining 50, Round 2 schools in 2008.

A key feature of the KMI is an evaluation program designed to gather quantitative and qualitative data from teachers and parents/caregivers across the two-years of the KidsMatter pilot phase. The evaluation program measured a range of features of the KMI, including parents/caregivers' and teachers' reports of student mental health status, teachers' self-reported knowledge and competence in delivering social and emotional learning programs, KMI Project Officers' reports of events related to implementation, stakeholders' engagement with the KMI, students' perspectives, and school leaders' reports. In this chapter we focus upon the first area of data collection, namely students' mental health status. We present the three measures of students' mental health status that we gathered as part of the evaluation of the KMI: 1) the Strengths and Difficulties Questionnaire (SDQ, Goodman 2005); 2) the Flinders Student Competencies Scale (SCS), which is a purpose built set of items based upon social and emotional constructs identified in the literature; and 3) school teacher and leadership staff identifications of students in their school considered to be 'at risk' of social, emotional and behavioural difficulties. We employ correspondence analysis to identify relationships between the three measures of mental health status employed in the KMI evaluation.

Conceptualising and measuring mental health

Mental health can be conceptualised as consisting of two domains, namely a) the absence of dysfunction (impairment) in psychological, emotional, behavioural and social spheres, and b) the presence of optimal functioning in psychological and social domains (Kazdin, 1993; Roeser, Eccles, & Strobel, 1998). As represented in a recent World Health Organisation report, "Mental Health is not simply the absence of mental disorder or illness, but also includes a positive state of mental well-being" (WHO/AFRO, 2005).

A commonly used instrument for measuring child and adolescent Mental Health status is the Strengths and Difficulties Questionnaire (Goodman, 2005). The SDQ operationalises the difficulties' subscales in order to calculate a total mental health difficulties score (the pro-social scale of the SDQ is excluded in calculating the total mental health score). Thus, in most investigations of child and adolescent mental health, the SDQ attends to one conceptual dimension of mental health – mental health difficulties.

To provide a broader measure of mental health, we created a scale that specifically measured positive expressions of mental health. We reviewed the literature (e.g. Konu & Rimpela, 2002; Krosnick, 1999; Levitt et al., 2007) and developed the Flinders Student Competencies Scale (SCS). The items in the SCS were submitted for iterative feedback to our clients, and trialled with teachers and parents at local schools (not KidsMatter schools).

We also determined that teachers and leadership staff, as a result of their professional training and school-based experience, were in a position to identify students exhibiting social, emotional and behavioural difficulties. These observable difficulties might operate as indicators of students who could be 'at risk' of developing mental health difficulties. It was not intended that such teacher/staff judgements should take the place of more formal clinical assessments, however, it seemed reasonable to recognise the role that teachers and leadership staff play in the early identification of students exhibiting signs of social, emotional or behavioural difficulties. Therefore, as a third measure of student mental health, we asked teachers and leadership staff in KidsMatter schools to nominate students in their school who were 'at risk' of exhibiting social, emotional or behavioural difficulties.

In summary, three measures of students' mental health, as displayed in Table 1 were gathered for almost 5000 students enrolled in the 100 KidsMatter primary schools across Australia.

Table 1. Instrumentation

Measure	Completed by
Goodman's (2005) Strength and Difficulties Questionnaire (SDQ),	Parent/caregiver and teacher of (up to) 76 stratified, randomly selected students in each KMI school
Flinders Student Competency Scale (SCS)	Parent/caregiver and teacher of (up to) 76 stratified, randomly selected students in each KMI school
Non-clinical assessment of students at risk of social, emotional or behavioural difficulties	teachers and leadership staff in each KMI school

The three sources of data permit us to address the following research question:

What are the proximities between teachers' and parent/caregivers' ratings of students' mental health status using the SDQ, the SCS, and teacher/leadership staff non-clinical assessment of students 'at risk'?

Method

Ethics

We submitted all data collection instruments, including the SDQ, the Flinders SCS, collection of demographic data including the teacher/staff "at-risk" nomination, and other questionnaire items to the Flinders University Social and Behavioural Research Ethics committee and also to the relevant ethics committee in each school jurisdiction in each Australian State (30 ethics applications in total). In particular, we outlined stringent procedures for de-identification of data, with clear separation of student enrolment lists from student ID codes and data. The goodwill demonstrated by each of the ethics jurisdictions to support research into the national KidsMatter project greatly facilitated the efficient processing and approval of the extensive ethics applications.

The KidsMatter schools

A request for expressions of interest to take part in the KMI (pilot phase) was sent to all Australian primary schools. One hundred schools for the pilot phase were selected from the large pool of applicants through negotiation between the funding bodies, the Australian Principals Association Professional Development Council, and the evaluation team located at Flinders University. The aim of the selection process was to ensure a diverse, representative sample based on the schools' State, location (metropolitan, rural or remote), size, and sector type.

The selected schools ranged in size from 11 students with one staff member to 1085 students with 100 staff. There were individual school populations that had no students with English as a Second Language (ESL), through to populations with 94 per cent ESL. Some schools had no Aboriginal or Torres Straight Island (ATSI) students, and some had more than 75 per cent ATSI students.

The students

Based on school enrolment lists, up to 50 students aged 10 years (some schools had fewer than 50, 10-year-old students enrolled) were randomly selected from each of the 100 KidsMatter schools. A stratified random sampling procedure, based on gender and age, was developed to select equal numbers of male and female students, turning 10 years of age in 2007, to participate in the evaluation. The 10 year-old students were targeted as it was not clear at the beginning of the KMI whether all schools would be able to roll out the KMI to all year levels in the first instance. It was agreed that if a staged roll out was required, due to school resourcing, structural or other issues, then the first group to be targeted by the KMI would be the 10 year-old students. Hence these students were the first focus of evaluation. In addition, up to 26 additional students of all ages were selected from each school to ensure, through over-sampling, that students in identified subgroups of interest were included in the evaluation.

Phase One (of four phases) of the data collection is reported in this chapter, for which responses were received from the parents/caregivers and teachers of 4890 students (70% return rate). Since the KidsMatter schools formed a sample of convenience through a process of application and selection, rather than a random sample, they are not representative of other primary schools in Australia and therefore school or state weights are not applied to the analysis of the data reported herein.

Instrumentation: The parent/caregiver and teacher questionnaires

We designed two extensive questionnaires, one for each of the targeted students' parents/caregivers and the other for each of the students' teachers. The questionnaires contained items about school, family and child factors, along with the outcome measures of student mental health. Items were trialled with parent/caregiver and teacher groups. Both questionnaires contained a set of items to make up the Flinders SCS, which is the first data source of interest in this paper. These items were sourced from the five core groups of social and emotional competencies recommended by CASEL (2006), from a search of relevant literature (Levitt et al., 2007), from discussion with our clients, from our own research and practical experiences with schooling, families, and student wellbeing (Askell-Williams, Lawson, & Murray-Harvey, 2007; Russell, Hart, Robinson, & Olsen, 2003). The five core groups of indicators of students' social and emotional competencies identified by CASEL are self-awareness, self-management, social awareness, relationship skills, and responsible decision making. Items were developed to ask parents/caregivers and teachers to reveal the extent to which children were developing 10 indicators of competencies for mental health. These 10 items were presented as attitudinal or belief statements and required participants to respond using a seven-point Likert scale of Strongly Disagree (1) to Strongly Agree (7).

The second data source was Goodman's (2005) Strengths and Difficulties scale, parent/caregiver or teacher informant for 11 to 17 year old youths[2]. The SDQ is a brief behavioural screening questionnaire that examines five attributes, four of which are used as measures of mental health status (Levitt, et al. 2007). The four attributes relating to student mental health are measured using 20 items requiring responses on a three-point Not True (0), Somewhat True (1) and Certainly True (2) scale. The four attributes examine Hyperactivity, Emotional Symptoms, Conduct Problems, and Peer Problems.

The third measure was the school-staff identifications of students in their school who appeared to be 'at risk' of experiencing, emotional, social or behavioural difficulties. This 'at risk' identification formed a dichotomous variable, scored as 'not at risk' (0) and 'at risk' (1), and was a non-clinical assessment based upon teacher/staff professional judgement and classroom and school-based experiences with their students.

Preparation of the SCS and SDQ scales

Questionnaire data extraction and compilation was undertaken using Remark Office OMR Version 6 scanning software (Principia 2005). The statistical analyses conducted for this paper were undertaken with the use of the software program SPSS for Windows (SPSS 2001).

Results

Treatment of missing data

Following the scanning and cleaning of the parent/caregiver and teacher questionnaires, analysis of missing data revealed an acceptable range below 20 per cent across the scales. For the SDQ, and in accordance with Goodman's (2005) recommendation, when at least three of the five SDQ items in a scale were completed, then the remaining two scores were replaced by their mean. When more than three items were missing in an SDQ scale, then the missing items were not replaced. For items in the Flinders SCS, missing values were not

[2] Note that over the two years of the evaluation, the targeted 10 year-old students would turn 11. We considered whether to use the 11 to 17 youth version, or the child version of the SDQ, and decided on the Youth version. The wording of the two forms is very similar.

replaced. Missing values for the binary (yes/no) teacher/staff nominations were not replaced. Cases with missing data were excluded from the final analyses, resulting in a data file of 3921 cases.

Validity and normality

Preliminary descriptive analysis of the baseline data was undertaken in order to ascertain the distributional characteristics of the SCS and SDQ. It is clear from the Australian statistics on the prevalence of mental health difficulties that we could predict that most students would be assessed as relatively low on the difficulties scales on the SDQ, and relatively high on the Flinders SCS. This turned out to be the case, with the resulting distributions for the two scales exhibiting high skewness and kurtosis. These distributional characteristics argued against the use of parametric statistical techniques.

We used confirmatory factor analysis for asymptotically distribution-free data (CFA-ADF) using AMOS (in SPSS) to determine the factor structure of the SCS and SDQ (Garson, 2009; Tabachnick & Fidell, 2001).

Item inspection and CFA-ADF supported the homogeneity of the 10 items comprising the SCS to form a single measure of student competencies. Thus, an SCS variable was constructed by summing the 10 items. CFA-ADF showed the factor structure of the SDQ to be generally similar to that discussed by Mellor (2005) with an Australian sample. Thus we used the four subscales (excluding the pro-social subscale) of the SDQ intact as a total measure of mental health difficulties, as intended by Goodman (2005) (see CAEF, in preparation, for further details about the CFA-ADF of the SCS and SDQ).

Descriptive statistics

The Flinders Student Competency Scale (SCS)

Each individual child's social, emotional and behavioural competencies were measured by the 10 items that resolved into one distinct factor described as Student Competencies. Figure 1 displays the frequency of parents/caregivers' and teachers' responses to the Flinders SCS scale (0 – 70), categorised into eight equal categories, and clearly shows a highly skewed distribution. The overall positive response to these statements, as shown in Figure 1, suggests that parents/caregivers and teachers generally feel that students in the KidsMatter schools have, on average, relatively well-developed competencies for positive mental health.

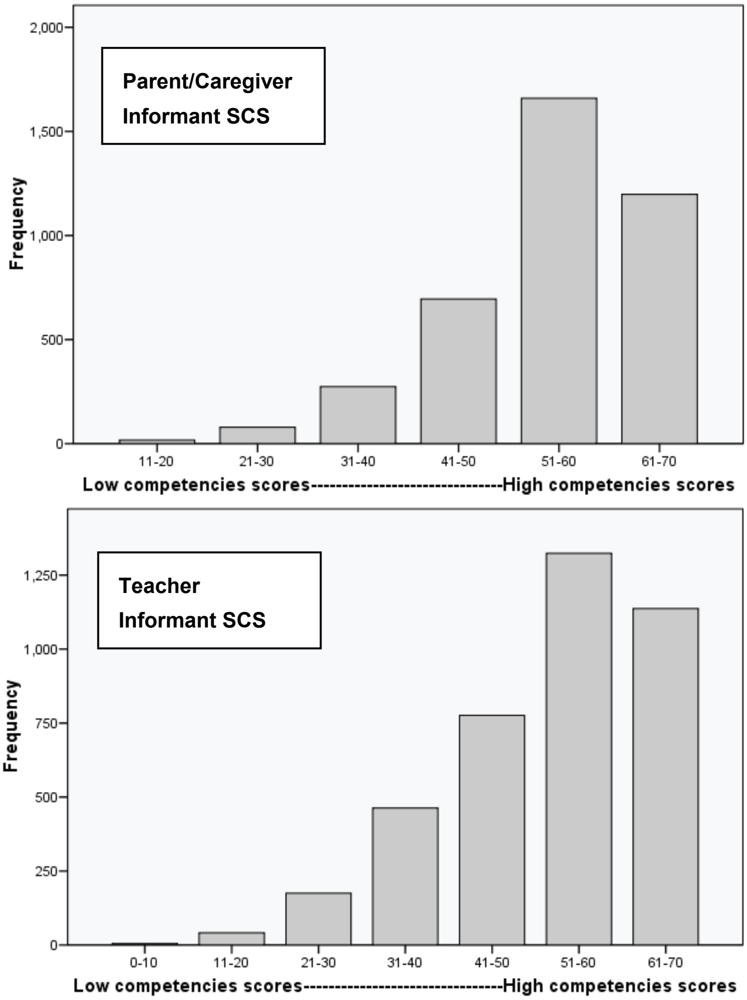


Figure 1. Distribution of SCS scores according to parents/caregivers and teachers

The Strength and Difficulties Questionnaire (SDQ)

Figure 2 displays the frequency of parents/caregivers and teachers responses to the SDQ. The profiles show a highly skewed distribution, but this time, in the reverse direction (compared to the SCS, as expected). The nature of the items and the scale-point labels for the SDQ suggest that this distribution might also be left-truncated. Within these limitations, the response to these statements, as shown in Figure 2, does suggest that both parents/caregivers and teachers generally feel that, on average, students in the KidsMatter schools are not experiencing extreme levels of mental health difficulties.

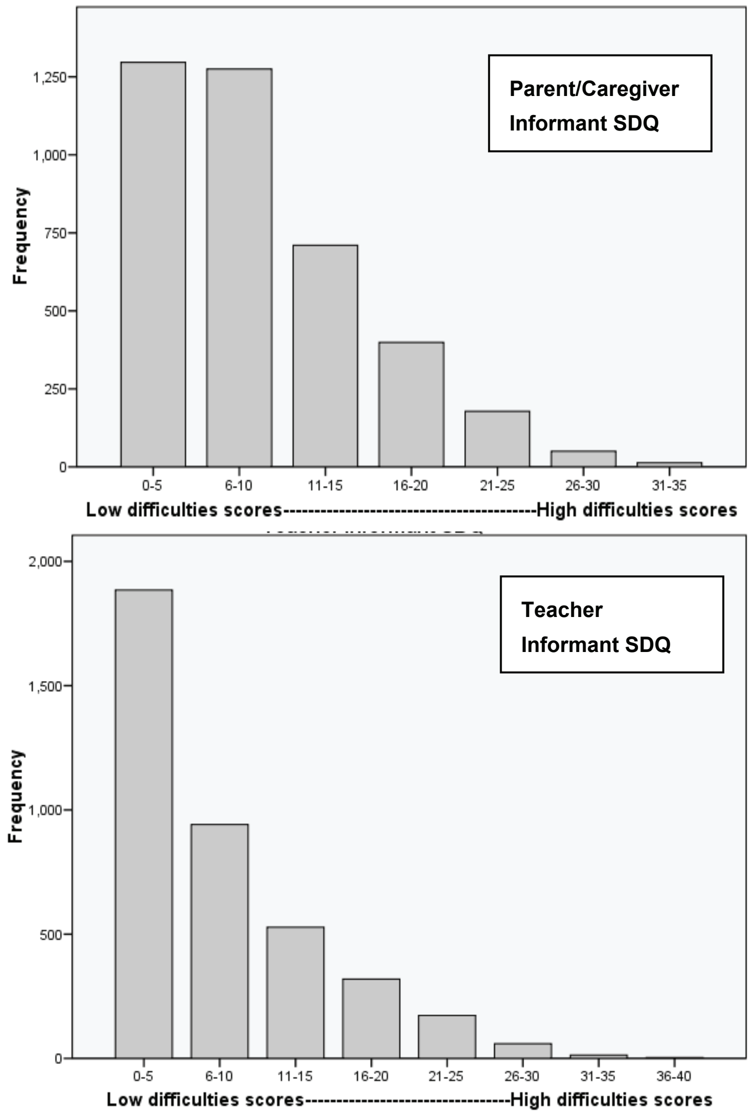


Figure 2. Distribution of SDQ scores (0=normal, 40=abnormal)

Teacher nominated students 'at risk' status

The final measure used in this study required teachers and school leadership staff to identify (using a de-identifying coding scheme) students in their school who were 'at risk' of experiencing, emotional, social or behavioural difficulties, based upon their professional judgement and school-based interaction with their students. This 'at risk' identification formed a dichotomous variable, scored as 'not at risk' (0) and 'at risk' (1).

Interestingly, Figure 3 shows the higher proportion of boys nominated as being 'at risk' in comparison to girls. The odds ratio (boys to girls) of 0.706 showed that boys were less likely than girls to be nominated "not at risk."

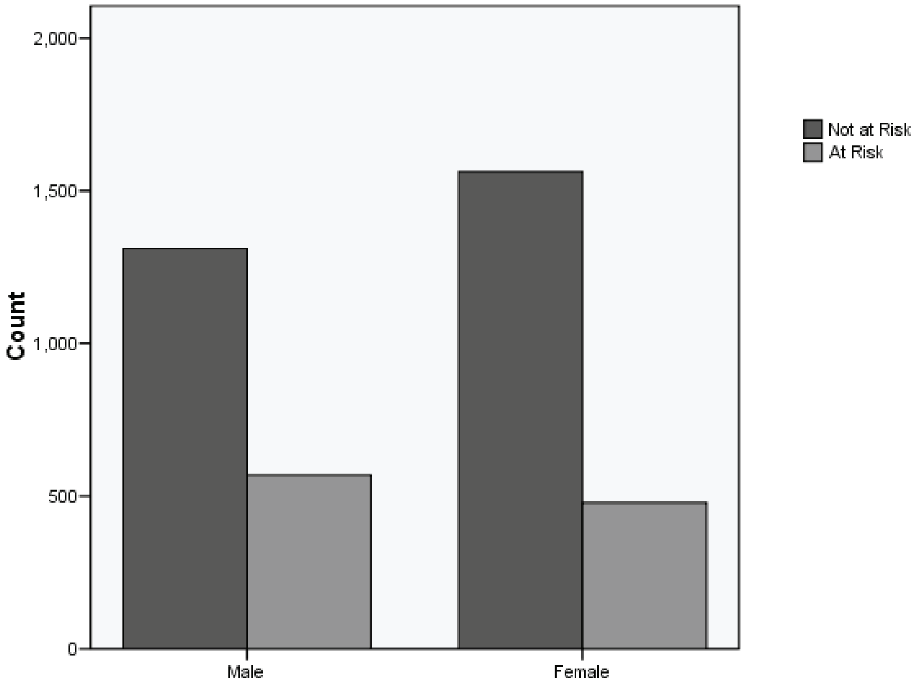


Figure 3. Teacher nominations of students' 'at-risk' status

Correspondence Analysis

Correspondence Analysis aims to find a low-dimensional representation of the dependence between predetermined categories in a two-way contingency table (Hair, Anderson, Tatham, & Black, 1995; van der Heijden & de Leeuw, 1985) Conceptually, Correspondence Analysis is similar to Principal Components Analysis. However, Correspondence Analysis has a substantial practical advantage for researchers in the social sciences, in that it can be used to analyse data that may not meet the distributional restrictions on data necessary for other statistical analyses (such as normally distributed data). Correspondence Analysis can work with count data (Greenacre, 1984; Nishisato, 1994; Weller & Romney, 1990) to produce low-dimensional solutions in graphical displays, which can provide comparisons between participants, between variables, and between participants and variables in their relative placement in shared low-dimensional space. In this study correspondence analysis is used to illustrate the relative proximity in perceptual space between teachers' and parents/caregivers' assessments of students' mental health. Correspondence Analysis is also used to identify proximities between the different indicators of mental health status employed in this study.

The first step in interpreting the Correspondence Analysis solution is to assess the significance of the Chi square statistic as a indicator of the viability of the analysis. In this case this value is highly significant, X^{22} (168) = 7183.201 $p < 0.001$. Table 2 contains information about the singular values, inertia and proportion of variance (within the variance explained by the solution) for each dimension.

Table 2. Summary statistics for Correspondence Analysis of SDQ, SCS, Risk

Dimension	Singular Value	Inertia	Proportion of Inertia Accounted for	Cumulative Proportion
1	0.535	0.286	0.625	0.625
2	0.277	0.077	0.168	0.793
3	0.196	0.038	0.083	0.877

The singular value indicates the relative contribution of each dimension to an explanation of the variance accounted for by the solution. The singular values can be interpreted as the correlation between the rows and columns of the Correspondence Analysis contingency table, and are analogous to the Pearson correlation coefficient. As in Principal Components Analysis, the first dimension explains as much variance as possible, the second dimension is orthogonal to the first and explains as much of the remaining variance as possible, and so on for higher dimensions. Hair et al. (1995) recommended that singular values of greater than 0.2 be included in the interpretation of the Correspondence solution. Thus, the Correspondence Analysis here provides a two dimensional solution to be interpreted, with Dimension 1 accounting for 63 per cent of the variance in the solution, and Dimension 2 accounting for 17 per cent of the variation accounted for by the solution.

Table 3 contains the indices of fit for all variables in the Correspondence Analysis solution (as well as long and short labels for the variables in the solution), displayed in Figure 4. Fit is determined by the proportion of variance in each variable accounted for by the dimension and is measured by the squared correlation (Clausen, 1998).

In a low dimensional solution, the squared correlations provide an index of fit of the representation of each variable (point) in the solution, which is equivalent to communalities in a principal components analysis (Clausen, 1998). Fit ranged from 0.40 to 0.98. Hair et al. (1995) recommend a cut off point of 0.5 to assess the fit of the

Correspondence Analysis solution to each variable. For the present solution, we adopted a slightly more generous cut off point of 0.4 to assess the fit of each variable, with a view to retaining all variables in the analysis for further interpretation.

Table 3. Correspondence Analysis Fit indices

	SHORT LABELS	FIT DIMENSION 1	FIT DIMENSION 2	TOTAL FIT
PARENT				
SDQ SCORE 0 to 5 (Low difficulties)	PSDQ5	0.76	0.09	0.85
SDQ SCORE 6 to 10	PSDQ10	0.45	0.11	0.56
SDQ SCORE 11 to 15	PSDQ15	0.47	0.32	0.79
SDQ SCORE 16 to 20	PSDQ20	0.69	0.07	0.77
SDQ SCORE 21 to 25	PSDQ25	0.77	0.09	0.86
SDQ SCORE 26 to 30	PSDQ30	0.54	0.14	0.68
SDQ SCORE 31 to 35	PSDQ35	0.36	0.18	0.54
SDQ SCORE 35 to 40 (High difficulties)	PSDQ40#			
TEACHER				
SDQ SCORE 0 to 5 (Low difficulties)	TSDQ5	0.80	0.14	0.94
SDQ SCORE 6 to 10	TSDQ10	0.00	0.46	0.46
SDQ SCORE 11 to 15	TSDQ15	0.41	0.47	0.89
SDQ SCORE 16 to 20	TSDQ20	0.75	0.05	0.80
SDQ SCORE 21 to 25	TSDQ25	0.77	0.08	0.84
SDQ SCORE 26 to 30	TSDQ30	0.62	0.25	0.88
SDQ SCORE 31 to 35	TSDQ35	0.35	0.37	0.72
SDQ SCORE 35 to 40 (High difficulties)	TSDQ40	0.15	0.25	0.40
PARENT				
SCS SCORE 0 to 10 (Low competency)	PSCS10#			
SCS SCORE 11 to 20	PSCS20	0.32	0.21	0.54
SCS SCORE 21 to 30	PSCS30	0.50	0.14	0.64
SCS SCORE 31 to 40	PSCS40	0.80	0.01	0.80
SCS SCORE 41 to 50	PSCS50	0.54	0.21	0.75
SCS SCORE 51 to 60	PSCS60	0.49	0.12	0.61
SCS SCORE 61 to 70 (High competency)	PSCS70	0.72	0.09	0.81
TEACHER				
SCS SCORE 0 to 10 (Low competency)	TSCS10	0.12	0.28	0.40
SCS SCORE 11 to 20	TSCS20	0.51	0.34	0.85
SCS SCORE 21 to 30	TSCS30	0.73	0.11	0.84
SCS SCORE 31 to 40	TSCS40	0.77	0.08	0.85
SCS SCORE 41 to 50	TSCS50	0.27	0.56	0.83
SCS SCORE 51 to 60	TSCS60	0.66	0.03	0.70
SCS SCORE 61 to 70 (High competency)	TSCS70	0.74	0.20	0.94
TEACHER NOMINATION				
At risk of social, emotional or behavioural difficulties	Not at Risk*	0.98	0.01	0.98
	At Risk*	0.98	0.01	0.98

no cases
*Supplementary variable - located in dimensional plane but not used to calculate dimensionality

Table 4 shows the relative contribution of each variable to the Correspondence Analysis solution. It can be seen that 18 of the 28 variables contribute more than an expected 4 per cent to the Dimension 1 solution. The most influential contributors to Dimension 1 are highlighted in orange, and include indicators both of difficulties and of capabilities.

The horseshoe shape (Guttman effect) (Greenacre, 1994) of the Correspondence Analysis solution, as displayed in Figure 4, is characteristic of ordered data (Weller & Romney, 1990). This characteristic occurs in the present case, as the categories of the SDQ and SCS are ordered from low to high and high to low, for the SDQ and SCS respectively.

Table 4. Contributions of variables to dimensions

	SHORT LABELS	CONTRIBUTION TO DIMENSION 1	CONTRIBUTION TO DIMENSION 2
PARENT			
SDQ SCORE 0 to 5 (Low difficulties)	PSDQ5	0.11	0.05
SDQ SCORE 6 to 10	PSDQ10	0.01	0.01
SDQ SCORE 11 to 15	PSDQ15	0.02	0.05
SDQ SCORE 16 to 20	PSDQ20	0.07	0.03
SDQ SCORE 21 to 25	PSDQ25	0.12	0.05
SDQ SCORE 26 to 30	PSDQ30	0.06	0.06
SDQ SCORE 31 to 35	PSDQ35	0.02	0.04
SDQ SCORE 35 to 40 (High difficulties)	PSDQ40#		
TEACHER			
SDQ SCORE 0 to 5 (Low difficulties)	TSDQ5	0.18	0.12
SDQ SCORE 6 to 10	TSDQ10	0.00	0.10
SDQ SCORE 11 to 15	TSDQ15	0.04	0.16
SDQ SCORE 16 to 20	TSDQ20	0.11	0.03
SDQ SCORE 21 to 25	TSDQ25	0.14	0.05
SDQ SCORE 26 to 30	TSDQ30	0.08	0.13
SDQ SCORE 31 to 35	TSDQ35	0.02	0.07
SDQ SCORE 35 to 40 (High difficulties)	TSDQ40	0.01	0.04
PARENT			
SCS SCORE 0 to 10 (Low competency)	PSCS10#		
SCS SCORE 11 to 20	PSCS20	0.02	0.04
SCS SCORE 21 to 30	PSCS30	0.09	0.10
SCS SCORE 31 to 40	PSCS40	0.12	0.00
SCS SCORE 41 to 50	PSCS50	0.04	0.06
SCS SCORE 51 to 60	PSCS60	0.01	0.01
SCS SCORE 61 to 70 (High competency)	PSCS70	0.08	0.04
TEACHER			
SCS SCORE 0 to 10 (Low competency)	TSCS10	0.00	0.03
SCS SCORE 11 to 20	TSCS20	0.08	0.19
SCS SCORE 21 to 30	TSCS30	0.17	0.10
SCS SCORE 31 to 40	TSCS40	0.16	0.06
SCS SCORE 41 to 50	TSCS50	0.03	0.20
SCS SCORE 51 to 60	TSCS60	0.04	0.01
SCS SCORE 61 to 70 (High competency)	TSCS70	0.16	0.16
TEACHER NOMINATION			
At risk of social,	Not at Risk*	0.00	0.00
emotional or	At Risk*	0.00	0.00
behavioural difficulties			

\# no cases

*Supplementary variable - located in dimensional plane but not used to calculate dimensionality

Note that distance from one SDQ point to another SDQ point, and distance from one SCS point to another SCS point can be interpreted. However, only relative positions from an SCS point to the group of SDQ points, and similarly, relative positions from an SDQ point to the group of SCS points, can be interpreted. In other words, the joint display of coordinates illustrates the relationship between a point from one set (e.g. SDQ) and all the points in the other set (e.g. SCS) (UNESCO, n.d.).

The clear pattern in Figure 4 of the low SDQ to high SDQ scores, relative to the high SCS to low SCS scores illustrates a predictable relationship between these two indicators of mental health. Thus the first point to be made about this Correspondence Analysis solution is the close proximity of each of the ordered categories of the SCS and the SDQ. Given the concerns raised about the item wording and scale labels used in the SDQ it is important to note that the SCS provides an alternative representation of the students' mental health status. This also suggests that Dimension 1 in this solution represents a dimension of mental health, positive on the left pole and negative on the right. As we move across this dimension from left to right we see a conceptually consistent placement of each of the SCS and SDQ scale categories.

A second noteworthy feature of the correspondence analysis solution is the high degree of concordance of the school staff's assessments and the parents/caregivers' assessment at each ordered category. For example, TSCS70 is in close proximity to PSCS70; TSDQ5 is in close proximity to PSDQ5, and so on for all categories of the SDQ and the SCS.

A third point of note in this analysis concerns the positioning of the school staff's non-clinical, but professional judgments about which students in their school were considered "at risk" or "not at risk" of social, emotional or behavioural difficulties. In the present Correspondence Analysis, "at-risk" status is treated as a supplementary variable, which means it is an additional variable that has a meaningful profile which the Correspondence Analysis projects onto the low-dimensional solution. The positions of the supplementary variables relative to the active variables can be interpreted (UNESCO, n.d.).

Figure 4 shows that the "Not at risk" supplementary variable is located at the expected positive pole of Dimension 1, that is, near high SCS and low SDQ. The supplementary variable "At risk" is located towards the middle of Dimension 1, suggesting that teachers are observing expressions of difficulties in students, but these difficulties are not necessarily only observed in those students assessed at the extreme right hand end of Dimension 1 (low SCS/High SDQ).

With regard to Dimension 2, at least two interpretations seem possible. The first is related to the relative contributions of the dimensions in explaining variance in the solution. The horseshoe effect observed in Figure 4 can indicate that the second dimension is a quadratic transformation of the first dimension, which can occur when the first dimension is dominant (Greenacre & Blasius, 1994). In such a case, the second dimension does not require substantive interpretation.

A second possibility is that this second dimension represents a set of judgments, similar for school staff and for parents/caregivers, that are not related to mental health. The similarity of the location of the 'at risk' and 'not at risk' judgments along the second (vertical) dimension supports such an interpretation. On this second dimension there is a close proximity, centered around the neutral point, between students who are rated at moderate levels of difficulty/competence and those who are rated as high, with the students rated as having most difficulty at the negative pole. At present we have no further information that would allow us to make progress in deciding whether this second interpretation is viable.

Conclusions

In this Chapter we have reported the use of the Correspondence Analysis procedure to permit investigation of count data that is characterised by features that would limit its analysis using other techniques. Correspondence Analysis can be used with the highly skewed, and possibly truncated, data that we have in this project. Correspondence Analysis is a well known technique in Europe, however its use in Australia has been very limited. We would like to acknowledge that our use of this technique has been supported and substantially influenced by the enthusiastic engagement with us of Professor John Keeves AM, engagement in consideration of both the broad data analysis issues and the specific details of the use of this technique. As has happened before, our problem also became his problem and he joined with us in learning about the new technique. We thank him for his collaboration and his teaching.

The analysis shows that the use of this technique can be quite rewarding. The technique allowed us to make meaningful comparisons among different measures of student mental health that were derived from conceptually distinct sources. We argued that mental health can be assessed in different ways, including an assessment of mental health difficulties, and an assessment of positive competencies for mental health that promote optimal functioning. The findings show quite similar patterns for the parallel categories of mental health status in the two scales. The Correspondence Analysis solution shows that the measurements of student mental health used in the KMI show predicted relationships within each scale and between scales. The solution also shows concurrence between different informants (parent/caregiver and teacher).

The findings also suggest that staff/teacher non-clinical professional judgments might provide a valuable source of information about students' mental health which has the potential to be useful for provoking strategies of early intervention. This is likely to be of importance for further practical work in the area of health promotion that is generated through schools.

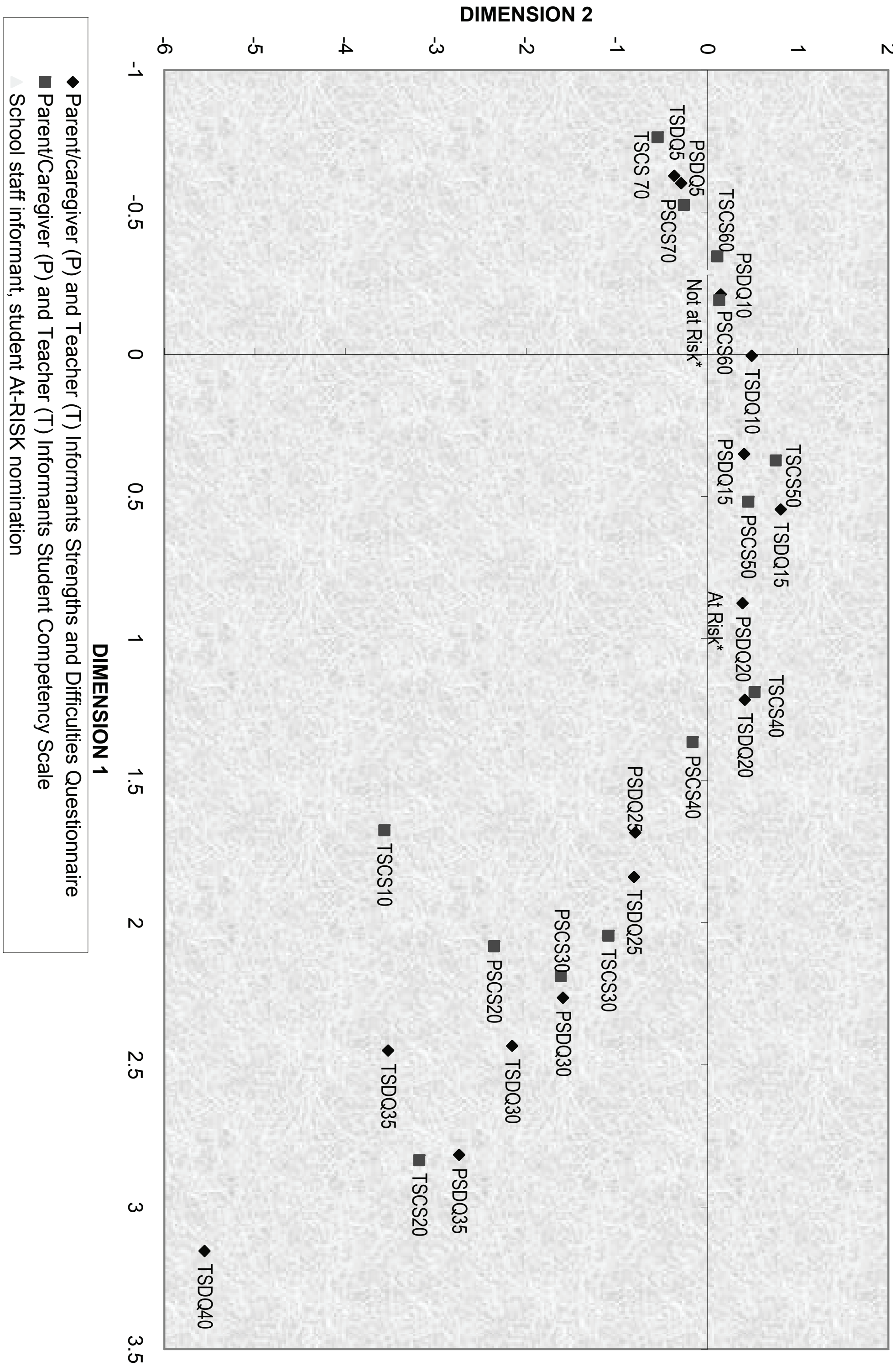


Figure 4. Two Dimensional plot of Correspondence Analysis solution

References

ARACY. (2008). Report card: The wellbeing of young Australians. Retrieved Nov 11th 2008, from http://www.aracy.org.au/AM/Common/pdf/report_card/report_card_A5_web.pdf

Askell-Williams, H., Lawson, M. J., & Murray-Harvey, R. (2007). Teaching and learning about mental illnesses: An Australian perspective. *International Journal of Mental Health Promotion, 9*(4), 26-36.

CAEF. (in preparation). *Evaluation of the KidsMatter Initiative*. Adelaide: Flinders University.

CASEL. (2006). What is social and emotional Learning: Skills and competencies. Retrieved 4th Aug 2008, from http://www.casel.org/basics/skills.php

Clausen, S.-E. (1998). Applied correspondence analysis: An introduction, *Sage University Papers Series on Quantitative Applications in the Social Sciences; 07- 121*. Thousand Oaks, CA: Sage.

Garson, G. D. (2009). Structural Equation Modeling. Retrieved March 30th 2009 http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm

Goodman, R. (2005). Strengths and Difficulties Questionnaire. Retrieved August 7th 2007 from http://www.sdqinfo.com/questionnaires/austral/aus10.pdf

Greenacre, M. (1984). Theory and applications of correspondence analysis. London: Academic Press.

Greenacre, M. (1994). Correspondence analysis and its interpretation. In M. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 3-22). London: Academic Press.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Kazdin, A. E. (1993). Adolescent mental health: Prevention and treatment. *American Psychologist, 48*, 127-141.

Konu, A., & Rimpela, M. (2002). Well-being in schools: A conceptual model. *Health Promotion International, 17*(1), 79-87.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology, 50,* 537-567.

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*(2), 163-191.

Mellor, D. (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist, 40* 215-222.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Hillsdale, NJ: Erlbaum.

Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health. *Educational Psychologist, 33*(4), 153-176.

Russell, A., Hart, C., Robinson, C., & Olsen, S. (2003). Children's sociable and aggressive behavior with peers: A comparison of the US and

Australia, and contributions of temperament and parenting styles. *International Journal of Behavioral Development, 27* 74-86.

Sawyer, M. G., Arney, F., Baghurst, P., & al., e. (2001). The mental health of young people in Australia: key findings from the child and adolescent component of the national survey of mental health and well-being. *Australia and New Zealand Journal of Psychiatry, 35*, 806–814.

Sawyer, M. G., Miller-Lewis, L. R., & Clark, J. J. (2007). The Mental Health of 13–17 Year-Olds in Australia: Findings from the National Survey of Mental Health and Well-Being. *Journal of Youth and Adolescence, 36*, 185–194.

Sawyer, M. G., Sarris, A., Baghurst, P., Cornish, C., & Kalucy, R. (1990). The prevalence of emotional and behaviour disorders and patterns of service utilisation in children and adolescents. *Australia and New Zealand Journal of Psychiatry, 24*, 323-330.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

UNESCO. (n.d.). Correspondence Analysis. Retrieved Feb 6th 2009, from http://www.unesco.org/webworld/idams/advguide/Chapt6_ 5.htm

van der Heijden, P. G. M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika, 50*(4), 429-447.

Weller, S. C., & Romney, A. K. (1990). *Metric Scaling: Correspondence Analysis.* Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-075. Newbury Park, CA: Sage.

WHO Europe. (2006, 8–11 June 2006). *Opening Address.* Paper presented at the Fifth workshop on practice of evaluation of the Health Promoting School - concepts, indicators and evidence, Sigriswil, Switzerland.

WHO/AFRO. (2005). Mental Health WHO/AFRO Overview. Retrieved Feb 6th 2009, from http://www.afro.who.int/ mentalhealth/overview.html

10

Identifying ranges of student mental health using Latent Class Analysis

Katherine L. Dix
Centre for the Analysis of Educational Futures, Flinders University

Introduction

The prevalence of child and adolescent mental health disorders over the last 50 years has been increasing (Maras and Kutnick, 1999; Zubrick, et al., 2000) to the extent that it is now of major concern in Australia (Richardson and Prior, 2005; Stanley, 2007; Sawyer, 2004; Sawyer et al., 2000). Reports from Sawyer and colleagues (Sawyer, Arney et al. 2001; Sawyer, Miller-Lewis et al. 2007; Sawyer, Sarris et al. 1990) record the prevalence of mental health disorders for Australian children and adolescents at approximately 13 to 21 per cent, according to self-report or parent information. In order to counter this trend, national initiatives, such as KidsMatter[1] (Slee et al., 2009), have been trialled with the intention of improving the mental health outcomes for Australian primary school children.

[1] KidsMatter (2006) is a primary school mental health promotion, prevention and early intervention initiative and is designed as a whole-school approach aimed to (a) improve the mental health and well-being of primary school students, (b) reduce mental health problems among students, and (c) achieve greater support and assistance for students experiencing mental health problems. KidsMatter was developed in collaboration with the Commonwealth Government Department of Health and Ageing, *beyondblue: the national depression initiative*, the Australian Psychological Society, Principals Australia and supported by the Australian Rotary Health Research Fund.

Effective evaluation of such initiatives and their subsequent regard by policy makers depends in part on the suitability of judgements made about students' mental health. Decisions about measuring student mental health were of central concern in designing the evaluation[2] of the KidsMatter Pilot Initiative and resulted in the use of multiple scales completed by multiple informants on multiple occasions in multiple settings (Askell-Williams et al., 2008; Dix et al., 2008; Gregory et al., 2008). The broad purpose of the study was to evaluate the effectiveness of KidsMatter Stage 1 Pilot Phase.

Teacher and parent assessments of students' mental health based upon data from the evaluation of KidsMatter were collected through the use of four different scales in 100 primary schools across Australia. A strength of the evaluation was that it did not rely upon just one instrument or one informant to assess student mental health. However, these multiple instruments from multiple informants provided a challenge when it came to classifying each child's mental health status at the beginning of the two year pilot study. While there were children for whom parents and teachers rated consistently across the four scales and in agreement with each other, there were other children for whom the reports from parents and teachers were contradictory. One simple explanation for such contradictions might be that the child acted differently at school than they did at home. An important challenge of the evaluation was to determine how best to bring these various assessments together in order to identify the appropriate mental health status of each child. A technique called Latent Class Analysis was considered to be the best approach and is the focus of this chapter.

Assessing student mental health

Four different measures of mental health were developed and administered to parents and teachers of 4970 mainly 10 year-old primary school students in 100 KidsMatter schools. The first measure of mental health was Goodman's (2005) *Strength and Difficulties Questionnaire* (SDQ). The SDQ was developed as a brief mental health screening instrument and is widely used in many nations, including Australia (Levitt et al., 2007). Perceptions of the child's mental health difficulties, in terms of, hyperactivity, conduct problems, emotional symptoms and peer problems are combined to give a score ranging between normal (0) to abnormal (40).

[2] The data for this paper were gathered as part of the evaluation of KidsMatter and the author is grateful to members of the consortium based in the Centre for Analysis of Educational Futures at Flinders University for their expertise in designing and undertaking this large evaluation.

The other three scales provided assessments of the five core groups of indicators of students' social and emotional competencies identified by the Collaborative for Academic, Social and Emotional Learning (CASEL, 2006), namely, self-awareness, self-management, social awareness, relationship skills, and responsible decision making, as well as students' optimism and problem solving capabilities. The first instrument, called the *Child Social and Emotional Competencies* (SEC) scale, was purposefully designed to give a measure of child protective factors. Teacher and parent views about the child's ability to maintain positive relationships, solve problems, consider others, and make responsible decisions ranged between low competencies (strongly disagree=1) to high competencies (strongly agree=7). Although the SEC scale is strictly a measure of competencies rather than mental health, a canonical correlation of -0.91 between SEC and the SDQ suggests that it is an effective measure of positive mental health attributes.

The second measure of mental health was the specifically developed, *Mental Health Strengths* scale (MHS) and, like the SEC, scores ranged between low strengths (1) to high strengths (7). The MHS provided teacher and parent perceptions of the child's positive mental health in terms of optimism and coping skills. The final scale for student mental health was the *Mental Health Difficulties* scale (MHD). Like the SDQ, the MHD scale placed those with few difficulties at the low end (1) and those with many difficulties at the high end (7). The MHD scale provided teacher and parent perceptions of the child's mental health difficulties in terms of poor behaviour and anxiety.

Confirming the scales

Each scale was first subjected to confirmatory factor analysis. Factor analytic procedures have been widely used, especially in the behavioural sciences, to assess the construct validity of a scale (for example, see Dix, 2007). However, it was also found that many of the items exhibited a skewed distribution, so usual methods involving factor analysis and other traditional techniques were considered to be inappropriate. For example, the outliers evident in the SDQ can distort correlations and the variance-covariance matrix, so the outliers may also distort the factor analysis.

Accordingly, the determination of scale validity and reliability required the use of distribution-free techniques in preference to using transforms to normalise data. Given the non-normal distributions and the large sample size, it was seen as appropriate to use asymptotically distribution-free (ADF) estimation provided in AMOS (Arbuckle,

2007) for each structural equation model (Browne, 1984; Garson, 2009; Hox, 1998; Kline, 1998; Tabachnick and Fidell, 2001). Asymptotically distribution-free estimation does not assume multivariate normality, and for this reason it was preferred over maximum-likelihood estimation methods.

In the interests of simplicity, it was desirable that the item structures were the same for parent and teacher scales, so scores were averaged (N=4,970). This had the additional benefit of reducing missing data, for which completeness was required. The remaining missing values were then replaced using the non-parametric median (rather than mean) of nearby points as an approximation for a complex sample, where design effects arise due to the nested nature of students within schools.

The criteria for rejecting items were based on a significance level of 0.05 and a minimum loading of 0.3. The indices selected for goodness-of-fit were, the Root mean square error of approximation (RMSEA≤0.08), the Standardised root mean square residual (SRMR≤0.1), the Tucker-Lewis index (TLI≥0.9), and the Comparative fit index (CFI≥0.9). These indices perform better than other indices under non-normal distribution conditions and are less sensitive to sample size (Fan, et al., 1999; Marsh, et al., 1988; Schumacker and Lomax 2004).

Table 1 presents a summary of the confirmatory factor analysis (ADF) in terms of the goodness of fit indices and number of items.

Table 1. Summary of the confirmatory factor analysis (ADF), with goodness of fit indices, assessment of normality, and number of items

Mental Health Scales	RMSEA	SRMR	TLI	CFI	Mardia's coefficient	No of items
Social and Emotional Competencies (SEC)	0.10	0.056	0.57	0.72	44.67	7
Mental Health Difficulties (MHD)	0.06	0.035	0.87	0.93	28.33	3
Mental Health Strengths (MHS)	0.06	0.035	0.87	0.93	28.33	3
Strengths and Difficulties Questionnaire (SDQ)	0.05	0.062	0.61	0.67	125.01	20

In addition, analysis of normality is also summarised in Table 1, with the reporting of the multivariate kurtosis value, known as Mardia's coefficient. Values of 1.96 or less mean there is non-significant kurtosis. Values greater than 1.96 mean there is significant kurtosis, which means significant non-normality (Garson, 2009). In all cases, Mardia's coefficient was well above the cut-off, and confirmed the need for non-parametric methods.

The structural models for the Child Social and Emotional Competencies scale (SEC) scale, the Mental Health Strengths (MHS) and Mental Health Difficulties (MHD) scales, and the Strengths and Difficulties Questionnaire (SDQ), along with the shortened versions of the items that comprise them, are presented respectively in Figures 1, 2 and 3. The factor loadings in each figure are well above the 0.3 cut-off and indicating that the items meaningfully reflect the concepts being measured, further suggesting that there is good internal scale validity.

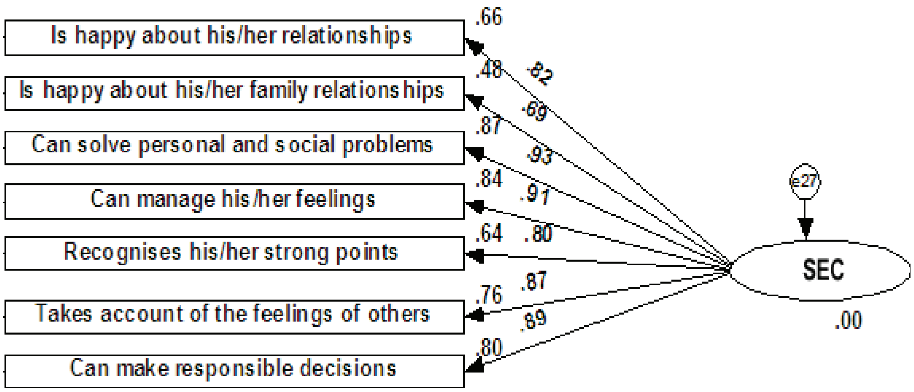


Figure 1. The Child Social and Emotional Competencies scale, showing for each item, the variance explained and the factor loading

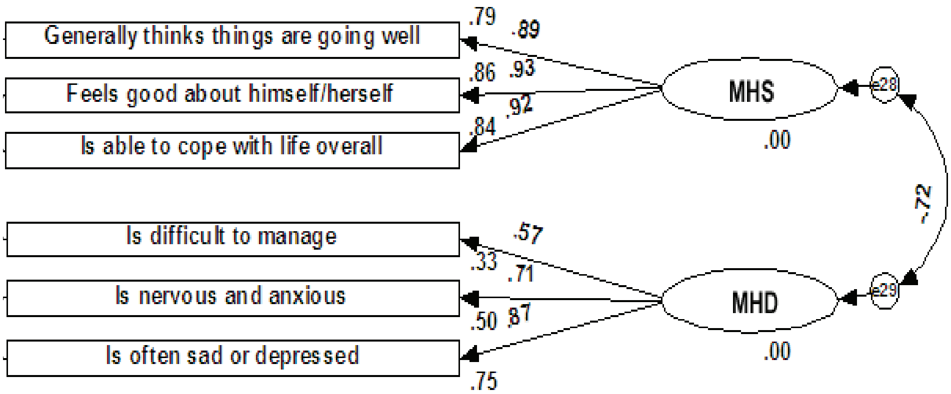


Figure 2. The Mental Health Strengths scale and Mental Health Difficulties scale, showing for each item, the variance explained, the factor loading and the correlation

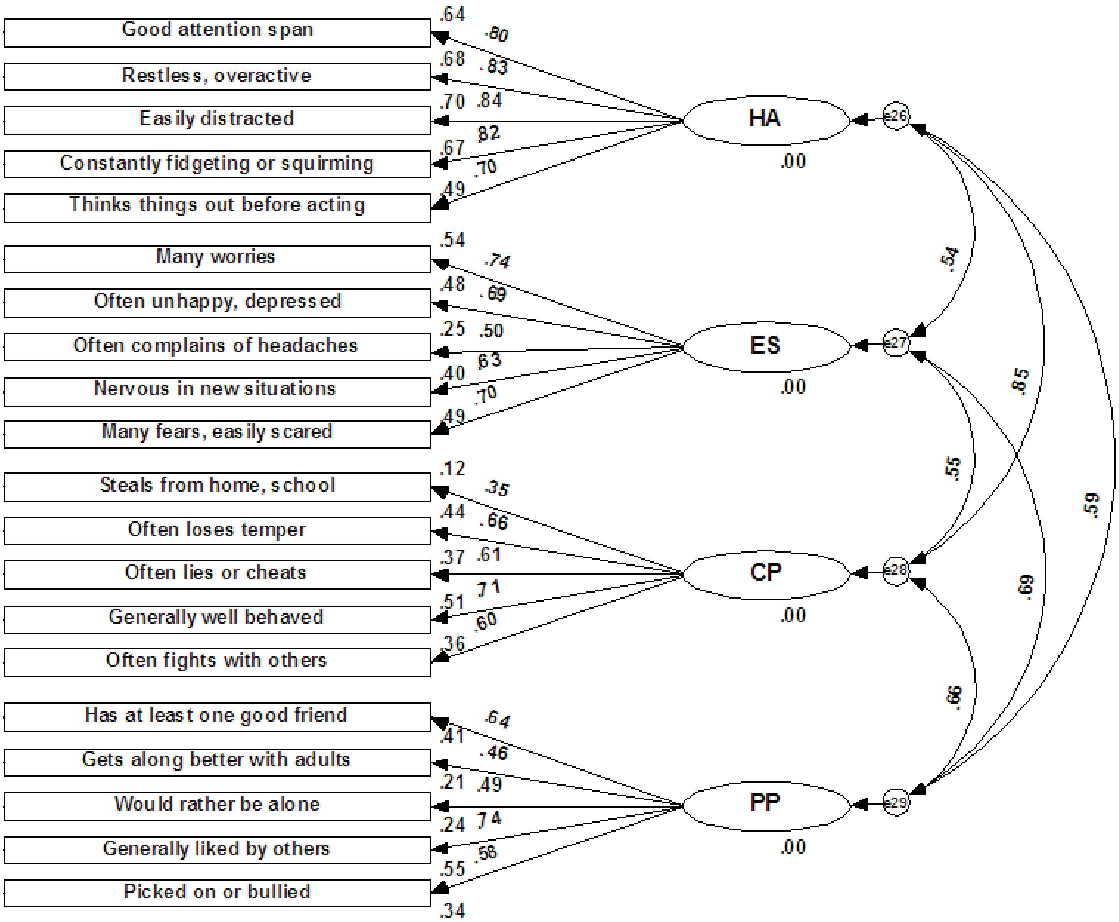


Figure 3. Total Strengths and Difficulties (SDQ) with subs-scales of hyperactivity (HA), emotional symptoms (ES), conduct problems (CP) and peer problems (PP), showing for each item, the variance explained, the factor loading and the correlations

Latent Class Analysis

Preliminary analysis of the SDQ and SEC were presented in an article by Dix et al. (2008), which concluded that there was reasonable agreement between parent and teacher ratings of the same child. Further evidence is presented in Figure 4 based on the canonical analysis (ADF) of parent and teacher reports for 4332 children[3], and shows a medium standardised correlation of 0.53 between parent and teacher ratings of the same child on the same scales.

[3] Some cases with missing data were removed from the ADF analysis, which requires complete data.

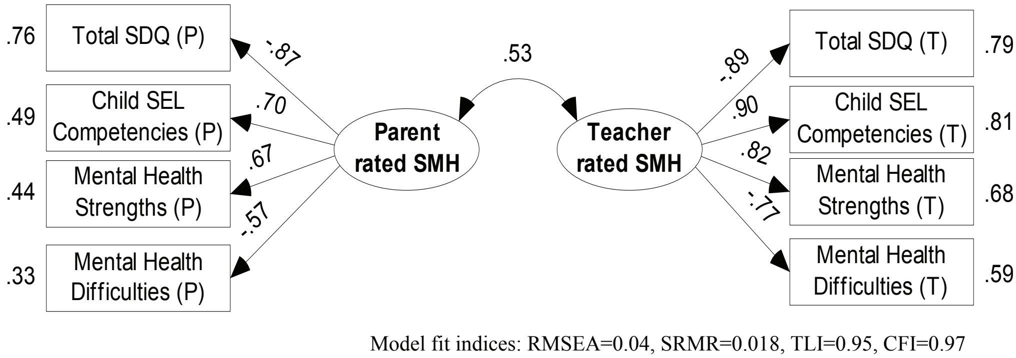


Figure 4. Canonical correlation of parent and teacher ratings of student mental health

Clearly, there were children for whom parents and teachers agreed, but there were also children for whom parent and teacher ratings were in contradiction between each other and between the different measures. The challenge, and the main focus of this discussion, was how then could a single score for student mental health be assigned to the student, particularly for those students where there was poor agreement between the multiple informants and multiple instruments?

One strategy found in the literature (Veenstra et al., 2008) was to select only students for whom there was agreement between parents and teachers. Based only on achieving agreement on the SDQ, this method would result in the unacceptable loss of 44 per cent of the data. Further losses in data would occur if the other measures of mental health were also applied in this way. Clearly, this was not a viable approach and alternative methods were considered.

Attention was drawn to the use of Latent Class Analysis (LCA), which is a statistical method for finding subtypes of related cases (latent classes) from multivariate categorical data, using the program MPlus version 5.2 (Muthén and Muthén, 2007). The benefits of using LCA, was its non-reliance on assumptions of normality, its ability to manage complex nested data and missing data, and because it has several advantages over conventional regression analyses that use total scores or cut-off scores (van Lier et al., 2003).

Preparation of the data was first necessary in order to change scale data into categorical data. Accordingly, Goodman's (1997) cut-point for parent and teacher rated SDQ ranges were differentially applied. For teachers, the cut-points were 'normal range' (0-11), 'borderline range' (12-15), and 'abnormal range' (16-40). For parents, the cut-points were more generous, with 'normal range' (0-13), 'borderline range' (14-16), and 'abnormal range' (17-40). The visual binning command in SPSS

(2001) was used. Appropriate cut-point ranges for the other measures of mental health (SEC, MHS, and MHD) were then determined by using the percentage of students in each of the normal, borderline and abnormal ranges. Once again, SPSS was used to undertake the analysis for each of the parent and teacher variables. Combined into this already complex categorisation process, the positively viewed scales (SEC and MHS) were reverse coded to align with the negatively viewed scales (SDQ and MHD). Accordingly, the resulting ranges for each of the four scales were labelled 'normal' (1), 'borderline' (2), and 'abnormal' (3).

The decision to use Goodman's labels for each category was not taken lightly and alternative labels that encompassed concepts of strengths and difficulties were considered. However, in the interests of avoiding confusion, alternative labels were not adopted, since the SDQ was being used and Goodman's cut-point percentages were being applied to the other three scales. The resulting data file contained the categorised variables from parents and teachers for each student at the beginning of the evaluation.

Preliminary latent class analysis in MPlus was conducted, taking into consideration missing data and clustering at the school level, using the four parent-rated and four teacher-rated scales of mental health. Three classes were requested and revealed that for one groups of students, exhibiting scores within the normal range, there was good separation of probability estimates and good agreement between parents and teachers. However, for the other two classes of students, the difference between raters was of more influence than the difference between students. One reason for this could be that students exhibit different behaviours at home than they do at school. A practical approach was taken and equivalent parent and teacher scales were averaged to form four combined measures.

Figure 5 presents the resulting probability estimates of the three classes for Time 1. The vertical axis can be interpreted as the probability of being in the 'normal' range. Accordingly, students with scores the 'abnormal' range have the lowest probability and are near the bottom of the graph. Conversely, students with scores in the 'normal' range have the highest probability of being considered to exhibit normal behaviours. Each probability graph also shows that combining parent and teacher data sets was highly effective in overcoming their differences, with good separation between the probabilities.

The MPlus input file was prepared and is presented here.

```
Title:
 Student Mental Health LCA: Occasion 1.
Data:
 File is SMH1.dat ;
Variable:
 names = schid sid SDQ SEC HMHS MHD;
 usevariables = sid SDQ SEC HMHS MHD;
 categorical = SDQc SDQ SEC HMHS MHD;
 missing are all(-1);
 cluster = schid;
 classes = class(3);
Analysis:
 Type=mixture;
 Type=missing;
 Type=complex;
Plot:
 type is plot3;
 series is SDQ (1) SEC (2) MHS (3) MHD (4);
Savedata:
 file is SMH1.txt ;
 save is cprob;
 format is free;
Output:
 tech11 tech14;
```

The most immediate interpretation of the probability typologies presented in Figure 5, is that the category of borderline is indicative of students 'at risk' of experiencing mental health problems, and not just an intermediate group half-way between the normal and abnormal groups, and it suggests that three groups are appropriate.

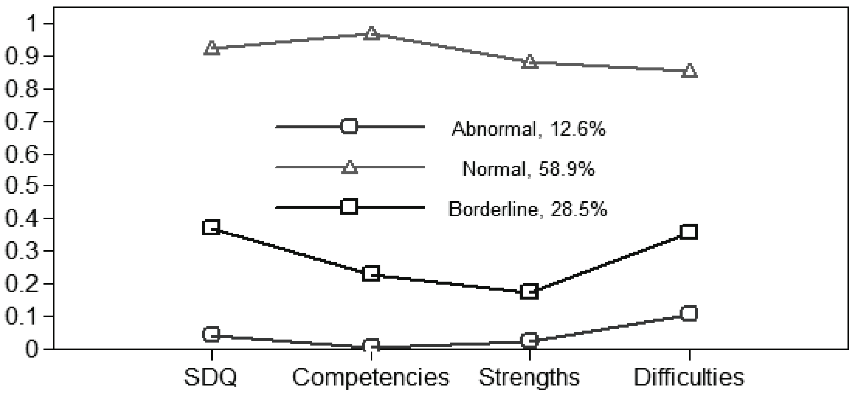


Figure 5. Latent class probabilities of being 'normal' on Time 1 (N=4970)

Although visually convincing, goodness of fit measures were considered in order to assess further whether the right number of classes was chosen. The Vuong-Lo-Mendell-Rubin test and the bootstrapped parametric likelihood ratio test were requested to compare the model with K classes (in this case 3 classes) to a model with K-1 classes (2 classes). The results are presented in Table 2 and show that on Time 1 the Vuong-Lo-Mendell-Rubin test had a probability-value of 0.058 and the Lo-Mendell-Rubin adjusted LRT test had a p-value of 0.059. If the probability-value is less than 0.05, the K model is superior and additional classes are added until the p-value for the statistic is greater than 0.05, at which point the previous model is accepted (Smith et al., Lo et al., 2001). The tests were marginal suggesting that two classes were possibly sufficient and that three classes might not really be needed. However, the bootstrapped parametric likelihood ratio test had a p-value of 0.000 and suggested that three classes were better than two classes. UCLA (2009) reported unpublished results that indicated the bootstrap method might be more reliable. Moreover, the three class model fitted the theoretical expectations. Accordingly, three classes were chosen and are identified in Figure 5 as normal, borderline and abnormal.

Table 2. Latent class analysis tests for number of mental health classes

Vuong-Lo-Mendell-Rubin Likelihood Ratio Test For 2 (H0) Versus 3 Classes	**Time 1**
H0 Loglikelihood Value	-61234.207
2 Times the Loglikelihood Difference	1180.335
Difference in the Number of Parameters	10
Mean	14.328
Standard Deviation	939.493
P-Value	0.0581
Lo-Mendell-Rubin Adjusted Lrt Test	
Value	1166.628
P-Value	0.0595
Parametric Bootstrapped Likelihood Ratio Test For 2 (H0) Versus 3 Classes	
P-Value	0.0000

From the output file produced from the latent class analysis, classes were appropriately coded and the file was saved. The reconstructed student data file contained, in addition to the student's identification number, the probability scores of being in each class, along with their final assigned group. This last variable achieved the definitive placement of each student into one category of mental health, be it normal, borderline or abnormal. It optimised the available data and overcame the challenge of contradictory reports from multiple informants on multiple scales of mental health. It did this for all students and also dealt with the problem of missing data by avoiding the need to impute.

Composite student mental health Status

Through the use of LCA a new composite measure of student mental health status was created, which could then be used in conjunction with other variables for subsequent more refined analyses.

One immediately useful outcome of the analysis, shown in Figure 5, was the percentage of students in each group. The classification using LCA of students in the groups of abnormal (13%), borderline (28%) and normal (59%) mental health status provides an initial examination of student mental health. The results for students considered to be abnormal come close those reported by Sawyer and colleagues, of 13 to 21 per cent, mentioned at the beginning of this chapter.

However, Goodman's SDQ was based on norms of only 10 per cent of the Australian child population being abnormal, and another 10 per cent being considered borderline (Mellor, 2005). In comparison, the sample of 4970 10 year-old children participating in this study and based on the averaged parent and teacher SDQ categories, resulted 10 per cent of children considered borderline, but 15 per cent considered abnormal. One further comparison, which is clearly to be avoided, considers the sample with cases of non-agreement between parents and teachers removed. Under these conditions, the normal group was grossly inflated with only 2 per cent considered borderline and 8 per cent considered abnormal. Table 3 summarises the comparisons and suggests that LCA inflates the borderline group but better accounts for the differences between parent and teacher reports on multiple measures.

Table 3. A comparison of frequencies of students in normal, borderline and abnormal categories according to different assessments of mental health

Mental Health Scales	Normal	Borderline	Abnormal
Goodman's SDQ Norms (2005)	80%	10%	10%
Sample SDQ with non-agreement removed	90.49%	1.76%	7.75%
Sample SDQ with parent and teacher averaged	75.95%	9.53%	14.51%
Student Mental Health Status using LCA	58.90%	28.50%	12.60%

Further comparison between the identification of students based only on the SDQ total difficulties (parent and teacher versions) and the composite measure of student mental health are profiled against each of the mental health scales in Figure 6. By doing so, a character profile of the normal, borderline and abnormal groups can be compared and judgements made about how effectively the profiles reflect theoretical expectations of each group.

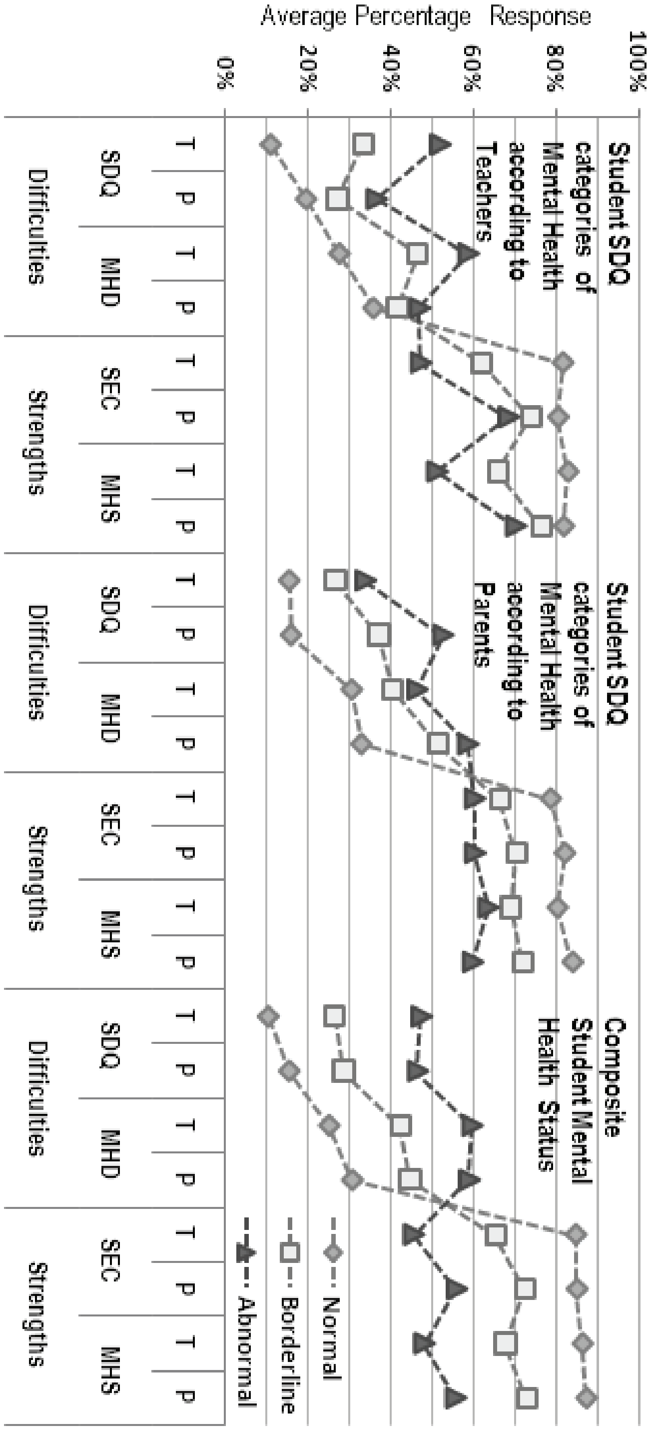


Figure 6. Profiles for the categories of mental health according to teacher and parent ratings on the SDQ only, compared with composite mental health status

One clear pattern across each method of categorisation, shown in Figure 6, is that students identified as being normal are reported as having the lowest difficulties and the highest strengths. Conversely, students identified as being abnormal tend to have the most difficulties and the fewest strengths. Those students identified as borderline, fall somewhere between the normal and abnormal groups. This general pattern matches our theoretical understanding of these groups. However, the agreement within measures provides a different perspective. While the profiles for the normal groups and, to a lesser extent, the borderline groups in each method of categorisation are similar, the profiles for the abnormal groups are markedly different. Those students identified as abnormal by parents and teachers based on SDQ difficulties alone, show characteristics of having more strengths than difficulties, similar in profile to the borderline groups. The separation between borderline and abnormal groups is less distinct for these SDQ-derived categories. In comparison, the general trend evident in the abnormal group, as defined by the composite measure of mental health, shown in the last set of Figure 6, presents a near-horizontal profile and achieves the greatest separation between the three groups. It would suggest that the composite measure better identifies students into their appropriate categories, since they reflect characteristics that more closely align to theoretical expectations. Accordingly, by using LCA and assessing student mental health status on dimensions of strengths as well as difficulties, in addition to bringing together multiple informant perspectives, results in the more accurate identification of students' mental health status. The importance of accurately assessing the number of children with mental health difficulties has ramifications for policy makers and the importance they place on initiatives such as KidsMatter. Potentially, it is the difference between whether 20 per cent of children stand to benefit from such initiatives, or twice that amount.

While the analysis presented in this chapter has resulted in a useful assessment of student mental health, there is need to further explore better ways to bring together multiple measures from multiple informants. For example, the use of categorical diagnostic constructs (normal, borderline, abnormal) resulted in loss of valuable information by regarding those who scored just below the diagnostic threshold as non-cases. Further analyses using MPlus will be undertaken that examine intact standardised scales at the subscale level to identify patterns in the differences between parent and teacher informants, in order to more accurately identify the underlying dimensions of students at risk of, or experiencing, mental health problems.

Acknowledgement

The author deeply acknowledges and wishes to thank John P. Keeves for his ongoing wisdom and brilliance of mind, in overseeing the statistical analysis presented in this chapter.

References

Arbuckle, J.L. (2007). *Amos User's Guide*. Chicago: SPSS.

Askell-Williams, H., Russell, A., Dix, K. L., Slee, P. T., Spears, B. A.., Lawson. M. J., Owens, L. D., and Gregory, K. (2008). Early challenges in evaluating the KidsMatter national mental health promotion initiative in Australian primary schools. *The International Journal of Mental Health Promotion,* 10, 35-44.

Browne, M.W. (1984), Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

CASEL. (2006). What is social and emotional Learning: Skills and competencies. [Online] http://www.casel.org/basics/skills.php [4 August 2008].

Dix, K.L. (2007) *Is School-Wide Adoption of ICT Change for the Better?* FUIIE Research Collection Number 17. Adelaide: Shannon Research.

Dix, K.L., Askell-Williams, H., Lawson, M.J. (2008). Different measures, different informants, same outcomes? Investigating multiple perspectives of primary school students' mental health. Paper presented at the Annual Conference of the Australian Association for Research in Education, Brisbane, December.

Fan, X., Thompson, B. and Wang, L. (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling,* 6, 56-83.

Garson, G.D. (2009). Structural Equation Modeling. *Statnotes: Topics in Multivariate Analysis.* [Online] http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm [5 May 2009].

Goodman, R. (1997). Scoring the informant-rated Strength and Difficulties Questionnaire. [Online] http://www.sdqinfo.com/ ScoreSheets/e1.pdf [17 March 2008].

Goodman, R. (2005). Strengths and Difficulties Questionnaire. [Online] http://www.sdqinfo.com/questionnaires/austral/aus10.pdf [7 August 2007].

Gregory, K., Lawson, M.J., Russel, A. and Dix, K.L. (2008) Issues in measuring student mental health. Paper presented at *From Margins to Mainstream: 5th World Conference on the Promotion of Mental Health and the Prevention of Mental and Behavioural Disorders,* 10-12 September, Melbourne. [Online] http://www.margins2mainstream.com/ppt-pdf/thu/Gregory_K_Thu_1030_ H2.ppt.pdf [23 March 2009].

Hox, J.J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar and M. Schader (Eds.). *Classification, Data Analysis, and Data Highways* (pp. 147-154). New York: Springer Verlag.

Kline, R.B. (1998). Principles and Practice of Structural Equation Modelling. New York: Guilford Press.

Levitt, J. M., Saka, N., Romanelli, L. H., and Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163-191.

Lo Y., Mendell, N.R., Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika,* 88, 767–78.

Maras, P. and Kutnick, P. (1999). Emotional and behavioural difficulties in schools: Consideration of relationships between theory and practice. *Social Psychology of Education*, 3, 135-153.

Marsh, H.W., Balla, J.R., and McDonald, R.P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin,* 103, 391-410.

Mellor, D. (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist*, 40, 215-222.

Muthén, L.K., and Muthén, B.O. (2007). *Mplus User's Guide* (5th Ed.). Los Angeles, CA: Muthén and Muthén.

Richardson, S. and Prior, M. (2005). What is to be done? In S. Richardson and M. Prior (Eds.). *No time to lose: The wellbeing of australia's children*. Victoria: Melbourne University Press.

Sawyer, M. (2004). The mental health and wellbeing of young people in Australia. In Proceedings of ACER Research Conference 2004: *Supporting Student Wellbeing*, Adelaide, South Australia. Camberwell, Victoria: Australian Council for Educational Research.

Sawyer, M. G., Arney, F., et al. (2001). The mental health of young people in Australia: Key findings from the child and adolescent component of the national survey of mental health and well-being. *Australia and New Zealand Journal of Psychiatry*, 35: 806–14.

Sawyer, M. G., Miller-Lewis, L. R., et al. (2007). The Mental Health of 13–17 Year-Olds in Australia: Findings from the National Survey of Mental Health and Well-Being. *Journal of Youth and Adolescence*, 36: 185–194.

Sawyer, M. G., Sarris, A., et al. (1990). The prevalence of emotional and behaviour disorders and patterns of service utilisation in children and adolescents. *Aust N Z J Psychiatry*, 24: 323-330.

Sawyer, M.G, Arney, F.M., Baghurst, P.A., Clark, J.J., Graetz, B.W., Kosky, R.J., Nurcombe, B, Patton, G.C., Prior, M.R., Raphael, B., Rey, J., Whaites, L.C. and Zubrick, S.R. (2000). *The Mental Health of Young People in Australia: The Child and Adolescent Component of the National Survey of Mental Health and Wellbeing*. Canberra, ACT: Australian Government Publishing.

Schumacker, R.E. and Lomax, R.G. (2004). *A Beginner's Guide to Structural Equation Modeling*, Second edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Slee, P.T., Lawson, M.J., Russell, A., Askell-Williams, Dix, K.L., H., Owens, L., Skrzypiec, G. and Spears, B. (2009) *KidsMatter Evaluation Final Report*. Adelaide: Flinders University.

Smith, G.W., Shevlin, M. (2008). Patterns of alcohol consumption and related behaviour in Great Britain: A latent class analysis of the alcohol use disorder identification test (AUDIT). Alcohol and Alcoholism Advance Access published May 22, 2008. [Online] http://alcalc.oxfordjournals.org/cgi/reprint/agn041v1 [1 July 2009].

SPSS (2001). *Statistical Package for the Social Sciences*. Chicago: SPSS.

Stanley, F.J. (2007). The importance of caring for children in Australian society, *Family Matters*, 76, 44-47.

Tabachnick, B. and Fidell, L.S. (2001). *Using Multivariate Statistics, Fourth Edition.* Needham Heights, MA: Allyn and Bacon.

UCLA (2009) *Mplus Data Analysis Examples: Latent Class Analysis*. UCLA: Academic Technology Services, Statistical Consulting Group. [Online] http://www.ats.ucla.edu/stat/mplus/dae/lca1.htm [26 June 2009].

van Lier, P.A.C., Verhulst, F.C. , and Crijnen A.A.M. (2003). Screening for disruptive behaviour syndromes in children: The application of latent class analyses and implications for prevention programs. *Journal of Consulting and Clinical Psychology*, 71, 353-363.

Veenstra, R., Lindenberg, S., Oldehinkel, A.J., de Winter, A.F., Verhulst, F.C., and Ormel, J. (2008). Prosocial and antisocial behavior in preadolescence: Teachers' and parents' perceptions of the behavior of girls and boys. *International Journal of Behavioral Development*, 32 (3), 243–251.

Zubrick, S.R., Silburn, S.R., Burton, P. and Blair, E. (2000). Mental health disorders in children and young people: Scope, cause and prevention. *Australian and New Zealand Journal of Psychiatry*, 34, 570-578.

11

Educational attainment and job status: The role of status inconsistency on occupational burnout

Lawrence J. Saha
The Australian National University

A. Gary Dworkin
The University of Houston

Introduction: The puzzling link between educational attainment and burnout

In a previous paper we examined the independent effects of education, work conditions and job stress on burnout in Australia (Saha and Dworkin, 2004). We found that both perceived job stress and role ambiguity in working conditions were positively related to burnout. We also found that the effect of educational attainment only significantly reduced burnout in our final analytical model for which a cluster of variables that we called "job context" had been added. These job context variables included characteristics of the respondent's employer, respondent's supervisory status and the respondent's social psychological factors of job stress and role ambiguity.

We have been puzzled by the unusual role played by education in that analysis, because at the bivariate level there appears to be no statistically significant relationship between educational attainment and burnout. Contrary to conventional wisdom, increasing levels of education do not seem to serve as a buffer against burnout, at least in the Australian data which we examined.

However, because of the unusual findings in our earlier paper, we decided to explore further what appears to be a complex relationship between educational attainment and burnout by taking into account the relationship between occupational status and education.

The relationship between occupational status and burnout is likely a complex one in and of itself. Burnout is often conceived as a malady of the workers in the human service professions, who themselves are most likely to be college educated. In a separate study, Dworkin (1987; 1997) found that education and burnout were not related in a sample of school teachers because of the lower variance in educational attainment among them, and therefore he deleted the variables from his analyses. So, what we wanted to know is why education became statistically related in the third of our analytical models in our analysis of a cross section of the Australian adult population. To understand our own previous finding, first we want to review how education and occupation are related to burnout, and then we will discuss how an old sociological and psychological concept, namely status inconsistency, is related to burnout.

The sociology of occupational burnout

Considerable research in the United States has maintained that people's identity is interlinked with the work they do. When asked to identify oneself on Kuhn and McPartland's (1954) "Twenty-statements test" (The "Who am I" Test), work activities came third after gender and race. This tendency identify with one's work is even more prevalent for individuals in professional and managerial occupations. Work is seen as a central life interest, and the conditions that make work activities unpleasant result in an existential crisis and create feelings of job burnout (Pines 1993a; Pines 1993b).

Research summarised by Clark (1986) indicated that over two-thirds of Americans, especially those in occupations that are considered to be of high status, including professional, technical, managerial and supervisory positions, report that their work is central to their life interests and sense of personal worth. At the time of the studies reported by Clark, European and Australian high status workers were more likely to resemble American blue-collar workers, for whom their jobs were a means to earning a living and were not a central source of identity. In fact, British workers saw work as less central to identity than did Americans, regardless of the status of their occupations.

The findings of an Australian survey of 2001, conducted some fifteen to twenty years after the studies cited by Clark (1986) and Hamilton and Wright (1986), do not seem to support the earlier view of Australian workers. When asked the extent to which they considered an array of statuses to be central to their core identity, no less than 80 per cent of all Australians, regardless of their occupation, cited their work and an equal per cent also cited their educational attainment level. Among professional, managerial, and technical workers the percentage identifying their job as central to their identity exceeded 90 per cent. Thus, over the time period between the studies summarised by Clark and the national survey in 2001, there has been a convergence of Australian and American work-related self-identities.

Education and occupation are closely related. Indeed, education has often been described as the main route for occupational mobility for individuals (Evans and Kelley, 2001). Since educational and occupational attainments are significant elements of people's self image, we have turned our attention to the consequences of a disjuncture in these two attainments. In other words, if individuals' occupational attainments are not consistent with their educational attainments, could it affect such job sentiments as burnout? Of course, the direction of the disjuncture is important. We imagine that individuals whose educational attainments are not converted into a prestigious job might perceive their job differently than individuals whose occupational attainments exceed their educational attainments.

But why should the experience of a form of status inconsistency lead to occupational burnout? To answer this question, we must first examine the concept of status inconsistency.

Status inconsistency and burnout

For a number of decades, from the 1950s to today, sociologists and psychologists have been interested in the imbalance between the attainments of individuals on a number of vertical hierarchies. Originally, Lenski (1954) listed four such hierarchies, namely income, occupation, education and ethnicity. Lenski and other sociologists have since found that when the status levels of these variables are not consistent for a person, a condition which has been labelled "status inconsistency" occurs[1]. Previous research has been based on the theory

[1] A number of terms have been used to describe this imbalance, each of them indicating a different perspective of the same phenomenon. Some authors prefer the term "status incongruence" while other who focus on a balance in hierarchies prefer the term "status crystallization".

that people who do not attain the same level on one or more hierarchies may show unusual behavioural patterns, or may experience other negative consequences.

A wide range of negative consequences of status inconsistency has been found in research. For example, Bonjean, Hill and McLemore (1967) have provided a review of early studies which reported that status inconsistency is related to political attitudes, differential association, stress, heart disease, and suicide. Lenski (1954) put forward one theory of why these inconsistencies bring about negative consequences. He argued that a person with an acute status inconsistency is a type of marginal person who feels or experiences different cross-pressures from various groups in society. When a person feels conflicting pressures due to inconsistencies, discomfort occurs. Lenski suggests that a society with many persons in a condition of status inconsistency is an unstable society.

After searching the literature and relevant websites, we conclude that, as far as we know, no one has ever examined the relationship between status inconsistency and occupational burnout. Furthermore, no one has examined the unique role that education plays in the various configurations of status inconsistency, and how these various configurations are related to burnout. This is the issue that we address in this paper.

The present chapter builds upon the model used by Saha and Dworkin (2004), which consisted variables in a three-stage model to address the relationship between education and burnout. In that paper we reported an anomaly: that education failed to account for a significant amount of variance in burnout until job context, including characteristics of the workplace and social psychological conditions of work, entered the equation. It appeared that certain kinds of employees find themselves in certain kinds of job contexts that lead to burnout, but that specific personal characteristics of the respondents and even the extent to which they identify with their job or their educational attainment, do not produce a statistically significant link between educational attainment and burnout.

Therefore, we want to focus our attention on the relationship between an education-occupation induced form of status inconsistency and its relationship to burnout. More specifically, we want to understand better the conditions under which educational attainment is or is not related to burnout.

We hypothesise that where individuals are unable to convert their educational attainments into equal occupational attainments, that job stress and job burnout might be higher than when individuals are successful in converting their educational attainments into equivalent occupations, or those whose occupational attainments are higher than their educational attainments. We hypothesise that those whose occupational and educational attainments are at parity with one another will have stress and burnout levels that were midway between the two extremes.

The data, analytical methods and variable descriptions

Our data are from the Australian component of the International Social Science Surveys (IsssA) collected in 2001. The simple random sample consists of panel respondents from previous IsssA surveys, and was originally drawn from the compulsory electoral roll. The survey was conducted by mail and had a response rate of about 60 per cent, or 1545 respondents. The sample was reasonably representative when compared to the national population statistics as reported by the Australian census of 1991.

In addition to detailed biographical information, the survey included a wide range of subject areas, including a set of items intended to measure perceived stress, job burnout, and a range of items related to work conditions. The method uses an analysis of covariance (ANCOVA) model, with Actor Traits, Sources of Self-Identity, Job Context, Job Stress and Role Conflict as five sets of independent variables. Burnout is the dependent variable.

Our technique for constructing the status inconsistency variable is similar to that originally adapted from Lenski's 1954 work by Hope (Hope, 1971). The technique has been criticised when several dimensions of inconsistency are used, as in the original Lenski model that included income, education, occupation, and ethnicity (see for example, Hornung (1977) and House and Harkins (1975)). Blalock (1967) and Whitt (1983) utilised interaction terms for inconsistency, while a lively debate on the advantage of computing separate inconsistency scores based on variations within each occupational category transpired between Wilson and Zurcher (1979) and Sandefur and Finley (1982).

We applied both the difference technique and the differences within occupational category techniques and obtained similar results in a preliminary analysis. For the present paper we reverted to the computation of z-score differences between the education and occupation rankings, considering status inconsistency to exist when an individual is a standard deviation above or below the group mean. This permits us to isolate the occupational groupings of individuals who are inconsistent viz. occupational and educational rankings.

In our previous analysis (Saha and Dworkin, 2004), we had created a three-stage, OLS regression model, with actor traits, sources of self-identity, job context (including stress), role conflict and role ambiguity to account for burnout. The purpose of this chapter is to isolate the mediating factor that results in a significant link between burnout and educational attainment. We focus on status consistency because we recognise that educational attainment is connected to occupational attainment, both empirically and especially in the minds of the public (Evans and Kelley, 2001). Not only do schools attempt to instil in their students the desire to continue their education and not drop out, but they warn of the consequences for employment opportunities that occur when students drop out of schools. Australians, Americans, and others, at least in the West, turn to additional schooling when they seek upward job mobility or when they experience job loss due to downturns in the economy.

Our interest in the concept of status inconsistency arose out of the anomaly found in our previous work (Saha and Dworkin, 2004) in which educational attainment did not have a significant impact on burnout until such job context variables as job stress and role conflict/role ambiguity were incorporated into the regression equation. Once these variables were included educational attainment had a small, but significant, negative effect on burnout ($B = -0.082$). Our review of the status inconsistency literature suggested to us that experiential gaps between the levels of educational and occupational attainment could affect burnout. However, there is reason to suspect that the relationship between status inconsistency and burnout might diverge from linearity, but that neither computing an interaction term nor simply taking the quadratic of the status inconsistency variable would suffice.

We can thus imagine that inconsistencies between occupational attainment and educational attainment will have perceptual consequences that may affect job attitudes. They may also affect attitudes about education. Mickelson (1990) described the paradox among African Americans in the inner-city who conclude that education may be the route to upward mobility, but not for their people.

In his critique of status crystallisation models, Broom (1960) argued that unless individuals are aware of inconsistencies in the elements of their status, it is unlikely that such conditions will be sources of negative sentiments. This implies that social structure must be experiential in order for it to impact attitudes and behaviours, including job burnout and job actions associated with burnout. Small, statistical differences in status rankings among indicators are unlikely to be felt and to prompt action. However, work places are often locations where comparative reference groups operate, permitting people to judge whether, given their investment in a job, they are relatively deprived (See Chafetz and Dworkin, 1986, Chapter 3, for a discussion of this observation). Thus, where there are significant gaps between educational attainment and occupational attainment we would expect that such gaps would result in awareness of relative advantage or disadvantage and affect attitudes toward work.

Our analysis examines the mediating role of status consistency on the relationship between educational attainment and burnout. Our strategy is to use the blocks of variables originally used in our previous work (Saha and Dworkin, 2004) as covariates and status consistency as the main effect in an ANCOVA analysis. Adjusted means will be determined for individuals whose educational attainment is higher than their occupational status (Code 1), individuals whose occupational status is higher than their educational attainment (Code 3), and others whose educational and occupational attainments are in parity (Code 2). The variables used in our analysis are described in Table 1.

As can be seen in Table 1, the covariates are organised as Actor Traits, Sources of Self-Identity, Job Context variables, Job Stress and Role Conflict. The Status Inconsistency and Burnout variables are described separately. The Actor Traits refer to variables that represent characteristics of the respondent that are brought to the work situation. This category includes the respondent's age, family income, employment status, and perceived social class membership.

While years of education is in the data set and is coded from 0 for no formal schooling to 18 for 18 years of schooling, representing a post-graduate education, this variable is not included among the covariates. Likewise, occupational status, coded from 10 for a farm worker to 100 for the highest ranked profession (with 0 for no occupation), is omitted from the covariates. The reason that neither education nor occupation is included is that they would be co-linear with Status Inconsistency, which is made up of the education and occupation scores.

The covariates that reflect sources of Self-Identity consist of four variables. They are included with the argument that the more that work is central to one's identity, the stronger is the relationship between stress and burnout. In this group of covariates we include three additional significant variables of centrality of identity.

Table 1. Description of variables in the analyses

Variable Label	Variable Description
	1. Covariates
Actor Traits	
Age	Coded in actual years at the time of the survey in 2001
Employed	Coded as 0=unemployed and 1=employed
Family Income	Coded as 5000=$5000 or less, to 100,000=$100,000 or more
Perceived Social Class	Response to the question: "In our society there are some social groups which are higher and some which are lower. Where do you think you are on this scale?" The codes are Low=1 to High=10.
Source of Self Identity	
Centrality of Education	Coded from 1=not at all important to 5=very important
Centrality of Job	Coded from 1=not at all important to 5=very important
Centrality of Honesty	Coded from 1=not at all important to 5=very important
Job Context	
Is a supervisor	Coded as 0=no and 1=yes
Is supervised	Coded as 0=no supervisor and 1=yes, have a supervisor
Public/Private	Coded as private/self=0 and government=1
Company Size	Coded as 0=no other employees to 1000=one-thousand or more employees
Job Stress	A four item factor analytically generated variable: 1) "I fear that the amount of stress in my job will make me physically ill", 2) "My work is more stressful than I ever imagined", 3) "I feel that I am usually able to handle the stress levels on my job", and 4) "I feel happiest and most satisfied when I am at work on my job". These were coded so that a high score represented high stress. The Cronbach's Alpha is 0.89.
Role Conflict	A six item factor analytically generated variable. The items are an abridged version of the measure of scale of Iwanicki and Schwab [1981]. The role conflict in the items reflects the experience of competing demands from their supervisors, and role ambiguity occurs when employees are uncertain of the expectations by supervisors. The Cronbach's Alpha is 0.77.
	2. Fixed Factor
Status Inconsistency	
Status Groupings	The JK Worldwide occupational status scores (from ISCO 1988) and the measure of the years of education were converted into standard (z) scores. We subtracted the z-score of occupational status from the z-score of educational status. The difference scores were trichotomised, with scores that were one or more standard deviations above the group mean being labeled as inconsistent higher education than occupation (Group 1); and those within plus or minus one standard deviation from the mean of zero being the consistents (Group 2); scores that were one or more standard deviations below the group mean being labelled as inconsistent higher occupation than education (Group 3). Separating the sample on the basis of z-scores permits the construction of statistically homogeneous groups. Groups 1 and 3 are status inconsistent.
	3. Dependent Variable
Occupational Burnout	Burnout, is a ten-item factor-analytically generated scale with an Alpha of 0.78. The scale was developed by Dworkin [1987] in which burnout is conceptualised as role-specific alienation in which the employee experiences powerlessness, normlessness, meaning-lessness, isolation, and estrangement. The dimensions of alienation were initially conceptualised by Seeman [1959, 1975].

The first variable is Centrality of Education Status ("Your Education – How Important?"), followed by Centrality of Job ("Your Job or Occupation"), and Centrality of Honesty ("Being honest – how important is that in describing who you are?"). Finally, job context represents the final group of covariates. These variables include company size, type of organisation (government or private), whether the respondent is supervised or not, and whether the respondent is a supervisor.

Job stress is a four item factor-analytically generated variable with a Cronbach's coefficient Alpha of 0.89. The items are as follows: 1) "I fear that the amount of stress in my job will make me physically ill", 2) "My work is more stressful than I ever imagined", 3) "I feel that I am usually able to handle the stress levels on my job", and 4) "I feel happiest and most satisfied when I am at work on my job". These were coded so that a high score represented high stress. The same applied to the factor variable.

Role Conflict/Role Ambiguity is a six item factor analytically generated variable with a Cronbach's coefficient Alpha of 0.77. The items are an abridged version of the measure of scale of Iwanicki and Schwab (1981). The role conflict in the items reflects the experience of competing demands from their supervisors, and role ambiguity occurs when employees are uncertain of the expectations by supervisors.

The dependent variable, Burnout, is a ten-item factor-analytically generated scale with an alpha of 0.78. The scale was developed by Dworkin (1987) in which burnout is conceptualised as role-specific alienation in which the employee experiences powerlessness, normlessness, meaninglessness, isolation, and estrangement. The dimensions of alienation were initially conceptualised by Seeman (1959; 1975).

Results

The analysis requires that we compute an ANCOVA in which the dependent variable is burnout, the fixed effect is membership in one of the three status inconsistency groupings, and the variables used as covariates to adjust for differences among the three groupings.

Table 2 presents the results of the ANCOVA. The adjusted R^2 is 0.418, with all the covariates playing a significant role. The first and most important finding is that status inconsistency is significantly related to burnout, with the covariates operating as controls. Thus it does make a difference whether a person has an imbalance between educational and occupational attainment, irrespective of the Actor Traits, Self-Identity Traits, Job Context and Role Conflict.

The second most important finding is that all covariate variables are independently related to burnout at a statistically significant level. Thus age, employment status, family income, and so on, do make a difference. Of the covariates, role conflict (F=237.0) is the strongest covariate, followed by centrality of the job to self-identity (F=46.66). These are followed by stress, age, being a supervisor, and the centrality of honesty to self-identity. The other covariates play a significant, but lesser role.

Table 2. Analysis of covariance of the effects of status inconsistency and covariates on burnout scores

Variable	Source of Effect	S.S.	df	M.S.	F	Sig.
Blocks	Corrected Model	484.764	15	32.318	56.040	0.000
	Intercept	64.268	1	4.268	111.444	0.000
1. Covariates						
Actor Traits	Age	16.107	1	6.107	27.930	0.000
	Employed	6.385	1	6.385	11.071	0.001
	Family Income	7.211	1	7.211	12.503	0.000
	Perceived Social Class	11.741	1	17.741	20.359	0.000
Source of Self Identity	Centrality of Education	2.152	1	2.152	3.732	0.054
	Centrality of Job	26.908	1	6.908	46.660	0.000
	Centrality of Honesty	13.851	1	13.851	24.019	0.000
Job Context	Is a Supervisor	14.893	1	14.893	25.826	0.000
	Is Supervised	2.485	1	2.485	4.308	0.038
	Public/Private	4.951	1	4.591	8.586	0.003
	Company Size	4.688	1	4.688	8.129	0.004
	Job Stress	18.187	1	18.187	31.537	0.000
	Role Conflict	136.678	1	136.678	237.006	0.000
2. Fixed Factor						
Status Inconsistency	Status Groupings	4.193	2	2.097	3.635	0.027
	Error	655.118	1136	0.577		
	Total	1139.935	1152			
	Corrected Total	1139.882	1151			
	$R^2 = 0.425$, Adjusted $R^2 = 0.418$.					

Our next question is whether the status groups are differently related to burnout. On the basis of our earlier discussion, we pose the following hypotheses: 1) the burnout scores will be highest among individuals where educational attainment exceeds occupational attainment (Code 1), and 2) the burnout scores will be the lowest where either educational attainment is at parity with occupational attainment (Code

2) or where educational attainment lags behind occupational attainment (Code 3).

To answer this question we examine the unadjusted mean transformed burnout scores (before ANCOVA) and the adjusted mean transformed burnout scores (after ANCOVA) for each of the three status groups. These results are given in Table 3.

Table 3. Adjusted and Unadjusted Mean Burnout Scores for Each Status Consistency and Inconsistency Type*

Group	N	Unadjusted Mean	SD	Adjusted Mean	SD
(Code 1) High Educational and Low Occupational Status	210	0.248	0.979	0.095	0.054
(Code 2) Educational and Occupational at Parity	820	-0.029	1.000	-0.011	0.027
(Code 3) Low Educational and High Occupation Status	122	-0.295	0.995	-0.154	0.072

*The adjusted means control for all 13 covariates. The effect size of the adjusted model is 0.418. Further, all groups are significantly different from each other at beyond the 0.05 level for both the unadjusted and the adjusted means.

In Table 3, a positive mean indicates an increased burnout level, while a negative mean indicates a decreased burnout level. The unadjusted mean for the respondents with high educational and low occupational status (Code 1) is 0.248, indicating burnout levels that are above the mean for the sample (as z-scores the mean is zero and the standard deviation is one). Individuals whose educational attainment is low and their occupational attainment is high (Code 3) had a substantially lower burnout mean score of -0.295 compared with the average of zero for the total sample. Finally, where educational and occupational attainments were at parity (Code 2), the obtained burnout mean was -0.029, which closely approximated the sample average.

Adjusting for the covariates attenuates the mean scores considerably, as we might expect, but the differences between the three groups remain statistically significantly different. High educational and low occupational status attainment results in a positive burnout mean score of 0.095 that is above the sample mean. Low educational attainment and high occupational attainment produces a mean burnout score of -0.154, significantly below the overall sample mean. Finally, when educational and occupational attainments are at parity, burnout scores are equal to -0.01, that is nearly at the whole sample mean. Thus, even after controlling for the covariates and thereby equilibrating the subgroups in terms of actor traits, job context, and social psychological characteristics, and role conflict, status inconsistency where educational exceeds occupational attainment (Code 1) results in significantly higher burnout scores than conditions where the attainments are at parity (Code 2) or where occupational attainment exceeds educational attainment (Code 3). In both the adjusted and the unadjusted conditions the differences between pairs of status groupings were significantly different from each other at beyond the 0.05 level.

Discussion

Burnout as a sociological construct has been viewed as a form of role-specific alienation that arises out of a perceived gap between expectations about a job and the actual job experience (Cherniss, 1980), (Dworkin, 1987; 1997) and (Pines, 1993b). Educational systems motivate students to continue their schooling through a promise that occupational attainment is closely coupled to educational achievement. Education and jobs are seen as inextricably linked (Evans and Kelley, 2001). When jobs are threatened, most people seek additional education in order to make them more competitive viz. other job candidates. Similarly, in credentialist societies additional academic credentials and certificates are interpreted as increased competence and competitiveness in the labor market (Collins, 1979). When job experiences fail to meet expectations, or when people realise that they are over-qualified for the jobs they have compare with their like-positioned coworkers, they may either question the extent to which more schooling leads to good jobs, or they may question whether they were somehow denied the legitimate career that their schooling was to provide. That is, they may either question whether more schooling means better jobs or whether the particular job they have is deficient relative to their skill level.

Ogbu (1978) and Mickelson (1990) report that some disadvantaged minorities conclude that while education is the route toward upward mobility, it does not work for their people. They question the group-specific viability of education. However, when job experiences do not characterise whole social groups, the dissatisfaction with employment situations are unlikely to be generalised to the educational system and more likely to reflect on a specific job and/or on self-blame for not taking advantage of the educational experiences.

There is no reason to believe that the respondents with higher levels of burnout have abandoned the societal maxim that education leads to better quality employment. In fact, these individuals are more likely to belong to statistical groups whose education exceeds the modal level for people with their occupational attainment. Inspection of the occupational categories of individuals who are status inconsistent negative (Code 1, where their educational attainment is greater than their occupational attainment) reveals that 57.1 per cent are in farming, unskilled or semi-skilled labour, or routine sales work with either college preparatory training or some years of college experience. The Code 1 individuals had higher burnout scores than did the individuals in Code 2, who were status consistent (parity between educational and occupational attainment). Code 3 individuals, with the lowest burnout scores, were status inconsistent, too, but their occupational attainments exceeded their educational attainments. Code 3 individuals were more likely than others to have high clerical, technical, administrative and managerial, and high professional job without the commensurate levels of education. Many had completed only a non-college preparatory level of education. They had, in a sense, beaten the credentialist system.

In order to confirm the role of status inconsistency in the operation of burnout, we examined the mean burnout Z-scores and mean years of education for each status inconsistency grouping within each occupational group. Table 4 presents these results.

Table 4. Mean Burnout Scores and Years of Education for Each Status Group within Each Occupational Category

(No values given for conditions in which there were fewer than ten cases)

Occupational Category	Status Group 1		Status Group 2		Status Group 3	
	Burnout Mean	Mean Yrs Education	Burnout Mean	Mean Yrs Education	Burnout Mean	Mean Yrs Education
Farm Worker	0.332	11.3				
Farm Owner	-0.290	11.9	-0.321	8.4		
Unskilled Worker	0.620	12.3	0.478	9.2		
Routine Service	0.650	14.1	0.307	8.9		
Semi-Skilled Worker	0.110	13.1	0.344	9.3		
Routine Sales	0.699	14.1	0.037	10.1		
Higher Service	0.123	14.1	-0.160	10.3		
Skilled Worker					0.132	10.6
Routine Clerical	0.711	15.8	0.085	10.8		
Higher Sales	-0.128	15.7	-0.228	11.5	-.755	7.8
Higher Clerical			0.238	12.2	-.119	8.2
Technical Worker	-0.074	18.0	-0.216	14.1	-0.403	9.1
Administrator/Manager			-0.385	14.5	-0.394	9.5
Higher Professional			0.202	16.1	-0.226	12.1

Wherever there were fewer than ten individuals in an occupation and status inconsistency sub-grouping, no results are presented. However, we aggregated those small samples across all occupational categories for that status consistency category. With only the exception of semi-skilled workers, respondents in status inconsistency group one (Code 1) had higher burnout scores (and of course, more education) than did their occupational counterparts in status inconsistency category two (Code 2). In all of the cases where the subgroup sample size permitted analysis for status inconsistency category three (Code 3), burnout scores were lower than for individuals within the occupation group who were in status inconsistency categories one or two. Where there were too few cases for an occupational grouping, as among respondents who were farm workers through routine clerical works in status inconsistency group three, we aggregated the burnout scores and years of education for those individuals. Their burnout scores were lower than a similar aggregation of those occupations in status inconsistency groups one and two. The mean burnout score for status inconsistency group Code 3 was -0.288 and a mean of 8.0 years of education. By contrast, Code 1 had a mean burnout score of 0.296 and a mean of 13.4 years of education, while Code 2 had a mean burnout score of 0.154 and a mean of 9.8 years of education. The differences among groups, each assessed with one-way ANOVA, were significant at beyond the 0.04 level for burnout and at beyond the 0.0001 level for education.

Conclusions

We think we have solved the puzzling link between education, occupation and burnout, about which we spoke at the beginning of this chapter. It is not educational attainment *per se* that influences burnout. Rather, educational attainment makes possible the attainment of occupations in which individuals are freer from role conflicts and job stress, both of which are due to a lower sense of control over work (powerlessness). When higher educational attainment fails to result in employment in an occupation which is commensurate with the investment in schooling, adverse results occur. One of these results is occupational burnout as we have defined it. When individuals are overqualified for the jobs they hold, by and large they experience alienation and powerlessness which make them disengage from their work. It ceases to be a source of satisfaction for them. Conversely, when an individual experiences an occupation higher than the education credentials they hold, they may continue to find their work satisfying, and indeed they may regard themselves as extremely fortunate. They are not nearly as prone to experience occupational burnout.

Therefore, the link between education and occupational burnout is not straightforward. It is not true to say that education has no relationship to burnout, as we were confronted with in our earlier paper (Saha and Dworkin, 2004). In one respect therefore, according to our findings, education may be seen as a cause of burnout. It is because societies, or at least most western societies, have so strongly coupled education and occupation that in some circumstances, where educational attainment does not pay off occupationally, that high educational attainment may have negative consequences. (see Saha (1992) for a discussion of this coupling effect in a socio-economic context.) In other words, high educational attainment does incur some risks.

In our chapter we do not address the reasons why status inconsistency occurs. Nor can we comment on why educational and occupational attainment are so strongly coupled in western societies such as Australia. More research of this kind in other western societies should be undertaken. Finally, although much research into the role of education and occupational attainment has been done, especially in the form of occupational attainment models, the study of the consequences for when this relationship does not occur has been less explored. We think our chapter makes a step in this direction.

References

Blalock, H.M. (1967). Status inconsistency, social mobility, status integration and structural effects. *American Sociological Review,* 32(5), 790-801.

Bonjean, C.M., Hill, R.J. and McLemore, D. (1967). *Sociological Measurement: An Inventory of Scales and Indices*. San Francisco, Chandler.

Broom, L. (1960). Social Differentiation and Stratification. In R.K. Merton, L. Broom and L.S. Cottrell Jr. (Eds) *Sociology Today: Problems and Prospects* (pp.429-441). New York, Basic Books.

Chafetz, J.S. and Dworkin, A.G. (1986). *Female Revolt: Women's Movements in World and Historical Perspective*. Totawa, NJ, Rowman and Allanheld.

Cherniss, C. (1980). *Professional Burnout in Human Service Organizations*. New York, Praeger.

Clark, N. (1986). Central life interests of Australian workers. *Work and Occupations,* 13, 67-75.

Collins, R. (1979). *The Credential Society*. New York, Academic Press.

Dworkin, A.G. (1987). *Teacher Burnout in the Public Schools*. Albany, State University of New York Press.

Dworkin, A.G. (1997). Coping with reform: The intermix of teacher morale, teacher burnout, and teacher accountability. In B.J. Biddle, T.L. Good and I.F. Goodson. (Eds) *International Handbook of Teachers and Teaching* (pp.459-498). Dordrecht, Kluwer Academic.

Evans, M.D.R. and Kelley, J. (2001). Why is education rewarded - Necessary skills or arbitrary credentialism. In M. D. R. Evans and J. Kelley, *Australian Economy and Society, 2001: Education, Work and Welfare*. Sydney, Federation Press.

Hamilton, R.F. and Wright, J.D. (1986). *The State of the Masses*. New York, Aldine.

Hope, K. (1971). Models of status inconsistency and social mobility effects. *American Sociological Review,* 40(June), 322-343.

Hornung, C.A. (1977). Social status, status inconsistency, and psychological stress. *American Sociological Review,* 42 (August), 623-638.

House, J.S. and Harkins, E.B. (1975). Why and when is status inconsistency stressful? *American Journal of Sociology,* 81(2), 395-412.

Iwanicki, E.F. and Schwab, R.L. (1981). A cross validation study of the Maslach Burnout Inventory. *Educational and Psychological Measurement,* 41, 1167-1174.

Kuhn, M.H. and McPartland, T.H. (1954). An empirical investigation of self-attitudes. *American Sociological Review*,19, 68-76.

Lenski, G.E. (1954). Status crystallization: A non-vertical dimension of social status. *American Sociological Review,* 19, 405-413.

Mickelson, R.A. (1990). The attitude-achievement paradox among black adolescents. *Sociology of Education,* 63(1), 44-61.

Ogbu, J.U. (1978). *Minority Education and Caste*. New York, Academic.

Pines, A.M. (1993a). Burnout. In L. Goldberger and S. Breznit, *Handbook of Stress: Theoretical and Clinical Aspects*, (pp.386-402). New York, Free.

Pines, A.M. (1993b). Burnout: existential perspectives. In W.B. Schaufeli, C. Maslach and T. Marek, *Professional Burnout: Recent Developments in Theory and Research*. Washington, D.C., Taylor and Francis.

Saha, L.J. (1992). The effects of socio-economic development on student academic performance and life plans: A cross-national analysis. *International Journal of Educational Development,* 12(3), 191-204.

Saha, L.J. and Dworkin, A.G. (2004). Work conditions and burnout: The mitigating effects of work status and education. In R. Ruzicka, J.H. Ballantine and J. Ruiz San Roman. (Eds) *Key Contexts for Education and Democracy in Globalising Societies*. Prague, Zeithamlova Milena, Ing - Agentura Action M: A5/2 - A5/7.

Sandefur, G.D. and Finley, N.J. (1982). Specifying models of status inconsistrency: A commentary on Wilson and Zurcher. *Social Forces* 60(June), 1168-1170.

Seeman, M. (1959). On the meaning of alienation. *American Sociological Review,* 24, 783-791.

Seeman, M. (1975). Alienation studies. *Annual Review of Sociology,* 1, 91-123.

Whitt, H. P. (1983). Status inconsistency: A body or negative evidence or a statistical artifact? *Social Forces*, 62(September), 201-263.

Zurcher, L.A. and Wilson K.L. (1979). Status inconsistency and the Hope technique, II: A linear hypothesis about status enhancement, status detraction, and satisfaction with membership. *Social Forces,* 57(June), 1248-1263.

12

Estimating the complexity of applied English language skills to the perceived ability of non-English speaking student nurses, using Rasch analysis

Ian R. Blackman and *M. M. Hall*
School of Nursing and Midwifery, Flinders University

> "Which medication did you say, Betadine or Pethidine?" Estimating the complexity of applied English language skills to the perceived ability of non-English speaking student nurses, using Rasch analysis: a pilot study.

Developing the rating scale content: language and nursing context related

Gathering participants to complete attitude rating scales at different times during their learning is often a useful method to measure educational performance, not only in terms of the student ability but performance of the curriculum too. Other reasons for attitude measurement are to:

- ascertain participant attitudes and dispositions (to the applied use of English) that underlie their thinking. These can either be positive or negative in nature;
- measure the intensity of participant attitude (ie: degree of self efficacy they possess in terms of English language skills usage); and

- estimate the consistency of their attitude toward some belief/value (execution of applied English in the nursing context).

Anderson (1997) has suggested that attitudes, interests and values were central to the educative process both as a means and an end and it was identified as a process where internalisation had taken place within others that implied that inner change had taken place. Measuring attitudes and values is difficult as they cannot be directly observed but can be inferred from what a person says or does. While Wolf (1997) pointed out the pitfalls of using rating scales as a method for obtaining data, self rating of their own language sophistication was the most efficient way to collect data without having to train assessors or employing multiple raters.

To identify what relationships existed between perceived self efficacy of student nurses who used English as an additional language and the complexity of the nursing tasks that were expected of them was constructed for estimation. The study sought to ascertain if there was a difference between the perceived complexities of vocationally related English language tasks, as understood by nurses who use English as an additional language at different levels of sophistication. The data were generated by administering a 45-item, four point (Likert-type) rating scale to a pilot group of 33 student nurses from different language backgrounds.

The partial credit model (part of the Rasch family) was selected to ascertain if there was an underlying dimension of the nursing-related English language tasks and the perceived self capacity of the students to engage in the nursing tasks of different complexities. Furthermore, the study sought to assess whether a scale of learning could be constructed, based on the estimation of the difficulty of the nursing related English language skills and tasks based on the responses of the participants.

In order to identify the numerous factors that influence the use of English in a vocational context to nursing and to inform the construction of the rating scale in the survey construction, a literature review was undertaken In order to identify variables that would become important in determining the validity of the scale that had been developed. (see unidimensionality described below).

Historically, the Australian nursing profession has been slow to respond effectively to the presence of overseas qualified nurses who use English as an additional language and have sought to work here. Registration authorities who regulate the nursing workforce have been unable to deal with the assessment of their English Language skills effectively, barring these Australian migrants from working within the nursing field on the pretext that their language skills are inferior and would therefore impose considerable risk to patient care. According to Blackman, Gonda, Hussin and Gaston (1996), nursing has struggled to identify effective English as an Additional Language (EAL) assessments that can ascertain the degree of sophistication of clinical English as a Second Language (ESL) skills held by “migrant” health care workers. An essential reason behind this according to Cigler (1985) and Abu-Saad et al. (1981) is the inability of Australian nursing professionals to test effectively Non-English Speaking Background (NESB) qualified nurses in an ESL context and the difficulty ESL teachers have in assessing overseas qualified nurses EAL language skills as they lack the technical or vocationally related context. This void has served only to keep bilingual and biculturally sensitive health care workers from the Australian health care system, which is itself struggling to provide culturally appropriate health care for its multicultural consumers. McNamara and Roever (2006) explain that language acts socially as a marker of national, racial and ethnic identity in that linguistic identity is shared by members of a speech group. They propose that language creates ‘in’ groups and ‘out’ groups. The problematic identity of the person seeking entry to the profession is known in advance and the language test, administered on the premise of functional requirements for successful participation in the workforce and the welfare of patients, works in favour of those who oppose the granting of such rights: the gatekeepers.

According to McNamara (1987), while there have been some improvements in ESL tests for NESB health professionals in the recent past, historically, such tests have lacked validity in that they do not test knowledge that is appropriate to the health care context. Canale et al. (1980) have been critical of the examination technique used for NESB health workers, in that multiple choice formats are inappropriate. Established English language testing services such as TESOL and IELTS are also inappropriate for determining whether a NESB qualified nurse possesses suitable clinical language sophistication to deem that the health care worker is competent to practise in the Australian health care system. Referring to the Canadian Official Languages Policy and the bilingual competency required of civil servants, McNamara and Rover (2006) argue that there is a disparity between the postulation of the language test about the use of official language and the reality of the workplace. A similar claim could be made for the nursing context.

In the past decade, the Australian health care system has been experiencing a shortage of registered nurses and one strategy to increase the availability of registered nurses for the workforce has been recruitment of overseas qualified registered nurses. Overseas qualified registered nurses who seek employment in the Australian health care system and who use EAL, are commonly required to gain additional nursing qualifications, through either a bridging program or a conversion program, as a pathway to qualification recognition in Australia. This requirement, Hawthorne (2002) argues, is largely confined to this group of nurses and that prior to commencing these programs a substantial proportion of these nurses will have completed English language courses. As these courses are generally located in the Australian tertiary sector, these nurses will have met the language criteria for entry established by these institutions. These criteria assume that the student will be sufficiently competent in English language skills to successfully complete the course. This is in contrast to English speaking background nurses who pass seamlessly into employment. She sees the introduction of competency based assessment by the nursing profession post 1990 as providing a structured pathway for EAL students to achieve qualification recognition by Australian Nurses' Registering authorities by meeting the competency standards set by the Australian Nursing and Midwifery Council. Hawthorne (2002) goes on to explain that competency based assessment of overseas qualified professionals was promoted by the National Population Council (NPC) as a strategy to overcome and prevent professional groups establishing unfair or inequitable assessment mechanisms to protect their own interests.

In their most recent study, Blackman, Hall and Darmawan (2006) continue to demonstrate that nurses who use English as an alterative language experience difficulties with academic achievement, and also struggle to show applied language competence in the clinical setting. This applies to both the general nursing setting and also in the mental health field (Blackman, 2001). Konno (2006 p. 91) reports a similar findings in that EAL overseas qualified registered nurses who undertake a conversion or bridging program experience difficulty adjusting to nursing practice in the context of the Australian Health Care system. This is attributed to communication and cultural issues. Menon (1992) attributes these difficulties to the complexity of the clinical language employed by Australian nurses. In the clinical context two forms of nursing language were identified. At times Australian nurses used a communication style which was clear, fairly direct and factual whilst at other times the communication style used was highly complex, comprised of conceptual ideas, idiomatic usage, technical jargon, and abbreviations, combined with miscellaneous ideas. Briguglio (1998) reported data from a study of the linguistic and educational needs of EAL students that the participants said they could comprehend formal registers, but that they had greater difficulty understanding the informal registers. The participants found fellow students more difficult to understand than the teaching staff who they perceived spoke more slowly and more clearly. The participants indicated that staff that had awareness of the participants linguistic and cultural needs provided better support.

Identifying relevant test items for the construction of a nursing language rating scale

A four-point Likert-type Scale was intentionally utilised for this survey and, although Anderson (1997) has cautioned that greater internal consistency/reliability may have been achieved had more response options been offered, the authors maintained that understanding complicated semantic differentials was difficult for NESB users and to employ an easier scale would improve accuracy of responses.

1	2	3	4
Very difficult	Difficult	Easy	Very Easy

A category typically located in numerous rating scales is a “not appropriate” or “not relevant” descriptor. This has been purposely omitted in this survey scale, as its inclusion changes the underlying dimension of the whole scale, and does not accurately measure this on a continuum, which our four-point scale sets out to do.

With reference to Table 1, a total of 45 specific English language nursing tasks were identified to be surveyed. The active language skills of speaking and writing are represented by 22 survey items, with the remaining passive language (listening and reading) skills identified by a further 23 items for measurement.

Problems with the construction and subsequent analysis of rating scales, overcome by Rasch scaling

Bond and Fox (2007) caution that in traditional rating scales construction and analysis, there are some assumptions commonly made which are inconsistent with the measurement of the underlying construct being estimated.

Table 1. Item Number and Nursing Language Task items submitted for rating by student nurses who use English as an additional language (EAL)

1	Understands information given during change of shift handovers	listening
2	Understands slang words (colloquialisms) used by patients/staff	listening
3	Understands medical abbreviations used by healthcare team	listening
4	Prioritises nursing actions from a selection of verbal orders/instruction	listening
5	Explains to another staff ember what duties are expected of them	speaking
6	Asks another staff member for help in giving care	speaking
7	Correctly pronounces the names of medication	speaking
8	Talking informally/chatting with colleagues	speaking
9	Advises other staff of tasks that need completing	speaking
10	Praises another staff member when they have done their work well	speaking
11	Provides accurate handover to others about allocated patients	speaking
12	Enunciates correctly the surnames of others	speaking
13	Introduces self to a patient	speaking
14	Understands the accents of other people	listening
15	Offers a range of choices or alternatives in nursing practices to patients	speaking
16	Takes a nursing history using acceptable questioning formats	speaking
17	Explains procedures to patients in simple language using non-medical terms	speaking
18	Converses with other students in nursing laboratories	speaking
19	Note the results of a patient's pathology test as it given over the telephone	listening
20	Provides information over the telephone about a patient	speaking
21	Asks a telephone caller to repeat information not understood at first	listening
22	Understands surnames given over the telephone	listening
23	Interprets names of medications given over the telephone	listening
24	Understands what the purpose/content of a telephone call	listening
25	Explains to a telephone caller how to contact another person	speaking
26	Identifies the correct medication from a written drug chart/prescription	reading
27	Understands details of patient care given in patient's case notes	reading
28	Understands memos arising from nursing administration	reading
29	Correctly interprets signs (eg: fire exit/emergency signs)	reading
30	Check patient's charts to ensure nursing tasks have been completed	reading
31	Collates appropriate documents to compile patient case notes a patient	reading
32	Understands a nursing care plan generated by another	reading
33	Interprets the correct meaning of data entered into patient's charts	reading
34	Identifies a specific medication from a range of stored medication	reading
35	Understands hospital menus/food names	reading
36	Understands contents of drug guide (MIMS)	reading
37	Using hospital computer system (EXELCARE	reading
38	Finding out hospital protocols re patient care	reading
39	Constructs a nursing care plan	writing
40	Uses objective language when completing a patient's case notes	writing
41	Completes an inventory of patient's clothes and valuables	writing
42	Charts data arising from patient care	writing
43	Takes personal notes about patient care regimes during handover	writing
44	Submits a detailed description of events associates with patient/staff injury	writing
45	Constructs an information sheet/letter for patient use upon discharge	writing

One incorrect assumption is that summing the total scores gained by participants, indicate an accurate measure of the complexity of the underlying construct. For example, if two participants, A and B, respond to any three items as shown in Table 1 with scores of 2, 3 and 4 and the other participant with score of 4, 4 and 1, respectively, both their cumulative scores would be 9. This practice assumes that all items are of equal difficulty and therefore the raw scores may be added. Rasch analysis suggests that all test items used in the questionnaire differ from each other according to each participant's perceived ability and therefore cannot be quantified in this manner. Moreover, Rasch analysis assumes that the threshold levels (spaces in between each of the response categories on the Likert-type scale) differ between each test item being estimated.

Another limitation of using participant raw scores as estimates of ability is that there is no mechanism for exploring the consistency of an individual's responses to the items answered in the questionnaire. The inclusion of inconsistent participant response patterns has been shown to increase standard error of threshold estimates and to compress the threshold range during the instrument calibration. It is more helpful, then, to use a method of analysis that can detect the participant's inconsistency in their responses (fit statistics) to the questionnaire and which can provide estimates of item thresholds and individual trait estimates on a common interval scale. Moreover, where respondents do not respond to all items of the questionnaire, the Rasch model is able to estimate their trait score, taking into account the level of item difficulty. This is not possible using raw scores or classical test theory.

Calibrating the nursing language scale

Why the partial credit model was selected

The English language nursing scale comprised 45 items that were used to survey participants' perceived ability in using the active language skills of speaking and writing and the passive skills of listening and reading within differing complexities of nursing tasks. Given that the scale used for their ability estimation was polytomous in nature, the partial credit model was engaged for the analysis. Unlike the rating scale model (which estimates a dichotomous model), the partial credit model has the advantage of not constraining threshold levels between the items used in the questionnaire, but allows response categories to vary across each of the items asked of the respondents. This model's approach estimates if the distances between the response categories are constant for each nursing language item and if the options for each item vary in the number of response categories given. The alternative approach (the rating scale model) uses only one set of item thresholds estimates and it applies this to all the items that make up the nursing language scale (Bond & Fox, 2007).

Are the questionnaire items valid and are they unidimensional?

While the nursing language survey instrument used in this study contained less than fifty items, a measurement of reliability of the survey instrument was indicated. An underlying assumption before the Rasch model could be legitimately employed was that test items used to construct the nursing language scale were required to reflect the same underlying construct; unidimensionality. Indeed, if any items were not seen to do this, those items needed to be reviewed by the researcher, changed or taken out of the measuring instrument altogether (Hambenton, 1991; Linacre, 1995; Smith, 1996, McNamara, 1996).

Goodness of fit statistics

This test for unidimensionality (Goodness of fit statistics) ensures test item validity and in terms of this study, this was done by employing the QUEST program (Adams et al. 1996). The two main goodness of fit indices used in the analysis were focused upon, namely the unweighted (or the outfit means square) and the weighted (or infit means square) index. These indices yielded excellent diagnostic information about discrepancies between the predicted and observed data, particularly in terms of the direction and size of the residuals for each of the nursing language items estimated. Typical fit values varied around a mean of zero and each item tested was labelled as either positive or negative, depending whether the observed values showed a greater variation in responses, than what was expected (greater variation showed a positive value and minimal variation as a negative value). In this way, the compatibility or the orderliness of the data obtained from the nursing language scale used in this study could be measured against the Rasch model requirements. Items might show an appropriate fit if the pattern for individual items fit the overall pattern, whereas ill-fitting items were seen to occur when items did not correspond to the overall pattern.

Overfitting items are said to occur when items are unpredictable or lack variability. As Wright and Stone (1979) point out, when the item and the person are close to one another, ie. on target, then the item contributes more to the measure of the person ability than when the item and person are far apart. The greater the difference between item and person fit statistics, the greater the number of items that are needed to obtain a measure of comparative precision and, as a result, the less efficient is the item. The outfit means square index is sensitive to outliers especially where the participants gained unexpectedly high scores on difficult nursing language items or achieved uncharacteristic low scores on easy language items. One or two of these large outliers can cause the fit statistic to become very large as when a very able respondent views an easy nursing language item as difficult or when another respondent with low ability views a difficult nursing language item as being easy. With reference to Figure 1 which represents the fit model of all the nursing language items responded to by the participants, it can be seen that three items "misfit" (items 39, 40 and 45).

It can be seen in Table 2 that three language items scored by respondents have fit values below 0.75 which represents over-fitting items and suggests also that they are not performing at all well in terms of accurately portraying unidimensionality of the nursing language survey tool, which all the other test items have done.

Table 2. Infit values of three reading items not fitting the Rasch Partial Credit Model

Item number	Nursing Language activity	Infit value
39	Writing a new nursing care plan	0.68
40	Making written entries into case notes	0.67
45	Constructs an information sheet/letter for patient use upon discharge	0.71

It is worth noting that these three misfitting items all relate to the participant's self-rated capacity for writing and represent no other language skill tested. The misfitting nature of these items could be related to the possibility that the participants did not have opportunities to undertake these skills prior to the survey being undertaken. Nevertheless these items have been removed form further analysis in this study.

Estimating the complexity of each of the nursing language tasks

Figure 2 reveals how each NESB student estimated the ease or difficulty in carrying out each of the nursing language tasks surveyed. The logit scale is plotted on the left of the histogram and the nursing language item numbers are found to the right and beside each item number, an additional dot point and number has been added (either a 0.1, 0.2 or a 0.3). These latter numbers reflect the threshold value for that item.

```
Item Fit  (N = 48  L = 45  Probability Level= 0.50)
---------------------------------------------------------------
INFIT
MNSQ   0.56   0.63   0.71   0.83   1.00   1.20   1.40   1.60
--------+------+-----+-----+-----+------+------+------+--------
1 item 1               .         | *       .
2 item 2               .         |    *    .
3 item 3               .  *      |         .
4 item 4               .         | *       .
5 item 5               .         |    *    .
6 item 6               .         |      *  .
7 item 7               *         |         .
8 item 8               .         | *       .
9 item 9               .         |   *     .
10 item 10             .         *         .
11 item 11             .    *    |         .
12 item 12             .         | *       .
13 item 13             .         |*        .
14 item 14             .         |  *      .
15 item 15             .    *    |         .
16 item 16             .  *      |         .
17 item 17             .         |       * .
18 item 18             .         |    *    .
19 item 19             .      *  |         .
20 item 20             .        *|         .
21 item 21             .         |    *    .
22 item 22             .         *         .
23 item 23             .     *   |         .
24 item 24             .       * |         .
25 item 25             .         | *       .
26 item 26             .         | *       .
27 item 27             .         |       * .
28 item 28             .         |  *      .
29 item 29             .         |      *  .
30 item 30             .         *         .
31 item 31             .    *    |         .
32 item 32             .         |   *     .
33 item 33             .    *    |         .
34 item 34             .         |      *  .
35 item 35             .         | *       .
36 item 36             .      *  |         .
37 item 37             .         |  *      .
38 item 38             .         |*        .
39 item 39        *    .         |         .
40 item 40      *      .         |         .
41 item 41             .         |*        .
42 item 42             .     *   |         .
43 item 43             . *       |         .
44 item 44             . *       |         .
45 item 45         *   .         |         .
===============================================================
```

Figure 1. Fit Statistics for nursing language items undertaken by NESB participants

Low threshold values (0.1 to 0.2) indicate less complexity of the nursing language task or the impression that the students perceive themselves as having greater ability in achieving these nursing language tasks while higher threshold values (up to 0.3) reflect greater task complexity and or that students perceive themselves as having less ability to complete these nursing language tasks.

Estimating the item thresholds

As mentioned above, the partial credit model assumes that threshold values will be different within each individual language item itself and across all other nursing language survey items. The assumption of equidistance between the categories or thresholds of the nursing language rating scale is not held by the Rasch's partial credit model. In relation to Figure 2, it can be seen that the three different threshold values for item 16 (Taking a nursing history using acceptable questioning formats) are well dispersed along the whole logit scale. Logit value for the first threshold level for item 16.1 occurs at the –4.0 logit, the second threshold for item number 16.2 is found at the logit value of –0.72 and the third threshold value for this test item is located at logit +3.53. The dispersion of all thresholds across the whole logit scale suggests this nursing language task is performing well in the sense that it is identifying a large range of perceived language abilities of the students surveyed. Conversely, when item 10 in Figure 2 is examined and the three thresholds levels for that item (Praises another staff member when they have done their work well) are examined it can be seen they are all confined very narrowly on the logit scale. The first threshold level 10.1 is located at –3.30 logits, the second threshold only marginally higher at –2.29 and the third threshold at logit 2.0. What this pattern suggests when there is little logit difference in the first and second thresholds, students had difficulty in judging whether this nursing language task was easy or difficult for them. Students could clearly differentiate whether the task was very easy or very hard but could not readily determine self-efficacy that was not confined to the extreme ends of the scale.

Item 10 remains a valid test item. Item 19 (Noting the results of a patient's pathology test as it given over the telephone) again has three thresholds. Note the first threshold level (19.1) occurs at logit scale –1.31 which is very near where most of all the other test items' second threshold levels generally commence. Note that the second threshold level for item 19 (19.2) occurs at +1.03 and the last threshold value at logit +2.93.

```
 5.0          |
              |
              |
              |
              |   5.3   8.3
              |  18.3  24.3  25.3  32.3
 4.0          |   4.3  15.3  20.3  39.3
              |  37.3  44.3
              |
              |  16.3  30.3
              |   2.2  22.3  28.3  38.3
              |  43.3
              |  17.3  23.3  41.3
 3.0          |  11.3  19.3  26.3  29.3  33.3  34.3
              |   7.3  35.3  36.3
            X |   6.3  21.3
            X |  42.3
              |  12.3  14.2
              |
 2.0        X |  10.3
            X |  13.3
            X |  19.2
          XXX |
              |   1.2  23.2
            X |   3.2
 1.0          |
            X |  44.2
           XX |   7.2  20.2  37.2  39.2  43.2
         XXXX |  31.2  45.2
          XXX |   4.2  11.2  15.2  27.2  38.2  40.2
         XXXX |
  .0     XXXX |   2.1  22.2  26.2
           XX |  24.2  34.2
         XXXX |  28.2  32.2  42.2
           XX |  18.2
           XX |   9.2  25.2  36.2
          XXX |   8.2  12.2  16.2
              |  17.2  41.2
-1.0        X |  30.2  35.2
            X |
            X |   1.1  19.1  20.1
           XX |   5.2  14.1  21.2
              |  33.2
            X |  11.1  23.1  29.2
-2.0          |  26.1  45.1
              |   6.2  39.1
              |   4.1   7.1  17.1  25.1  36.1  37.1  41.1
              |   5.1  24.1
              |  10.2   3.1  12.1  40.1  44.1
            X |   6.1   9.1  13.2  28.1  34.1
-3.0          |  27.1  33.1  43.1
              |  35.1
              |   8.1  18.1  32.1  42.1
            X |  10.1  15.1  38.1
              |
              |  21.1  29.1
              |  30.1
-4.0          |  16.1
              |  22.1
```

Figure 2. Item Estimates (Threshold values) for language nursing tasks estimated by non-English speaking background nursing student Item Estimates (Thresholds) X refers to one student

Clearly, it can be seen that the assumption of equidistance between the three different threshold levels of the different language nursing tasks being tested does not exist. This finding, which is the hallmark of partial credit scaling, suggests that participant ability estimates vary within the scale of one item itself and will vary with every other item according to students perceived ability and the complexity of the language nursing task being tested. This unique contribution allows a hierarchical map of participant ability estimates to be generated. Figure 4 does this quite clearly where all the valid nursing language items are placed on another logit scale. Again, items that are placed higher on the logit scale suggest that the students required greater ability to carry out these nursing language tasks than items located lower on the logit scale. The lower items are viewed as becoming increasingly easier for participants to undertake and lack complexity. Note items of average complexity are located at zero on the logit scale (items 17, 1, 8, 34 and 41). These nursing language test items include explaining procedures to patients in simple language using non-medical terms, understanding information given during change of shift handovers, talking informally or chatting with colleagues, identifying a specific medication from a range of stored medication and completing an inventory of patient's clothes and valuables, respectively. Nursing language test items 2, 20,19, 23, 4, 37 and 44 are rated as being the most complex requiring greatest ability and include Understanding slang words (colloquialisms) used by patients and staff, providing information over the telephone about a patient, noting the results of a patient's pathology test as it given over the telephone, interpreting names of medications given over the telephone, prioritising nursing actions from a selection of verbal orders or instructions, using hospital computer system (EXELCARE) and submitting a detailed description of events associated with patient or staff injury to management, respectively. It is noted that a significant number of complex nursing language tasks involve students communication in a non face to face context (e.g. telephone use) and as it relates to the passive skill of listening.

The most simplistic nursing language skills include tasks 21, 29, 27, 10 and 9 and include asking a telephone caller to repeat information not understood at first, correctly interpreting signs (e.g. fire exit and emergency signs), understanding details of patient care given in patient's case notes, praising another staff member when they have done their work well and advising other staff of tasks that need completing. For a complete understanding of Figure 3, the item numbers list there can be compared to the item numbers listed in Table 1.

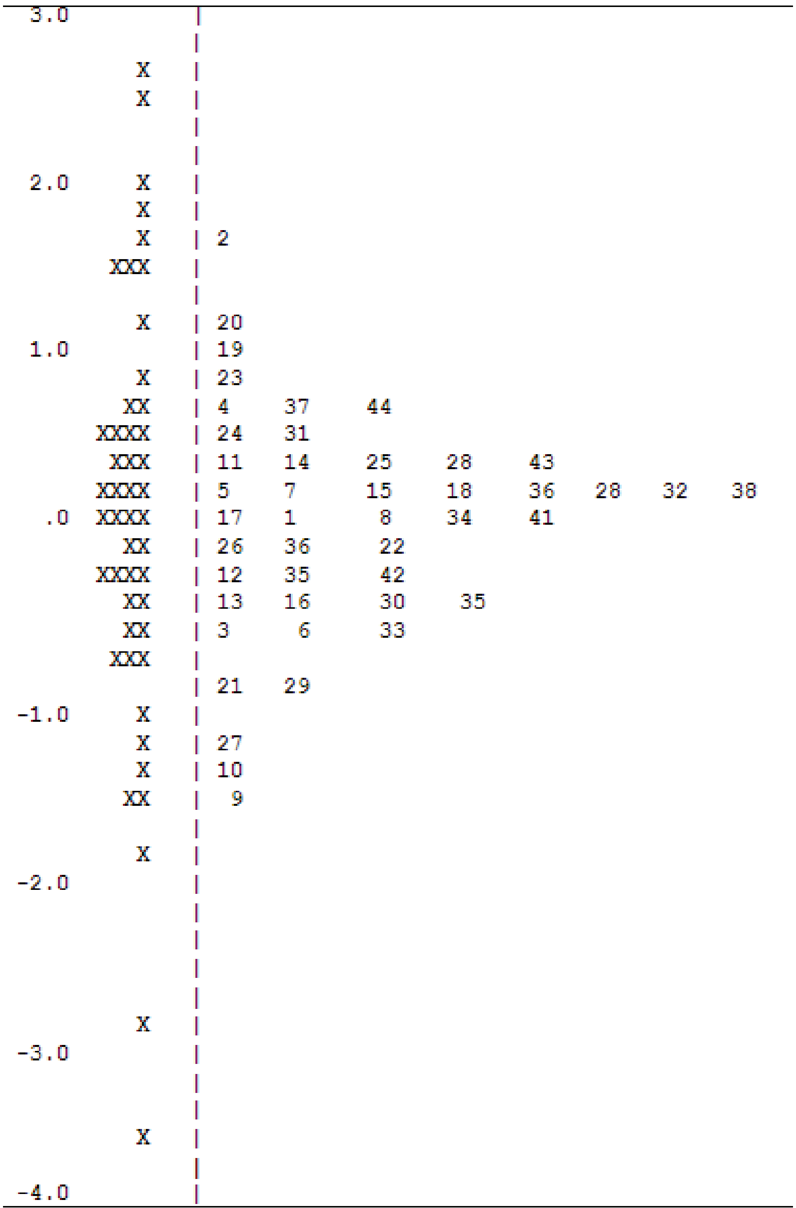


Figure 3. A Hierarchy of Language Nursing task item difficulty and participant ability estimates. X refers to one student

As cited above, nurse assessors are ill-prepared to measure the sophistication of English language complexity of students who use EAL, despite the fact students are required to demonstrate suitable clinical competence before they can practice as registered nurses in Australia. English language teachers are reluctant to assess applied English language proficiency in the nursing context for students who use EAL, because they are unfamiliar with the e clinical context of the nursing profession. Rasch analysis as used in this survey, has "bridged that gap" and provides direction as to how English language sophistication can be measured in students who use EAL. Training institutions already know that existing tests of language proficiency (TOEFL and IELTS) are unable to predict student achievement in their courses and more importantly, posses no validity in determining English language sophistication in the clinical setting. Blackman et al., (1996) suggested that ASLPR (Australian as a Second Language Proficiency Language Scale) now known as ISLPR (International Second Language Proficiency Rating Scales) would be most relevant for estimating student EAL proficiency in the clinical context. Rasch scaling has reliably identified nursing and language task complexity and can now be easily adapted for teaching and measurement both in the classroom and clinical settings. With the use of criteria-referenced language tests developed from the Rasch rating scales identified in this paper, language testing can be undertaken by non language teachers in the clinical area. Additionally, this Rasch study has demonstrated that different language tasks pose varied complexities for EAL students. These outcomes clearly suggest that course providers need to provide language support for EAL students based on the outcomes of this pilot study.

Conclusions

Rasch analysis offers the researcher a great deal for the development and analysis of attitude scales, and subsequently gives useful information to nursing educators about the readiness of nurses who use English as an additional language to take on the language tasks of reading, writing, listening and speaking. There are limitations to using traditional analytical procedures to analyse rating scales which are overcome when Rasch scaling is used to measure nursing language item difficulty and ability estimates of nursing student participants engaged in a learning process. By employing the partial credit model, the educational researchers are no longer confined by the notion that rating scale categories are static or uniformly estimated across each item being tested. Instead rating scales as used in this study are to be visualised as a continuum of participant nursing language ability which, when used on multiple occasions can be a valuable adjunct to see if learning has taken place. The objectivity of the nursing language ability Instrument can also be assessed in Rasch analysis by examining items for unidimensionality as indicated by their fit statistics.

The authors wish to acknowledge the encouragement and support of Professor John Keeves in their bid to develop and apply Rasch concepts in the measurement of nursing education.

References

Abu-Saad, H. & Kayser-Jones, J. (1981). Foreign nursing students in the USA Problems in their educational experiences. *Journal of Advanced Nursing*. 6. 397-403

Adams, R & Khoo S. (1996). *Quest: The Interactive Test Analysis System*,V.2.1. Australian Council for Educational Research. Melbourne.

Anderson, L.W. (1997). Measurement of Attitudes. In J.P. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook,* (2nd Edition. pp.380-390). Oxford: Pergamon.

Blackman, I., Hussin, V., Gonda, J. & Gaston, J. (1996). Migrant nurses: Their unique value. In G.Gray & P. Pratt (Eds.) *Issues in Australian Nursing: Volume 4,* (pp.129-143). Melbourne: Livingstone.

Blackman, I. (2001). A predictive model identifying latent variables which influence undergraduate student nurse achievement in mental health nursing skills. *International Education Journal.* .4, 2. 53-64

Blackman, I., Hall. M. & Darmawan, I.G.N. (2006). Predicting undergraduate nurse achievement and clinical competence. *International Education Journal,* 8(2), 222-236.

Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental Measurement in the Human Science.* New Jersey: Lawrence Erbaum Associates.

Briguglio, C. (1998). Non-English speaking background (NESB) students' perceptions of their linguistic and educational needs. 1997 Equity and Access Grant. Curtin University of Technology, Perth.

Canale, M. & Swain, M. (1980), Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,*1(1), 1-47.

Cigler, M. (1985). English for specific purposes: thoughts on the state of the art. *TESOL News,* 8(1), 30-35.

Hambenton, R.K. (1988). Principles and selected applications of item response theory. In R.L. Linn (Ed.). *Educational Measurement* (3rd Edition, pp.146-200). London, American Council on Education, Collier MacMillan.

Hambenton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item Response Theory.* Newbury Park: Sage.

Hawthorne, L. (2002). Qualifications recognition reform for skilled migrants in Australia, applying competency-based assessment to overseas qualified Nurses. *International Migration*, 40(6), 58-90.

Konno, R. (2006). Support for overseas qualified nurses in adjusting to Australian nursing practice, a systematic review. *International Journal of Evidence Based Health Care*, 4(2), 83-100.

Linacre, J.M. (1995). Prioritising misfit indicators. *Rasch Measurement Transactions, 9,* 422-423.

Menon, M. (1992). Migrant nurses: Cross cultural understanding or misunderstanding? In G. Gray & R. Pratt (Eds.) *Issues in Australian Nursing 3,* (pp.321-337) Melbourne*:* Livingstone.

McNamara T.F. (1987, April). *Assessing the Language Proficiency of Health Professionals: Recommendations for the Reform of the Occupational English Test*. Report submitted to the Council on Overseas Professional Qualifications, University of Melbourne.

McNamara, T.F. (1996). *Measuring Second Language Performance*. London and New York: Longman.

McNamara, T.F. Roever, C. (2006). *Language Testing: The Social Dimension*, Oxford: Blackwell.

Smith, R. (1996). A comparison of methods for determining dimensionality in Rasch measurement*. Structural Equation Modelling,* 3, 25-40.

Wolf, R.M. (1997). Rating scales. In J.P. Keeves (Ed) *Educational Research, Methodology and Measurement: An International Handbook*, (2nd Edition, pp.4923-4930), Oxford: Permagon.

Wright, B.D. & Stone, M.H., (1979). *Best Test Design: Rasch Measurement*. Chicago: Mesa Press.

13

Investigating the influences on reading literacy in primary school students

Ratna Rintaningrum
School of Education, Flinders University

In today's global society, children who do not learn to read early enough, who do not learn to read with comprehension and fluency, who do not learn to read for pleasure, are likely to meet difficulties in their lives. Moreover, their chances of succeeding in an academic environment, getting satisfying jobs that are rewarding financially, are, in practice, slight. However, if these children are equipped with an adequate reading literacy, they are able to participate fully in their communities and in the larger society.

Reading is one of the most fundamental aspects students acquire as it is the foundation for learning across all subjects. Reading literacy is vital when looking for a job or engaging in higher education. Reading literacy plays an essential role in improving the individual's life by "enabling economic security and good health and enriches societies by building human capital, fostering cultural identity and tolerance, and promoting civic participation" (Jalal & Sardjunani, 2005, p.1).

People have different purposes when they are reading. They read for fun, for literary experience and for retrieving information (Mullis, Kennedy, Martin & Sainsbury, 2006). This indicates that there are relationships between each of the purposes of reading and the types of text employed; therefore, this implies that readers are required to employ different strategies to understand the content, organisation and style of each particular text genre (Kirsch & Mosenthal, 1989; Weaver

& Kintsch, 1996). The approaches or strategies students employ in acquiring reading comprehension have an impact on their reading performance.

Different students have different levels of skill in acquiring reading comprehension. There are a number of reasons why some children perform better in reading literacy than others do. This indicates that a number of factors appear to operate to influence student achievement in reading. The results of prior research show that both student and school factors influence student achievement in reading (Adams & Wu, 2002; Elley, 1989; Elley, 1994; Kotte et al., 2005; Lietz, 1996; OECD, 2001; 2003b; Purves, 1973).

This chapter reports the results of a study that investigates factors influencing reading literacy achievement at the Primary School Students of South Australia. Actually, a study, on the factors influencing student achievement in literacy in South Australia, has been conducted across Grade 3, Grade 5, and Grade 7 over a span of three cohorts for the Basic Skills Testing Program target populations. However, in order to gain detail of the results of the study and to examine factors influencing literacy achievement at each Grade, the results of the study are reported in parts. This chapter reports only on the factors influencing reading literacy achievement at Grade 3 students of Primary School of South Australia and reports the result of analysis at the student level only.

Review of the literature

Literacy and language are fundamental to all forms of personal learning, intellectual growth in school or out of school. In today's global society, a literate population, particularly in Australia, is the most valuable asset for a nation's social and economic development. In contrast, being illiterate in a society is a serious threat both to economic performance and to social cohesion as concluded by the Organization for Economic Cooperation and Development (OECD, 1995, p.13):

> In recent years, adult literacy has come to be seen as crucial to the economic performance of industrialised nations. Literacy is no longer defined merely in terms of a basic threshold of reading ability, mastered by almost all those growing up in developed countries. Rather, literacy is now seen as how adults use written information to function in society. Today, adults need a higher level of literacy to function well; society has become more complex and low-skill jobs are disappearing. Therefore, inadequate levels of literacy among a broad section of the population potentially threaten the strength of economics and the social cohesion of nation.

With the rapid changes in the development of technology, literacy skills play an important role in the increased competitiveness and productivity that the national economy demands. Literacy is also essential for "innovation, for mobility and adaptation to change" (Lo Bianco & Freebody, 2001, p.vii).

In order to improve the quality of its people, a country needs to maximise the potential of its human, social, and material resources. People who are able to read and write are crucial to this effort. One way to improve the quality of people's lives is to make high quality schooling an imperative goal and all students must receive high quality teaching in order to raise the quality of learning for all people.

School quality is being questioned and it becomes a fundamental issue in most countries. This issue of quality is associated with the society and market demands. In order to respond proactively to this demand, the school and its stakeholders must work together to fulfil what society and the market expect from schools. As the main player in the process of learning, teachers are central to school improvement efforts directed at students' learning outcomes and achievement progress (National Inquiry into the Teaching of Literacy, 2005). Teaching is a process of transferring knowledge (Grave, 2000) and this requires people who have a high level of competence in their profession. Therefore, in order to improve the efficiency and the effectiveness of schooling, people who choose teaching as their career need to have high professional standards in their pre-service and in-service education and teaching practices as improving school quality depends on those people.

> Since the central aim of schooling is to generate and maintain efforts towards ongoing improvements in teaching and learning, equipping teachers with evidence-based teaching skills that are demonstrably effective in meeting the developmental and learning needs of the students for whom they have responsibility is vital. (National Inquiry into the Teaching of Literacy, 2005, p.4)

There is no aspect of teaching anywhere in the world that is more important than teaching literacy (i.e. reading, writing, speaking and listening, and viewing) as literacy is the foundation for other learning. Individuals who are able to demonstrate higher levels of literacy are more likely to succeed not only in their schooling, but are able to be employed and also are able to participate in social and economic activity. The emergence of sophisticated means of information and communication technologies and the rapidly changing nature of the tools, such as the computer, mobile phone, video cameras, text scanner, voice recognition and language translation technology slowly but surely demands people who are competent in complex multi-literacies (Cope & Kalantzis, 2000; Lo Bianco & Freebody, 2001).

Heckman (2000, 2005, cited in National Inquiry into the Teaching of Literacy, 2005, p.4), Nobel Prize winning economist, has supported these assertions and has given an overview of the economic aspects of human skills formation. He argued that literacy competence was an essential area of learning investment in the young, being a skill that begets many other skills (he calls it an index of “self-productivity”), because it constitutes a “key part of our capacity to increase our capacity”.

Since competence in reading and writing is the foundation of educational provision, the findings from local and international evidence-based research have identified effective pedagogical practices that maximise the learning and achievement progress of all students in literacy, and especially for those experiencing reading difficulties. Estimating the percentage of children who have learning difficulties and reading difficulties is difficult since there are no standard definitions of such difficulties. Nonetheless, it has been estimated that 80 per cent of students nominated by teachers as having learning difficulties are identified as having problems in mastering reading skills (Louden et al., 2000).

A study of reading literacy conducted by the Organisation for Economic Co-operation and Development (OECD) during 2000 and 2003 has indicated that, on average, the performance of 15-year-old students in Australian schools is higher than that of the majority of their counterparts in other OECD countries. However, 12 per cent (ACT, WA) to 28 per cent (NT) perform at a low achievement level and are not developing the literacy skills needed for further education, training and work, particularly indigenous students (35%) and males (17%) (National Inquiry into the Teaching of Literacy, 2005). The concept of ‘reading literacy’ in the OECD Program for International Student Assessment (PISA) emphasizes skill in using written information in situations that students may encounter in their life both at and beyond school. Thus, reading literacy is defined as:

> … understanding, using and reflecting on written texts in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society. (OECD, 2003, p.108)

Similarly, between 1975 and 1998, it has been reported that achievement in reading comprehension of 14-year old Australian students, with few exceptions, has remained constant (National Inquiry into the Teaching of Literacy, 2005). Rothman (2002, p.ix) noted that

> for some groups, there has been improvement, most notably for students from language backgrounds other than English. For other groups, however, results indicated a significant achievement gap. The most significant gap is between indigenous Australian students and all other students in Australian schools.

The report from the Australian Bureau of Statistics (ABS, 1997) suggested that approximately 20 per cent of 15-74 year-old Australians have been identified as having very poor literacy skills, with an additional 28 per cent who could be expected to have trouble in using many of the printed materials that might be encountered in daily life. A survey conducted by Masters and Forster (1997b) in the *National School English Literacy Survey* (1996) indicated that 27 per cent of students at Year 3 and 29 per cent of students at Year 5 did not meet minimum performance standards of reading required for effective participation in further schooling. Additionally, it was reported that in 2003, approximately 10 per cent of students at Year 3 and 12 per cent of students at Year 5 and Year 7 did not achieve the minimum *National Benchmarks for Reading* (MCEETYA, 2005).

By any criterion, these outcomes are unacceptable in terms of the educational, psychosocial wellbeing and life chances of these Australians, as well as the economic and social future of the nation. Dr Reid Lyon, Chief of the Child Development and Behavior Branch of the National Institute of Child Health and Human Development (National Institutes of Health, Bethesda, United States) noted:

> In our longitudinal studies, we have learned that school failure has devastating consequences with respect to self-esteem, social development, and opportunities for advanced education and meaningful employment. Nowhere are these consequences more apparent than when children fail to learn to read. Why? Simply stated, the development of reading serves the major foundational academic ability for all school-based learning. Without the ability to read and write, the opportunities for academic and occupational success are limited. Moreover, because of its importance, difficulty in learning to read crushes the excitement and love of learning, which most children have when they enter school. (Lyon, 2003, p.1)

In parallel with the central aim of schooling, it is vital to build educational fences by building capacity in teacher expertise and professionalism by ensuring that teachers are equipped with adequate professional development and teaching strategies that they need in teaching their students. Quality teaching requires deep knowledge of content and extensive knowledge of how students learn that content. It also requires pedagogical content knowledge; that is, knowledge about how to teach the content. In terms of teaching reading,

> quality teaching involves knowledge of how students learn to read, knowledge of how to assess reading proficiency and growth, and knowledge of how to use assessment information to apply the appropriate strategies from a repertoire of practices that are demonstrably effective for teaching students to read. (National Inquiry into the Teaching of Literacy, 2005, p.21)

In order to sustain quality teaching in reading, therefore, ongoing professional development is essential.

Some factors, such as students' background knowledge, motivation, gender, home background, home language, and geographical location may affect reading comprehension achievement. From the point of view of school systems, communities, teachers, students, and families, the factors influencing literacy achievement have been argued to be these factors that were statistically related to both success and failure in literacy learning (McDermott & Gospondinoff, 1981; Elley, 1994; Lietz, 1996).

Gender

A variety of international and local assessment programs of reading literacy consistently show gender differences in the subject, favouring female students (Elley, 1994; Masters, 1997; OECD, 2001; Mullis et al., 2003). The report of national surveys of literacy achievement among Australian school students in Years 3 and 5 have indicated that the mean literacy achievements of girls was higher at these year levels than those of boys and the differences were greater for writing and speaking than for reading (Masters & Foster, 1997). The results of national tests in some states in Australia such as Western Australia, Victoria and New South Wales suggested that there were significant gender differences in favour of girls in average literacy achievement (Collins, Kenway & McLeod, 2000). Moreover, the initial PISA results showed that the pattern of gender differences was consistent across countries. In every country, on average, girls reached a higher level of performance than boys (OECD, 2001). However, recent work by Lietz (2006 not in reference list) has suggested that there are aspects of these studies that warranted further critical examination.

Home language

In many studies, there were marked differences in school literacy achievement between students who were from a Non-English Speaking Background (NESB) and those from an English Speaking Background (ESB). Many studies have also shown the effects of home language on reading achievement and on school success (Turner, 2000). Moreover, ACER has reported that more than 50 per cent of students from a non-

English speaking home background living in Remote or Very Remote areas scored at only proficiency Level 1 or below in Reading in PISA 2000, while 26 per cent of students with an English speaking background performed at those levels (Cresswell & Underwood, 2004). Additionally, the results of PISA 2000 indicated that students whose home language differed from the language of the assessment did not perform as well, on average, as students who spoke the language of the assessment at home (Lemke, et al., 2001). The existence of the same language regularly used at home as in the testing program helped students to achieve well in English reading basic skills (Rosenthal, Baker & Ginsburg, 1983). Moreover, the results of comparing the reading and writing achievement of the NESB and ESB groups by the National Assessment of Educational Progress (NAEP) indicated that students who spoke a home language other than English, that was tested, were not learning to read and write as well as their English speaking background counterparts (Mullis & Jenkins, 1990). A number of previous studies had shown that students' linguistic development in a home language other than English had often advanced this relationship as providing the source of their lower performance at school (Rosenthal, Baker & Ginsburg, 1983; Au, 1998). This view argued that the reason why the students performed poorly was because they did not fully understand the lessons taught in the English-speaking classrooms (Rosenthal et al., 1983).

Geographical location

The studies about rural student achievement both in Australia and internationally have become of increasing interest. The interest in these studies is linked to the results of many rural studies in Australia that show the differences in educational performance between students from non-metropolitan areas and students from metropolitan areas (Cheers, 1990; HREOC, 2000; Doolan & Zimmer, 2002). Students from the non-metropolitan areas were "less likely to participate in schooling, more likely to be absent, less likely to complete the compulsory school years, less likely to complete Year 12 and less likely to participate in tertiary education and training" (Sidoti, 2000, cited in Doolan & Zimmer, 2002). In 2003, rural students performed with lower mean scores (515 and 516) than their urban counterparts (528 and 529) in both mathematical and scientific literacy, respectively (ABS, 2003). Some studies had suggested that geographical location significantly contributed to creating educational disadvantage for country students (National Board of Employment Education and Training, 1990; Doolan & Zimmer, 2002) when accessing suitable senior secondary and post-compulsory schooling. Although schools in rural and remote Australia were often equipped with adequate educational facilities, school children from these areas suffered under the effects of other factors.

School card holders

In South Australia, students from low-income families in financial need may be eligible for a so-called 'school card' benefit, which entitles them to receive assistance to pay school fees, obtain books and materials, and to participate in extracurricular activities such as attending camps and excursions. School card has in the past been used as an indicator of low socioeconomic status (SES), with disadvantaged schools being identified by the proportion of students who received the school card, although not without controversy (Rothman, 1997). Within the Department of Education, Training and Employment, measures have been taken recently to ensure that the provision of a school card was more related to family need. Throughout the analyses of the data, school card has been used as a measure of low socioeconomic status.

Research questions

The research questions guiding this study are: 1) Do girls perform better than boys in reading literacy achievement? 2) Do students who live in metropolitan area achieve higher performance in reading than students who live in rural area? 3) Do Non-Aboriginal Torres Strait Islander (NON_ATSI) students have better performance in reading than ATSI students? 4) Do students with Non-School Card Holder (NON-SCARD) achieve higher performance in reading than students with School Card? 5) What are student levels factors influencing Reading (literacy) achievement at Grades 3?

Methods and procedures

Instruments used in the study

The instruments available for this study are basically those developed by the Basic Skills Testing Program for assessing student achievement in Literacy and Numeracy. The Basic Skills Testing Program (BSTP) instruments consist of three major sections: (a) Student Questionnaire, (b) Literacy test, and (c) Numeracy test. However, in further discussion, the Numeracy test is not discussed in this section as this study simply analyses factors influencing student achievement in Literacy and Reading, in particular.

Literacy tests

The Literacy tests consist of two sub-tests tested at each grade level, namely: (a) the Language sub-test, and (b) the Reading sub-test. However, this study focuses only on the Reading sub-test.

The type of item employed in the Literacy tests are, in the main, multiple-choice items with four options, and the students have to choose the correct alternative by filling in the bubbles next to their answer. For the Reading sub-tests, the students are required to go through reading materials that are provided. These reading materials not only consist of texts but they also include pictures and diagrams that provide information about various things.

Student questionnaire

Students who participated in the Basic Skills Testing Programs (BSTP) were required to fill in the questionnaire before doing the tests. The participants had to provide information about their gender, age, and race, language spoken in the home, whether or not they were born in Australia and length of stay in Australia by putting a tick in the box provided.

Data sets

The main set of secondary data, namely South Australia Basic Skills Testing Program data obtained from the Department of Education and Children's Services (DECS) in South Australia are used in this study. The following section provides a brief description of the information contained in the data set.

South Australia BSTP data

The South Australia BSTP data set provides information that was collected each year from students of Grade 3, Grade 5 and Grade 7 in the Government Schools throughout South Australia who were required to undertake the BSTP. This study reports only the data set of Grade 3 students.

Student data

The BSTP data made available for these analyses are the years 2000 to 2002. There were about 9745 students from Grades 3 in approximately 632 primary schools in South Australia that participated in the Basic Skill Tests (BST). However, when descriptive statistics were run across the selected variables to check for missing data and to examine the frequency distribution, it was found that there were too many missing data for certain variables. Consequently, the data needed to be cleaned up in order for meaningful data to be available for use in the further analyses. After data were cleaned up, the number of students at Grade 3 reduced to 7887 students.

Preparation of the data for analysis

This section describes the predictor variables constructed from the South Australia BSTP data and the School Information data sets presented above.

Construction of the student-level variables

The results of the construction of South Australia BSTP data sets involve the student-level variables with the following names and acronyms in parentheses: Logarithm of distance from General Post Office (GPOLOG), Isolation index (ISOLAT), Metropolitan or urban school (METRO), Male or female student (GENDER), Australian or Torres Straight Island student (NON-ATSI), Student with a disability (NON-DISABIL), English language spoken at home (H_LANG), Language background English (LBE), Student school card holder (NON-SCARD), and Reading Rasch scaled score (READING). The information is summarised in Table 1. Very limited information about the student-level variables is available from other studies of South Australian BSTP data that are collected every year.

Proposition to be tested

The propositions advanced in this section relate to the data collected through the student questionnaire and achievement scores in reading obtained from the administration of the literacy tests. It has been suggested that the technique of path analysis can be employed to investigate factors (obtained from the questionnaire) that influence student achievement (obtained from the achievement test scores), and the examination of causal relationships using regression analysis (Keeves, 1988). It has been further argued that a causal relationship "is indicated by a unidirectional arrow from the determining variable to the variable dependent on it" (Keeves, 1988, p.724). The investigation of the following propositions is undertaken by employing path models. It is argued that the investigation of path models examining causal relationships such as the ones proposed in this study can be most readily carried out by employing the Partial Least Squares Path Analyses (PLSPATH 3.1) Program (Sellin & Keeves, 1994).

In this study, a causal path model is employed to analyse reading literacy achievement at Grade 3 level.

Table 1. Student-level variable for the study under investigation

Latent variable	Manifest variable	Source	Coding
RURAL	Logarithm of distance from GPO (GPOLOG), Isolation index (ISOLAT), Metropolitan or urban school (METRO)	DECS	0=metro, 1=rural
GENDER	Sex of student (SEX),	DECS	0=male, 1=female
NON-ATSI	Australian or Torres Straight Island student (ATSI)	DECS	0=ATSI, 1=non-ATSI
NON-DISABIL	Student disability (DISABIL)	DECS	0=disabled, 1=non-dis
LANG	English language spoken at home (H_LANG), Language background English (LBE)	DECS	0=non-English, 1=Eng
NON-SCARD	A school card holder (SCARD)	DECS	0=scard, 1=non-scard
READING	Standard score Rasch Scaled (SSCORE)	DECS	

Student level model for reading literacy achievement

Student level model presented in Figure 1 can be tested by employing path analysis. At Grade 3 level of the model, it is hypothesised that reading literacy achievement (READING) is directly influenced by six variables of interest, namely: (a) geographical location or urban rural (RURAL); (b) racial background of student (NON-ATSI); (c) sex of student (GENDER); (d) student disability, whether students are disabled or not (NON-DISABIL); (e) speaking English in the home (LANG); and (f) student is not a school card holder (NON-SCARD). In addition to the direct effect that the six variables have on reading achievement, the indirect effects of several variables are also hypothesised to be operating to influence reading achievement. Table 2 summarises the results.

Results of the analysis

Student's achievement in reading (READING)

It is hypothesised in the path model presented in Figure 1 that the student's achievement in reading (READING) is influenced by six Latent Variables (LVs), namely RURAL, NON-ATSI, GENDER, NON-DISABIL, LANG, and NON-SCARD. Of these six LVs, the inner model results indicate that there are four LVs that have significant effects on READING, namely RURAL, NON-ATSI, GENDER and NON-DISABIL. In addition, of these four LVs, two LVs also have indirect effects on READING. Since the focus of this study is to examine factors influencing reading achievement at the student level, the results of the inner model are presented in more detail in the following section. It is worth noting, however, that LANG and NON-

SCARD do not have significant effects on the achievement outcome in this study.

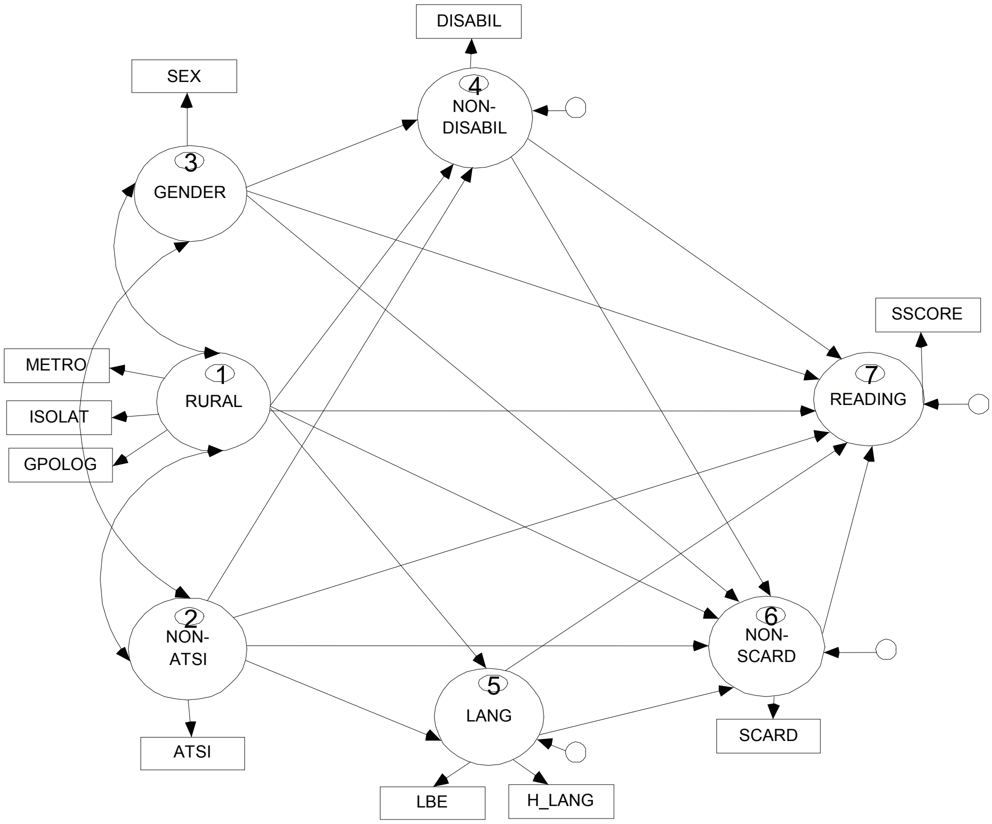


Figure 1. Student-level model for reading

The locality of school (RURAL)

The locality of a school or geographical location (**RURAL**) has a significant effect on reading achievement (**READING**) with a path coefficient of -0.12 and with a corresponding correlation of -0.13. The minus sign indicates that students from rural areas do not perform as well as students from metropolitan areas. Furthermore, the value of its tolerance is quite low (0.04). The results in Table 2 and Figure 2 indicate that **RURAL** has only a direct effect on **READING**. The results of this analysis indicate that geographical location or the locality of the school has a significant effect on student achievement. Students from metropolitan areas are more likely to have better performance in reading than students from non-metropolitan areas. This finding is consistent with the reports of many rural studies in Australia that show significant differences in educational performance between students from non-metropolitan areas and students from metropolitan areas (Cheers, 1990; HREOC, 2000; Doolan and Zimmer, 2002). However, this finding is not consistent with the finding indicated by educational research in Western Australia which reported that there is no significant difference between students attending rural schools and those attending metropolitan schools at the primary level in literacy abilities and skills (Ministerial Review of Schooling in Rural Western Australia, 1994).

Table 2. Inner model results for student-level factors influencing reading achievement at Grade 3

Variable	Beta	Correlation	Delta	Tolerance
NON-DISABIL	2 Pred-LVs	R-square = .013		
NON-ATSI	.08	.08	.01	.00
GENDER	.08	.08	.01	.00
LANG	2 Pred-LVs	R-square = .055		
RURAL	.16	.14	.03	.01
NON-ATSI	.19	.17	.03	.01
NON-SCARD	2 Pred-LVs	R-square = .020		
RURAL	-0.11	-.12	.01	.01
NON-ATSI	.07	.08	.00	.01
READING	4 Pred-LVs	R-square = .104		
RURAL	-0.12	-.13	.01	.04
NON-ATSI	.11	.15	.01	.05
GENDER	.10	.12	.01	.01
NON-DISABIL	.23	.25	.05	.01

Racial background of the student (NON-ATSI)

The second LV that has a direct, as well as an indirect effect on reading achievement is the racial background of the student (**NON-ATSI**). The results in Table 2 show that the path coefficient for the **NON-ATSI** is 0.11 with the corresponding correlation of 0.15 and a tolerance value that is quite low (0.05). The results in Table 3 and Figure 2 indicate that **NON-ATSI** also has an indirect effect on **READING** through **NON-DISABIL** even though the size of the indirect effect of **NON-ATSI** is relatively low (0.03). The results of this analysis indicate that Aboriginal or Torres Strait Islander (**ATSI**) students do not perform as well as non Aboriginal or Torres Strait Islander (**NON-ATSI**) students in reading. In addition, the influence of racial background on student's reading achievement is indirectly mediated by whether or not the students are disabled.

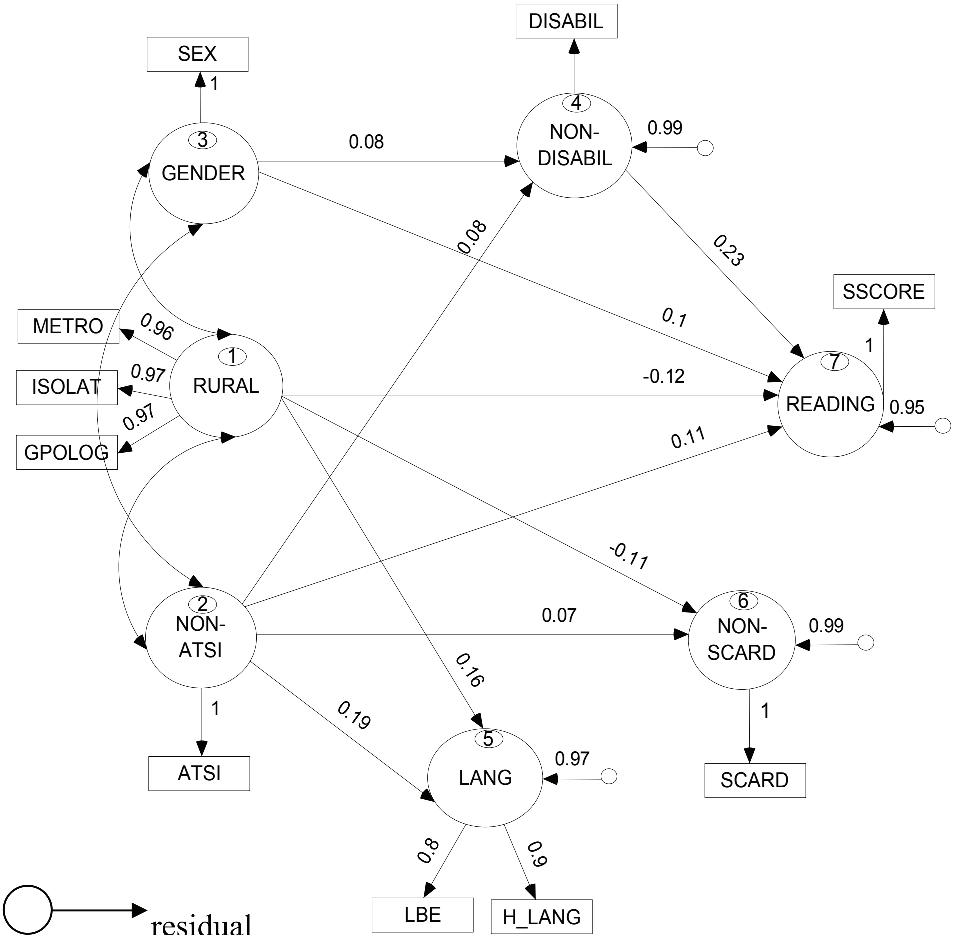


Figure 2. Student-level factors influencing student achievement in reading at Grade 3

Sex of Student (GENDER)

The results for the inner model indicate that the sex of student (**GENDER**) has direct as well as indirect effects on reading achievement (**READING**). Table 2 shows that the path coefficient for **GENDER** is 0.10 with the corresponding correlation of 0.12 and the tolerance value that is very low (0.01). In addition, **GENDER** also has indirect effects on **READING** through **NON-DISABIL**. Table 3 shows that the size of indirect effect of **GENDER** is 0.02. The results in Tables 2 and 3 suggest that there are significant gender differences in favour of girls in average reading achievement. In general, female students perform better than their male counterparts in reading. In addition, the influence of sex of student on a student's reading achievement is indirectly conveyed by whether or not the student is disabled. This finding is consistent with the results of national tests in some States in Australia such as Western Australia, Victoria and New South Wales (NSW) which suggested that there were significant gender differences in favour of girls in average literacy achievement (Collins, Kenway and McLeod, 2000). Moreover, the initial PISA results reported that in every country, on average, girls reached a higher level of performance than boys (OECD, 2001).

Table 3. Inner model effects for student level factors influencing reading achievement at Grade 3

Variable	Direct	Total	Indirect	Correlation	Fit
NON-DISABIL		R-square = .013			
NON-ATSI	.08	.08	-	.08	-
GENDER	.08	.08	-	.08	-
LANG		R-square = .055			
RURAL	.16	.16	-	.14	-
NON-ATSI	.19	.19	-	.17	-
NON-SCARD		R-square = .020			
RURAL	-.11	-0.11	-	-0.12	-
NON-ATSI	.07	.07	-	.08	-
READING		R-square = .104			
RURAL	-0.12	-0.11	.00	-0.13	-
NON-ATSI	.11	.13	.03	.15	-
GENDER	.10	.12	.02	.12	-
NON-DISABIL	.23	.23	-	.25	-

Student Disability (NON-DISABIL)

In this study, student disability (**NON-DISABIL**) has a direct effect on **READING**. The results in Table 2 indicate that **NON-DISABIL** has a significant effect on **READING** with the path coefficient of 0.23 and the corresponding correlation of 0.25. The tolerance value for **NON-DISABIL** is 0.01. Among the other LVs, **NON-DISABIL** has the greatest path coefficient and correlation coefficient with **READING**. This indicates that out of the LVs that influence reading achievement, **NON-DISABIL** has the strongest influence on **READING**. However, Table 3 indicates that **NON-DISABIL** has only a direct effect on **READING**. The results of this analysis indicate that student disability has a significant effect on student achievement. Non-disabled students are more likely to have better performance ratings in reading than students with a disability.

Discussion and conclusions

Generally, the results of student-level factors influencing reading achievement presented in this report are consistent with several past studies conducted in both Australia and overseas. PLS analysis was undertaken to examine the predictive power of the proposed model presented in Figure 1. This analysis showed that this model was slightly less powerful than the refined model referred to as Figure 2. However, the results of the analysis showed that all variables that were predicted to influence Reading Literacy Achievement (**READING**), namely **RURAL, NON-ATSI, GENDER, NON-DISABIL, LANG***,* and **NON-SCARD,** there are four LVs that had significant effects on **READING**, namely **RURAL, NON-ATSI, GENDER** and **NON-DISABIL.**

First, consistent with the results reported in this section, few studies have reported that students who attend metropolitan schools perform better in reading than students from rural areas. The NSELS study, conducted by Masters and Foster (1997), found that students at Grades 3 in major urban areas outperformed their counterparts in rural areas in reading (literacy). Similarly, Afrassa and Keeves (1999) found that students at Grades 3 in urban areas performed better than their peers in country areas in literacy with relatively small differences. Moreover, Elijio (2004) conducted a study using the data base PIRLS (Progress in International Reading Literacy Study), an international study of students' reading achievement, found that in the Eastern Europe students from rural communities performed at a lower level than the students from urban communities.

If locality has an influence on educational outcomes, there must be certain factors behind the locality that also contribute to the lower or higher reading (literacy) achievement of students. The first feature is the 'average home background situation of students in a school' or, in other words, the 'social and educational peer-environment' (Elijio, 2004, p.13). The second feature is related to the idea that 'the quality of education is not the same in cities or towns and country areas' (Elijio, 2004, p.13). Although this assumption is of concern, it is meaningful. Social and economical differences between the circumstances surrounding lives in metropolitan areas and non-metropolitan areas can easily influence the quality of education provided, both with regard to the school facilities and also with regard to the teachers' qualifications.

Second, several Australian studies have investigated the effects of Aboriginality on academic achievement and have reported findings that were consistent with the results of this study. Reading literacy study conducted by Masters and Foster (1997) reported that Aboriginal students had very low average levels of English literacy achievement, (3 to 4 year levels below students in the main sample in the National School English Literacy Survey). In South Australia, Afrassa and Keeves (1999) reported on the lower average literacy and numeracy levels of ATSI students at Grades 3. Moreover, Year 3 ATSI students were also found to have made little or no progress over the following two years. Another study (Rothman, 1999a) reported that Aboriginal students had relatively high rates of absence from school and this evidence would appear to be a factor in the lower literacy achievements of these groups.

Consequently, if Aborigines were to have equality and opportunity to gain access to education, something needed to be done. As suggested by Hughes (2007, p.3):

> Education reform is the key to ending Aboriginal deprivation. The number of 'homeland' Aborigines in SA is so small that providing decent, mainstream education would not be a financial burden.

Third, gender is a factor that influences both literacy achievement and numeracy achievement, with girls scoring higher on tests of reading comprehension and boys scoring higher on tests of mathematics (Rothman and McMillan, 2003). More importantly, Keeves (1995) reported that gender differences in achievement are influenced more by societal and curricular factors than by genetic factors, as has sometimes been assumed. Understanding gender gap in educational outcomes and understanding literacy and numeracy remains the primary goal of schooling in Australia. Masters (2003, p.1), the ACER's chief executive, said that

> it is imperative that researchers continue to examine the literacy and numeracy skills of Australian school students and understand better why some students achieve higher levels than other students.
>
> Understanding the influences on students' achievement in literacy and numeracy remains a major topic for education research in Australia. In turn, ensuring that all students, regardless of background are literate and numerate must be a primary goal for Australian educational policy makers.

Fourth, for student disability, the National Education Longitudinal Study of 1988 (NELS:88), conducted by the NCES (1997) investigated the influence of student disabilities on educational outcomes. The findings of NCES studies were consistent with the results presented in this chapter. NCES (1977) reported that students who identified themselves as disabled had higher levels of participation in remedial English and mathematic programs than did their non-disabled counterparts (32.1 per cent versus 18.7 per cent and 42.9 per cent versus 20.3 per cent, respectively). Moreover, disabled students achieved lower scores on reading and mathematics proficiency tests, and were more likely to drop out of school than their non-disabled counterparts. Consequently, these groups of students needed to have special education services and they needed to be served in more inclusive environments that might affect their awareness in a special program (NCES, 1997). It is worth noting, however, that not all children with disabilities need special school services. For example, a child with a purely physical disability who received the proper medical services for that disability might not require special services.

However, before accepting this potentially important result, it would seem essential to test these relationships (factors influencing reading literacy achievement) with more rigorous maximum likelihood estimation procedures rather than a more exploratory partial least squares regression analytical procedure.

John, you are really true teacher, helpful but encourage someone to work hard and keep smiling. Sometimes, you also make someone worry. "Keep working and keep writing", your words that I have never forgotten.

References

Adams, R. and Wu, M. (2002). *Program for International Student Assessment (PISA): PISA 2000 Technical Report.* Paris: OECD.

Afrassa, T. M. and Keeves, J. P. (1999). Factors influencing Years 3 and 5 students' basic skills test performance: A two level HLM analysis. Paper presented at the Research Expo 1999. March 20, 1999, Adelaide.

Cheers, B. (1990). Rural Disadvantage in Australia. *Australian Social Work*, 43(1), 5-13.

Collins, C., Kenway, J., and McLeod, J (2000). *Factors Influencing the Educational Performance of Males and Females in School and their Initial Destinations after Leaving School.* A Project Funded by the Commonwealth Department of Education, Training and Youth Affairs: Deakin University, University of South Australia.

Cresswell, J. (2004). *Location, Location, Location: Implications of Geographic Situation on Australian Student Performance in PISA 2000: ACER Research Monograph No 58*. Australian Council for Educational Research.

Development, OECD (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment.* Paris: OECD.

Development, OECD (2003b). *Reading for Change: Performance and Engagement Across Countries: Results from PISA 2000.* Paris: OECD.

Development., OECD (2001a). *Knowledge and Skills for Life: First Results from PISA 2000.* Paris: OECD.

Doolan, W. (2002). Australian rural students face severe disadvantage [Electronic Version]. Retrieved February 6, 2007 from http://www.wsws.org.

Elijio, A. (2004). Reading achievements in urban and rural communities: comparative analysis of the equity in education. Retrieved April 20, 2007, from http://www.iea.nl/fileadmin/user_ upload/IRC2006/

Elley, B. W. (1989). *Review of Research on Reading Literacy*. New Zealand: University of Canterbury.

Elley, W. B. (1994). *The IEA Study of Reading Literacy. Achievement and Instruction in Thirty-Two School Systems*. Oxford: Pergamon Press.

Jalal, F. and Sardjunani, N. (2005). *Increasing Literacy in Indonesia.* Paper commissioned for the EFA Global Monitoring Report 2006, Literacy for Life: UNESCO.

Keeves, J. P. (1988). Path analysis. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (1st Edn) (pp.723-731). Oxford: Pergamon.

Keeves, J. P. (1995). *The World of Learning: Selected Key Findings from 35 Years of IEA Research.* The Hague: The International Association for the Evaluation of Educational Achievement (IEA).

Kirsch, I. S., and Mosenthal, P.B. (1989). Understanding documents. A monthly column appearing in *The Journal of Reading*. Newark, DE: International Reading Association.

Kotte, D. & Lietz, P., & Lopez, M. M. (2005). Factors influencing reading achievement in Germany and Spain: Evidence from PISA 2000. *International Education Journal*, 6(1), 113-124.

Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y.Y., Roey, S., Williams, T., Kruger, T., and Bairu, G (2001). *Outcomes of Learning: Results From the 2000 Program for International Student Assessment of 15-year-olds in Reading, Mathematics, and Science Literacy*. U.S. Department of Education, National Center for Education Statistics (NCES).

Lietz, P. (1996). *Reading Comprehension Across Cultures and Over Time.* Munster/New York: Waxmann.

Lietz, P. (2006). Issues in the change in gender differences in reading achievement in cross-national research studies since 1992: A meta-analytic view. *International Education Journal*, 7(2), 127-149.

Literacy, N. (2005). Teaching Reading [Electronic Version]. Retrieved January 6, 2007 from www.dest.gov.au/schools/ literacyinquiry.

Lo Bianco, J. (2001). *Australian Literacies Informing National Policy on Literacy Education*. Language Australia.

Lyons, T., Cooksey, R., Panizzon, D., Parnell, A., and Pegg, J. (2006). *Science, ICT, and Mathematics Education in Rural and Regional Australia.* University of New England.

Masters, G. N. and Foster, M. (1997). *Mapping Literary Achievement: Results of the 1996 National School English Literacy Survey: Management*

Committee for the National School English Literacy Survey. Australian Council for Educational Research, Department of Employment, Education, Training and Youth Affairs.

McDermott, R. (1981). Social Contexts for Ethnic Borders and School Failure. In G. Trueba and K.H. Au, (Ed.), *Culture and the Bilingual Classroom: Studies in Classroom Ethnography* (pp. 212-230). Rowley, MA: Newbury House.

Ministerial Review of Schooling in Rural Western Australia. (1994). S*chooling in Rural Western Australia's Report*. Chairman, Derrick Tomlinson.

Mullis, I. V. S., Kennedy, A. M., Martin, M. O., and Siansbury, M. (2006). *PIRLS 2006 Assessment Framework and Specifications* (2nd Edition) Progress in International Reading Literacy Study.

National Board of Employment Education and Training. (1990). *Toward A National Education and Training Strategy for Rural Australians*. Commonwealth of Australia.

National Center for Education and Statistics (NCES). (1997). Profiles of Students with Disabilities as Identified in NELS:88. Retrieved April 20, 2007, from http://nces.ed.gov/

Organisation for Economic Co-operation and Development. (2003). *Literacy Skills for the World of Tomorrow-Further Result from PISA 2000*. Paris: OECD.

Purves, A. C. (1973). *Literature Education in Ten Countries: International Studies in Education II*. Stockholm: Almqvist and Wiksell.

Rosenthal, A. S., Baker, K., and Ginsburg, A. (1983). The effect of language background on achievement level and learning among elementary school students. *Sociology of Education,* 56(4), 157-169.

Rothman, S. and McMillan, J. (2002). *Influences on Achievement in Literacy and Numeracy*. Melbourne: ACER and the Commonwealth Department of Education, Science and Training.

Sellin, N. (1995). Partial least square modelling in research on educational achievement. In W. Bos and R.H. Lehmann, (eds) *Reflection on Educational Achievement: Paper in Honour of T. Neville Postlethwaite* (pp. pp. 256-267). Waxmann, New York.

Sidoti, C. (2000). Speech. Human Rights and Equal Opportunities Commission. Retrieved 20 December, 2006, from www.hreoc.gov.au/news_info/speeches/human_rights/sp_16.html

Turner, S. (2000). Comment on poor school funding, child poverty and mathematics achievement. *Educational Researcher*, 29(5), 15-18.

Weaver, C. A. and Kintsch, W. (1996). Expository text. In R. Barr, M. L. Kamil, P. Mosenthal and P. D. Pearson (Eds.), H*andbook of Reading Research* (Vol. 2, pp.230-245). Mahwah, NJ: Lawrence Erlbaum Associates.

14

Student level factors influencing mathematics achievement: A path analysis with latent variables from a cluster sample of Grade 8 Ethiopian students

Tilahun Mengesha Afrassa
Data Management, Department of Education and Children's Services South Australia

Factors influencing mathematics achievement

Previous research studies in Ethiopia have indicated that there are substantial differences between students in their achievement levels in school mathematics. Researchers have identified some of the student level factors that influence mathematics achievement. Derese, Wagner and Alemayehu (1990), Endalkachew (1990), Seleshi (1995) and Tadesse (1993), argued that students' attitudes towards mathematics were among the student level factors influencing mathematics achievement in Ethiopian schools.

Another student level factor that was found to influence the achievement of Ethiopian students was home background. Berhanu (1986), Damtew (1972), and Gennet (1991) reported that parent's level of education influenced the students' school performance. Furthermore, Teshome (1993) and Tilaye (1999) found that students from higher socioeconomic status homes were likely to perform better in school.

Berhanu (1986), Eshetu (1988), Teshome (1993), Behutiye and Wagner (1995) also reported that performance at school was negatively related to the size of family. The involvement of students in home activities was also found to be a factor that influenced negatively the achievement of Ethiopian students (Damtew, 1972; Daniel, 1995; Derese, Wagner & Alemayehu, 1990; and Gennet, 1991).

In this investigation from the findings of previous research studies, a model of student level factors influencing mathematics achievement at the Grade 8 level was developed and PLSPATH (Sellin, 1990) was chosen as an appropriate multivariate procedure to examine the hypothesised model. This chapter discusses the results obtained when the hypothesised model was tested by employing PLSPATH in the analysis of data from the Ethiopian Mathematics Study.

Data collection

The sample employed for data collection in this study was 1200 Grade 8 students who were selected from 40 government and non-government schools (30 students from each school) in the Addis Ababa Region of Ethiopia. The schools and the students were selected using a random sampling procedure with a probability proportional to the size of the school (pps sampling). All schools that were approached except one, agreed to participate and for the school that declined to take part a replacement school was randomly selected. In addition, the students who responded to the tests and questionnaires were selected at random by the investigator, with students who were absent from school at the time of testing being replaced by other students also selected at random.

The general information about the students' background was collected from those 1200 students who participated in this study using a General Information Questionnaire. Three types of information were collected from the students. The first type of information was about themselves, such as their date of birth, sex, place of birth and whether or not they spoke Amharic at home with their parents. The second type of information was about their schools and the learning of mathematics. This information included the number of students in their mathematics class, periods of mathematics each week, number of hours they studied mathematics each week, number of hours they gave to mathematics homework, number of hours they devoted to all homework, and was about their liking of mathematics as well as their mathematics test results. The third type of information collected from the students was about their home background, such as their father's occupation and education, mother's occupation and education, number of brothers and sisters, number of books at home, father's and mother's place of birth, and languages most often used in their home.

The purpose of the questionnaire was to obtain background information about each student in order to develop variables that would help to explain differences between students in their achievement in mathematics. Therefore, the three types of information that were collected from the students are considered in the following sections in order to examine which variables can explain differences between students in their mathematics achievement at the Grade 8 level in Ethiopian schools.

Method

In this study partial least squares path analysis procedures are employed to identify the student level factors that influence mathematics achievement. Partial least square path (PLSPATH) analysis "is a general technique for estimating path models involving latent constructs indirectly observed by multiple indicators" (Sellin, 1992, p. 398). It is useful in modelling educational and social systems for the purpose of causal explanation. Thus the PLSPATH procedure can be employed as a method of analysing path models that involve latent (indirectly observed) and manifest (directly observed) variables. The path model includes an inner model that specifies the hypothesised relationships among the latent variables (LVs) and an outer model that specifies the relationships between the LVs and the manifest variables (MVs) that are their indicators (Sellin, 1992).

PLSPATH identifies the optimal linear relationships between variables and provides estimates of the parameters of the model. Bukowski, Hoza and Boivin (1993) and Jacobs (1991) have argued that PLSPATH was an ideal procedure, because it provided an index of the adequacy of the model, showed the strength of each individual path in the model, and examined the direct and indirect relationships among variables. In a causal model certain variables are singled out as causes and others as effects. Consequently, the strengths of particular paths in the model show how strongly the linked variables are causally related to each other. PLSPATH provides a number of advantages that are most appropriate for this study. It is useful for displaying graphically the pattern of causal relationships among sets of observed and unobserved variables that influence the mathematics achievement level of students. PLSPATH is technically simple, quick to operate and does not require strict distributional assumptions.

Because PLSPATH employs a least squares regression procedure in analysis it does not require that variables are normally distributed for the analysis. However, this computer program retreats from formal significance testing because in so many data sets the assumptions of normality are violated, or the degree of departure from normality can not be determined because the samples employed can not be considered to be simple random samples.

Construction of PLSPATH model

The starting point in employing PLSPATH is to draw a path diagram of the model to be analysed (Falk, 1987). The diagram includes both the outer and the inner models and the hypothesised links between them. In this study the hypothesised model of student factors that are argued to influence mathematics achievement of Grade 8 students in Ethiopia has been developed prior to analysis. The model specifies the variables included in the analysis and its interrelationships are hypothesised. In the path diagram, the MVs (drawn as boxes) or observed variables form the outer model, while the LVs (drawn as ellipses) formed the inner model. The MVs and LVs employed in this study are shown in Figure 1 and in Table 1. In total, 21 MVs and 10 LVs are included in the outer and inner models respectively. The acronyms chosen for the MVs and LVs are intended to reflect their item content. The reader must keep in mind that these acronyms are employed throughout this paper.

Outer model

Figure 1 shows the outer model relationships of the hypothesised student level factors thought to influence the achievement in mathematics of students in Ethiopia. In developing the outer path model the investigator has the choice between two modes of weight estimation, called the outward mode and the inward mode (Sellin, 1992). Sellin states that the outward mode indicators assume that the MVs reflect the corresponding latent construct. An example of the outward mode is the set of views about mathematics teaching and schooling scales that are assumed to reflect the underlying views of each student. Inward indicators assume that MVs form or produce a latent construct as presented in Figure 1. PLSPATH employs a canonical analysis approach for estimating the outward latent constructs. The criterion for the minimum level of a factor loading for the inclusion of a MV in this study was chosen to be 0.30 (Campbell, 1996). However, other researchers have proposed other criteria such as 0.55 (Falk & Miller, 1992), or 0.40 (Harman, 1976; Keeves, 1992; Pedhazur, 1982).

Manifest variables with loadings below the predetermined cutting point are dropped from further analysis. PLSPATH uses a regression procedure to calculate the weights of the MVs that form a LV in the inward mode, and variables with weights below 0.07 are removed from further analysis in this large sample with 1200 cases as they indicate that the observed variables do not contribute greatly to form the LV (Sellin & Keeves, 1997; Sobolewski & Doran, 1996).

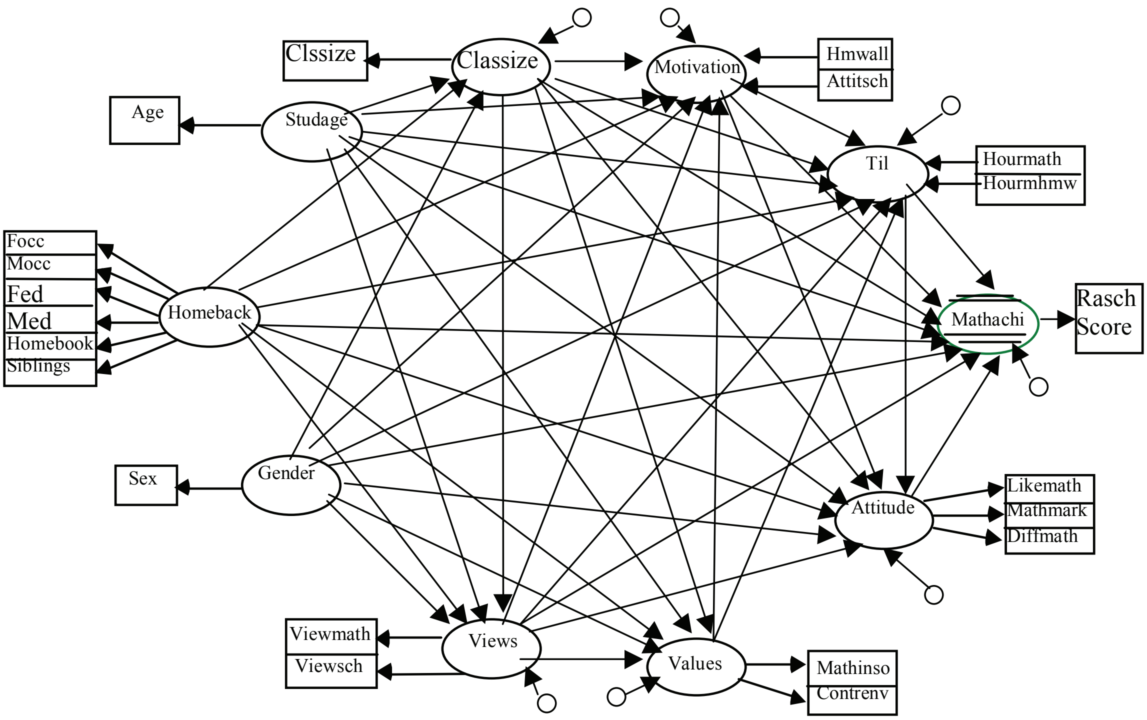


Figure 1. Outer and inner model relationships of the hypothesised model of student level factors influencing mathematics achievement

A simple measure of significance employed in this study is that the size of a path needs to be greater than twice its standard error. If the estimate of a path coefficient is less than twice its standard error, it is deleted from the final model. In order to obtain a robust model, the original model was successively refined to include only significant paths. In Figure 1 the outward mode is depicted by the arrows pointing from the LVs to their respective MVs, whereas the inward mode is shown by the arrows pointing from the MVs to the corresponding LVs.

The exogenous variables in this study with the exception of *Home background*, are defined with inward estimation, while the endogenous variables, except *Time in Learning* and *Motivation*, which employ the inward mode, are constructed with the outward mode including those constructs that consist of only one MV (see Figure 1).

Table 1. Latent and manifest variables employed in the PLSPATH analysis

Latent Variables	• Manifest Variables (Explanations in *italics*)
Homeback	Home background of student *Measuring the socioeconomic status of student's parents* • Focc (Father's occupation) • Mocc (Mother's occupation) • Fed (Father's level of education) • Med (Mother's level of education) • Homebook (Number of books in Home) • Siblings (Number of siblings)
Gender	Sex of student *Identifying whether the student is male or female* • Sex (Sex of student)
Ethnicity	Ethnic background of the student *Identifying the ethnic background of the student* • Fcntry (Father's country of birth) • Mcntry (Mother's country of birth) • Cntry (Student's country of birth) • Enghome (English spoken at home)
Studage	Age of student *Identifying student's age* • Age (Age of student)
Classize	Number of students in the class *Identifying the number of students in the student's classroom* • Clssize (Number of students in the class)
Views	Student's views about mathematics teaching and school and school learning *Measuring student's views on mathematics teaching, school, school learning* • Studpart (Student's participation)
Motivation	Motivation of student *Measuring the student's level of motivation* • Motiv1 (I need to do well in mathematics to get the job I want)[a] • Motiv2 (I need to do well in mathematics to please my parents)
Til	Time in learning *Assessing time used by student to learn and do mathematics home work* • Hourmath (Number of hours in a week used by student in learning mathematics) • Hourmhmw (Number of hours in a week used by student to do mathematics homework)
Attitude	Student's attitudes towards mathematics *Examining the attitudes of a student towards mathematics* • Likemath (Mathematics is student's best liked subject) • Mathmark (Best mark of the students is mathematics) • Diffmath (Student's attitudes towards ease of learning mathematics)
Mathachi	Mathematics achievement of student *Showing the mathematics achievement level of student* • Rasch Score (Rasch estimated scores)

[a]An example of an item is given in parenthesis.

Inner model

The inner model shows the causal links between the LVs. The positions of the LVs in the inner model are based on theoretical considerations.

Therefore, in this study of student level factors influencing student achievement in mathematics in Ethiopia, the findings of previous studies are employed as a starting point to hypothesise the causal links between constructs in the inner model (see Figure 1). In specifying the final structure of the inner model, it is recommended that direct paths with ß<0.07 are best removed, because such values showed an insignificant effect in the prediction of a relationship between two LVs. Hence, the larger the ß value the larger the effect in the path model. This estimation process is repeated successively until all nonsignificant paths have been removed.

The other criterion that is used to assess the strength of the final path model was the maximum variance explained (R^2) of the outcome variable, *Mathematics Achievement* (Mathachi). R^2 shows the variance of a construct explained when the preceding predictor variables are included in the analysis. Thus, the larger R^2, the more of the variance is explained.

Magnitude of effect and statistical significance

This study employs two approaches in the examination of the findings of analysis for significance. The first approach is to consider the size of an effect as assessed by a standardised regression coefficient or a factor loading, and the conventions employed are already discussed above. The second approach is carried out after the model is trimmed referred to here as the interim model, that is identified using the first approach. This involves undertaking jackknife repeated replications of the data under analysis by dropping one school group at a time. The mean value of each of the parameters and their standard deviation are calculated from the data provided. From the formula given by Ross and Wilson (1997, p. 666) the jackknife standard error associated with this mean can be calculated from the standard deviation, as a more appropriate error for the estimated parameters than is given either on the basis of either schools or students as the unit of analysis. Where an estimated parameter exceeds twice its jackknife standard error, which corresponds approximately to a five per cent level of statistical significance, it is considered to be of sufficient magnitude to be reported in the final model.

Trimming the model

The modification, trimming or deletion of variables and paths in the path model involves the removing of all paths not contributing to the LVs. The deletion or the removal of paths includes both the outer and inner model. As a result of the trimming procedure some manifest and latent variables are removed from further analysis.

Results of the PLSPATH analysis

Tables 2 and 3 show the outer model and the inner model results for the PLSPATH analysis. Twenty one MVs and ten LVs are included in the model. Thus the results of the PLSPATH analyses are presented in two parts. The first part gives the results for the outer model and the second part discusses the inner model.

Outer model results

In the following discussion the weights and the factor loadings, jackknife mean, jackknife standard errors, and the communality values of each MV within a construct are discussed with respect to the LV to which that MV contributes. The weights (ßs) are initially considered significant if their values are ß≥0.07, while the factor loadings (*l*s) are initially significant if *l* ≥0.30. The index employed for measuring the strength of the outer model is the average of the communalities of the MVs (Falk, 1987). Furthermore Falk argued that the higher the average of the communalities the better the outer model, and an average value of the communalities of 0.30 is generally be considered too low.

The other important assessment of significance considered in this study involves statistical significance testing. A variable is considered to be statistically significant if its estimated value is greater than twice its standard error. Variables that fail to satisfy this criterion are removed from the refined model.

In Tables 2 and 3 effects are recorded that satisfy the requirements of 0.30 for loadings or 0.07 for path coefficient and weights but may be less than twice their standard errors. These effects are marginal and are not recorded in the final model presented in Figure 2. However, they are of interest because they contribute to indirect effects in the model that are discussed in the text.

Home background (*Homeback*): Table 2 indicates that *Homeback* is constructed in the outward mode and reflected in five MVs. One MV, namely *Mocc* was deleted from the analysis, since the factor loading (for outward mode) of this variable was 0.14 which was below the critical value of 0.30 (Campbell, 1996). This indicates that Mocc failed to contribute in a meaningful way to the latent variable *Home background.* Hence, apart from *Mocc* all other MVs are well suited to reflect a student's home background or socioeconomic status. The communalities in Table 2 show that all the remaining MVs contribute to this construct. Communality can be defined as the squared loading for a MV. It is a measure of the explained variance of a particular MV with respect to the LV it reflects (Sellin, 1990). Consequently, the variance explained by each MV is from 12 per cent by *Siblings* to over 70 per cent by *Fed* show that all the MVs contribute to this construct.

Gender: The sex of the student is considered to indicate *Gender*. Thus, this LV comprises a single MV. Because it is in unity mode the loading and the communality are each unity.

Student age ***(Studage):*** The age of the student is used for this LV which comprises just a single MV.

Table 2. Outer Model Results

Variable	Weight/Loading[a]	JKMEAN	JKSE	Communalities
Homeback[o]				
Focc	0.38	0.41	0.06	0.15
Mocc	Deleted			
Fed	0.84	0.83	0.02	0.70
Med	0.83	0.83	0.02	0.69
Homebook	0.71	0.70	0.06	0.50
Siblings	-0.34	-0.37	0.06	0.12
Gender[u]				
Sex	1.00	1.00	0.00	1.00
Studage[u]				
Age	1.00	1.00	0.00	1.00
Classize[u]				
Clssize	1.00	1.00	0.00	1.00
Views[o]				
Viewmath	1.00	1.00	0.00	1.00
Viewsch	Deleted			
Values[o]				
Mathinso	0.81	0.80	0.06	0.66
Contrenv	0.78	0.79	0.02	0.61
Motivation[i]				
Attitsch	1.00	1.00	0.00	1.00
Hmwall	Deleted			
Time in Learning (Til)[i]				
Hourmath	0.98	0.98	0.06	0.97
Hourmhmw*	0.18	0.18	0.12	0.04
Attitude[o]				
Likemath	0.53	0.57	0.06	0.28
Mathmark	0.51	0.57	0.06	0.26
Diffmath	0.81	0.76	0.06	0.65
Mathachi[u]				
Rasch sore	1.00	1.00	0.00	1.00
Mean Communalities				0.65

i = Inward mode; o = Outward mode; u = Unity mode; a = weight for inward mode and loading for outward mode; JKSE= jackknife standard error; JKMEAN = Jackknife mean value;
* = deleted from the final model because of insignificance of estimate

Table 3. Inner Model Effects (On)

Variable	$JKMEAN_1$	$JKSE_1$	Di	Id	Total	r	R^2	$JKMEAN_2$	$JKSE_2$
Classize									
*Homeback**	0.08	0.06	0.08	-	0.08	0.08			
Views							0.01	0.01	0.01
*Homeback**	0.09	0.06	0.09	-	0.09	0.09			
Gender	0.07	0.02	0.07	-	0.07	0.07			
Values							0.07	0.07	0.02
Homeback			-	0.03	0.03	0.03			
Gender			-	0.02	0.02	0.02			
Views	0.27	0.06	0.27	-	0.27	0.27			
Motivation							0.22	0.22	0.06
Homeback			-	0.03	0.03	0.02			
Gender			-	0.02	0.02	0.03			
Views	0.23	0.06	0.23	0.09	0.33	0.33			
Values	0.35	0.06	0.35	-	0.35	0.41			
Til							0.16	0.16	0.09
Homeback	0.23	0.06	0.23	0.02	0.25	0.25			
Classize	0.31	0.12	0.31	-	0.31	0.32			
Attitude							0.21	0.21	0.06
*Homeback**	0.07	0.06	-	0.02	0.09	0.09			
Gender			-	0.02	0.02	-0.08			
Views	0.09	0.02	0.09	0.15	0.24	0.24			
Values	0.20	0.06	0.20	0.10	0.30	0.34			
Motivation	0.28	0.02	0.28	-	0.28	0.40			
Mathachi							0.13	0.14	0.06
Homeback	0.13	0.06	0.14	0.03	0.17	0.19			
*Gender**	-0.08	0.05	-0.08	0.003	-0.08	-0.10			
Studage	-0.12	0.02	-0.11	-	-0.11	-0.16			
Classize	-0.17	0.09	-0.18	0.04	-0.14	-0.14			
Views			-	0.04	0.04	0.05			
Values			-	0.05	0.05	0.10			
Motivation			-	0.05	0.05	0.02			
Til	0.12	0.06	0.11	-	0.11	0.11			
Attitude	0.20	0.06	0.17	-	0.17	0.21			
Mean R^2							0.12		

Di = Direct effect; Id = Indirect effect; r = Correlation; R2 = R-square; JKMEAN1 = jackknife mean of the direct effects; JKMEAN2 = jackknife mean of the R-square; JKSE1 = jackknife standard error of the direct effects; JKSE2 = Jackknife standard error of the R-square;
* = deleted from the final model because of insignificance of estimate

Class size *(Classize)***:** The number of students in class is used for this LV, *Classize*.

Views about Mathematics *(Views)*: This construct operates in the outward mode and two MVs, namely, *Viewmath* and *Viewsch* reflect this construct. The former is a scale measuring the student's *Views* about the way their mathematics teachers teach mathematics and the latter was a scale measuring their *Views* about their schools and school learning. However, the latter was dropped from further analysis because the loading was below the critical value of 0.30. Therefore, this LV involved only one MV namely, *Viewmath*.

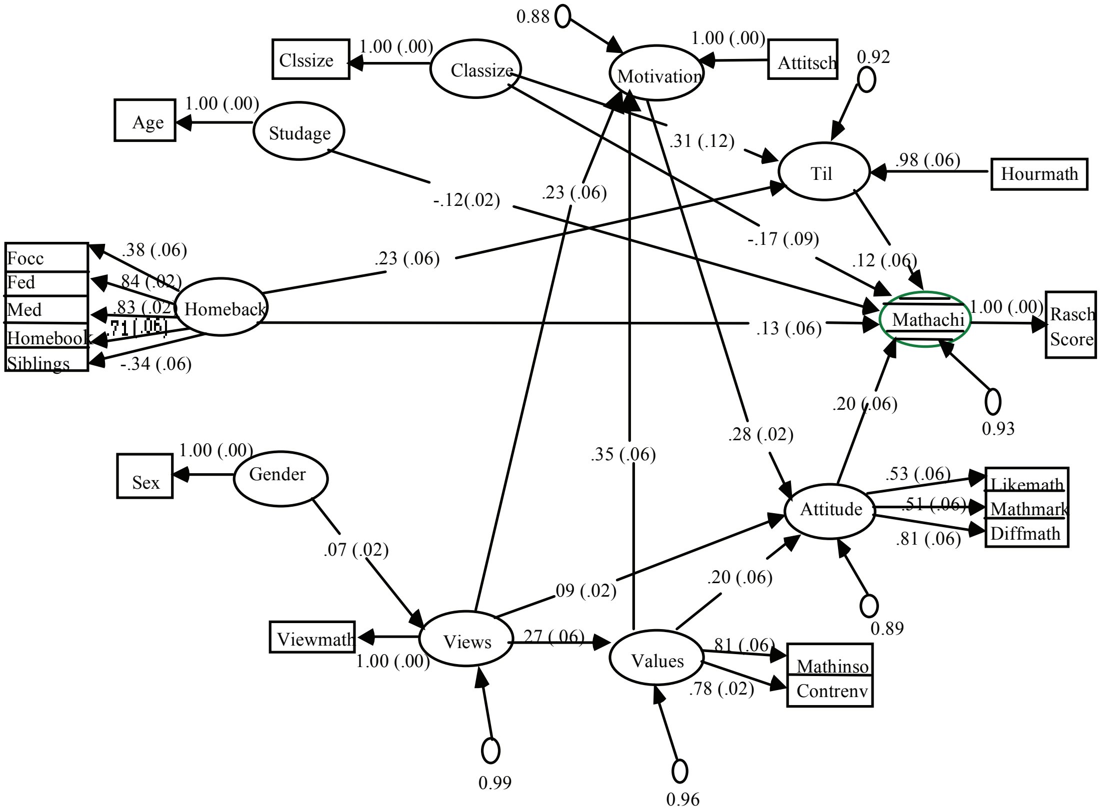


Figure 2. Estimates of magnitudes of student level factors influencing Mathematics Achievement (standard errors of estimates are recorded in parenthesis)

Values about Mathematics *(Values*)**:** *Mathinso* is a scale measuring student's Attitudes towards the Place of Mathematics in Society and *Contrenv* is a scale measuring student's Attitudes towards Control of the Environment. Both MVs reflect this LV and the LV operate in outward mode. It can be seen in Table 2 that these two MVs have high loadings (>0.70) on the latent variable *Values*. Therefore, it is possible to conclude that these two MVs are strong and sound contributors to the LV *Values*.

Motivation towards school learning (*Motivation*)**:** Two MVs, namely, *Attitsch* (a scale measuring student's Attitudes towards School and School Learning) and *Hmwall* (Hours taken by the student to do all his/her homework) were hypothesised to form this LV in the inward mode. The former MV involved the student's attitudes towards their schools and school learning and the latter MV involved the time taken by a student to do all his or her homework. However, the latter variable was dropped from further analysis because the weight fell below the critical value of 0.07. Thus, this LV was formed by only one MV namely, *Attitsch.*

Time in Learning *(Til)***:** The MVs *Hourmath* (0.98) and *Hourmhmw* (0.18) formed this LV. *Hourmath* involves the number of periods the student have in a week to learn mathematics, while *Hourmhmw* was the number of hours taken by the student to do his or her mathematics homework in a week (see Table 2). However, the latter MV was deleted from the refined model because it was not significant.

Attitudes towards Mathematics *(Attitude*)*: Likemath, Mathmark,* and *Diffmath* (a scale measuring student's attitudes towards Difficulty of Learning Mathematics) contribute through the reflective mode to this construct. The analysis shows that *Diffmath* (0.81) is the strongest contributor to reflecting the LV *Attitude*, while *Likemath* and *Mathmark* have effects given by the loadings of 0.53 and 0.51 respectively.

Mathematics Achievement *(Mathachi*)**:** The Rasch scaled mathematics test scores of students are used to indicate this construct. Thus, this LV comprises a single MV.

In summary, for the outer model, among the 21 hypothesised MVs that were hypothesized to contribute to the ten constructs, four MVs were removed from further analysis because of lack of fit in the path model and the remaining 17 MVs contributed substantially to the ten LVs. The average of the communalities of the MVs, which is considered as a measure of the strength of the outer model, is 0.65, and satisfactorily high.

Inner model results

Table 3 shows the jackknife mean, jackknife standard error, direct, indirect and total effects, correlations, and R^2. There are ten LVs in the inner model, and the results obtained from the analyses of these LVs are presented in Figure 2.

Among the ten LVs, *Homeback*, *Gender* and *Studage* are antecedents, which mean that they are exogenous variables and not influenced by any other LV. Therefore, the discussion in this section considers only those seven LVs that are endogenous variables that are hypothesized to be influenced by other LVs in the model. The criterion employed for measuring the strength of the inner model is the average multiple R^2 for the endogenous variables in the model (Falk, 1987). Falk argued that the larger the average multiple R^2 the better was the inner model.

Class size *(Classize*) provides information about the number of students in the mathematics classes which are obtained from each student. Three variables were hypothesised to influence this construct, however, none of them were found to be of sufficient magnitude for inclusion in the interim model.

Views about mathematics (*Views*): The *Views* of students about mathematics were hypothesised to be influenced by four LVs. Among these factors only *Gender* and *Homeback* are found to be of significant magnitude for inclusion in the interim model (see Table 3 and Figure 2). However, the direct effect of *Homeback* was not twice its standard error. Its standard error that was estimated by the jackknifing procedure was 0.06. Consequently, the path was not considered to be significant, and it was removed from the refined model.

Values about mathematics (*Values*): The values of students about mathematics were hypothesised to be influenced by *Homeback*, *Gender*, *Studage*, *Classize* and *Views*. Among these factors only *Views* influenced *Values* directly. Students who express stronger *Views* value mathematics more than those students who express weaker *Views* about mathematics. Table 3 shows that *Homeback* (0.03) and *Gender* (0.02) only weakly and indirectly influence this particular LV. These two factors indirectly influenced the values of students through their *Views*, which is a mediating variable.

Motivation towards mathematics (*Motivation*): The *Motivation* level of students towards mathematics was hypothesised to be influenced by *Homeback*, *Gender*, *Studage*, *Classize*, *Views* and *Values*. *Views* (0.23) and *Values* (0.35) showed a direct influence on *Motivation*. Table 3 indicates that *Homeback* (0.03) and *Gender* (0.02) indirectly influenced this LV through *Values*. Therefore, these results reveal that students' *Motivation* towards mathematics is directly and largely influenced by their *Views* and their *Values* and only indirectly and weakly by *Homeback* and *Gender*.

Time in Learning (*Til*): *Homeback*, *Gender*, *Studage*, *Classize*, *Views*, *Values* and *Motivation* were factors which were expected to influence *Til*. However, the results of the analysis reveal that only *Homeback* and *Classize* influenced this construct (see Table 3 and Figure 2). Table 3 indicates that *Homeback* (0.02), also influence *Time in Learning* indirectly through *Classize.* Thus the total effect of Homeback was 0.25 while the total effect of *Classize* was 0.31. Therefore, these results reveal that the time in which students are involved in learning mathematics is influenced largely by their home background and the number of students in the class.

Attitudes towards Mathematics (*Attitude*): The attitudes of students towards mathematics were hypothesised to be influenced by *Homeback*, *Gender*, *Studage*, *Classize*, *Views*, *Values*, *Motivation* and *Til*. The results of the PLSPATH analysis indicate that *Views* (total effect =0.24) show both direct and indirect influences on *Attitude* (see Table 3 and *Figure 2*). Those students who express stronger views about mathematics also held more positive attitudes towards the subject (see Table 3 and Figure 2). Likewise *Values* have a direct effect (0.20) and indirect effect (0.10) on attitude with students expressing stronger values also demonstrating more positive attitudes towards mathematics than those students who express weaker values. While, *Motivation* influences the criterion variable only directly (0.28), *Gender* (0.02) weakly influenced indirectly this particular LV of *Attitude* through *Values*, *Views* and *Motivation*.

Therefore, this analysis shows that students' *Attitudes* towards mathematics are directly influenced by their *Views*, *Values*, and *Motivation* and their *Attitude* towards mathematics is only weakly and indirectly influenced by the *Gender* of the students.

Mathematics Achievement (Mathachi): The mathematics achievement level of students was hypothesised to be influenced by nine LVs. The result of the PLSPATH analysis demonstrate that Home background (*Homeback*) influences *Mathachi* directly (0.14) and indirectly (0.03) through *Attitude*. The total effect is 0.17, while the correlation is 0.19. The analysis shows that students from a family of higher socioeconomic status background have higher achievement levels in mathematics than their classmates from a lower socioeconomic background.

Gender is another factor that influences Mathematics Achievement directly and indirectly. The direct effect was -0.08 and the indirect effect was very weak (0.003), while the total effect remained as -0.08. Thus *Gender* negatively influenced mathematics achievement at the Grade 8 level in Addis Ababa. The negative sign indicates that boys are higher achievers in mathematics than girls. However, the path was deleted from the final model because the path was not significant, its estimate was marginally below twice its standard error (see Table 3, *se* = 0.05). This finding is consistent with the findings reported by Kassahun and Kedir (2006) and Seleshi (2001). The authors reported that gender difference in mathematics achievement favouring boys began to be observed at primary school level. While the findings reported by National Organisation for Examinations (2000, 2001) and Adugna (2004) contradicts the present finding, the authors reported significant gender differences in mathematics achievement at primary school level in favour of boys. However, the authors did not control for other student level variables that influence the achievement of students in mathematics.

Time in Learning is one of the LVs that affect positively the achievement of mathematics students in Ethiopia. This variable influences mathematics achievement directly (0.11). This observation suggests that students who spend more time in learning mathematics are higher achievers in mathematics than those students who spend less time in learning mathematics.

Attitude towards Mathematics (Attitude) influences mathematics achievement only directly and its effect is 0.17. The evidence shows that students who hold more positive attitudes towards mathematics are likely to achieve at a higher level in mathematics than their classmates who express less favourable attitudes. This observation is also consistent with previous studies undertaken in Ethiopia (Andualem, 2007; Endalkachew, 1990; Kebede, 2007 Tadesse, 1993; Seleshi, 1995).

Views, *Values* and *Motivation* also influenced mathematics achievement of students indirectly (see Table 3). The factors that have indirect paths on *Mathematics Achievement* that are mediated by *Attitude* are *Views* (0.04), *Values* (0.05) and *Motivation* (0.05).

The overall strength of the model is given by the mean of the R^2 values of the endogenous variables which is 0.12. It must be acknowledged that the variance explained both in the criterion variable of mathematics achievement (0.13) and in the model as a whole is not large. However, there are five factors that are found to have recognisable direct effects. In part the low proportion of the variance explained in the criterion variable must be considered to be consequence of the fact that the students involved in this investigation are a selected group who were more uniform in their views, values and attitudes than students would be if the complete age cohort remained at school. Thus the participating group from which the sample was drawn suffered from selection bias in so far as it is estimated that the group only represents approximately 30 per cent of the cohort from which it was derived.

Conclusions

In this study the data obtained through a General Information Questionnaire from a sample of Ethiopian Grade 8 students was analysed by employing path analysis procedures. Twenty-one MVs and nine LVs are hypothesised to influence the mathematics achievement levels of Ethiopian students. The results of the path analysis show that four out of 21 MVs had to be removed from further analysis, because these variables did not contribute to the outer model. While the results of the inner model analyses indicated that among the nine LVs hypothesised to influence the mathematics achievements of Ethiopian students, five of them prove to be factors that influenced mathematics achievement substantially, at the Grade 8 level in Ethiopia. These factors that influenced mathematics achievement were *Home Background, Student Age, Class Size, Time in Learning* and *Attitudes* to*wards Mathematics,* while *Views, Values* and *Motivation* influence Mathematics Achievement only had an indirect influence.

In conclusion, the main findings of the present study are:

a) students from a family of higher socioeconomic status background have higher achievement levels in mathematics than their classmates from a lower socioeconomic status background;

b) there is only a marginally significant sex difference effect found on mathematics achievement;

c) younger Grade 8 students are higher achievers in mathematics than older students;

d) students from smaller class size groups are higher achievers in mathematics than students from larger class groups;

e) students who spend more time in learning mathematics are higher achievers in mathematics than those students who spend less time in learning mathematics; and

f) students who express more positive attitudes towards mathematics are likely to achieve at a higher level in mathematics than their classmates who express less favourable attitudes.

Therefore, mathematics teachers need to provide more help and support to their students who are from lower socioeconomic status backgrounds and to older students to help them to improve their level of mathematics achievement. They also need to help students to develop more positive attitudes towards mathematics and to spend more time in learning mathematics, because the results of this study show that the attitudes of students towards mathematics and the time spent on learning the subject are found to be factors that influence the mathematics achievement of students.

Educational officers, school principals and government officials need to find ways and means of reducing the number of students in mathematics classes, since the present findings show that students from smaller class groups are higher achievers in mathematics than students from larger class groups. Likewise, school principals need to consider both time spent in the classroom and time spent in homework if they are to raise the levels of achievement in mathematics in their schools. Further research needs to be undertaken to obtain information on the optimal number of students in mathematics classes in Ethiopia.

References

Adugna Gangaru (2004). Gender Difference in Mathematics Achievement at Second Cycle Primary School.. Unpublished MA Thesis, Addis Ababa University, School of Graduate Studies.

Andualem Tamiru (2007). Perception of Gender Difference in Mathematics Achievements: the Case of Jimma University. Unpublished MA Thesis, Addis Ababa University, School of Graduate Studies.

Behutiye, N. and Wagner, J. (1995). Some aspects of Primary School Pupils Cognitive Development and their Relationship to Scholastic Achievement. Comprehensive Report. Addis Ababa,:Curriculum Evaluation and Educational Research Division, Institute for Curriculum Development and Research, Addis Ababa.

Berhanu Beyene. (1986). An Investigation of the Efect of Family Background on Academic Success of Students in Menelik II Secondary School and its Psychological Implications. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Bukowski, W. M., Hoza, B. & Boivin, M. (1993). Popularity, Friendship, and Emotional Adjustment During Early Adolescence. *New Directions for Child Development,* 60, 23-37.

Campbell, J. R. (1996). *PLSPATHPATH Primer.* (2nd ed.), New York: St. John's Press.

Cheung, K. C. and Keeves, J. P. (1990). The Problems with Current Analytical Procedures. I*nternational Journal of Educational Research*, 14 (3), 233-244.

Damtew Gebru. (1972). Certain Social Factors Affecting Scholastic Achievement of fifty Grade 5 Pupils from Menelik II School. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Daniel Desta.(1995). School attendance and achievement among boys and girls: the case of selected grade six students in Addis Ababa. *Proceedings of the National Workshop on Strengthening Educational Sesearch,* Addis Ababa University, Institute of Educational Research, pp. 253-264.

Derese Mengistu, Wagner, J. and Alemayehu Minas. (1990). *Factors Affecting Achievement of Lower Primary School Pupils.* Addis Ababa: Institute of Curriculum Development and Educational Research.

Endalkachew Wolde Mariam. (1990). Factors that Affect Students' Academic Performance in some selected Senior Secondary Schools in Addis Ababa. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Eshetu Alemu. (1988). Comparative study of Family Back-ground of High Achievers and Low Achievers at Entoto Technical, Vocational and

Academic School. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Falk, F R. (1987). *A Primer for soft modeling*, Berkeley: University of California, Institute of Human Development.

Falk, F R. and Miller, N. B. (1992). *A Primer for Soft Modeling,* Akron, Ohio: The University of Akron Press.

Gennet Zewdie. (1991). Women in Education: A study of Academic Performance and Participation of Female Students in the High Schools of Addis Ababa. Unpublished manuscript, Addis Ababa: Addis Ababa University, Faculty of Education, Department of Business Education.

Harman, H H. (1976). *Modern Factor Analysis,* Chicago, IL: University of Chicago Press.

Jacobs, J.E. (1991). Influence of Gender Stereotypes on Parent Child Mathematics Attitudes. *Journal of Educational Psychology*, 83 (4), 518-527.

Kassahun Tilaye and Kedir Bedru (2006). Girls' Performance in Mathematics in Upper Primary Schools of Addis Ababa. Addis Ababa.

Kebede Tefera (2007). Comparative Study of Private and Government Secondary School Students' Attitudes toward Mathematics and their Achievement: the Case of Adama City. Unpublished MA thesis, Addis Ababa University, School of Graduate Studies.

Keeves, J. P. (1992). Changes in Science Education and Achievement , 1970-1984, Oxford: Pergamon Press.

National Organisation for Examinations (2000). Annual Magazine No1, May 2000. Addis Ababa: BSSP.

National Organisation for Examinations (2001). Annual Magazine No2, May 2001. Addis Ababa: BSSP.

Noonan, R. D. & Wold, H. (1988). Partial Least squares. In John P Keeves eds., *Educational Research Methodology and Measurement: An International Handbook.* Oxford: (pp. 710-716), Pergamon Press.

Pedhazur, E. J. (1982). Multiple Regression in Behavioral Research: Explanation and Prediction, New York: Holt, Rinehart, and Winston.

Ross, K. N. & Wilson, M. (1997). Sampling Errors in Survey Research. In J P Keeves (eds.), *Educational Research, Methodology and Measurement . An International Handbook (*2nd ed.), Oxford: (pp. 663-670), Pergamon Press.

Seleshi Zeleke. (1995). Gender Differences in Mathematics Achievement as a Function of Attitudes in Grades 8 through 11 (in Northern Shewa Region). Unpublished Masters thesis, Addis Ababa University, School of Graduate Studies.

Seleshi Zeleke. (2001). Gender Differences in Mathematics Performance in Elementary Grade: Implications for Women's Participation in Scientific and Technical Education. *East Africa Social Science Review,* 17, 109-127.

Sellin, N. & Keeves, J. P. (1997). Path analysis with latent variables. In J P Keeves, ed., *Educational Research, Methodology, and Measurement. An International Handbook* (2nd ed.), Oxford: Pergamon, pp. 633-640.

Sellin, N. (1989). PLSPATHPATH Version 3.01 Application Manual, Hamburg: West Germany.

Sellin, N. (1990). *PLSPATHPATH Version 3.01 Program Manual,* Hamburg: West Germany.

Sellin, N. (1992). "Partial Least Squares Path Analysis". In J P. Keeves, ed., *The IEA Technical Handbook*. The Hague, The Netherlands: The International Association for the Evaluation of Educational Achievement (IEA), pp.397-412.

Sobolewski, S. J. & Doran, R. L. (1996). Replication of a Path Analysis Model of Secondary Physics Enrollments: 20 Years Later. *Journal of Research in Science Teaching,* 33(5), 501-512.

Tadesse Getaneh. (1993). In-School Factors Affecting the Attitudes of Students towards Learning in selected Senior Secondary Schools in Eastern Shoa. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Teshome Abebe.(1993). A Comparative study of Academic Achievement of Students brought up under SOS Children's village and under Parental Care at Higher 23 Secondary School. Unpublished BA thesis, Addis Ababa University, Faculty of Education.

Tilaye Kassahun (1999). The Problems of Secondary School Dropouts in East Gojjam. *The Ethiopian Journal of Education* XIX (1), 65-93.

15

Mismatch between cognitive and affective outcomes of mathematics learning in Macao: Seeing school system through PISA

Kwok-cheung Cheung
University of Macao, China

The Problem

The PISA index of *sense of belonging at school* (*BELONG*) was derived from students' reported agreement that school is a place where: (i) I feel like an outsider (or left out of things); (ii) I make friends easily; (iii) I feel like I belong; (iv) I feel awkward and out of place; (v) Other students seem to like me; and (vi) I feel lonely (OECD, 2004a, p.309). After item response theory (IRT) scaling, positive values indicate positive feelings about the students' school. Hong Kong and Macao, both special administrative regions of People's Republic of China, ranked bottom (*BELONG* index = -0.061) on this affective index of schooling (see OECD, 2004a, p.368 for the *BELONG* league table; see also OECD, 2005, pp.53-69, for a discussion of the IRT scaling procedures). This was in stark contrast with the favorable achievement of the Hong Kong and Macao students, whose mean mathematical performance scores are 550 and 527 respectively, whereas the OECD mean is set at 500 (see Ho, 2003 & 2004 and Wong, 2003 & 2004 for an analysis of the Hong Kong results).

One special feature of the Macao educational system was that public government schools, compared with the private schools, played only a minor role (i.e. 2 out of 39 schools in the total school sample) in the basic education system (see Lo & Cheung, 2005, p.13 for details). It was a common understanding that the apparent mismatch between cognitive and affective outcomes of mathematical learning could not be appropriately resolved without taking into account the school level effect of *School Type* (variable name is *SC03Q01* in the PISA database; see OECD, 2005, p.331 for the codebook of the database). This was because affluent families generally preferred to send their children to study in the more popular private schools.

A simple Hierarchical Linear Modeling (HLM) analysis of *BELONG* on *SC03Q01* sets the context of the problem in the present study (for an explication of the HLM modeling or multilevel modeling, see Goldstein, 1995; Kreft & Leeuw, 1998; Raudenbush & Bryk, 2002).

$$\begin{aligned} BELONG &= G_{00} + G_{01} * SC03Q01 + U_0 + R \\ &= -0.611\ (0.021) + 0.070\ (0.027) * SC03Q01 + Uo + R; \end{aligned}$$

(standard error of G_{00} and G_{01}enclosed in brackets; Uo and R are residual and error terms; *SC03Q01* grand centred; restricted maximum likelihood estimation)

There was a school type effect of the magnitude 0.070 on *BELONG* favouring the private school. In Macao, the general public's perception was that those students enrolling in public government schools were generally from families of low social, economic and cultural status. One hypothesis needed to be addressed in the present study was that although the two public government schools were generally well-resourced, because of the low school autonomy and low student achievement teachers' morale and students' sense of belonging to the school might be seriously hampered, and this had a bearing on the mismatch between the cognitive and affective outcomes of schooling in Macao's basic education system. By analysing relevant variables in the PISA 2003 database, this paper sought to unveil the intricate relationships between the cognitive and affective aspects of mathematics learning for the 15 year-olds in Macao.

The data

At the student level, apart from taking *BELONG* as one major unfavourable affective outcome of schooling, a set of five plausible values of mathematical literacy (*PV1MATH* to *PV5MATH*) was used to generate the corresponding score point estimates indicative of students' cognitive outcome of schooling (OECD, 2005, p.71-80). Exploratory data analyses revealed that gender of the student (*ST03Q01*), and mean school economic, social and cultural status of the home (*ESCS*) had substantial influences upon mathematical literacy of 15 year-olds in Macao (Lo & Cheung, 2005, p.29-37). As data was drawn from a sample survey of students from all secondary schools in Macao, ensuing analyses would be appropriately weighted by *W_FSTUWT*, i.e. the student sample weight (OECD, 2005, p.19-30). Table 1 detailed the descriptive statistics of the student level variables used in the present study.

Table 1. Descriptive statistics of student level variables

Variable	Label	Mean	SD	Min	Max
ST03Q01	Gender	1.50	0.50	1.00	2.00
BELONG	Sense of belonging to the school	-0.62	0.74	-3.38	2.22
ESCS	Economic, social and cultural status of the home	-0.87	0.88	-3.48	2.22
PV1MATH	Maths literacy plausible value 1	522.34	89.16	237.69	795.48
PV2MATH	Maths literacy plausible value 2	522.82	88.96	217.67	779.83
PV3MATH	Maths literacy plausible value 3	522.09	87.88	213.30	765.10
PV4MATH	Maths literacy plausible value 4	523.46	88.82	203.96	769.47
PV5MATH	Maths literacy plausible value 5	523.25	88.62	213.54	794.63
W_FSTUWT	Sample weight (student level)	5.24	3.35	1.00	19.20

Number of students = 1250, consisting of 631 boys and 619 girls respectively

At the school level, four variables were used to explain the school-level outcomes of schooling or cross-level relationships in the multilevel analyses. These variables were: school type (*SC03Q01*), teacher morale (*TCMORALE*), school autonomy (*SCHAUTON*), and mean economic, social and cultural status of the home (*MEANESCS*). *TCMORALE*, *SCHAUTON and MEANESCS* were chosen as school level explanatory variables because of their potential influences upon the within-school affective processes pertaining to the sense of belonging of the school. Table 2 detailed the descriptive statistics of the school level variables used in the present study.

Table 2. Descriptive statistics of school level variables

Variable	Label	Mean	SD	Min	Max
SC03Q01	School Type	1.95	0.22	1.00	2.00
TCMORALE	Teacher Morale	-0.65	0.91	-2.81	1.65
SCHAUTON	School Autonomy	1.60	0.38	-0.52	1.69
MEANESCS	Mean *ESCS*	-0.86	0.49	-1.59	0.75

Number of schools = 39, consisting of 2 public government schools and 37 private schools

The specified model

Guided by a conceptual model of school learning explicated at both the learner and institutional levels, hierarchical linear modeling of the student and school level variables was undertaken (Cheung, 1993). As a result of a number of exploratory runs, the following multilevel statistical model was proposed for testing the hypothesis raised in the present study.

Student level model:

$$Y = B_0 + B_1*\mathit{ST03Q01} + B_2*\mathit{BELONG} + B_3*\mathit{ESCS} + R \quad (1)$$

School level model:

$$B_0 = G_{00} + G_{01}*\mathit{TCMORALE} + G_{02}*\mathit{MEANESCS} + U_0 \quad (2)$$

$$B_1 = G_{10} + U_1 \quad (3)$$

$$B_2 = G_{20} + G_{21}*\ \mathit{SCHAUTON} + U_2 \quad (4)$$

$$B_3 = G_{30} + G_{31}*\mathit{SC03Q01} + U_3 \quad (5)$$

Combined model:

$$\begin{aligned} Y &= B_0 + B_1*\mathit{ST03Q01} + B_2*\mathit{BELONG} + B_3*\mathit{ESCS} + R && (6)\\ &= G_{00} && \text{[overall effect]}\\ &+ G_{01}*\mathit{TCMORALE} + G_{02}*\mathit{MEANESCS} && \text{[school level effects]}\\ &+ G_{10}*\mathit{ST03Q01} + G_{20}*\mathit{BELONG} + G_{30}*\mathit{ESCS} && \text{[student level effects]}\\ &+ G_{21}*\mathit{SCHAUTON}*\mathit{BELONG} + G_{31}*\mathit{SC03Q01}*\mathit{ESCS} && \text{[interaction terms]}\\ &+ U_0 + U1*\mathit{ST03Q01} + U_2*\mathit{BELONG} + U_3*\mathit{ESCS} + R && \text{[residual terms]} \end{aligned}$$

(Y is the set of five plausible values of mathematical literacy; B_0 is expected school mathematical literacy when *ST03Q01*, *BELONG* and *ESCS* take on values of zero after they have been centered around their respective grand means; *TCMORALE*, *MEANESCS* and *SCHAUTON* grand-centred; restricted maximum likelihood estimation)

At the student level model, three student level variables were used to predict mathematical literacy scores. Five plausible values have been drawn randomly from the posterior score distribution to account for the measurement errors of the achievement measures. The three student level variables were: (1) gender of student (*ST03Q01*), (2) sense of belonging to the school (*BELONG*), and (3) social, economic and cultural status of the home (*ESCS*). Three school level variables have been used to predict the random intercept and slopes of the student level model. The four school-level variables were: (1) teacher morale of the school *(TCMORALE*), (2) school autonomy (*SCHAUTON*), and (3) mean school economic, social and cultural status of the home (*MEANESCS*), and (4) type of the school (*SC03Q01*). After a number of exploratory runs to arrive at a parsimonious model, the combined HLM model was specified in order to test the hypothesis of the present study. To simplify the interpretation of the fixed and random effects, all student level predictor variables have been centered at the respective grand means of the Macao student population.

As a result of this model specification, there are eight fixed effects specified for the HLM model: G_{00}, G_{01}, G_{02}, G_{10}, G_{20}, G_{21}, G_{30}, G_{31} (see equation (2), (3), (4) and (5) of the school level model). All predictors specified in the above student level models have been centered at their respective grand means. There are five random variance components specified for R, U_0, U_1, U_2 and U_3 (see (1) of the student level model, and (2), (3), (4) and (5) of the school level model). In the HLM output, these random components are expressed as: (1) *sigma squared* of R (assumed to be constant across the 39 schools in the Macao population), and (2) *tau* variance and covariance matrix as well as *tau* correlation matrix of B_0, B_1, B_2, B_3.

Modeling results

The HLM computer program (version 6.02) can take care of the set of 5 plausible values of mathematical literacy in a single computer run and produces average results with appropriate weighting and these are summarised below:

(1) s*igma squared* of R = 6374.11

(2) *tau* variance/covariance matrix

B_0	715.17			
ST03Q01 (B_1)	35.96	62.66		
BELONG (B_2)	-4.48	36.23	82.86	
ESCS (B_3)	21.44	-17.55	-22.09	22.65

(3) *tau* correlation matrix

B_0	1.000			
ST03Q01 (B_1)	0.170	1.000		
BELONG (B_2)	-0.018	0.503	1.000	
ESCS (B_3)	0.168	0.466	-0.510	1.000

(4) Reliability of random coefficients B_0, B_1, B_2, and B_3 are estimated to be 0.784, 0.072, 0.183, and 0.062 respectively. They are based on 30 of 39 schools that had sufficient data for computation.

(5) Final estimation of fixed effects (with robust standard errors) and variance components are summarised in Table 3 and 4.

(6) Substituting the HLM modeling results into combined model, equation (6) becomes:

Y = 527.31(4.643) [overall effect]

+ 25.66 (12.930)**MEANESCS* + 19.04 (5.689)**TCMORALE* [school level effects]

+ 6.12 (3.170)**ESCS* + 23.84 (5.886)**ST03Q01* [student level effects]

+ 21.80 (7.678)**SC03Q01*ESCS* - 9.54 (4.951)* *SCHAUTON*BELONG* [interaction terms]

+ U_0 + U1**ST03Q01*+ U_2**BELONG* + U_3**ESCS* + R [residual terms]

This combined model equation describes how the school and student level variables affect mathematical literacy.

Table 3. Estimates and significance testing of fixed effects

Fixed effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTERCEPT, B_0					
INTRCEPT2, G_{00}	527.3	4.643	113.574	36	0.000
TCMORALE, G_{01}	19.04	5.689	3.347	36	0.002
MEANESCS, G_{02}	25.66	12.930	1.985	36	0.054
For ST03Q01 slope, B_1					
INTRCEPT2, G_{10}	23.84	5.886	4.050	38	0.000
For BELONG slope, B_2					
INTRCEPT2, G_{20}	3.05	4.356	0.699	37	0.489
SCHAUTON, G_{21}	-9.54	4.951	-1.927	15	0.073
For ESCS slope, B_3					
INTERCPT2, G_{30}	6.12	3.170	1.930	37	0.061
SC03Q01, G_{31}	21.80	7.678	2.839	28	0.009

Note: Fixed effects are based on all the data in the sample.

Table 4. Estimates and significance testing of variance components

Random effect	Standard deviation	Variance component	d.f.	Chi-square	p-value
U_0	26.743	715.166	27	137.21145	0.000
U_1	7.916	62.657	29	20.67965	>.500
U_2	9.103	82.862	28	25.63499	>.500
U_3	4.760	22.652	28	26.85354	>.500
R	79.838	6374.109			

Note: The chi-square statistics reported above are based on 30 out of 39 schools that had sufficient data for computation. However, variance components are based on all the data in the sample.

Discussion of results

First, there is an overall Macao region effect of 527.31 points. Second, on top of this overall effect, an additional of 25.66 and 19.04 score points are added if students are located one standard deviation unit above the mean *ESCS* and *TCMORALE* scales. Schools with higher mean level of economic, social and cultural status of the home and/or higher level of teacher morale are shown to be associated with higher level of mathematical literacy. Third, a marginally significant 6.12 score points may be added if a student is located one standard deviation unit above the ESCS scale, and an impressive 23.84 score point difference between the two sexes (see Figure 1 for the pattern of interaction effects between *ST03Q01* and ESCS on mathematical literacy).

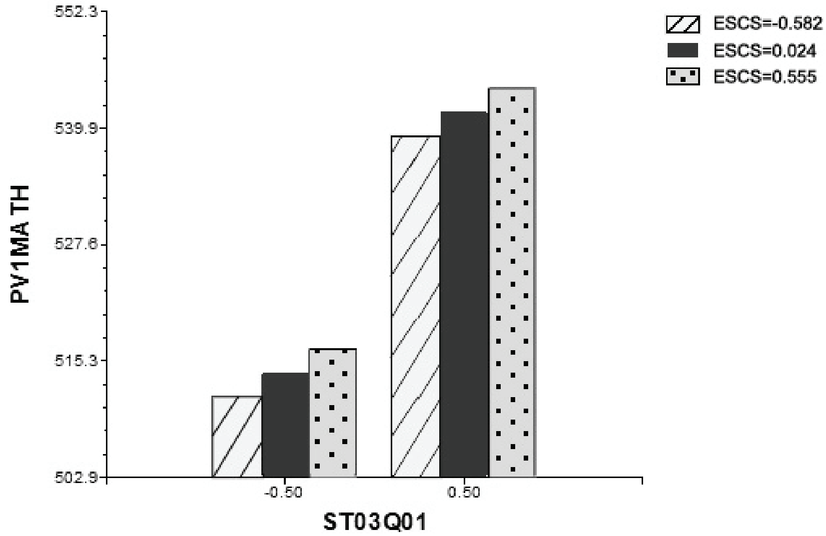


Figure 1. Interaction effects between gender of student and ESCS on mathematical literacy

Fourth, there was an impressive cross-level interaction effect of 21.80 score points between type of school and economic, social and cultural status of the home. There are only two public government schools in Macao. Figure 2 reveals that not only students from the public government schools generally score low, but this inferiority is especially so for those students coming from homes of higher ESCS. Figure 3 further shows that schools having higher level of mathematical literacy are associated with higher teacher morale as perceived by the school principals. These schools are mainly established, well-resourced private schools that can attract high ability students from all levels of ESCS in Macao.

Fifth, there is another minor cross-level interaction effect of -9.54 score points between school autonomy and students' sense of belonging to the school. There are three schools of low school autonomy that were associated with increasing mathematical literacy scores with increasing level of student's sense of belonging to their schools. This interesting result showed that when a school's autonomy is low, mathematical literacy can still be increased when students are having higher sense of belonging to the school, shown in Figure 4. This phenomenon is especially conspicuous in the two public schools in Macao.

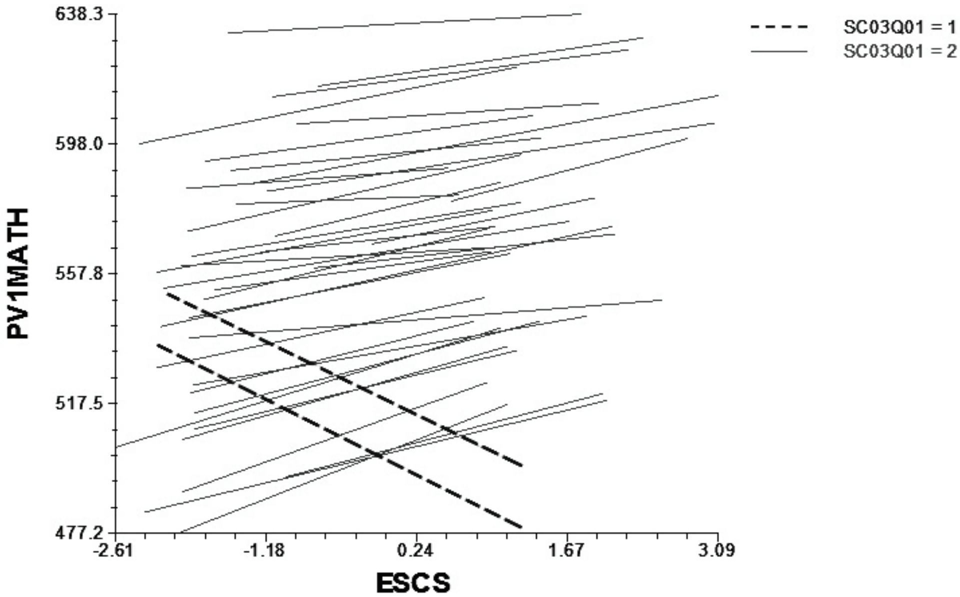


Figure 2. Cross-level effect of school type on the relationship of ESCS with mathematical literacy

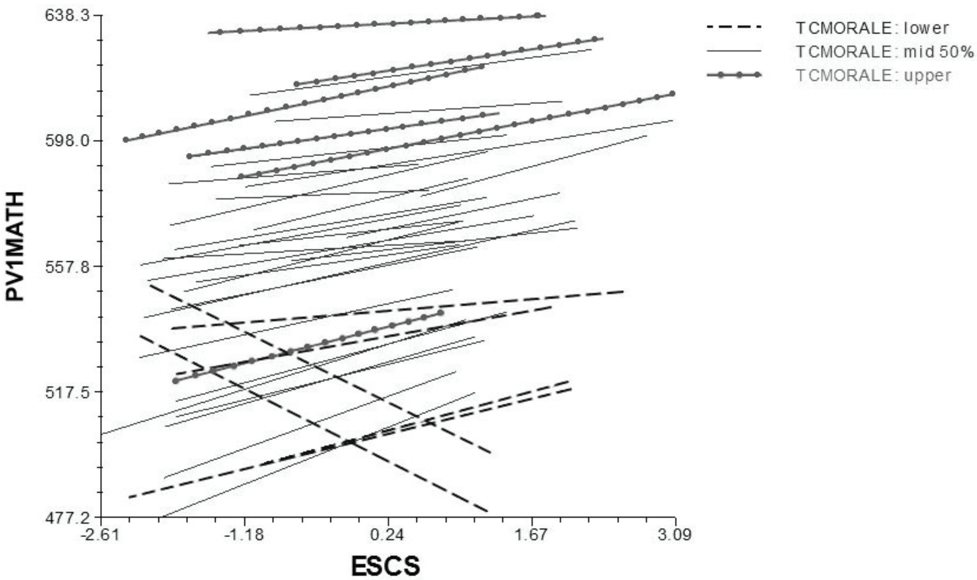


Figure 3. Cross-level effect of teacher morale on the relationship of ESCS with mathematical literacy

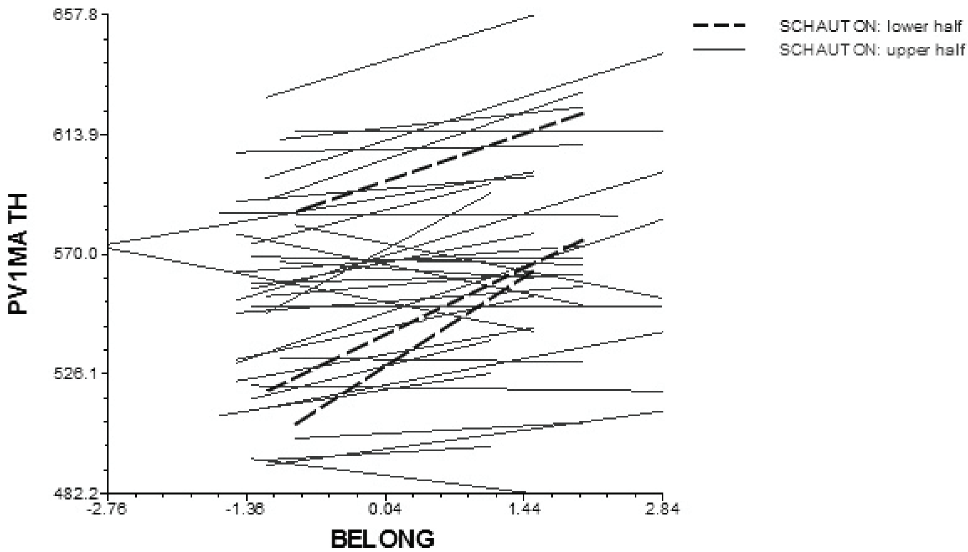


Figure 4. Cross-level effect of school autonomy on the relationship of *BELONG* with mathematical literacy

Discussion of findings

The combined HLM model results uncover a number of findings pertinent to a discussion of the mismatch between cognitive and affective outcomes of mathematics learning in Macao. Points worthy of attention are stated as follows:

While 15-year-old Macao students score relatively highly on the mathematical literacy tests in the PISA 2003 international assessment, their sense of belonging to the school cannot be regarded be satisfactory. The present study found that there was a significant negative school type effect on both cognitive (i.e. mathematical literacy) and affective (i.e. sense of belonging to the school) outcomes of schooling. One speculation is that because of the public government school students' low self-esteem and self-concept of their academic achievement, their sense of belonging to the school cannot be high. This is especially so for students coming from homes of higher ESCS, who normally will either further their studies abroad or otherwise are able to afford the expensive school fees to study in the better resourced private schools. The findings confirmed that in the two public government schools, students of higher ESCS scored much lower than their peers of lower ESCS. Officials in charge of education in Macao should feel relieved by learning that students of lower ESCS studying in the two public government schools were actually performing better than at least one third of other private schools, and that it was the students of higher ESCS that had lowered substantially the mean mathematical literacy performance of the public government schools.

Gender of student is another important student level variable that has strong effects on mathematical literacy. However, this study showed that for each of the two gender students of higher ESCS performed consistently better than peers coming from the lower ESCS. While an explanation of within-school gender gap in mathematical literacy cannot be established in the present study, this gap was found to be quite consistent across the 39 schools in Macao. More analyses are needed to be conducted to unveil the cause of mathematical gender differences in the future.

Teacher morale as perceived by the principal and mean ESCS of the home, both school level variables are found to have significant effects on mathematical literacy of the school. In Macao, there are a few renowned private schools that cater for students coming from homes of higher ESCS. These schools are willing to accept good students from homes of low ESCS. Teachers in these schools are very well-paid, highly qualified and experienced. It is therefore not surprising to find that teacher morale was found to be associated with high level of mathematical literacy in these renowned private schools.

Similar to that of a small number of poorly-run private schools, teacher morale of the two public government schools was rated by the principals as not very satisfactory. On the one hand, government teachers face hardships to elevate both the cognitive and affective outcomes of the students. On the other hand, because of the relatively lower school autonomy in public government schools than the more established renowned private schools, public government school teachers may find it difficult or are reluctant to change the school ethos and classroom climate for more effective teaching. When teachers' self-efficacy is hampered their morale cannot go up. One point worthy of the principal's attention is that if teacher morale as perceived by them is low, it is less likely that students' sense of belonging to the school can be high. The finding that when a school's autonomy is low its mathematical literacy can still be increased by having students with high sense of belonging to the school is really good news to both teachers and principals. Since teachers of the two public government schools are on the whole very experienced and fully qualified, students' sense of belonging may have considerable impact on mathematical literacy if the principals can have more positive views on their teachers. In light of the analyses results, principals may be convinced that their teachers are actually doing a good job and this is especially so for students coming from homes of lower ESCS.

Summary

This study is a first attempt to unveil the mismatch between cognitive and affective outcomes of mathematics learning in Macao. As illustrated in the above analyses, because Macao possesses a unique privatised educational system other systems in the region or aboard may find it hard to make a straight forward comparison. Because of school autonomy in the majority of private schools, there are oppositions to standardised assessment and therefore there is as yet no public examination system that caters for the whole of primary and secondary education to allow educational practitioners to gauge the quality of school provision. In 2003, Macao finally reached out and took a bold step by participating for the first time in an international PISA assessment. The assessment data collected is not only valuable for comparative education purposes but also throws light on the quality and equity aspects of schooling in the 39 secondary schools in Macao. The HLM model established in this study is only an initial step towards more adequate analyses and international comparisons in the future.

References

Cheung, K.C. (1993). The environment and its effects on learning – Product and process modeling for science achievement at the sixth form level in Hong Kong. *School Effectiveness and School Improvement*, 4(4), pp. 242-264.

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Ho, E.S.C. (2003 & 2004). Accomplishment and challenges of Hong Kong education system: What we have learned from PISA. *Educational Journal*, 31(2) & 32(1), 1-30.

Kreft, I. & Leeuw, J.D. (1998). *Introducing Multilevel Modeling*. London: Sage.

Lo, L.F. & K.C. Cheung (2005). *2003 Macau-PISA Final Report – Assessment of 15-year-old Macau Students.* Macao: The Macau-PISA Centre, Faculty of Education, University of Macau.

OECD (2004a). Learning for tomorrow's world – First results from PISA 2003. Paris.

OECD (2004b). *First Results from PISA 2003 – Executive Summary.* Paris.

OECD (2005). *PISA 2003 Data Analysis Manual – SPSS Users*. Paris.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edition.). Thousand Oaks, CA: Sage.

Wong, P.K.M. (2003 & 2004). Mathematical literacy of Hong Kong's 15-year-old students in PISA. *Educational Journal*, 31(2) & 32(1), 91-120.

16

Factors influencing mathematics achievement in the Gaza Strip and the West Bank: Evidence from TIMSS 2003

Petra Lietz and *Dieter Kotte*
Australian Council for Educational Research

Background

A number of national and international studies that have been conducted in the Gaza Strip and the West Bank since 1999 have shown performance differences by students in schools run by different authorities. More specifically, schools under the authority of the United Nations Relief and Works Agency for Palestine Refugees (UNRWA) have performed at a slightly but significantly higher level than their peers in schools run by the Palestinian National Authority government (AEC, 2006; Matar et al., 2000). In order to put the analyses reported in this chapter into context, a brief overview of the school system in the Gaza Strip and the West Bank is included.

The school system in the Gaza Strip and the West Bank consists of one to two years of pre-school education, ten years of basic compulsory education between the ages of six and 16, and two years of secondary education. The net enrolment ratio in primary and secondary education is 90.9 (equal for boys and girls) and 83.8 (boys 81.7, girls 86.0) per cent respectively. Particularly the secondary enrolment rate is relatively high when compared with some other Arab countries (e.g. Morocco (35.7%), Saudi Arabia (52.7%) and Tunisia, 64.5%) and about the same as in Egypt (80.8%) and Jordan (79.9%), Lietz et al., 2008). This reflects the positive attitudes towards education as a basis to improve their future by the Palestinians (Dickerson, 1974) as well as the results of specific programs to decrease school dropout rates especially of students from poorer homes (Arafat, 1989)

The total number of schools in the Gaza Strip and the West Bank is 2415 of which about 70 per cent are under the authority of the Ministry of Education and Higher Education (MoEHE), 23 per cent are run by UNRWA and seven per cent are privately operated (Lietz et al., 2008). Decisions regarding curricula and textbooks are made centrally by the MoEHE and are implemented across all schools, regardless of authority.

UNRWA's involvement in the provision of pre-primary and basic education in the area dates back to 1949 when UNRWA was established by resolution #302 (IV) of the General Assembly of the United Nations in order to implement the relief and works program for Palestinian refugees in the aftermath of the 1948 Arab-Israeli conflict. Since then, the resolution has been confirmed repeatedly and currently extends to 30 June 2011.

In addition to pre-primary and primary education UNRWA also provides vocational and technical training as well as teacher training, making education its largest field of activity. In terms of school education, UNRWA offers free education to a total of nearly 500,000 Palestinian refugees in about 700 elementary and five secondary schools (all located in Lebanon) in the school year 2005/2006 (UNRWA, 2008). About three-quarters of these schools are located in the Gaza Strip and the West Bank. UNRWA also provides services in the areas of health, relief and social services as well as microfinance across its five locations – in addition to the West Bank and the Gaza Strip - including Jordan, Lebanon and Syria.

On the one hand, it is sometimes argued that UNRWA (e.g. Kushner, 2005) also has a political agenda, through emphasising the "right to return" of Palestinian refugees to originally Arab villages that have been replaced by Israeli cities and farms. As a consequence, it is claimed that refugees served by UNRWA, "have no sense of permanency and for the most part (some living in Jordan being the exception) without citizenship" (Kushner, 2005, p.18). Unlike Jordan (Besson, 1997), neither Syria nor Lebanon offers citizenship to Palestinian refugees. Likewise, the Palestinian Authority (PA) is reported not to support a more permanent status for these refugees in order not to impede the possibility of their return.

On the other hand, UNRWA is sometimes regarded as having been established by the Western powers to lessen the refugees' political rights (Al-Husseini, 2000). This, it is argued, was done by UNRWA attempting, particularly in the beginning, to resettle the refugees in the host countries, thus lessening their desire to return. This, however, was met by refugee opposition which led to UNRWA's focusing more on education than it had done previously.

In 2008, UNRWA's funding mainly originated from the European Commission and a number of individual governments, namely the Netherlands, Norway, Sweden, the United Kingdom, and the United States, with the latter being the largest contributor (UNRWA, 2008). UNRWA's funding is non-recurrent as the donor countries decide from year to year how much money they make available. This insecurity regarding funding has led, among other things, to about 80 per cent of primary schools and about 70 per cent of pre-schools run by UNRWA operating in double shifts (UNRWA, 2008).

The above articles as well as reports of earlier studies of mathematics and science conducted by the International Association for the Evaluation of Educational Achievement (IEA) (Burstein, 1993; Keeves, 1966; Lokan et al, 1996; Robitaille & Garden, 1989), were used as a guide when screening the hundreds of variables measured in TIMSS in an attempt to identify those that might be related to the school authority on the one hand and mathematics achievement on the other hand.

Data

Data used in the analysis reported in this chapter were collected as part of TIMSS 2003 as the third cycle of assessment in the framework of the Trends in International Mathematics and Science Study (TIMSS), which is conducted by the International Association for the Evaluation of Educational Achievement (IEA).

The Palestinian National Authority was one of a total of 48 countries which participated in TIMSS 2003 at the Population 2 level for which the definition was as follows (Martin et al., 2004, p.4):

> … all students enrolled in the upper of the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was indeed the eighth grade in most countries.

The international data for the Palestinian National Authority for Population 2 consisted of information for 5,357 students and 150 schools. The sample was explicitly stratified by two school sizes, namely large (N=22) and very large schools (N=123) and implicitly stratified by regions (Gaza Strip, West Bank), school type (public, private, UNRWA) and gender (boys, girls, mixed) for a total of 16 implicit strata (IDSTRATI).

As was the case in the first two testing cycles, two sets of instruments were used in TIMSS 2003. The first set covered the achievement tests, which used a rotated booklet design in order to provide as broad curriculum coverage as possible without overburdening individual students. In addition, background questionnaires for students, teachers, school principals and national research coordinators were administered that had been developed based on the framework put forward originally by Keeves (1974). This framework specified the intended curriculum as mandated at the system level, the implemented curriculum as taught by teachers in classrooms, and the attained curriculum in terms of the curriculum as learnt by students.

Thus, information obtained in this TIMSS 2003 provided a rich set of data that could shed light on the issue of how UNRWA and government schools in the Gaza Strip and the West Bank differ in terms of the provision and outcomes of education.

Analyses

Three types of analyses were undertaken in this study. First, preliminary analyses comprising descriptive analyses and correlational analyses were conducted. The descriptive analyses were aimed at identifying any variables that had excessive missing data. Here, missing data in excess of 20 per cent was applied as cut-off point as a sample was considered not to be representative of the intended target population if information was obtained from less than 80 per cent of originally selected participants (Martin et al., 2007). Correlational analyses were undertaken to identify whether those variables that were identified from the literature review were actually related to mathematics achievement on the one hand and type of school authority on the other hand.

The following two types of analyses, namely hierarchical linear modeling (HLM) and logistic regression analysis were undertaken with somewhat different aims in mind. First, a two-level HLM model was examined to elucidate the extent to which student- and school-level variables contributed to explaining differences in students' mathematics achievement. In this analysis, a variable indicating whether a school was run by UNRWA or by the government was included as a predictor at the school-level. As a consequence, the analysis allowed the estimation of effects of other important student and school-level factors on achievement while controlling for school authority. Moreover, it allowed the examination of whether relationships between student-level predictors and mathematics achievement differed depending on school authority.

Second, logistic regression analysis was used to identify that set of variables that distinguished best between UNRWA and government schools. In other words, this analysis served the aim of identifying whether important differences existed between the average score profiles on a set of variables for UNRWA and government schools. Logistic regression rather than discriminant analysis was chosen to take into account the fact that most of the predictor variables were ordinal and nominal rather than interval or ratio scaled (Hair et al., 2005; Tabachnick and Fidell, 2001).

Both HLM and logistic regression analysis have their advantages and disadvantages. The advantage of logistic regression is that it can specifically identify those characteristics on which UNRWA and government schools differ. The disadvantage of logistic regression analysis stems from the fact that it is a single-level analytical technique which requires the aggregation of student-level variables to the school level, resulting in aggregation bias (Cheung & Tsoi, 1990; Sellin, 1990). The advantage of HLM stems from its ability to process data at the appropriate level, hence overcoming the necessity to aggregate or disaggregate data. Its disadvantage is that it requires the outcome variable to be measured at the lowest level which meant that school authority, as a school-level variable, could not be used as a dependent variable in the analysis.

In summary, the two types of multivariate analyses served two different yet somewhat complementary aims. The aim of the HLM analyses was to identify those factors that were important in explaining differences in students' mathematics performance in the Gaza Strip and the West Bank, including school authority. The aim of the logistic regression analysis was to identify those variables on which UNRWA and government schools differed considerably.

Results

Preliminary analyses

As first step, a new variable (UNRWA) was computed based on the variable identifying the implicit school stratum (IDSTRATI). This new variable was assigned a '0' if the school was run by the government or a '1' if a school was operated by UNRWA. Privately run schools were excluded from the analyses.

With the exception of mathematics performance, which was measured on a continuous scale and the school authority variable as a dichotomous variable, all other variables were measured on an ordinal scale. Hence, Spearman rank-order correlation coefficients were calculated as the appropriate measure of association (Thorndike, 1997).

In the next step, all correlations coefficients of student and school variables with each other and with mathematics achievement were inspected to identify possible variables and factors for the subsequent HLM and logistic regression analyses. Following this, factor analyses were used to compose latent constructs of single manifest variables that were considered to measure a common underlying construct such as, for example, attitudes towards homework or towards mathematics.

Table 1 (student level) and Table 2 (school level) provide details concerning the variables and factors used in the HLM analysis.

Table 1. Descriptive statistics and specifications of key student-level variables and factors

Variable/ Factor	Specification/Coding	Min.	Max.	Mean	Std. Dev.
BSBGBOOK	number of books at home: 1=0-10, 2=11-25, 3=26-100, 4=101-200, 5=more than 200	1	5	2.30	1.13
BSBGFMED	mother's highest education level	1	8	4.56	1.95
BSBGHFSG	student's expected education level	1	5	3.91	1.46
BSBGMFED	father's highest education level	1	8	3.99	1.70
BSBGSEX	gender of student: 1=female, 2=male	1	2	0.45	0.50
BSDAGE	student age in years	10.33	18.33	14.12	0.85
BSDGEDUP	parents' highest educ. level: 1=no more than primary school, 2=lower sec. school, 4=post-sec./vocational school, 5=university/high sch	1	5	3.38	1.23
BSMMAT01	1st plausible value, Mathematics score	66.97	677.35	390.83	90.72
BSMMAT02	2nd plausible value, mathematics	102.52	675.41	390.65	91.72
BSMMAT03	3rd plausible value, mathematics	83.39	689.59	389.83	92.85
BSMMAT04	4th plausible value, mathematics	56.02	685.55	389.94	92.66
BSMMAT05	5th plausible value, mathematics	58.06	707.13	391.19	91.30
DISTRACT	level of distraction from school work; factor score comprising BSBGWKPJ (work at a paid job), BSBGUSIN (use internet), BSBMEXTO (extra lessons); positive values denote higher degree of distraction	-0.97	3.98	-0.04	0.98
EASYMATH	ease of learning Mathematics; factor score comprising BSBMTCLM (maths more difficult), BSBMTTOP (not understand new topic), BSBMTSTR (maths no strength); positive values denote that students find learning Mathematics easier	-1.82	1.91	0.00	1.00

Variable/ Factor	Specification/Coding	Min.	Max.	Mean	Std. Dev.
HOMEWORK	frequency of Mathematics homework given per week by teacher; higher values denote homework given more frequently (BSBMHWMA recoded)	0	5	4.14	1.28
LIKEMATH	attitude towards Mathematics; factor score comprising BSBMTWEL, BSBMTMOR, BSBMTENJ, BSBMTQKY, BSBMAHDL, BSBMAOSS, BSBMAUNI, BSBMAJOB, BSBMAGET; positive values denote a more positive attitude towards Mathematics	-4.12	1.40	0.03	0.99
LIKESCH	attitude towards school in general; factor score comprising BSBGALBS, BSBGATTB, BSBGATCS, BSBGATSB; positive values denote a more positive attitude towards school	-4.37	0.92	0.01	0.99
MATHMECH	frequency of mechanical tasks in Mathematics; factor score comprising BSBMHGCT, BSBMHEFR, BSBMHWSG, BSBMHBHC, BSBMHCAL; positive values denote higher exposure towards mechanical tasks in Mathematics	-2.72	3.07	-0.01	1.00
MATHREAS	frequency of reasoning in Mathematics; factor score comprising BSBMHEXP, BSBMHSCP, BSBMHROH, BSBMHLSP, BSBMHWPO; positive values denote higher exposure to reasoning tasks in Mathematics	-4.23	1.69	0.05	1.00
PALEHOME	ethnic origin; factor score comprising BSBGPLHO, BSBGFBRN, BSBGBORN; positive values denote greater ties to Palestine	-4.56	0.75	0.01	0.99
PCUSAGE	frequency of using PC at school; factor score comprising BSBMOINF, BSBSOINF, BSBGOREP; positive values denote higher exposure to PC-related work at school	-2.01	1.68	0.00	1.00
POSSESS	home possessions; factor score comprising BSBGPS02, BSBGPS04, BSBGPS06, BSBGPS09, BSBGPS11, BSBGPS13; positive values denote more possessions	-2.09	1.44	0.00	1.00
PROBLSCH	degree of experiencing problems at school; factor score comprising BSBGSTOL, BSBGHURT, BSBGMADE, BSBGMFUN, BSBGLEFT; positive values denote more problems experienced at school	-0.90	2.94	-0.02	1.00

Notes: variable names are taken from IEA TIMSS' codebooks (e.g. BSBGMFUN)
derived variables and factor are denoted in *italic* (e.g. *LIKESCH*)
variables included in factors showed a factor loading of α>0.40

Hierarchical linear modelling

As can be seen, eleven factors were extracted from student-level variables that were correlated with mathematics achievement. In addition, five factors at the school level (*ATTEND, BCDGWKYR,* *SCHCLIMA, TCHQUAL, UNRWA*) and three variables at the teacher level (BTBGAGE, BTDMSTUD, *L_TEACH*) that correlated substantially with mathematics achievement emerged from the preliminary analyses. As one mathematics teacher per school (both n=145) completed a teacher background questionnaire, teachers and schools were combined at level-2 in the current analysis.

Table 2. Descriptive statistics and specifications of school-level variables and factors

Variable/ Factor	Specification/Coding	Min.	Max.	Mean	Std. Dev.
ATTEND	degree of school attendance: smaller values denote more regular school attendance (BCDGSP, BCDSST; recoded)	-9.00	-1.00	-3.90	1.92
BCDGWKYR	N of weeks of schooling per year	30	43	34.21	1.72
BTBGAGE	teacher's age: 1=under 25, 2=25-29, 3=30-39, 4=40-49, 5=50-59	1	5	2.83	1.09
BTDMSTUD	class size (N of students) for Mathematics instruction: 1=1-24, 2=25-32, 3=33-40, 4=41 or more	1	4	3.30	0.88
L_TEACH	limitations of teaching resources; factor score comprising BTBGLT10, BTBGLT11, BTBGLT12, BTBGLT13	-2.17	2.31	-0.27	0.98
SCHCLIMA	school climate; factor score comprising BCBGFP01, BCBGFP02, BCBGFP03, BCBGFP04, BCBGFP07, BCBGFP08, BCBGFP10, BCBGFP11	-3.63	1.52	-0.02	0.98
TCHQUAL	perceived instructional quality; factor score comprising BCBGCHTS, BCBGCHSD, BCBMSORM	-3.63	1.89	0.02	0.97
UNRWA	school type: 0=government, 1=UNRWA	0	1	0.35	0.48

Notes: variable names are taken from IEA TIMSS' codebooks (e.g. BCBMSORM)
derived variables and factor are denoted in *italic* (e.g. *SCHCLIMA*)
variables included in factors showed a factor loading of α>0.40

In order to compare effect sizes within the model, all variables were standardised. Then, all relevant factors and variables were entered into a two-level HLM model, using the appropriate weight at each level (i.e. total student weight, *totwgt*, at the student level and school weight, *schwgt*, at the school level) in the analysis. The initial model was based on previous IEA studies (Elley, 1994; Keeves, 1991; Lundberg & Lynnakylä, 1993; Postlethwaite & Wiley, 1991) and informed by results of exploratory path analyses of the current data (PLSPATH, Sellin, 1990) for of two single-level models, one containing only student-level variables and the other containing only school/teacher variables. In an iterative process, the HLM model was refined whereby any fixed effect with a value of $\gamma<|0.05|$ and a p-value >0.01 (for a discussion of cut-off criteria see Kotte, 1992; Lietz, 1996) was considered to be not sufficiently substantive and, therefore, removed.

The fixed effects in the final two-level HLM model of factors influencing mathematics achievement in the Gaza Strip and the West Bank are shown in Table 3 and illustrated in Figure 1.

Table 3. Fixed effects on mathematics achievement in Palestine

Fixed Effect	**Coefficient (γ)**	**Std. Error**	**T-ratio**	**P-value**
Level 2/school-level effects				
TCHQUAL	0.15	0.05	2.78	0.007
Level 1/student-level effects				
BSDAGE	-0.17	0.03	-6.60	0.000
LIKEMATH	0.20	0.03	6.22	0.000
MATHREAS	0.09	0.02	3.57	0.001
PCUSAGE	-0.14	0.02	-6.04	0.000
DISTRACT	-0.26	0.03	-8.49	0.000
EASYMATH	0.21	0.02	9.95	0.000
interaction effect by UNRWA	0.04	0.02	1.69	0.095

Notes: 19 level-2 units lost due to missing data.
The maximum number of level-1 units (students) = 4715
The maximum number of level-2 units (schools) = 126

Overall, it is apparent that the stronger and more diverse factors influence mathematics achievement at the student level. Here, six factors emerge which, in the sequence of their relative strength, are: DISTRACT (γ=-0.26), EASYMATH (γ=0.21), LIKEMATH (γ=0.20), the student's age (γ=-0.17), PCUSAGE (γ=-0.14) and MATHREAS (γ=0.09). The remaining five student-related factors which also correlated with student achievement (see above) did not emerge as influential factors once the other factors in the model were taken into account.

The negative effect of DISTRACT on mathematics achievement (γ=-0.26) reflects the fact that those students who work more at a paid job or who spend more time on the internet perform at a lower level in mathematics. The other negative effect of students' age (γ=-0.17) reflects that younger students, in the same grade, perform at a higher level than their peers, probably because they did not experience interruptions of their schooling or because they started to school earlier because they were regarded as relatively bright.

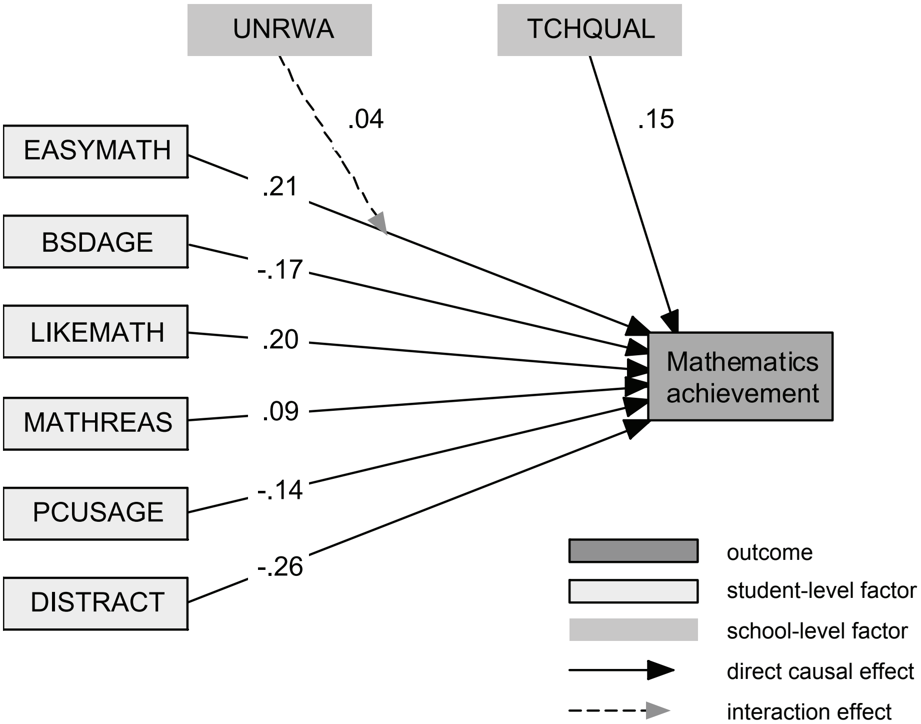


Figure 1. Two-level HLM model of Mathematics achievement for 15-year-old students in Palestine

In contrast to findings in other countries (Kotte 1992, Lietz 1996), the more general liking of school (LIKESCH) did not contribute to explaining performance differences over and above the contribution of the subject-specific attitude, here, the liking of mathematics (LIKEMATH) in Palestine.

At level-1, results indicate that students with a more positive attitude towards mathematics, who perceive mathematical tasks as easy and are occupied with more reasoning tasks during mathematics lessons perform at a higher level in mathematics than students with a more negative attitude toward mathematics, who consider mathematics to be difficult and are exposed to fewer reasoning tasks. As regards the negative impact of PC usage on achievement, it can be speculated that the introduction of PCs as part of the ordinary Mathematics instruction does not always happen in the most productive way: equipping schools with PCs is one thing, assisting teachers to use them as an effective instructional tools is another.

At level-2, only one direct effect emerges. Here, greater teaching quality (TCHQUAL) has a moderate, positive impact on student achievement in mathematics (γ=0.15).

In this context, it interesting to note the interaction effect – even if it is fairly small (γ=0.04) and of borderline significance – which the placement of students in government or UNRWA schools has on the students' perceptions of how easy Mathematics is for them. Since the interaction effect is positive, it indicates in UNRWA schools, even greater differences exist between students' mathematics performance depending on how easy they consider mathematics to be as a subject. The differences in the mathematics of students in government schools, in contrast, are much less dependent on how easy or difficult they perceive mathematics to be as a subject.

As regards the thrust of the current analysis, it is interesting to note that no direct effect emerges between school type and achievement. In other words, while a significant achievement difference was initially observed between students in UNRWA (mean score=380; SE=3.8) and students in government schools (mean score=404, SE=4.6), no significant difference according to school type emerges in an analysis in which other substantive factors affecting mathematics achievement are taken into account.

Logistic regression analysis

It will be recalled that the current analyses were designed to serve two purposes. First, effects of various factors, including type of school, on mathematics performance in Palestine were to be investigated in a hierarchical linear model. Second, in a logistic regression analysis, type of school was used as an outcome variable in order to examine factors that might discriminate between UNRWA and government schools.

Only student-related variables which correlated with school type (government/UNRWA) at $r \geq |0.10|$ were included in the logistic regression analysis. The cut-off criterion for school-related variables was set slightly higher at $r \geq |0.30|$ due to the smaller number of schools than students in the sample. Table 4 lists these variables and their correlation coefficients with type of school. As can be seen, three student-level and twelve school-level variables were identified for inclusion in the subsequent logistic regression analysis.

For the purpose of the logistic regression analysis, all variables were aggregated to the school level. Of the total number of 145 schools in the Palestinian sample, 96 were government schools, 42 UNRWA schools and seven private schools that latter being excluded from this analysis. Due to some missing data on variables in the analysis, 14 schools were excluded resulting in 124 schools (90 government; 34 UNRWA) remaining in the analysis.

The main results of a logistic regression analysis are similar to a regression analysis and can also be interpreted in a similar fashion (Hair et al., 2005; Tabachnick & Fidell, 2001). In Table 5, the main results of the analysis are summarised including the B values and their standard error, the Wald test of significance, degrees of freedom and significance level.

Table 4. Variables correlating with school type (government/ UNRWA) in Palestine

Variable	IEA TIMSS Variable Label	Correlation (r)
Student level		
BSBGCSCH	use a computer at school	-0.33
BSBGPS04	home possessions: dictionary	0.17
BSBSHWGO	work in group experiment or investigation	-0.13
School level		
BCBGTENR	total school enrollment: all grades	0.72
BCBGEENR	total school enrollment: eighth grade	0.57
BCBGHIGG	highest grade at school	-0.56
BTBMPSEM	major area of study: Mathematics education	-0.47
BTBMSTUD	number of students in the class	0.43
BTDMSTUD	class size for Mathematics instruction	0.39
BTBMPSMA	major area of study: Mathematics	0.35
BTBMTEEX	how often is a test given in Mathematics	0.35
BTBGFEDC	level of formal education completed	0.34
BTBGYTTR	years of teacher training	0.34
BCBGCMPS	computers available for student instruction	-0.34
BCBGCHTS	teachers job satisfaction	-0.31

Notes: cut-off criteria for student-level variables' correlation with type of school were $r \geq |0.10|$ and for school-level variables $r \geq |0.30|$; all correlations were highly significant with $p \leq 0.001$

First it is noteworthy that only three of the original 15 variables entered in the analysis emerged as variables on which UNRWA (coded '1') and government schools (coded '0') differed significantly. Of these three variables, the one indicating whether or not teachers had a degree in education mathematics had the strongest effect (B=-3.20). As the sign of the effect was negative and the predictor had been coded so that '0' indicated not having education mathematics as a major area of study and '1' indicating teachers without such a major it meant that teachers in UNRWA schools did not have such a major. The second largest effect emerged for the variable indicating the number of computers available for instruction. Again, the effect was negative meaning that UNRWA schools had fewer computers available for instruction than government schools. The third significant effect was associated with total school enrolment whereby government schools had significantly lower total enrolment than UNRWA schools, probably because the latter used their school buildings for more than one shift of students per day.

Table 5. Summary results of the logistic regression analysis (n=124)

	IEA TIMSS variable label	B	Std. Error	Wald	df	Sig.
Intercept		-4.40	0.55	64.10	1	0.00
BCBGCMPS	computers available for instruction	-0.19	0.03	38.92	1	0.00
BCBGTENR	total school enrollment: all grades	0.01	0.00	84.35	1	0.00
BTBMPSEM	major area of study: Education - Mathematics	-3.20	0.47	47.00	1	0.00

Overall model fit:	χ2-value:	358.70;	df=3;	Sig. 0.00
Pseudo R^2-values:	Cox&Snell:	0.39		
	Nagelkerke:	0.74		
Classification results:	93% correctly classified; 3 of 34 UNRWA and 6 of 90 government schools incorrectly classified			

Notes: Total number of schools in the PSE sample was 145 of which 96 were government schools, 42 UNRWA schools and 7 private schools. The latter were excluded from the current analysis. 14 schools had to be excluded due high amount of missing data on a predictor, leaving 124 schools in the analysis.
- School weight (schwgt) was used as a weighting variable in the analysis.
- A backward deletion approach was used: First all 15 variables identified in Table 4 were entered in the analysis. Based on results, the least significant variable was removed first, the analysis re-run, then the next least significant variable was remove until only significant predictors in the model remained.

In terms of overall fit of the model, the chi-square value of 358.70 (df=3) was highly significant ($p \leq 0.000$) and indicated that the model was better than the comparison model without any predictors which, based on the initial percentages (UNRWA=27%; government=73%) would have assumed that all schools were government schools. As a consequence, given the variables in the model, group membership for 93 per cent of the schools could be predicted correctly, leaving only six government schools and three UNRWA schools to be misclassified. The Cox & Snell and Nagelkerke pseudo R-Square values indicated that between 39 and 74 per cent of the variability between UNRWA and government schools was explained by the three variables in the model. Thus, the model could be considered as satisfactory.

Conclusions

Given that previous bivariate analyses had shown slight yet significant differences in mathematics performance between UNRWA schools and government schools in the Gaza Strip and the West Bank, this chapter set out to shed light on a couple of related questions by way of multivariate analysis. While the first question inquired about reasons for the previously reported differences in mathematics achievement between government and UNRWA schools in Palestine, the second question asked which variables differentiated between UNRWA and government schools.

As regards the first question, a two-level hierarchical linear model analysis revealed several student- and school-related factors which influenced mathematics achievement among 8th graders in Palestine. Negative effects on performance were found if students were distracted from school work through working at a paid job or using the internet, for students who were older than their fellow students and those students who reported using PC at school more often than their peers. In contrast, those students performed at a higher level in mathematics who liked this subject, reported being involved in a greater number of reasoning tasks and who found mathematics to be an easy subject. In addition, a positive effect of teaching quality on mathematics performance emerged. Here, students in schools that reported greater satisfaction of teachers with their job and a greater desire of students to learn performed at a significantly higher level than students in other schools.

No direct effect of type of school on achievement in mathematics emerged once the significant student and school-level factors were taken into account. However, an interesting interaction effect emerged of school type on the relationship between the perception of mathematics as an easy subject and achievement. While small, the effect was such that in UNRWA schools the differences between higher and lower performers in mathematics depending on whether students perceived mathematics as an easy subject was greater in UNRWA schools than they were in government schools. In other words, in government schools, it did not make as much of a difference in terms of mathematics performance whether or not students thought mathematics was an easy subject than in it did in UNRWA schools.

In order to address the second question, a logistic regression analysis was employed to identify variables that were able to differentiate between UNRWA and government schools. Here, of an initial 15 student and school-related variables, only three emerged as discriminating significantly between UNRWA and government schools. Differences were such that UNRWA schools were larger, had teachers without a major in education-mathematics and fewer computers. Government schools, in turn, were smaller, had more computers available for instructional purposes and had teachers with a major in education-mathematics. These results were somewhat surprising, given that the initial impetus for the analysis emanated from the small but significant higher mathematics performance of students in

UNRWA schools. Yet, the analyses showed that students in government schools enjoyed those conditions that are usually thought to be more conducive to learning, namely more instructional resources such as computers, smaller schools and teachers that are more specifically trained.

However, it would be too simplistic to advise the Palestinian National Authority who is responsible for the government schools to hire more teachers trained in education-mathematics, abolish computers and make schools larger. However, an argument could be made to see how the available computers might be used more fruitfully to support mathematics instruction. In addition, it maybe worth investigating to which special content students training to become teachers who take education-mathematics as a major are exposed that might benefit their subsequent mathematics teaching.

Given the very challenging social, economic and political difficulties under which the people in the West Bank and the Gaza Strip have to operate, such advice might sound out of place given the real-life situation of many Palestinian families, the students and the schools. Thus, the statistical analyses and results described can only serve as a starting point for educators and administrators in a situation that is characterized by uncertainty in all areas of life.

To John, with sincere appreciation for his introduction to and support during the never-ending journey that is educational measurement.

References

Assessment and Evaluation Center (AEC). (2006). *National Report on the Results of TIMSS 2003*. Ramallah: Ministry of Education and Higher Education.

Al-Husseini, J. (2000). UNRWA and the Palestinian nation-building process. *Journal of Palestine Studies, 29*(2), 51-64.

Arafat, S. (1989). Formal education in UNRWA. *Journal of Refugee Studies, 2*(1), 108-112.

Besson, Y. (1997). UNRWA and its role in Lebanon. *Journal of Refugee Studies, 10*(3), 335-348.

Burstein, L. (1993). *The IEA study of Mathematics III: Student Growth and Classroom Processes.* Oxford: Pergamon.

Cheung, K.C., Keeves, J.P., Sellin, N. & Tsoi. S.C. (1990) The analysis of multilevel data in educational research: Studies of problems and their solutions. *International Journal of Educational Research, 14*(3). Oxford: Pergamon.

Cheung, K.C., & Tsoi, S.C. (1990). Model specification and related issues. In K.C. Cheung *et al.* 1990 *op. cit*. 233-244.

Dickerson, G. (1974). Education of the Palestinian refugees: The UNRWA/Unesco program. *Journal of Palestine Studies, 3*(3), 122-130.

Elley, W. B. (1994). *The IEA Study of Reading Literacy. Achievement and Instruction in Thirty-Two School Systems.* Oxford: Pergamon Press.

Forsythe, D. (1971). UNRWA, the Palestine refugees and world politics: 1949-1969. *International Organisation*, 25(1), 26-45.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2005). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ.: Prentice Hall.

Keeves, J. P. (1966). *Evaluation of Achievement in Mathematics*. Hawthorn, Vic.: The Australian Council for Educational Research.

Keeves, J. P. (1974). The IEA science project. Science achievement in three countries: Australia, the Federal Republic of Germany, and the United States. In UNESCO 1974, *Implementation of Curricula in Science Education*. Cologne: German Commission for UNESCO.

Keeves, J.P. (Ed.) (1991). *The IEA Study of Science III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.

Kotte D. (1992). *Gender Differences in Science Achievement in 10 Countries - 1970/71 to 1983/84*. Frankfurt: Peter Lang.

Kushner, A. (2005). *UNRWA: A Hard Look at an Agency in Trouble*. Brookline, MA: Center for Near East Policy Research.

Larkin, A. I., & Keeves, J. P. (1984). *The Class Size Question: A Study at Different Levels of Analysis*. ACER Research Monograph No. 26. Hawthorn, Vic.: The Australian Council for Educational Research.

Levy, A. (1977). *Handbook of Curriculum Evaluation.* Paris: International Institute for Educational Planning (IIEP).

Lietz, P., Wagemaker, H., Neuschmidt, O., & Hencke, J. (eds.) (2008). *Educational Issues in the Middle-East North Africa Region.* The Hague, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Lietz P. (1996). *Reading Comprehension Across Cultures and Over Time*. Münster/New York: Waxmann.

Lokan, J., Ford, P., & Greenwood, L. (1996). *Maths & Science on the Line. Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study.* Melbourne, Vic.: The Australian Council for Educational Research.

Lundberg, I., & Linnakylä, P. (1993). *Teaching Reading Around the World.* Hamburg: International Association for the Evaluation of Educational Achievement.

Martin, M.O., Mullis, I.V.S. & Kennedy, A.M. (2007) *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., & Chrostowski, S.J. (Eds.) (2004). *TIMSS 2003 Technical Report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Matar, M., & Bisharat, K. (2008). Palestinian analysis: Factors related to mathematics teachers and their effect on the achievement of students studying in government and UNRWA schools in Palestine. Chapter 8 pp. 89-102). In P. Lietz et al. (eds), 2008 op.cit.

Matar, M., Al-Khaleeli, O., & Al-Janazreh, A. (2000). *Level of achievement for 4th graders in Arabic language, mathematics, and science* (primary report within the Education for All Study). Palestine: Ministry of Education and Higher Education.

Pedhazur, E. (1982). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Robitaille, D.F., & Garden, R.A. (1989). *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics*. Oxford: Pergamon.

Sellin, N. (1995). Partial least squares modeling in research on educational achievement. In W. Bos & R.J. Lehman (eds). *Reflections on Educational Achievement* (pp. 256-267). Münster/New York: Waxmann.

Sellin, N. (1990). On aggregation bias. In In K.C. Cheung *et al.* 1990 *op. cit*. 257-268.

Tabachnick, B.G., & Fidell, L.S. (2001). *Multivariate Statistics* (4th edn). Boston: Allyn & Bacon.

Thorndike, R.L. (1997). Correlational methods. In J.P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed., pp. 484-493). Oxford: Elsevier Science.

United National Relief and Works Agency for Palestine Refugees in the Near East (July 2007). *Program Budget 2008-2009*. Available from http://www.un.org/unrwa/finances/pdf/ProgBudget08-09.pdf last accessed 06/10/2008.

UNRWA (2008). Official website: www.un.org/unrwa/english.html Last accessed 23/10/2008.

17

A longitudinal study of single sex school reform

Shirley M. Yates
School of Education, Flinders University

Since the 1970s many countries have witnessed the changeover of several schools from single sex (SS) to coeducation (CE) through mergers and amalgamations of SS institutions or by the admission of one gender to schools which were the exclusive domain of the other (Mael, 1998). These reforms have been prompted by a range of social concerns (Lee & Bryk, 1986; Tyack & Hansot, 1990), economic crises (Bednall, 1993) and legal reasons such as Title IX in the United States (Hayes & Moses, 1992), but there is very little evidence of the efficacy of this trend. Reformations from single to mixed sex education have been investigated in only a handful of studies (Jones & Thompson, 1981; Marsh, Smith, Marsh & Owens, 1988; Payne & Newton, 1990; Yates, 2002; 2004a) predominantly at the secondary level (Mael, 1998). No significant gender differences in achievement in newly coeducational settings have been found (Marsh, 1989; Marsh et al., 1988; Yates, 2001; 2003) and effects on student self-concept (Marsh, 1989; Marsh et al, 1988; Yates, 2004b), optimistic and pessimistic explanatory style (Yates, 2000) evident only during the transitional phase. However, boys' and girls' perceptions of the climate of newly coeducational learning environments and their educational and career aspirations in the coeducational setting have yet to be investigated at individual and group levels (Payne & Newton, 1990). Furthermore, with the exception of Yates (2004b), the inherent nested multilevel organisational structure of the schools in which the conversions took place and the particular situational and environmental variables that may be operating across cohort groupings and grade levels in those schools have not been considered (Mael, 1998). Appropriate statistical methods must be employed to allow for the interdependence between various factors operating within the changed learning environments (Raudenbush & Bryk, 1988) and investigations of gender differences at the individual student level to be undertaken.

Single sex and coeducation

Research reviews in Australia, United States, Canada, New Zealand, Ireland and the UK have indicated that there is little evidence of consistent advantages of either SS or CE (Dollison, 1998). Overall, the efficacy of each school type for boys and girls is at best equivocal (Mael, 1998) but there is dearth of systematic longitudinal studies of SS and CE learning environments (American Association of University Women (AAUW), 1998a; Pollard, 1999) and of the changes from one to the other. Separate and mixed sex schools have been compared and evaluated in relation to academic achievement, attitudes, curriculum access, selection of non-stereotypical subjects, classroom discipline, social interaction, student self esteem, self-concept and post school success (Mael, 1998; Jackson & Smith, 2000). Some studies suggest SS schools and classes have formal and informal social benefits (Datnow, Hubbard & Woody, 2001; Pollard, 1999; Sadker & Sadker, 1995) as well as academic benefits for boys and girls but equally others suggest CE contexts are preferable (see: Mael, 1998). While the majority of studies have focused on girls (Datnow et al., 2001; Mael, 1998), some suggest "boys only" SS schools maintain male societal prerogatives (Goodman, 1991; Ruhlman, 1996) and promote inequality (McGough, 1991). Although not all researchers concur with this view (e.g., Hawley, 1993; Riesman, 1991), male dominance in CE classrooms at all levels (Barba & Cardinale, 1991; Jones, 1989; Lockheed, 1984; Sadker & Sadker, 1985) and the fostering of gender inequities in CE settings are well documented by a plethora of studies over the last three decades (see AAUW 1992; 1998b for reviews; Mael, 1998). In mixed gender classrooms boys monopolise linguistic interactions (Baker, 1986; Becker, 1981; Sadker & Sadker, 1985), are more likely to exhibit off-task and inappropriate behaviour (Collins, Kenway & McLeod, 2000; Freeman, 2004), to be disengaged (MacDonald, Saunders & Benfield, 1999) and to receive more academic attention and support from teachers (Becker, 1981; Sadker & Sadker, 1985). In CE classrooms girls feel intimidated, are less likely to participate and receive less teacher time (AAUW, 1992; Mael, 1998; Mahony, 1985; Pollard, 1999).

A recent study of a school-wide gender equity program found gender differences in experiences, behaviour and treatment of Grade 7 boys and girls were not only regarded as reflecting inherent or natural differences between genders, but also considered to be equitable and fair by students and teachers alike (Spencer, Porche & Tolman, 2003). However, even if gender inequities are tolerated, they nevertheless have adverse effects on girls including lowered self esteem, learned helplessness and lowered expectations for academic success, particularly in gender stereotyped subjects (Dweck & Repucci, 1973; Gardner, Mason & Matyas, 1989).

If gender inequities are the accepted "norm" in CE classrooms, then how do students fare when SS schools become CE? Questions of gender equity are particularly salient when girls enter the previously exclusively male environments of SS boys' schools. The introduction of CE into any SS school is usually accompanied by a flurry of activity in relation to the physical environment but while it is relatively straightforward to make cosmetic changes to facilities, buildings and sporting fixtures, it is less easy to legislate for gender equity within the classroom and schoolyard (Sadker & Sadker, 1995; Sadker, 1999). Inevitably, the introduction of mixed sex education into a unisex school is also marked by substantial differences in gender ratios, at least for the first few years following the changeover (Henderson, 1993; Yates 2004b). Thus, boys will continue to outnumber girls inside and outside the classroom for some time after the initial introduction of CE into a SS boys' school. In such situations, gender ratio differences may have important implications for girls, as previous studies have found girls' achievement at elementary and secondary levels to be inversely related to the number of boys on the school campus (Sadker, Sadker & Klein, 1991).

Gender differences in student educational and career aspirations

Students' educational intentions, occupational aspirations and attitudes towards school play a significant role in determining whether they will complete secondary school and whether they will participate in post secondary education (Khoo & Ainley, 2005). While an overall decline in student attitudes towards school has been evident from the 1980s (Ainley, Reed & Miller, 1986; Ainley & Sheret, 1992; Freeman, 2004; Marks, 1998), females are more likely to be satisfied with school (Ainley et al., 1986; Ainley & Sheret, 1992; Marks, 1998), to have higher educational aspirations (Bae, Choy, Geddes, Sable, & Snyder, 2000) and to enrol in a tertiary institution immediately after finishing secondary school than their male counterparts (Bae et al., 2000). Although evidence of the relationship between attitudes towards school and achievement outcomes has been quite mixed (Ainley, 1995; Ainley & Sheret, 1992; Mok & Flynn, 1997), the recent *Program for International Student Assessment* (PISA) studies have indicated that countries which have been found to perform well emphasise strategies and approaches for teaching heterogeneous groups of students (Organisation for Economic Co-operation and Development (OECD), 2004). The general consensus of opinion is that academic achievement is higher for students in SS schools (Riordan, 1990; 1993; Lee & Bryk, 1986; Young & Fraser, 1990), particularly for girls (Rowe, 1999), with SS institutions considered to have more serious and studious educational climates (Finn, 1980; Koepke, 1991; Lee & Bryk, 1986). The finding that SS students have higher achievement has been confirmed in large scale studies of secondary schools in the United States (Lee & Byrk, 1986), Australia (Foon, 1988) and Ireland (Cairns, 1990), but these academic advantages in achievement are related to adolescents' higher educational and occupational aspirations in SS schools (Trice, Naudu, Lowe & Jaffee, 1996). However, most studies of the relationships between SS schooling and career aspirations have focused on girls at the postsecondary level (Mael, 1998; Rubenfeld & Gilroy, 1991). While it is generally agreed that women from SS colleges in the United States have higher career aspirations and accomplishments (Rice & Hemmings, 1988; Riordan, 1990), the relative effects of single and mixed sex education on the educational and career aspirations of boys or girls at the secondary level remains largely unknown (Kenway & Willis, 1986; Marsh, 1989; 1991). Further, gender differences in educational and occupational aspirations have not been explored in SS schools which have introduced CE.

School and classroom climate

Much of the controversy about SS and CE has centred on issues of gender equity, without taking into account the content, practice, organisation, climate and culture of particular educational settings (Wahl & Campbell, 1998). In any school, learning takes place in social contexts inside and outside the classroom (Hofman, Hofman & Guldemond, 2001). Student perceptions of school climate (Fraser, 1994; Rothman & McMillan, 2003), particularly for girls in CE schools (Fullarton, 2002) and classroom psychosocial characteristics are related strongly to students' cognitive and affective outcomes (Haertel, Walberg & Haertel, 1981; Fraser, Walberg, Welch & Hattie 1987; Fraser, 1998). A meta-analysis of 12 studies involving 17,805 students in 823 classes in eight subject areas across four nations found student achievement was enhanced in classrooms which students felt had greater Cohesiveness, Goal Direction and Satisfaction and less

Disorganisation and Friction (Haertel et al., 1981). The concepts of Cohesiveness, Goal Direction, Satisfaction, Disorganisation and Friction have their origins in Moos (1974) scheme for classifying human environments along Relationship, Personal Development and System Maintenance and Change dimensions. Cohesiveness, Satisfaction and Friction are classified within the Relationship Dimensions which identifies the nature and intensity of personal relationships within an environment and assesses the extent to which people are involved in that environment and support and help each other (Fraser, 1998). Personal Development embraces Competitiveness and Difficulty and is concerned with self-enhancement and personal development of class members. Goal Direction and Disorganisation belong to Moos (1974) third dimension of System Maintenance and Change Dimensions which involves the extent to which the environment is orderly, responsive to change, clear in expectations and maintains control. These dimensions have been studied in many different learning environments (Fraser, 1998) including the environments of various school types (Docker, Fraser & Fisher, 1989; Dorman & Fraser, 1996; Fisher & Fraser, 1991) but have not been measured in a changeover from SS to CE.

Single sex and coeducational schooling: The shifting scene

CE schooling continues to be the subject of bitter controversy in the United States (Tyack & Hansot, 1990) and has been debated hotly over a considerable period of time in many countries (Cocklin & Battersby, 1987; Jackson & Smith, 2000; Kauermann-Walter, Kreienbaum & Metz-Gockel, 1990) but contemporary concerns about the education of boys (Gilbert & Gilbert, 1998; Murphy & Ivinson, 2000), their increasing disengagement and disenchantment with school (Commonwealth of Australia, 2002; Marks, 1998; Rowe & Rowe, 2002) and the continuing gender segregation and unequal distribution of males and females in courses of study and careers (Sadker, 1999) has led recently to reconsiderations of the widespread adoption of CE schooling in several countries and a resurgence of interest in single gender education (Datnow et al., 2001; Leder & Forgasz, 1997; Parker & Rennie, 1995; Rennie & Parker, 1997; Sadker & Zittleman, 2004; Weaver-Hightower, 2005). Although boys dominate CE classrooms (Sadker et al., 1991), their experiences of school overall are significantly less positive (Cresswell, Rowe & Withers, 2002) in terms of enjoyment (Gentry, Gable & Rizza, 2002), perceived usefulness of the curriculum and teacher responsiveness (MacDonald et al., 1999), particularly for low achieving boys. Furthermore, boys are more likely than girls to exhibit greater externalising behaviours in the classroom (Collins, Batten, Ainley & Getty, 1996; Hinshaw, 1992; Sawyer, Arney, Baghurst, Clark, Graetz, Kosky et al., 2000), to engage in risk-taking behaviours and to drop out of school early (Marks, Fleming, Long & McMillan 2000). Although some consider the provision of SS schools and SS classes in CE schools a means of addressing issues of gender equity, particularly for girls (Australian Education Council, 1993; Parker & Rennie, 1995: Pollard, 1999), others have injected a note of caution into the reinstatement of single gender education provisions stating that it is not the SS context *per se* which is significant but factors such as class size, committed and well trained teachers and a strong academic focus which are crucial (Sadker & Zittleman, 2004; Rowe, 2002).

While the pros and cons of unisex and mixed sex education have been debated extensively (Mael, 1998), the advocacy of one educational type over the other continues to be contentious, with further research evidence for and against both types of arrangements and the conversion of one to the other required. In keeping with the call for longitudinal studies of the long-term socio-emotional effects of school type (Mael, 1998) it is timely to consider gender differences in elementary and secondary students' experiences of newly CE learning environments and the differential effects of these experiences on student educational plans, occupational aspirations and educational achievement over time. Such studies must include considerations of the nested nature of the data as in any classroom or school students share certain background, environmental and experiential characteristics (Osborne, 2000) and are more similar to each other than to students sampled randomly from a school region, sector or national population (Osborne, 2000). Further, for the majority of their time at school students are grouped within classes where they are taught by particular teachers and share learning experiences which affect their development in a variety of ways (Raudenbusch & Willms, 1991).

Design of the study

The study was conducted in an independent school in metropolitan Adelaide, South Australia during and after its transition from a SS school for boys to a CE school. This single campus school offering education from preschool to senior secondary levels, was founded as a CE school in 1847, became an SS school for boys in 1884 and remained an SS boys' institution until 1999 when CE was reintroduced. Mixed gender education was phased into the school over two years, with 43 girls commencing in the secondary Grades 7 to 12 at Time 1 (T1) and a further 24 girls admitted to the elementary Grades 3 to 6 the following year at Time 2 (T2) bringing the total number of girls in the Grades 3 to 12 in that year to 92. Although for privacy reasons it was not possible ascertain which type of school the girls came from at T1 and T2, anecdotal information indicated the girls came from a variety of SS and CE schools, with a very small number at T1 coming from a private SS girls' school in Adelaide that closed at the end of 1998. Girls also entered the preschool at T1 and the Preparatory, Grade 1 and Grade 2 levels of the elementary school at T2, but were not included in the study because of difficulties with the collection of self report questionnaire data from young children. Grade 12 students were also not included at T1, as they were completing their education that year. Career aspirations, plans to stay on at school, to undertake further education, perceptions of the learning environment and educational achievement were measured in all students from Grades 3 to 11 present in the school at T1 and Grades 3 to 12 over the following three years (Time 2, Time 3, Time 4) (T2, T3, T4). Self-report measures were used to gauge student impressions and experiences of the newly CE environment inside and outside the classroom and when pooled provide a rich source of their perceptions of life in the CE school. The study was conducted over four years to allow for developmental changes in student attitudes, aspirations and achievement and to explore interrelationships between these variables as they operated at the individual and school levels. Student aspirations, attitudes and achievement develop throughout their schooling, with earlier aspirations and attitudes often influencing later achievement and with earlier achievement influencing later aspirations and attitudes (Khoo & Ainley, 2005; Rothman & McMillan, 2003).

Aims

The aims of the longitudinal study were to investigate gender differences during and after the introduction of CE into the SS boys' school in students':

1. educational and occupational aspirations;
2. perceptions of the coeducational learning environment; and
3. educational achievement.

Method

Participants

All elementary and secondary students from Grade 3 attending the school in each of the four years participated, with the exception of Grade 12 students at T1. All boys in Grades 3 to 7 and all boys and girls in Grades 8 to 11 participated at T1. All students in Grades 3 to 12 participated at T2, T3 and T4. Girls comprised 9% of the sample at T1, 15% at T2, 20% at T3, 24% at T4 and 17.5% overall. Table 1 presents participants by gender and overall total for the four occasions.

Table 1. Participants by gender at T1, T2, T3 and T4

Time	T1	T2	T3	T4	Total
Boys	441	512	491	489	1933
Girls	43	92	123	151	409
Total	484	604	614	640	2342

Instruments

Educational and Career Aspirations Questionnaire: Students' educational and occupational aspirations were measured with the *Educational and Career Aspirations Questionnaire* (ECAQ) (Yates, 2001) which consisted of a single page on which students recorded their date of birth, gender and grade level, rated their plans for staying at school until Grade 10, Grade 11 or Grade 12 (Grade 12 is the final year of schooling in South Australia) and their expectations of no further education, one to two years, three to four years or more than 4 years of further education after leaving school. Students nominated their intended occupation after leaving school and described the occupation, type of duties they expected to perform and the organisation in which they expected to work. Wording of ECAQ items was adapted from the IEA First International Science Study questionnaire (Comber & Keeves, 1973), with the wording of the occupational aspirations changed slightly for different grade levels.

School Learning Environment Questionnaires: Student perceptions of the CE learning environment were measured with *My School Inventory* (MSI) (Yates, 2001) adapted from *My Class Inventory* (Fisher & Fraser, 1981; Fraser, Anderson & Walberg, 1982) or *School Learning Environment Inventory* (SLEI) (Yates, 2001), modified from the *Learning Environment Inventory* (Anderson & Walberg, 1974; Fraser et al., 1982). Students rated their agreement or disagreement with a number of statements about aspects of the school learning environment in the MSI and SLEI.

Educational Achievement: Student educational achievement was measured with *Word Knowledge Test 1* (WKTest1) *Word Knowledge Test 2* (WKTest2) or *Word Knowledge Test 3* (WKTest3) (Thorndike, 1973). Each test is composed of 40 word pairs which students rated as either the same or opposite in meaning. The tests have been found to be highly predictive of achievement (Fullarton, Lokan, Lamb & Ainley, 2004), can be administered readily, complied to form a single scale with the common items linking procedure and provide a reliable estimate of educational achievement across grade levels. Further, as some students were presented with the same test on more than one occasion, scoring of responses at T2, T3, T4 was anchored to T1 so that any practice effects would be taken into account and a more accurate measure of educational gain estimated.

Procedure

The questionnaires and *Word Knowledge* tests were administered to all students in their classrooms at the same time on the same day in October at T1, T2, T3 and T4. MSI was completed by students in Grades 3 to 7 and SLEI by students in Grades 10 to 12, with students in Grades 8 and 9 completing the MSI and SLEI to provide a common group for equating purposes. Table 2 shows the appropriate Word Knowledge test and school learning environment questionnaire administered to students in the various grades.

Table 2. Word Knowledge Tests and School Learning Environment Questionnaires

WKTest1	Grades 3-7	WKTest2	Grades 8-10	WKTest3	Grades 11-12
MSI	Grades 3-9	SLEI	Grades 8-12		

Analyses

Student responses at T1, T2, T3 and T4 were entered into a Statistical Package for the Social Sciences file, with data matched across the four occasions by a unique student ID assigned at T1. Student plans for further education after leaving school in the ECAQ were coded from 0 to 4, with 0 representing no plans for further education, and 4 representing 4 or more years of further education. Future career aspirations were coded on a 6-point scale, adapted from Brooks, Jones & Zubrzycki (1965), with 1 representing the highest and 6 the lowest occupational levels. Where it was difficult to categorise a student's nominated occupation, the classification was based on the written description of the work, expected duties and type of organisation in which s/he expected to work.

Reliability and validity of the MSI, SLEI, WKTest1, WKTest2 and WKTest3 was established through the one parameter item response theory or Rasch model (Bond & Fox, 2001) which scales the data in such a way that interval scale data are obtained for the variable formed (Wolf, 1994, p. 4926). The Rasch model (Rasch, 1966), which overcomes any sample-item interdependence problems (Hambleton, 1989), postulates estimates of item difficulty are independent of the persons whose performances are used to estimate them and estimates of the performance of persons are independent of the particular items that are attempted (Kline, 1993). For each instrument, all items were analysed with Quest (Adams & Khoo, 1994) which calculates infit mean square (IMS) values for each. Items with IMS values within a predetermined range of 0.83 and 1.20 were considered to fit the Rasch model and were retained while non-fitting items with IMS values outside this range were deleted. One non-fitting item was deleted from SLEI, three from WKTest1, and one item from WKTest2.

The Rasch scaling procedure calibrated the questionnaires and tests to bring them to common interval scales, with the calibration of WK tests based on students who answered all items. The three Word knowledge tests contain some common item pairs, with 13 word pairs common to WKTest1 and WKTest2, 20 pairs common to WKTest2 and WKTest3 and one word pair common to all three tests. A single Word Knowledge (WK) scale of educational achievement was formed from WKTest1, WKTest2 and WKTest3*,* with the tests linked by the common items. MSI and SLEI have three common scales of Cohesiveness, Friction and Satisfaction measuring Moos (1974) Relationships Dimensions and two common scales of Competitiveness and Difficulty assessing students' perceptions of the Personal Development Dimensions within the school. Five separate scales of Cohesiveness, Competitiveness, Difficulty, Friction and Satisfaction were therefore formed from the designated MSI and SLEI items, with each scale linked by responses of Grade 8 and 9 students who completed both questionnaires. The combined Cohesiveness scale contained 13 items, Competitiveness scale 12 items, Difficulty scale 14 items, Friction scale 15 items and Satisfaction scale 14 items. Scores for student responses, referred to as case estimate scores, were estimated through concurrent equating for WK and the five school learning environment scales for all students from Grades 3 to 11 at T1, and Grades 3 to 12 at T2, T3 and T4. Scoring of WK at T2, T3 and T4 was anchored to those students who answered all items at T1.

Relationships between student gender, the five learning environment scales and educational achievement were examined with analysis of variance, structural equation modelling (SEM) and hierarchical linear modelling (HLM). Univariate analyses of variance were undertaken with data from the CE transitional period at T1 and T2, with the SEM and HLM analyses based on case estimate scores from all students pooled across all four years of the study (N = 2342). SEM examines a series of dependent relationships simultaneously and is particularly useful when one dependent variable becomes an independent variable in subsequent dependent relationships (Hair, Anderson, Tatham, & Black, 1995). HLM analysis was used to account for the hierarchical structure of the data and enabled estimation of the influence of variables operating at the individual and group levels. HLM permits examination of direct effect of various potential predictors at both the student Level-1 and group Level-2 and allows for the modelling of cross-level interaction effects.

SEM analyses were carried out with the Analysis of Moment Structure program (AMOS) (Arbuckle & Wothke, 1999). Two path models were developed, with the first model, presented in Figure 1, designed to investigate relationships between gender, student perceptions of Cohesion, Competitiveness, Friction, Satisfaction and Difficulty of schoolwork in the CE learning environment and educational achievement. The second model, presented in Figure 2, explored the relationships between student career aspirations and expectations to undertake further education after leaving school on their perceptions of the learning environment and their educational achievement. In Model 1 and Model 2 variables are presented within rectangular boxes. Arrows between the boxes show only the significant relationships between the variables, with the values along each line reflecting estimates of standardised direct effects. AMOS also calculates significant indirect effects for each variable which are added to the direct effects to produce total effect scores. Although these indirect effects contribute to estimates of the standardised total effects, they will not be considered as for the most part they were less than 0.10 and very small.

Development of each model and the ordering of the variables from left to right within them were guided by previous research evidence, results of the analyses of variance and the aims of the study. For Model 1, student perceptions of school climate have been shown to be related strongly to educational achievement (Fraser, 1994; Rothman & McMillan, 2003), particularly for girls (Fullarton, 2002). Furthermore, Cohesiveness, Friction and Satisfaction have been shown to be strongly predictive of achievement at the class level (Haertel et al., 1981), with Cohesiveness shown to be a significant gender difference in the analysis of variance and Satisfaction gender related in previous studies (Ainley et al., 1986; Ainley & Sheret, 1992; Marks, 1998). Thus, in this model the gender variable was entered first followed by the Cohesiveness, Competitiveness and Satisfaction variables shown in the upper section of the model and the Friction and Difficulty variables in the lower section of the model, with educational achievement designated as the outcome variable. For the second model previous research has highlighted the role of gender in educational and career aspirations (Bae et al., 2000) and the significant influence of aspirations on achievement (Khoo & Ainley, 2005), at least in SS schools (Trice et al., 1996). It was therefore decided to retain the same order for the learning environment variables as Model 1, but to precede them with Career aspirations and plans for Further education. Career aspirations were entered before Further education as they were likely to exert an influence on students' expectations of undertaking further education. Intention to stay at school until the senior years was also entered into the model but was dropped when it was found not to be significant.

HLM analyses were undertaken to investigate group effects on relationships between all of the variables as students moved through the school. Within group comparisons were made over time at the student Level-1, and between cohorts and grades at Level-2 with HLM5 (Raudenbush, Bryk & Congdon, 2000). Two separate models were developed, with students grouped at Level-2 by cohort in Model 3 presented in Figure 3 and by grade at T1, T2, T3 and T4 in Model 4 presented in Figure 4. Educational achievement (shown as WK) was designated as the outcome variable in each model. Cohort groups were comprised of the same students clustered over the four occasions by their initial grade at T1. Thus Cohort 3 was composed of students who were in Grade 3 at T1, Cohort 4 comprised students in Grade 4 at T1, Cohort 5 those in Grade 5, and so on through each of the Grade levels. Grade groupings consisted of students in that grade level at T1, T2, T3 and T4. Thus cohort analyses always involved the same students while Grade level analyses involved different students who were in that grade level on each of the four occasions. Student plans to stay on at school, to undertake further education, career aspirations and perceptions of cohesiveness, competitiveness, difficulty, friction and satisfaction were considered in each model but only the significant effects shown in Figure 3 and Figure 4 were retained. In Models 3 and 4 coefficients and standard errors are presented for each significant variable which is enclosed within an ellipse. Significant student Level-1 variables are shown in the lower rhombus shape in each model and cohort or grade Level-2 variables in the upper rhombus shape.

Results

Analyses of variance

Results of the univariate analyses of variance for the WK scale and five school learning environment scales, conducted with mean case estimate scores, are presented in Table 3. Differences between means were examined across Grade levels, over time and between boys and girls. While a steady, significant pattern of growth in student achievement was evident across all grades during the transition to CE, significant differences were found across Grade levels for perceptions of Cohesiveness, Friction, Satisfaction and Difficulty, across time for Friction, Satisfaction and Competitiveness and between the boys and girls for Cohesiveness only. In the testing of statistical significance the student rather than the class was used as the unit of analysis. While elementary students are grouped in classes within the school, students at the secondary level change classes several times a day according to subjects being studied. Although it was necessary to focus the analyses at the student level, it is recognised that this assumption may not be completely appropriate at the elementary level, where the class is probably the operational unit.

Table 3. Educational progress and school environment scales by Grade level, time and gender

Variable	Grade	Time	Gender
Educational progress	$F = 24.62$***	ns	ns
Satisfaction	$F = 59.45$***	$F = 38.88$***	ns
Friction	$F = 23.60$***	$F = 10.06$***	ns
Difficulty	$F = 26.20$***	ns	ns
Competitiveness	ns	$F = 3.83$*	ns
Cohesiveness	$F = 4.91$***	ns	$F = 4.91$*

*** $p < 0.000$, ** $p < 0.01$, * $p < 0.05$, ns = not significant

Structural equation modelling

Figure 1 indicates that although there was no significant direct relationship between Gender and Educational achievement (referred to as Progress on the model), gender is directly and significantly related to Cohesiveness, Friction and Satisfaction, with girls more cohesive and more satisfied with coeducational school life than boys but with boys perceiving a higher level of friction in the coeducational environment than girls. Educational achievement is directly influenced by Satisfaction and Friction, with Satisfaction and Friction themselves mediated by other variables. More cohesive students report less friction and are more satisfied with their life at school as are those who are less competitive. Conversely, students who perceive higher levels of interpersonal friction are less satisfied with school life, as are those who find their schoolwork more difficult.

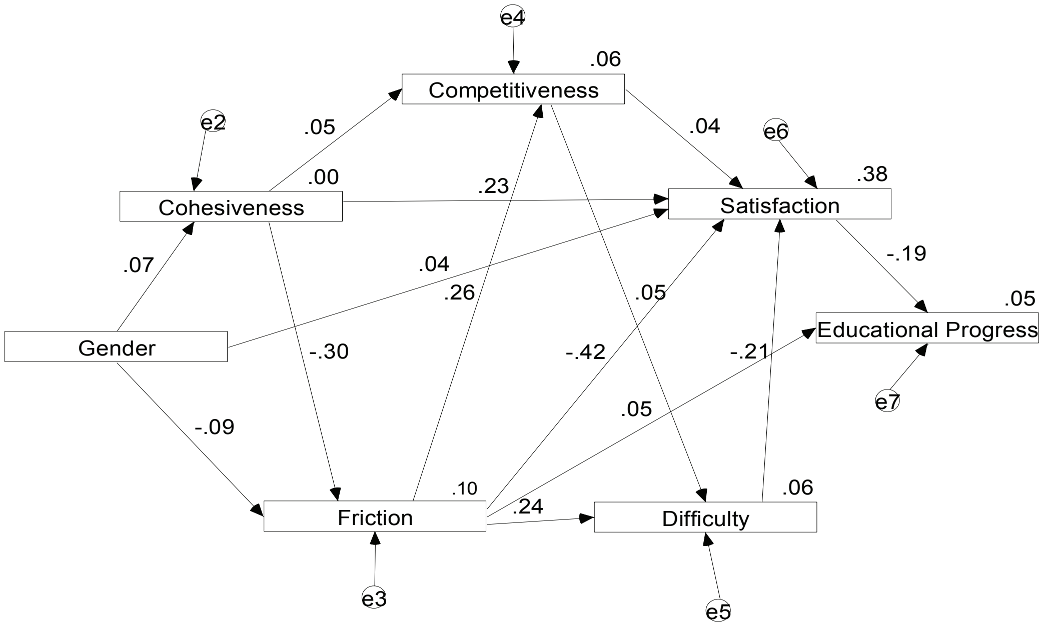


Figure 1. Model 1 Relationships between gender, CE learning environment factors and educational achievement

Figure 2 shows that although there were no significant relationships between student Gender, Career Aspirations and plans for Further Education there were strong direct relationships between Career Aspirations, Further Education plans and Educational Achievement, with Career Aspirations also having an additional indirect effect through Further Education. While the path model shows the same intricate web of direct and indirect relationships between the remaining variables as in Model 1, the slightly stronger negative value for the direct effect between Satisfaction and Educational Progress reflects decreasing satisfaction with school life across grade levels (Yates, 2005).

Hierarchical linear modelling analyses

Students' educational achievement (WK) increased significantly over time in Model 3 (Figure 3) and Model 4 (Figure 4), with a greater rate of increase for Cohort 3 (C3) in Model 3, but a decrease for Grade 3 (G3) in Model 4. Grade 4 (G4) had significantly lower WK scores in Model 2. Satisfaction (SAT) was related negatively to Educational achievement in both models, with more satisfied cohort groups in Model 3 making better progress over time. Further education plans (FED) and competitiveness (COMP) have significant positive effects on educational achievement at Level-1 in both models, with the exception of Cohort 4 (C4) and Grade 7 (G7). However, at the cohort

Level-2 Further education plans relate negatively to WK in Model 3. In Model 4 students who intended to stay at school longer had higher WK scores. Moreover, these scores were higher in Grades where schoolwork was not perceived as difficult and students intended to stay on at school longer. Student career aspirations and perceptions of cohesiveness and friction were not significant factors in either model.

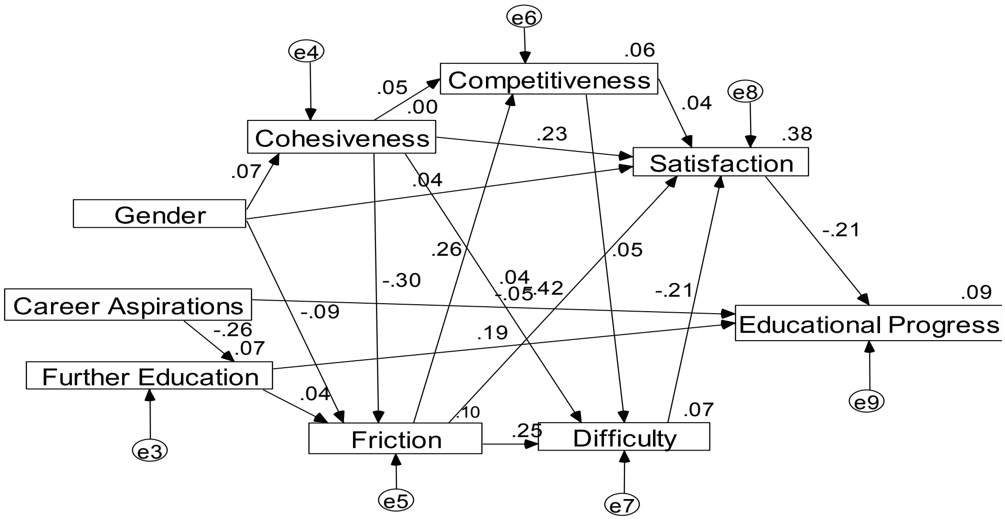


Figure 2. Model 2 Relationships between gender, aspirations, CE earning environment factors and educational achievement

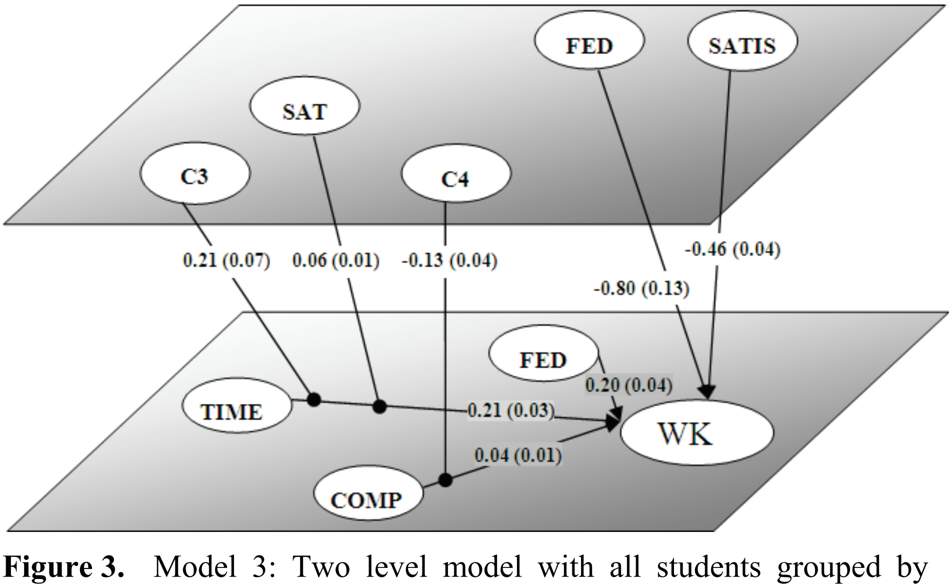


Figure 3. Model 3: Two level model with all students grouped by Cohort

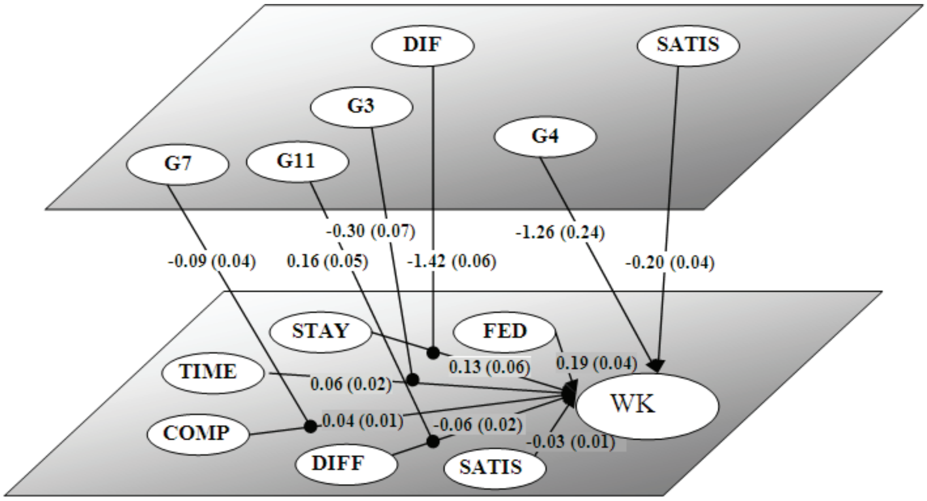


Figure 4. Model 4: Two level model with all students grouped by Grade

Discussion

Dictionary definitions of the word *reform* invariably carry connotations of an improvement or correction of something that is inherently wrong and a return to a former and better state. The last three decades have witnessed numerous reforms of SS schools in several countries, but whether the CE institutions that have taken their place are better for boys and girls have received scant research attention. The admission of girls to environments which have been the exclusive domain of boys, or of boys to all female environments raise issues of gender equity over the short and longer term. Effects of these profound transformations on student aspirations, perceptions and achievement are of particular significance currently as the search for solutions to problems presented by boys and their disaffection with school life is occurring in several countries (Collins et al., 2000; Commonwealth of Australia, 2002; OECD, 2004). Shared experiences of school life affect students in a variety of ways, but whether these experiences should take place in an educational setting which is segregated or integrated by gender is still open to further scrutiny (Mael, 1998). This study makes an important contribution to this quest as it captured perceptions of elementary and secondary level boys and girls during a period of radical school reform and explored gender differences in students' further education plans, career goals, and educational achievement over a four year period.

The first aim of this study focused on gender differences in student intentions to stay at school, to pursue further education and their occupational aspirations after leaving school in the context of the introduction of CE into an SS school. The SEM analyses show clearly that although gender was not a significant factor, students' aspirations had an influence on their further education plans, with both factors related directly to educational achievement. Since career aspirations had been reverse coded, the negative result indicates that higher occupational aspirations were associated with greater expectations of undertaking further education and higher educational progress; relationships had been identified previously as being associated with students from SS schools (Trice et al., 1996). While independent school students are more likely than students from either government or Catholic schools to complete Grade 12 and undertake tertiary education (Long, Carpenter & Hayden, 1999), the HLM analyses indicate that although elementary level students are less sure of their further education plans, it is the further education plans of cohort groups that are so influential.

While gender was not a significant direct factor in student educational achievement, elements of the CE learning environment did play significant mediating roles, with student perceptions of Cohesiveness, Friction and Satisfaction in particular being gender related in the SEM models. Girls were more cohesive, satisfied with life in the CE school and perceived less friction in the CE learning environment than boys. While these differences may simply reflect a "natural order" of accepted and acceptable gender inequities identified by teachers and students in the Spencer et. al., (2003) study, the SEM analyses revealed significant interrelationships between the Relationship Dimensions of Cohesion, Satisfaction and Friction and the Personal Development Dimensions of Competitiveness and Difficulty. Clearly these two dimensions did not operate independently of each other in the CE setting. Relationship dimensions of Cohesiveness, Friction and Satisfaction have been identified previously as being significantly related to student achievement at the classroom level (Haertel et al., 1981). Findings from this study indicate that these Relationship dimensions extend beyond the classroom door to the wider school environment, and furthermore interact together and with the Personal Development dimensions of Competitiveness and Difficulty in a manner that has not been described in previous studies.

The picture of interrelatedness of the Relationship Dimension variables changes somewhat when the hierarchical nature of the data were taken into account through the HLM analyses. Student perceptions of Cohesiveness and Friction were no longer significant factors at either the student or cohort and grade level group levels indicating that over time, these perceptions are not influenced significantly by group clustering effects. By contrast, student Satisfaction with life in the CE school was related directly to educational achievement for cohorts and grades, with more satisfied cohort groups making better progress over time. Girls' greater satisfaction with school life has been reported previously (Ainley et al., 1986; Ainley & Sheret, 1992; Marks, 1998) but the negative relationship between Satisfaction and Educational Progress in the HLM analyses confirms trends of decreasing satisfaction and increasing achievement scores as students proceed through the Grade levels (Yates, 2001, 2002, 2005). This decline in satisfaction has been evident in several other studies (see Gentry, Gable & Rizzo, 2002), but the HLM shows that these effects are stronger for particular cohort groups as they move through the school.

CE was phased into this school over a two year period with a small numbers of girls entering the secondary grades in the first year and the elementary grades the following year. Although the number of girls in the school increased steadily across the four years of the study, boys continued to outnumber girls inside and outside of the classroom throughout the entire time period. Fears of gender inequities in CE classrooms cited by so many studies and the concern that the disproportionate number of boys may have an effect on the achievement of the girls (Sadker et al., 1991) was not borne out in any of the analyses. Although there were clear indications of significant increases in educational achievement for all students, particularly for cohort groups over time, gender *per se* was not a significant factor. Boys were not disadvantaged academically by the inclusion of girls in their school and equally, girls were not disadvantaged by entering a learning environment that had been formerly and exclusively male dominated. The finding that academic performance was not affected adversely confirms results from an earlier study of secondary school students (Marsh et al., 1989) but extends the findings to include elementary level students. Furthermore, results show that the effects are much stronger for cohorts over time than across grade levels, with the notable exception of students in Grade 4 who had significantly lower achievement on all four occasions. Similar patterns of achievement for 4^{th} graders have been shown in the United States *National Assessment of Educational Process* assessments in reading from 1992 - 2003 and science from 1995 – 2003 (Wirt, Choy, Rooney, Hussar, Provasnik, & Hampden-Thompson, 2005).

The last thirty years have witnessed an unprecedented number of reforms of SS schools across several countries yet despite the pervasiveness of these reforms and the profound changes they have brought to the lives of countless numbers of children, the reforms have passed almost without notice. Results from this study would suggest that the introduction of CE into this SS boys' school was beneficial overall and that the current resurgence of interest in SS classrooms as a means of addressing gender issues is contraindicated, at least for boys. The intricate and detailed analyses of the results from this longitudinal study have shed some interesting light on the SS/CE reform in one school, but as this study was conducted in an independent school with an avowed academic focus there is a need to capture further reforms of boys' SS schools particularly in the government sector, to extend these studies to include SS girls' schools and to investigate the views of teachers.

Acknowledgements

This study was supported by two Australian Research Council Small Grants and two Flinders University Small Grants. Assistance from Professor John Keeves, John Harris, Dianne Harris and Dr Ngurah Darmawan is gratefully acknowledged.

References

Adams, R. J. & Khoo, S. K. (1994). *Quest The interactive test analysis system.* Hawthorn, Victoria: ACER.

Ainley, J. (1995). Student views of their schools. *Unicorn. 21*(3), 3-16.

Ainley, J. Reed, R. & Miller, H. (1986). *School organisation and the quality of schooling*. ACER Research Monograph No. 29. Melbourne: Australian Council for Educational Research.

Ainley, J. & Sheret, M. (1992). *Progress through high school: A study of senior secondary schooling in New South Wales.* ACER Research Monograph No. 43. Melbourne: Australian Council for Educational Research.

American Association of University Women (1992). *How schools shortchange girls*. Washington, DC: American Association of University Women Educational Foundation.

American Association of University Women (1998a). *Separated by sex: A critical look at single sex education for girls*. Washington, DC: American Association of University Women Educational Foundation.

American Association of University Women (1998b). *Gender gaps: Where schools still fail our children.* Washington, DC: American Association of University Women Educational Foundation.

Anderson, G. J. & Walberg, H. J. (1974). Learning environments. In H. J. Walberg (Ed.), *Evaluating educational performance: A sourcebook of methods, instruments, and examples.* Berkeley: McCutchan.

Arbuckle, J. L. & Wothke, W. (1999). *Amos 4.0 User's Guide*. USA: Small Water Corporation.

Australian Education Council (1993). *National action plan for the education of girls, 1993-1997.* Carlton, Victoria: Curriculum Corporation.

Bae, Y., Choy, S., Geddes, C., Sable, J & Snyder, T. (2000). *Trends in educational equity of girls and women* (NCES 2000-030). Washington, DC: U.S. Department of Education Office of Education, National Center for Educational Statistics.

Baker, D. (1986). Sex differences in classroom interactions in secondary science. *Journal of Classroom Interactions, 22*, 212-218.

Barba, R. & Cardinale, L. (1991). Are females invisible students? An investigation of teacher-student questioning interactions. *School Science and Mathematics, 12,* 40-53.

Becker, J. (1981). Differential treatment of males and females in mathematics classes. *Journal for Research in Mathematics Education, 12*, 40-53.

Bednall, J. (1993). Let's test the assumptions and examine some realities. In D. C. Hollinger & R. Adams (Eds.), *Single-sex schooling: Proponents speak*. (pp. 21-26). Washington, DC: US Department of Education.

Bond, T. G & Fox, C. M. (2001). Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, N.J: Lawrence Erlbaum.

Brooks, L., Jones, F. L. & Zubrzycki, J. (1965). An occupational classification for the Australian workforce. *Australian and New Zealand Journal of Sociology, 1*. Supplement.

Cairns, E. (1990). The relationship between adolescent perceived self-confidence and attendance at a single sex secondary school. *British Journal of Educational Psychology, 60*, 207-211.

Cocklin, B. & Battersby, D. (1987). Gender, the hidden agenda: A case study in educational decision-making. *New Zealand Journal of Educational Studies, 22,* 59-70.

Collins, C., Batten, M., Ainley, J. & Getty, C. (1996). Gender and school education: A project funded by the Commonwealth Department of Employment, Education, Training and Youth Affairs. Australian Council for Educational Research.

Collins, C., Kenway, J. & McLeod, J. (2000). *Factors influencing the educational performance of males and females in school and their destination after leaving school.* Canberra, ACT: Commonwealth Department of Education, Training and Youth Affairs.

Comber, L. C. & Keeves, J. P. (1973). Science Education in Nineteen Countries. International Studies in Evaluation I. New York: Wiley.

Commonwealth of Australia (2002). *Boys: Getting it right Report on the inquiry into the education of boys.* Canberra: The Parliament of the Commonwealth of Australia. www.aph.gov.au/house/committee/edt/eofb/index.htm

Cresswell, J. Rowe, K. & Withers, G. (2002). *Boys in school and society*. Melbourne: Australian Council for Educational Research.

Datnow, A., Hubbard, L. & Woody, E. (2001). Is single gender schooling viable in the public sector? Lessons from California's pilot program.

Docker, J. G., Fraser, B. J. & Fisher, D. L. (1989). Differences in the psychosocial work environment of different types of schools. *Journal of Research in Childhood Education, 4,* 5-17.

Dollison, R. A. (1998) A comparison of the effect of single-sex and coeducational schooling arrangements on the self-esteem and mathematics achievement of adolescent females. Alabama: Education Document Reproduction Service http://ericae.net/ericdc/ ED422188.htm.

Dorman, J. P. & Fraser, B. J. (1996). Teachers' perceptions of school environment in Australian Catholic and Government secondary schools. *International Studies in Educational Administration, 24*(1), 78-87.

Dweck, C. & Repucci, N. (1973). Learned helplessness and reinforcement responsibility in children. *Journal of Personality and Social Psychology, 25*, 109-116.

Finn, J. D. (1980). Sex differences in educational outcomes: A cross-national study. *Sex Roles, 6,* 9-25.

Fisher, D. L. & Fraser, B. J. (1981). Validity and use of My Class Inventory. *Science Education, 65,* 145-156.

Foon, A (1988). The relationship between school type and adolescent self esteem, attribution styles and affiliation needs: implications for educational outcome. *British Journal of Educational Psychology, 58*, 44-54.

Fraser, B. J. (1994). Research on classroom and school climate. In D. Gabel (Ed.), *Handbook of research on science teaching and learning.* (pp. 493-541) New York: Macmillan.

Fraser, B. J. (1998). Science learning environments: Assessment, effects and determinants. In B. J. Fraser & K. G. Tobin (Eds.), *The international handbook of science education* (pp. 527-564). Dordrecht, The Netherlands: Kluwer.

Fraser, B. J., Anderson, G. J. & Walberg, H. J. (1982). *Assessment of learning environments: Manual for Learning Environment Inventory (LEI) and My Class Inventory (MCI)* (third version) Perth: Western Australian Institute of Technology.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145-252.

Freeman, C. E. (2004). *Trends in educational equity of girls and women: 2004*. US Department of Education. National Center for Educational Statistics. http://nces.ed.gov/pubs2005/200516.pdf.

Fullarton, S. (2002). *Student engagement with school: Individual and school level influences.* Longitudinal Surveys of Australian Youth Research Report No. 27. Melbourne: Australian Council for Educational Research.

Fullarton, S., Lokan, J., Lamb, S., & Ainley, J. (2004). *Lessons from the Third International Mathematics and Science Study.* TIMSS Australia Monograph No. 4. Melbourne: Australian Council for Educational Research.

Gardner, A., Mason, C. & Matyas, M. (1989). Equity, excellence, and "just plain good teaching." *The American Biology Teacher, 51*, 72-77.

Gentry, M., Gable, R. K. & Rizza, M. G. (2002). Students' perceptions of classroom activities: Are there grade-level and gender differences? *Journal of Educational Psychology, 94*(3), 539-544.

Gilbert, R. & Gilbert, P. (1998). *Masculinity goes to school*. London: Routledge.

Goodman, E. (1991, June 22) VMI: Repository of reaction. *The Washington Post,* p. A23.

Haertel, G. D., Walberg, H. J. & Haertel, E. H. (1981). Socio-psychological environments and learning: A quantitative synthesis. *British Educational Research Journal, 7,* 27-36.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis*. Englewood Cliffs, New Jersey: Prentice Hall Inc.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. *Education Measurement.* (3rd edition), New York: Macmillan.

Hawley, R. (1993). The case for boys' schools. In D. K. Hollinger & R. Adamson (Eds.), *Single sex schooling: Proponents speak* (pp. 11-14). Washington, DC: US Department of Education.

Henderson, D. (1993). Hurray for girls. *Times Educational Supplement*, October 1, n4031 pR12(2), A14567508.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127-155.

Hofman, R H., Hofman, W. H. & Guldemond, H. (2001), Social context effects on pupils' perceptions of school. *Learning and Instruction, 11*(3), 171-194.

Jackson, C. & Smith, I. D. (2000). Poles apart? An exploration of single-sex and mixed sex educational environments in Australia and England. *Educational Studies, 26*(4). 409-422.

Jones, M. (1989). Gender bias in classroom interactions. *Contemporary Education, 60*, 216-222.

Jones, M. B. & Thompson, D. G. (1981). Classroom misconduct and integration by sex. *Journal of Child Psychology and Psychiatry, 22,* 401-409.

Kauermann-Walter, J., Kreienbaum, M. A., & Metz-Gockel, S. (1990). Formal quality and discrete discrimination. *Western European Education, 22*, 66-79.

Kenway, J. & Willis, S. (1986). Feminist single-sex educational strategies: Some theoretical flaws and practical fallacies. *Discourse, 7,* 1-30.

Khoo, S. T. & Ainley, J. (2005). *Attitudes, Intentions and Participation*. Longitudinal Surveys of Australian Youth Research Report 41. Melbourne: Australian Council for Educational Research.

Kline, P. (1993). Rasch scaling and other scales. *The handbook of psychological testing.* London: Routledge.

Koepke, M. (1991). A school of their own. *Teacher Magazine*. pp. 44-47.

Leder, G. H. & Forgasz, H. J. (1997). Single-sex classes in a coeducational high school: highlighting the parents' perspectives. *Mathematics Education Research Journal, 9*(3), 274-291.

Lee, V. E. & Bryk, A. S. (1986). Effects of single-sex secondary schools on student achievement and attitudes, *Journal of Educational Psychology*, *78,* 381-395.

Lockheed, M. (1984). *Final report: A study of sex equity in classroom interactions*. Washington, DC: National Institute of Education.

Long, M., Carpenter, P. & Hayden, M. (1999). *Participation in education and training 1980-1994.* Longitudinal Survey of Australian Youth Research Report No. 13. Melbourne: Australian Council for Educational Research.

MacDonald, A., Saunders, L. & Benfield, P. (1999). *Boys' achievement, progress, motivation and participation: Issues raised by the recent literature*. Slough, UK: National Foundation for Educational Research.

Mael, F. A. (1998). Single-sex and coeducational schooling: Relationships to socioemotional and academic development. *Review of Educational Research, 68*(2), 101-129.

Mahony, P. (1985). School for the boys? Coeducation reassessed. London: Hutchinson.

Marks, G. (1998). *Attitudes to school life school: Their influences and their effects on achievement and leaving school.* Longitudinal Surveys of Australian Youth Research Report No. 5. Melbourne: Australian Council for Educational Research.

Marks, G., Fleming, N., Long, M. & McMillan, J. (2000). *Patterns of participation in Year 12 and higher education in Australia: Trends and issues.* Longitudinal Surveys of Australian Youth Research Report No. 17. Melbourne: Australian Council for Educational Research.

Marsh, H. W. (1989). Effects of attending single-sex and coeducational high schools on achievement, attitudes, behaviours and sex differences. *Journal of Educational Psychology*, *81*, 70-85.

Marsh, H. W. (1991). Public, Catholic single-sex and Catholic coeducational high schools: Their effects on achievement, affect, and behaviors. *American Journal of Education*, *99*, 320-356.

Marsh, H. W., Smith I. D., Marsh, M., & Owens, L. (1988). The transition from single-sex to coeducational high schools: Effects on multiple dimensions of self-concept and on academic achievement. *American Educational Research Journal, 25,* 237-269.

McGough, M. (1991). Romeboys. *The New Republic*, pp 13-16.

Mok, M. & Flynn, M. (1997). Quality of school life and students' achievement in the HSC: A multilevel analysis. *Australian Journal of Education, 41*, 169-188.

Moos, R. H. (1974). *The social climate scales: An overview*. Palo Alto, California: Consulting Psychologists Press.

Murphy, P. & Ivinson, G. (2000). *Construction of knowledge, social identities and pedagogy in single and mixed sex groupings.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Organisation for Economic Co-operation and Development (2004). *Issues for discussion.* Paris: OECD.

Osborne, J. W. (2000). Advantages of Hierarchical Linear Modeling. *Practical Assessment, Research and Evaluation 7*(1).

Parker, L. H. & Rennie, L. J. (1995). For the sake of girls? Final report of the Western Australian single-sex education pilot project 1993-1994. Report for the Education Department of Western Australia, Perth.

Payne, M. A. & Newton, E. H. (1990). Teachers' and students' perceptions of the major advantages and disadvantages of coeducational secondary schooling. *Australian Journal of Education, 34*, 67-86.

Pollard, D. S. (1999). Single sex education. *WEEA Digest.* Newton, MA: Education Development Centre.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19,* 49-57.

Raudenbush S. W. & Bryk, A. S. (1988). Methodological advances in analysing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education 1988-1989*. Vol 15 (pp. 423-475). Washington, DC: American Educational Research Association.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2000). *HLM5 Hierarchical linear and nonlinear modelling.* Scientific Software International.

Raudenbush, S. W. & Willms, J. D. (Eds.), (1991). Schools, classrooms and pupils International studies of schooling form a multilevel perspective New York: Academic Press

Reisman, D. (1991). A margin of difference; the case for single sex education. In J. R. Blau (Ed.), *Social roles and social institutions*. Boulder, CO: Westview Press. pp. 241-257.

Rennie, L. J & Parker, L. H. (1997). Perceptions of single- sex and mixed-sex mathematics classes. *Mathematics Education Research Journal. 9,* (3), 257-273.

Rice, J. K. & Hemmings, A. (1988). Women's colleges and women's achievers: An update. *Journal of Women in Culture and Society, 13,* 546-559.

Riordan, C. (1990). *Girls and boys in school: together or separate?* New York: Teachers College Press.

Riordan, C. (1993). The case for single sex schools. In D. K. Hollinger & R. Adamson (Eds.), *Single sex schooling: Proponents speak* (pp 47-53). Washington, DC: US Department of Education.

Rothman, S. & McMillan, J. (2003). *Influences on achievement in literacy and numeracy.* Longitudinal Surveys of Australian Youth Research Report 36. Melbourne: Australian Council for Educational Research.

Rowe, K. J. (1999). *VCE Data Project (1994-1999): Concepts, issues, directions & specifications.* A research project conducted for the Board of Studies, Victoria. Centre for Applied Educational Research, The University of Melbourne.

Rowe, K. J. (2002). The importance of teacher quality. *Issue Analysis 22*. Sydney, NSW: Centre for Independent Studies. http://www.cis.org.au.

Rowe, K. J. & Rowe, K. S. (2002). What matters most: evidence-based findings of key factors affecting the educational experiences and outcomes for girls and boys throughout their primary and secondary schooling. Supplementary submission to the House of Representatives Standing Committee on Education and Training: Inquiry into the Education of Boys. Melbourne, Victoria: Australian Council for Educational Research and Department of General Paediatrics, Royal Children's Hospital.

Rubenfeld, M. I. & Gilroy, F. D. (1991). Relationship between college women's occupational interests and a single sex environment. *Career Development Quarterly, 40,* 358-361.

Rulman, M. (1996). *Boys themselves: a return to single sex education*. New York: Henry Holt and Company.

Sadker, M. P. (1999). Gender equity: Still knocking at the classroom door. *Educational Leadership, 56,* 22-26.

Sadker, M. P. & Sadker, D. M. (1985). Is the OK classroom OK? *Phi Delta Kappan, 66,* 358-361.

Sadker, M. P. & Sadker, D. M. (1995). *Failing at fairness: How America's schools cheat girls.* New York: Touchstone Press.

Sadker, M. P., Sadker, D. M. & Klein, S. (1991). The issue of gender in elementary and secondary education. *Review of Research in Education, 17,* 269-334.

Sadker, D. M. & Zittleman, (2004). Single-sex schools: A good idea gone wrong? The *Christian Science Monitor*, p.9.

Sawyer, M. G., Arney, F. M., Baghurst, P. A., Clark, J. J., Graetz, B. W., Kosky, R. J., Nurcombe, B., Patton, G. C., Prior, M. R., Raphael, B. Rey, J. Whaites, L. C. & Zubrick, S. R. (2000). *Child and adolescent component of the National Survey of Mental Health and Wellbeing: The mental health of young people in Australia.* Canberra: Commonwealth Department of Health and Aged Care.

Spencer, R., Porche, M. V. & Tolman, D. L. (2003). We've come a long way -- maybe: New challenges for gender equity in education. *Teachers College Record, 105*(9), 1774-1807.

Thorndike, R. L. (1973). *Reading comprehension in fifteen countries.* New York, Wiley: and Stockholm: Almqvist & Wiksell.

Trice, A. D., Naudu, A., Lowe, M. & Jaffee, H. (1996). Men's college classroom environments. *Education, 116*(4), 572-576.

Tyack, D & Hansot, E. (1990). *Learning Together: A History of Coeducation in American Schools.* New Haven, CT: Yale University Press.

Wahl, E & Campbell, P. (1998). Of two minds: Single sex/coeducation and the search for gender equity. *New York Law School Journal of Human Rights, 14.*

Weaver-Hightower, M. (2005). Dare the school build a new education for boys? *Teachers College Record*. http://www.tcrecord.org/PrintContent.asp?ContentID=11743.

Wolf, R. M. (1994). Rating scales. In T. Husén & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education.* (Second edition). (Vol. 8, pp. 4923-4930), Pergamon: Elsevier Science.

Wirt, J., Choy, S., Rooney, P., Hussar, W., Provasnik, S. & Hampden-Thompson, G. (2005). *The condition of education 2005.* US Department of Education: National Center for Educational Statistics. http://nces.ed.gov/programs/coe.

Yates, S. M. (2000). *Student optimism and pessimism during the transition to co-education*. Annual conference of the Australian Association for Research in Education, Sydney, New South Wales, www.aare.edu.au/00pap/yat00439.htm.

Yates, S. M. (2001). Students' achievement and perceptions of school climate during the transition from single sex to coeducation. International Education Journal, Educational Research Conference 2001 Special Issue.2, (4), 317-328.

Yates, S. M. (2002). Stepping from single to mixed sex education: Boys' progress and perceptions during the restructuring. International Education Journal, Educational Research Conference 2002 Special Issue 3, (4), 23-33.

Yates, S. M. (2003). Gender differences in perceptions of school climate: A longitudinal study of school reform. Joint conference proceedings of the New Zealand and Australian Associations for Research in Education, Auckland, New Zealand. http://www.aare.edu.au/03pap/ yat03139.pdf.

Yates, S. M. (2004a). Aspirations, progress and perceptions of boys from a single sex school following the changeover to coeducation. International Education Journal Special Educational Research Conference 2003 Issue, 4, (4), 167-177.

Yates, S. M. (2004b). Grade and cohort effects on boys' self-concept and achievement during the transition from single sex to coeducation. Proceedings of the 3rd International Biennial SELF Conference, Berlin. http://self.uws.edu.au/Conferences/2004_Yates.pdf

Young, D. J. & Fraser, B. J. (1990). Science achievement of girls in single-sex and coeducational schools. Research in Science and Technological Education, 8, 5-20.

18

Factors influencing science performance in Indonesia: What PISA 2006 tells us

I Gusti Ngurah Darmawan and *Firenius Sitinjak*
School of Education, the University of Adelaide

Introduction

No longer is science in schools being viewed merely as an important initial step in the training of professional scientists and technologists. The importance of science and technology and the need to develop these life skills are now widely recognised.

> People often have to draw appropriate conclusions from evidence and information given to them; they have to evaluate claims made by others on the basis of the evidence put forward and they have to distinguish opinion from evidence-based statements. Often the evidence involved is scientific, but science has a more general role to play as well since it is concerned with rationality in testing ideas and theories against evidence. Of course this does not deny that science includes creativity and imagination, attributes that have always played a central part in enhancing human understanding of the world. (OECD, 2006, p.21)

The learning of science has been and is a major interest of researchers in many countries. In the Indonesian context, the government has focused a lot of attention on science, and through the Indonesian Ministry of Education, has implemented several strategies to improve the teaching and learning of science.

In a broader context, around the world many researchers have investigated factors that may explain student achievement in science. These factors include gender (Young and Fraser, 1994; Weinburg, 1995; Lee and Burkam, 1996; DeMars, 1998; Dimitrov, 1999), socioeconomic status (Lamdin, 1996; Ripple and Luthar, 2000; Sirin, 2005; Mark, 2006), attitude towards science (Hough and Piper, 1982; Fredman, 1997; Kind, Hones and Barmby, 2007), school sector (Jaminez and Lockheed, 1995; Goldhaber, 1996), school location (Young, 1998; Webster and Fisher, 2000), ability grouping (Slavin, 1987; Kulik and Kulik, 1982, 1987, 1992), and teacher quality (Rivkin, Hanushek, and Kain, 2005; Akiba, LeTendre and Scribner, 2007). Most of the studies, however, have a major weakness because they only examine single level factors in relation to science achievement. In fact, examining student achievement in science would appear to need a multilevel model which not only examines individual factors but also explores the group relationships among and between these characteristics.

Furthermore, most of the studies to evaluate determinants of achievement in science, using either simple or multilevel models, have been conducted in developed countries with mature education systems while few studies have been done in developing countries. Factors that influence science achievement in developed countries may not have the same influence in developing countries such as Indonesia. This study addresses this incongruence by investigating both student and school factors influencing school science achievement in Indonesia in order to inform national and district policies.

A Glance at the education system in Indonesia

Indonesia is a very diversely populated nation with 33 provinces and more than 300 local languages. It is a country with more than two hundred million people scattered over more than 13,000 islands. Indonesia is located in Southeast Asia and shares geo-political boundaries with the Philippines to the north, Malaysia and Singapore to its northwest, Papua New Guinea to the east, Australia in the southeast, and the Indian Ocean to its south.

The existence of Indonesian public education can be traced back to a system of village schools established by the Dutch colonial government in 1906. The number of public schools was approximately 3,500 in the year 1913, together with a similar number of private and religious schools. However, very few native children at this time were permitted to study in the Dutch schools up to university level and the racial discrimination was very obvious (Kristiansen & Pratikno, 2006). During the Japanese occupation in the early 1940s, the education systems were changed slightly. The racial discrimination present in the Dutch era was not encouraged during the Japanese occupation. Native children had more access to education compared to the Dutch regime. Several Indonesian identities were allowed at public schools such as Indonesian language and Indonesian national anthems (Kopong, 1995).

The education conditions changed gradually after Indonesian people proclaimed their independence on August 17th, 1945. Following Indonesian's independence, the development of national education was given more attention. Schools were opened to people regardless of their race, ethnicity, language, religion, and culture. This policy was in line with the 1945 constitution, which stipulated that every citizen has the right to obtain education and that the government has the responsibility to provide one national education system (Kristiansen & Pratikno, 2006).

Over the last six decades, the Indonesian education system has been centralised. Most educational policies have been centrally planned in Jakarta (Kopong, 1995). The school curriculum was also centrally planned and developed in Jakarta. Currently, the situation is gradually changing following an era of reform and in response to the local autonomy law of 1999. According to Kristiansen and Pratikno (2006), there has been a dramatic change in Indonesian education systems since 2001. Education systems have been decentralised through a form of educational autonomy. Managerial and financial responsibilities for all levels of public education have been decentralised from the central government, which is mainly located in Jakarta, to local government at the district level.

School Level Curriculum (SLC) or *Kurikulum Tingkat Satuan Pendidikan* (KTSP) is an operational curriculum which is set down and provided by each of the schools. SLC is mandated through Law No 20/2003 concerning the National Education System and National Regulation No 19/2005 concerning the National Education Standardization. The arrangement of SLC by the schools started in 2006/2007 based on Content Standardization and Competence Standardization for basic and Middle Education as produced by Education Minister N0 22/2006 and No 23/2006, and the Guidance of SLC Development issued by Education Standardization Board.

Education in Indonesia includes kindergarten, primary school, junior high school, senior or vocational high school, and tertiary education. Primary education consists of six years and junior high school has three years. This is followed by three years of general or vocational senior high school. After completing senior high school, students may go to tertiary education to get their first degree which may be at university, polytechnics or academics. They may then progress to a master's program and follow that with a doctorate program.

PISA 2006: An overview

The Program for International Student Assessment (PISA) was set up by the Organisation for Economic Cooperation and Development (OECD) to monitor literacy in the fields of Reading, Mathematics and Science and to conduct testing programs at three-year intervals. This program has received strong support from most highly developed countries of the world as well as many developing countries with well over 50 countries being actively involved in this ongoing monitoring work. The first data collection took place in 2000, the second in 2003 and the third in 2006. The fourth cycle is in progress for 2009.

PISA measures students' ability to complete tasks relating to real life rather than limiting the assessment to the understanding of subject specific knowledge. The assessment of scientific literacy has particular importance in 2006, where it is the major domain being assessed, and it is defined as follows:

> Scientific literacy is an individual's scientific knowledge and use of that knowledge to identify questions, to explain scientific phenomena, and to draw evidence-based conclusions about science-related issues, understanding of the characteristic features of science as a form of human knowledge and enquiry, awareness of how science and technology shape our material, intellectual, and cultural environments, and willingness to engage in science-related issues, and with the ideas of science, as a reflexive citizen. (OECD, 2006, p.13)

In contrast to the earlier definitions, the PISA 2006 definition of scientific literacy has been expanded by explicitly including attitudinal aspects of students' responses to issues of scientific and technological relevance. This definition of scientific literacy consists of four interrelated aspects:

- scientific knowledge,
- scientific competencies,
- situations and context, and
- attitudes towards science.

The first two of these aspects are directly involved in the development of science curricula as well as in tests of achievement in the field of science. The third aspect serves to ensure that in the development of a science curriculum attention is paid to locating in the real world those aspects of science that are taught, learnt and assessed in a wide range of settings and that are relevant to the needs, interests and aptitudes of the students involved. The fourth aspect signifies the important role people's attitudes play in their interest, attention, and response to science and technology.

Multilevel model

In behavioural and educational sciences, the multilevel model analysis has become a standard method for analysing data structured hierarchically (Afshartous & Leeuw, 2005; Berkhof & Snijders, 2001; Frees & Kim, 2006). Multilevel models are proven tools in social research for modelling complex, hierarchical systems. According to (Hox, 2002, p. 1):

> Generally, the individuals and the social groups are conceptualised as hierarchical system of individuals and groups, with individuals and groups defined at separate levels of this hierarchical system. Naturally, such systems can be observed at different hierarchical levels, and variables may be defined at each level. This leads to research into the integration between variables characterizing individuals and variables characterizing groups, a kind of research that is now often referred to as 'multilevel research'.

Luke (2004, pp.7-8) defined the multilevel model as "a statistical model applied to data collected at more than one level in order to elucidate relationships at more than one level". He added that the statistical modelling has been labelled with various terms, including hierarchical linear models, random coefficient model, mixed-effects models and covariance structure models. The aim of a multilevel model is to predict values of some dependent variables based on a function of independent variables at more than one level.

Johnson (1995) argued that educational research data gathered in natural settings are multilevel in nature since factors affecting behaviour operate at several levels, including student, class, school and system levels. Any variables from other levels, such as student and class levels, excluded from a single analysis probably make an important contribution to the variation of outcome variables. Therefore, it is necessary to investigate educational issues using a multilevel model in order to acquire a deeper and more comprehensive understanding of the issue.

Luke (2004) suggested the following two main reasons for using a multilevel model: theoretical and statistical reasons. The former relates to the reality that many phenomena people want to study are multilevel in nature. As a consequence, people should use theories and analytic techniques that are also multilevel. Otherwise, they will encounter considerable limitations. The second reason for using a multilevel model is to achieve a statistical perspective. Some researchers have tended to use single level statistical techniques for their data, even if their data and hypotheses are multilevel in nature. One approach that has been used is to disaggregate group level-information to an individual level. Thus, all predictors in a multiple regression model are tied to the individual analysis. This leads to some problems such as the violation of the basic assumptions of multiple regressions. This means that traditional statistical methods such as the one that analyses data using single level techniques are questionable when dealing with multilevel data. Based on these two reasons, it is clear that a multilevel statistical model is needed to deal with multilevel data.

Hypothesised model

Testing of hypotheses in multilevel models can be carried out using multilevel data analyses software such as HLM 6 for Windows (Raudenbush et al., 2004). The HLM program was initially developed to find a solution for the methodological weakness of educational research studies during the early 1980s that was the failure of many analytical studies to attend to the hierarchical, multilevel character of much of educational field research data (Bryk and Raudenbush, 1992). This failure came from the fact that "the traditional linear models used by most researchers require the assumption that subjects respond independently to educational programs" (Raudenbush and Bryk; 1994, p. 2590). In practice, most educational research studies select students as a sample that are nested within classrooms, and the classrooms are in turn nested within schools, and schools within geographical locations. In this situation, the students selected in the study are not independent, but rather nested within organisational units and ignoring this fact results in the problems of "aggregation bias and misestimated precision" (Raudenbush and Bryk, 1994, p. 2590).

Figure 1 shows the two-level model proposed for testing in this study. The names, codes and description of the predictor variables tested for inclusion at each level of the two-level model have been provided in Table 1.

Data used

Regardless of grade or school type in which students were enrolled and whether they enrolled for full time or part time, PISA 2006 covered students who were aged between 15 years 3 months and 16 years 2 months at the time of the assessment and who have completed at least 6 years of formal schooling (OECD, 2007). Around 400,000 students in 57 countries: 30 OECD member countries and 27 non-OECD member countries were involved in the PISA 2006 study. Students were selected in a two-stage stratified sample, schools were selected first and, within each selected school, students were randomly sampled. Experts from the PISA Consortium performed the sample selection process for each participating country and monitored it closely in those countries where they selected their own samples. (OECD, 2007).

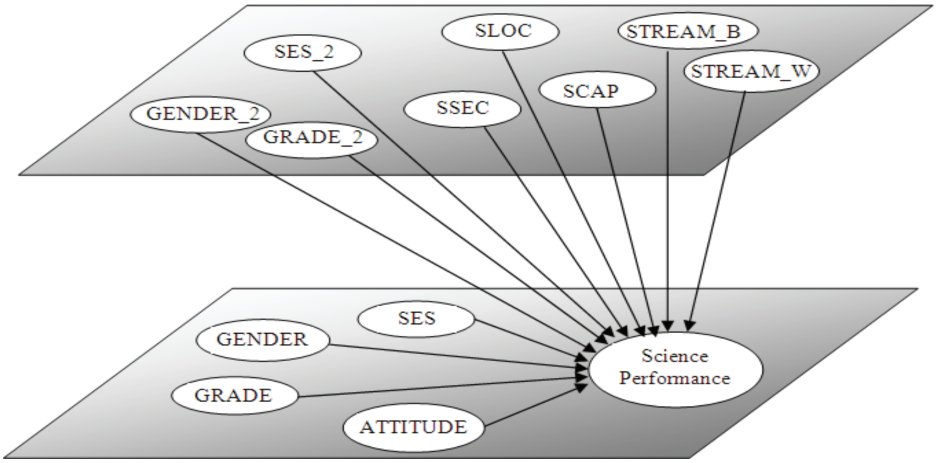


Figure 1. Hypothesised two-level model of Science Performance (direct effects only)

Table 1. List of variables

Level	Variable Code	Descriptions
Level 1	**Student Level**	
Outcomes	SCIENCE	Science Performance
Student Background	GRADE	Students' grade, min = 7, max = 11
	GENDER	Students' sex, female = 0, male = 1
	SES	Socio Economic Status, composite scores of parent's education, parent's occupation, and home possessions
	ATTITUDE	Students' attitudes towards science, composite scores of attitudinal items
Level 2	**School Level**	
School Characteristics	SSEC	School Sector, Public = 0, Private = 1
	SLOC	School's community, Village= 1…Large City = 5
	SCAP	School's resources, composite scores of human, material and educational resources
	STREAM_B	Ability grouping between classes, for all subjects = 1… not for any subject = 3
	STREAM_W	Ability grouping within class, for all subjects = 1… not for any subject = 3
Group Composition	GRADE_2	Average grades
	GENDER_2	Proportion of male students
	SES_2	Average SES

In the Indonesian context, the sampling strategy was conducted based on the framework set by OECD. The number of responding schools taken as a sample was 352. The number of the 15 year old students assessed as a sample was 10,647 (OECD, 2006).

Analyses

The multilevel models were built step by step. The first step was to run a model without explanatory variables, which is also called the 'null model'. The null model was fitted to provide estimates of the variance components at each level (Bryk and Raudenbush, 1996). The null model can be stated in an equation form as follows.

Level-1 model: $Y_{ij} = \beta_{0j} + r_{ij}$

Level-2 model: $\beta_{0j} = \gamma_{00} + u_{0j}$ (1)

where: Y_{ij} is the science performance of student i in school j

The second step undertaken was to estimate a level-1 model, that is, a model with student-level variables as the only predictors in Equation 1 above. This involved building up the student-level model or the so-called 'unconditional' model at level-1 by adding student-level predictors to the model, but without entering predictors at any of the other levels of the hierarchy. At this stage, a step-up approach was followed to examine which of the four student-level variables (listed in Table 1) had a significant (at $p \leq 0.05$) influence on the outcome variable, Science Performance. All four variables (GRADE, GENDER, SES, and ATTITUDE) were found to be significant and therefore were included in the model at this stage.

The final step undertaken was to estimate a level-2 model, which involved adding the level-2 or class-level predictors into the model using the step-up strategy mentioned above. At this stage, the level-2 exploratory analysis sub-routine available in HLM6 was employed to examine the potentially significant level-2 predictors (as found in the output) in successive HLM runs. Following the step-up procedure, three class-level variables (SCAP, STREAM_W and SES_2) were included in the model for the intercept. In addition, three cross-level interaction effects were included in the model (GENDER and GRADE_2, GENDER and SES_2, as well as ATTITUDE and SES_2).

The final model for science performance at levels 1 and 2, presented in Table 2 and Figure 2, can be denoted as follows:

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(\text{GRADE}) + \beta_{2j}*(\text{GENDER}) + \beta_{3j}*(\text{SES}) + \beta_{4j}*(\text{ATTITUDE}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{SES\_2}) + \gamma_{02}*(\text{SCAP}) + \gamma_{03}*(\text{STREAM\_W}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}*(\text{GRADE\_2}) + \gamma_{22}*(\text{SES\_2}) + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$
$$\beta_{4j} = \gamma_{40} + \gamma_{41}*(\text{SES\_2}) + u_{4j} \quad (2)$$

Table 2. Final model for science performance

Fixed Effect	Final estimation of fixed effects			d.f.	Approx. P-value
	Coefficient	Std Error	T-ratio		
For INTRCPTl, BO					
INTRCPT2, GOO	379.38	1.80	210.54	348	0.000
SES_2, GOI	29.82	3.12	9.53	348	0.000
SCAP, G02	7.23	1.86	3.87	348	0.000
STREAM_W, G03	-5.29	2.10	-2.51	348	0.013
For GRADE slope, Bl					
INTRCPT2, GI0	16.42	1.49	11.01	90	0.000
For GENDER slope, B2					
INTRCPT2, G20	8.68	1.31	6.62	31	0.000
GRADE_2, G21	6.42	1.96	3.27	117	0.002
SES_2, G22	-5.41	1.89	-2.86	349	0.005
For SES slope, B3					
INTRCPT2, G30	1.19	0.64	1.85	265	0.065
For ATTITUDE slope, B4					
INTRCPT2, G40	6.39	0.67	9.42	36	0.000
SES_2, G41	2.23	0.95	2.33	96	0.022

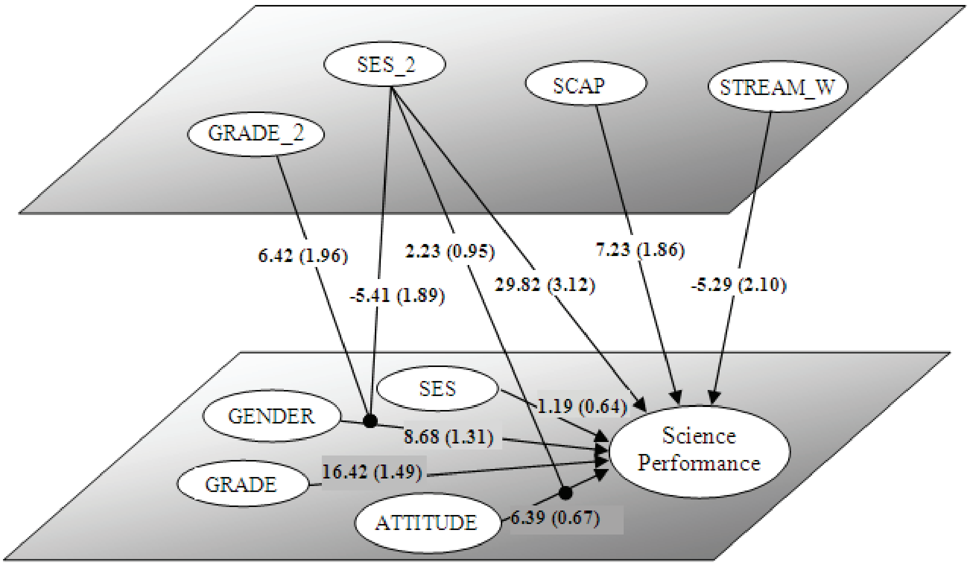


Figure 2. Final two-level model of science performance

Direct effects

All four Level-1 variables have significant effects on science performance. Grade makes a positive contribution to science performance. It means that the higher the grade of the students the higher their science performance. This finding is in line with the argument that the higher the grade the more content knowledge has been acquired by students. A similar pattern was also reported by Young and Fraser (1994).

The effect of gender on science performance is also positive. In other words, male students perform better than female students. This finding is similar to Linn and Hyde's (1989) study. In their meta-analysis they concluded that males outperform females in science. In another study it was noted that the difference between boys and girls might exist when there is a grouping of sub-topics in science or a grouping of students' ability (Dimitrov, 1999; DeMars 1998). De Mars (1998) concluded that if all ability levels were considered, the interaction between gender difference and science performance was small. When only the highest ability students were considered, male students scored higher on the multiple-choice section, whereas female students scored higher on the constructed-response section. The effect of gender difference on science performance varies according to the level and ability of the students.

SES has a positive impact on science performance. This means that the higher the index of SES of the students, the higher the level of science performance. In this study the index of SES consists of three main categories, namely the index of occupational status of the parents, the index of highest educational level of the parents and the index of home possessions. The finding is in contrast to the findings of Mark (2006) that students' performance cannot be accounted for by socioeconomic status. The socioeconomic background in his study was measured by combining the fathers’ and mothers’ occupation and education. It seemed that there might have been some influence of home possessions on science performance as shown in this study. The finding in this study seemed to be similar to a study conducted by Sirin (2005). Sirin's meta-analysis study found that there was a medium to strong relationship between SES and the students' achievement in the United States. This meta-analysis study was conducted by reviewing the literature on SES and academic achievement in journal articles published between 1999 and 2000. The different of geographical background did not seem to be an issue both in Sirin's study and in this study.

Attitude toward science is found to have a positive impact on science performance. This means that the higher the level of students' attitude, the higher the level of science performance. The finding is similar to the studies of Freedman's (1997) and Hough and Piper's (1982). In Freedman's study attitude toward science is developed-through hands-on, activity-based laboratory instruction. Teaching that makes science more exciting and encourages students (e.g., laboratory work) is considered to have a positive influence on attitude toward and achievement in science.

Ability grouping within class has a negative impact on science performance. This means that grouping within classes has a negative influence on science performance. As there are three options in grouping within classes, namely no grouping for any subject, grouping for some subjects and grouping for all subjects, it seems that, based on the PISA 2006 study, it would be better if the grouping is ignored in science classes. This finding is in contrast to a meta-analysis study conducted by Lou et al. (1996). Although they argued that within-class grouping is beneficial for students in facilitating student learning, particularly in large classes and especially in mathematics and science courses, the case in Indonesia might be different. It could be that the cooperative learning which it is hoped can be developed through grouping does not help in the case of the PISA 2006 study. Low ability students who are expected to benefit most when placed in mixed-ability groups do not achieve higher results.

School capacity is found to have a positive impact on science performance. It means that the greater the availability of the school resources the higher the science performance. There are three parts of school capacity, namely human resources, material resources and educational resources. The finding supports the conclusion of Heyneman and Loxley (1983) that for a few of the world school systems, such as those in Europe, North America and Japan the influence of factors such as school resources and teacher quality on academic achievement is less than that of family background or other characteristics of students. However, in most Asian countries, the effects of school variables are far greater than that of family background.

The factor of school capacity in Indonesia might be represented through material and resources at schools. Although there are many items that can be grouped as material and resources, the library and laboratory can be grouped as important resources in supporting students learning. The availability of libraries is not only important to support students' learning but also to develop students' information skills as part of their education. The laboratory has been given a central and distinctive role in science education, and science educators have suggested that there are rich benefits in science learning from using laboratory activities (Hofstein and Lunetta, 2002). As science teaching and learning should use inquiry skills, the laboratory is especially important to support science teaching and learning. However, in the case of Indonesia, there is a lack of such support.

Interaction effects

The so-called 'interaction effects' are at the core of the additional information that HLM produces as a result of the concept of 'slope as outcome' analysis (Bryk and Raudenbush, 1992). In general, a cross-level interaction effect relates three variables to one another, namely, the outcome variable, its level-l predictor and a level-2 variable that is considered to influence the effect of the level-l predictor on the outcome variable. In order to illustrate the interaction effect and the detail, which could be obtained from HLM output, parts of the equations for the final model involving the variables are set at zero since neither of the two variables are involved and there is no loss in generality.

The average grade (GRADE_2) has a positive effect on the slope of gender (GENDER). As it can be seen in Figure 3, in general, male students outperformed female students. However, the difference is larger in schools where more students in higher grades participated in the study. This result suggests that the gap in science performance is bigger in the upper level compared to the lower level of schooling.

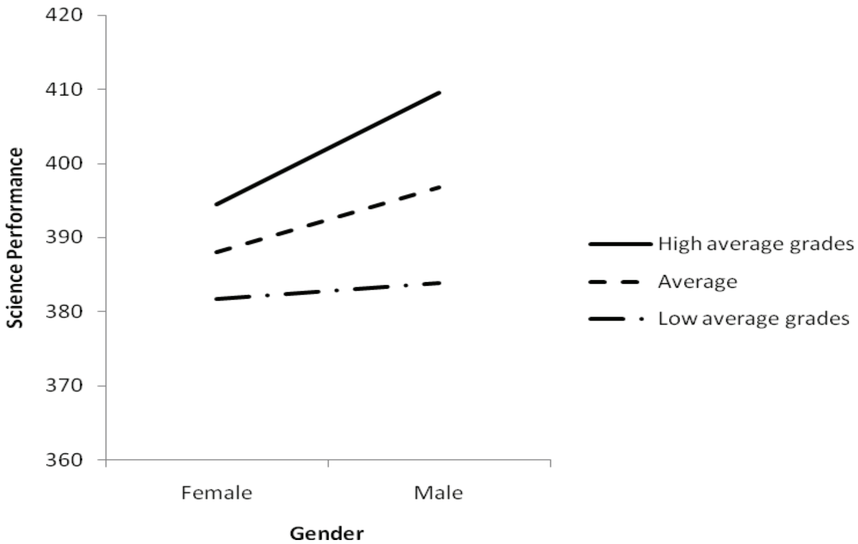


Figure 3. Cross-level interaction effects of GRADE_2 and GENDER on science performance

The effect of average SES (SES_2) on the slope of gender (GENDER) is shown in Figure 4. It can be seen in the figure that even though male students consistently performed better than female students, the difference is more apparent in schools that have lower average SES. This might be due to the fact that families with high socio-economic status are more likely to give equal educational aspiration and motivation to both their sons and daughters. In contrast, for families with a low socio economic status, when it comes to educational support, males seem to be in a better position.

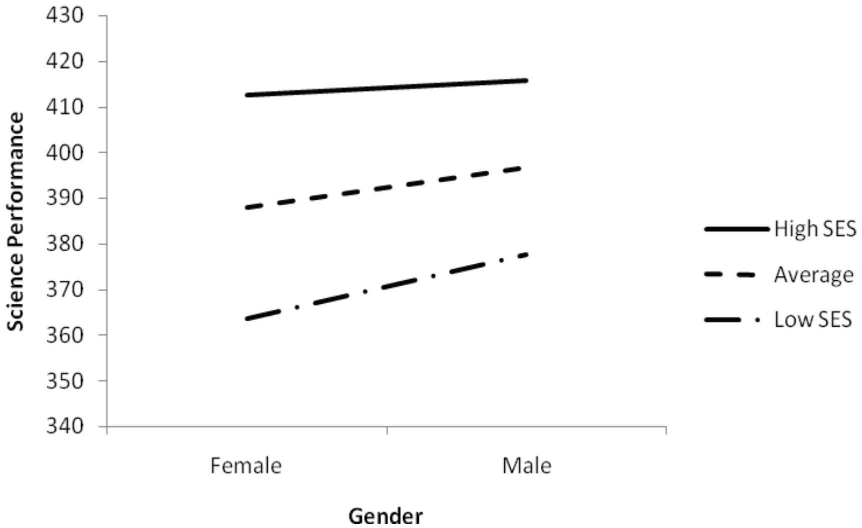


Figure 4. Cross-level interaction effects of SES_2 and GENDER on science performance

From Figure 5, it can be seen that students in schools with higher average SES tend to perform better than their counterparts in schools with lower average SES. In addition, it can also be argued that students' attitude toward science has a positive effect on science performance. This effect is stronger in schools that have higher average SES.

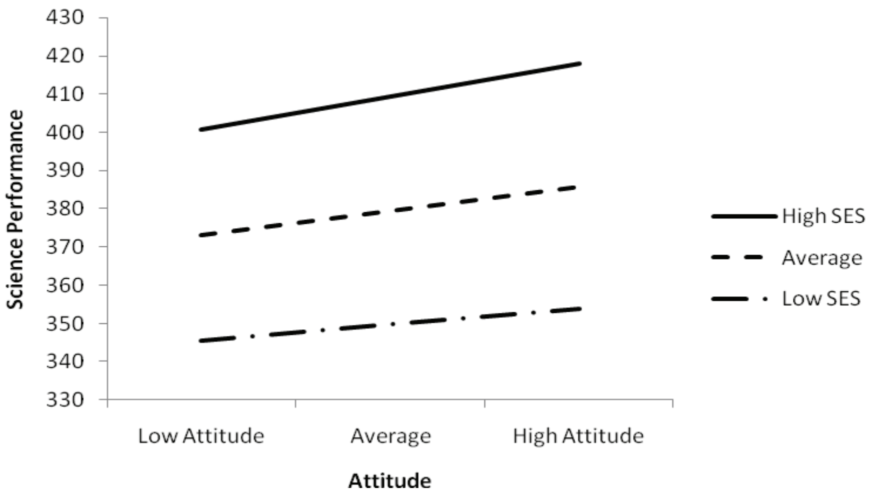


Figure 5. Cross-level interaction effects of SES_2 and ATTITUDE on science performance

Estimation of variance explained

In addition to the cross level interaction, it is also of interest to examine the variance components at both between student (Level 1) and between school (Level 2) levels. The results of the final estimation of variance components for final model and the results of the analyses of the variance components obtained from the null models are presented in Table 3.

The results in Table 3 show that, 57.1 per cent and 42.9 per cent of the variance of student scores are at the Levels 1 and 2 respectively. These percentages of variance of student scores at the various levels of the hierarchy are the maximum amounts of variance available at those levels that could be explained in subsequent analyses.

Table 3. Estimation of Variance components

Estimation of Variance Components:	between students (n = 10,647)	between schools (n=352)	Total
Null Model	2,286.48	1,724.51	4,010.99
Final Model	2,159.90	883.75	3,043.65
Variance available	57.1%	42.9%	100%
Variance explained	5.5%	48,8%	
Total variance explained	3.1%	20.9%	24%

In addition, the results in Table 3 show that the variables included in the final model explain 5.5 per cent of 57.1 per cent variance available at Level-1 and that is equal to 3.1 per cent (that is, 30.4 × 96.1) of the total variance explained at the Level-1. Similarly, the variables included in the final model explain 48.8 per cent of 42.9 per cent variance available at Level 2 and that is equal to 20.9 per cent. Thus, the total variance explained by the variables included in the final model is 24 per cent, which leaves 76 per cent of the total variance unexplained.

In summary, the results in Table 3 indicate that the model developed in this study explains almost half of the between-schools (Level-2) variance but explains only a small amount of the between students (Level-1) variance. The large amount of variance left unexplained at Level-1 strongly indicates that there are other important Level-1 factors influencing the students' science performance that have not been included in the models developed in this study. Therefore, there is a clear need for a further study to develop models that are the most appropriate for explaining students' science performance and which maximize the total variance explained at Level-1.

Conclusions

In this study, data from 10.647 students from 352 schools in Indonesia are used to examine various factors that may influence science performance. It should be noted that the discussions in this article are based on preliminary results of rich and complex data that need further examination before drawing conclusions or making policy recommendations. Nevertheless, this article has shed some light in explaining science performance among 15-year old Indonesian students.

In this study it was found that science performance is influenced by seven variables directly. The seven main effects are the direct effects from school capacity, ability grouping within classes, average SES, grade, gender, SES, and attitude toward science. In addition, three cross level interaction effects are also found. The three cross-level interaction effects involve average grade and gender, average SES and gender, and average SES and attitude.

Since grade, gender, attitude toward science, SES, school capacity and grouping within class influence science performance Walbergs' theory of educational productivity (Walberg, 1981, 1984) matches this study. The theory holds that there are nine factors which contribute to variance in students' achievement and affective outcomes: student ability, age, and motivation; the quality and quantity of instruction; and the psychological climate of the home, the classroom social group, the peer group outside the classroom, and the mass-media (especially television viewing). He concluded that the psychological attributes of individual students and their proximate environments directly and indirectly influenced cognitive, behavioral, and attitudinal outcomes.

It might be argued that in order to improve science performance, it is important to develop better school resources. The school resources capacity greatly influences science performance, followed by SES. The improvement of school resources may lead to improvement of science performance. The development of better schools is urgently required. From the questionnaires it can be seen that generally schools are hindered by the following factors and shortages: (a) science laboratory equipment, (b) instructional material, (c) computers for instruction, (d) computer software for instruction, (e) library materials, (f) audio-visual resources and particularly (g) Internet connectivity. It seems that the problems here are related to funding.

The findings might provide a clearer picture of science performance in Indonesia if parents had been involved in the questionnaire. As the PISA 2006 study did not involve the gathering of questionnaire information from the parents, the limitation in the type of the data appears to have influenced findings considerably. Moreover, the study poses many challenges in the collection and analysis of data, which are multilevel in nature, operating at the student and school level, where appropriate methods of multilevel modeling are not widely known. The management of these challenges, together with the practical implications of the study, indicates that this investigation provides a substantial and significant contribution to understanding the forces that operate in this and similar situations.

Acknowledgement

I would like to express my gratitude to Professor John P. Keeves who gave me the possibility to enter the world of probability. His guidance, stimulating suggestions and encouragement have helped me in my research, during the writing of my PhD thesis, and beyond.

References

Akiba, M., LeTendre, G. K., and Scribner, J. P. (2007). Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. *Educational Researcher,* 36 (7), 369-387.

Afshartous, D., & Leeuw, J. D. (2005). Prediction in multilevel models. *Journal of Educational and Behavioral Statistics,* 30(2), 109-139.

Bedi, A. S., and Garg, A. (2000). The effectiveness of private versus public schools: the case of Indonesia. *Journal of Development Economics,* 61, 63-494.

Bell, J. F. (2001). Investigating gender differences in the science performance of 16-year-old pupils in the UK. *International Journal of Science Education,* 23(5), 469-486.

Berkhof, J., & Snijders, T. A. B. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics,* 26(2), 133-152.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Model.* Newbury Park, CA: Sage.

DeMars, C. E. (1998). Gender Differences in Mathematics and Science on a High School Proficiency Exam: The Role of Response Format. *Applied Measurement in Education*, 11(3), 279-299.

Dimitrov, D. M. (1999). Gender Differences in Science Achievement: Differential Effect of Ability, Response Format, and Strands of Learning Outcomes, School Science and Mathematics. Online: http://www.questia.com/googleScholar. viewed 12/09/2008.

Fleming, M. L., and Malone, M. R. (1983). The relationship of students' characteristics and students' performance in science as viewed by meta-analyses research. *Journal of Research in Science Teaching,* 20, 481-495.

Freedman, M. P. (1997). Relationship among laboratory instruction, attitude towards science, and achievement in science knowledge. *Journal of Research in Science Teaching,* 34(4), 343-357.

Frees, E. W., & Kim, J. S. (2006). Multilevel model prediction. *Psychometrika,* 71(1), p79-104.

Germann, P. J. (1988). Development of the attitude towards science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching,* 25(8), 689-703.

Goldhaber, D. D. (1996).Public and private high school: is school choice an answer to the productivity problem? *Economics of Education Review,* 15(2), 93-109.

Heyneman, S. P., and Loxley, W.A. (1983). The effect of Primary School Quality on Academic Achievement across Twenty nine High and Low Income Countries. *The American Journal of Sociology,* 88(6), 1162-1194.

Hofstein, A., and Lunetta,V. N. (2003). The laboratory in science education: Foundations for the twenty-first century. *Science education,* 88(1), 28.

Hough, L. W., and Piper, M. K. (1982). The relationship between attitude towards science and science achievement. *Journal of Research in Science Teaching,* 19(1), 33-38.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. New Jersey: Lawrence Erlbaum Associates.

Jimenez, E., and Lockheed, M. E. (1995). *Public and private secondary education in developing countries: a comparative study.* Online: http://eric.ed.govIERICWebPortal/custom/portiets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED398122&ERICExtSearch_SearchType_O=no&accno=ED398122, viewed 12/09/2008.

Johnson, T. G. (1995). *Multilevel models of student achievement in science.* Unpublished Ph.D Thesis, Flinders University of South Australia, Adelaide.

Keeves, J. (2004). Monitoring the learning and teaching of science in a changing world. *International Education Journal.* 5(3).

Kind, P., Jones, K., and Barmby, P. (2007). Developing Attitude towards Science Measure. *International Journal of Science Education,* 29(7), 871-893.

Kopong, E. (1995). *Culture and schooling: An exploratory study of curriculum decentralisation in Indonesia.* Unpublished Ph.D Thesis, Flinders University of South Australia, Adelaide, Australia.

Kristiansen, S., & Pratikno. (2006). Decentralising education in Indonesia. *International Journal of Educational Development,* 26, 513–531.

Kulik, C. L., and Kulik, J. (1982). Effect of ability grouping on secondary school students. A meta-analysis of evaluation finding. *American Educational Research Journal,* 19, 415-428.

Kulik, J. A., and Kulik, C. L. (1987). Effect of ability grouping on student achievement. *Equity and Excellence,* 23, 22-30.

Kulik, J. A., and Kulik, C. L. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, 36(2), 73.

Lamdin, D. J. (1996). Evidence of Student Attendance as an Independence Variable in Education Production Function. *Journal of Educational Research,* 89(3), 155-162.

Lee, V. E., and Burkam, D. T. (1996). Gender Differences in Middle Grade Science Achievement: Subject Domain, Ability Level, and Course Emphasis. *Science & education,* 80(6), 613.

Linn, M. c., and Hyde, J. S. (1989). Gender, Mathematics, and Science. *Educational Researcher,* 18(8), 17-27.

Lou, Y., Abrami, P. C., Spence, J. c., Poulsen, C., Chambers, B., and d' Apollonia, S. (1996). Within-class grouping: a meta-analysis. *Review of Educational Research,* 66(4), 423-458.

Luke, D. A. (2004). *Multilevel modeling*. London: SAGE Publications.

Mark, G. N. (2006). Are between- and within-School Differences in Student Performance Largely due to Socioeconomic Background? Evidence from 30 Countries. *Educational Research,* 48(1), 21-40.

OECD. (2006). *Assessing Scientific, Reading, and Mathematical Literacy. A Framework for PISA 2006.* Online: http://www.oecd.orgldataoecd/63/35/37464175.pdf, 20/03/2008.

OECD. (2007). *PISA 2006. Science Competencies for Tomorrow's World. Volume I Analysis.* Online: http://www.pisa.oecd.org/ document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,0 O.html, viewed 07/04/2008.

Raudenbush, S. W., & Bryk, A. S. (1994). Hierarchical linear models. In T. Husén & T. N. Postlethwaite (Eds.), *International Encyclopedia of Education: Research and Studies* (2nd ed., 2590-2596). Oxford: Pergamon Press.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., and Congdon, R. T. Jr. (1996). *HLM: Hierarchical Linear and Nonlinear Modeling,* Chicago: Scientific Software International, Inc.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling* . Lincolnwood, IL: Scientific Software International.

Ripple, C. H., and Luthar, S. S. (2000). Academic Risk among Inner-City Adolescents: The Role of Personal Attribute. *Journal of School Psychology,* 38(3), 277-298.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica,* 73(2), 417-458.

Schreiber, J. B. (2002). Institutional and student factors and their influence on advanced mathematics achievement. *The Journal of Educational Research, 95*(5), 274-286.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research,* 75(3), 417-453.

Slavin, R. E. (1987). Achievement effect of ability grouping in secondary schools: a best evidence synthesis. *Review of Educational Research,* 60(3), 47l-499.

Walberg, H. J. (1984). Improving the productivity of America's school. Education Leadership, 41, 19-27.

Webster, B. J., and Fisher, D. L. (2000). Accounting for variation in science and mathematics achievement: a multilevel analysis of Australian data Third International Mathematics and Science Study (TIMMS). *School Effectiveness and School Improvement,* 11(3), 339-360.

Weinburgh, M. (1995). Gender Differences in Student Attitudes toward Science: A MetaAnalysis of the Literature from 1970 to 1991. *Journal of Research in Science Teaching, 32(*4), 387-398.

Young, D. J. (1998). Rural and urban differences in student achievement in science and mathematics: a multilevel analysis. *School Effectiveness and School Improvement,* 9(4), 386418.

Young, D., and Fraser, B. J. (1994). Gender Differences in Science Achievement: Do School Effects Make a Difference. *Journal of Research in Science Teaching,* 31(8), 857 - 871.

19

An examination of change over time in the approaches to learning in a sample of sojourner students from Confucian heritage cultures using hierarchical linear modelling

Bobbie Matthews
Flinders University

Introduction

A longitudinal study was undertaken to investigate whether students from Confucian heritage cultures (CHC) changed or in some way modified their approaches to learning in a Western social, cultural and educational environment. Multilevel modelling, using hierarchical linear modelling (HLM) (Bryk, Raudenbush & Congdon, 2000) was employed to examine the data. Thus, analyses were undertaken at the intra and inter student levels to assess the degree of change that had occurred. HLM was also able to indicate an absence of change.

In order to ensure an interval scale of measurement, the raw score data that were obtained from the SPQ were scaled and equated using the QUEST program (Adams & Khoo, 1993). The Rasch model was employed to produce interval scales on which all items in a particular scale and all participants in the study were placed. This procedure enabled the examination of the performance of individual students to be examined over time, rather than the mean performance of a sample of students. This was particularly useful as there was considerable variability between different students who provided responses on the five occasions of measurement. Therefore, if a student responded on at least two of the five occasions, the data provided were included.

An original sample of 153 cases was examined on five occasions over a period of two years in Australia. At the completion of the investigation a total of 573 case records were available for analysis. Only responses from sojourner students who said that they intended to return to their home country on completion of their studies were included in the data set. Data on the students' approaches to learning were gathered using a bilingual (Chinese and English) version of Biggs' (1987a, 1987b) *Study Process Questionnaire* (SPQ). On the first occasion this was done by the investigator personally, but on subsequent occasions, the data were collected on line, using the internet and a university server. There were three occasions of measurement in the first year and two in the second year of the study.

The initial impetus for differentiating between various learning approaches was based on a study undertaken by the Swedish team of Martön and Säljö (1976a, 1976b). Their report described an investigation undertaken with university students who were asked to read an academic paper and then were questioned as to what they had learned and how this learning had come about. The students' responses indicated that they tended to use two principal ways of processing the information they read. Some tried to memorise details or key words in order to answer subsequent questions and tended to focus on the word or sentence level. Others attempted to understand the message imparted by the passage globally, focused on themes and principal ideas, and tried to process the content for meaning. These approaches and associated reading strategies were called 'surface' and 'deep' approaches respectively.

Based on these ideas and findings, Biggs (1987a, 1996) described three distinct learning approaches, namely surface, deep and achieving. Each approach was a composite of a motivation that encouraged a particular type of learning as well as a strategy for the implementation of the motivation. Together they formed an approach to learning. The surface approach was characterised by limited, but directed aspirational goals. The deep approach was based on the desire to learn as much about a topic as possible and the achieving approach could be recognised by a competitive attitude to learning that led to ego enhancement. Based on these ideas, Biggs (1987a) developed and validated two instruments, one aimed at measuring approaches to learning in adolescents (the *Learning Process Questionnaire*, LPQ) and one aimed at measuring approaches to learning in adults (the *Study Process Questionnaire*, SPQ).

The SPQ consisted of 42 items whereby each of the six subscales comprised seven items. There was no consideration of gender difference on the SPQ questionnaire; although the gender of participants was recorded. The SPQ was designed to measure the following concept (an example from each scale is included): (a) Surface Motivation ("Whether I like it or not, I can see that further education is for me a good way to get a well paid or secure job."); (b) Surface Strategy ("I think browsing around is a waste of time, so I only study seriously what's given out in class or in the course outlines."); (c) Deep Motivation ("I find that at times studying gives me a feeling of deep personal satisfaction."); (d) Deep Strategy ("I find that I have to do enough work on a topic so that I can form my own point of view before I'm satisfied."); (e) Achieving Motivation ("I would see myself basically as an ambitious person and want to get to the top, whatever I do."); and (f) Achieving Strategy ("I try to work consistently throughout the term and review regularly when the exams are close.") Participants responded to each item using a five-point Likert-type scale that ranged from (1) "Never or only rarely true of me." to (5) "Almost or almost always true of me."

Table 1 presents a summary of the approaches to learning specified by Biggs (1987a).

Table 1. Motivations and strategies in student approaches to learning

Approach	Motive	Strategy
SA: Surface	**Surface Motivation** (SM) is instrumental: to meet requirements minimally; a balance between working too hard and failing.	**Surface Strategy** (SS) is reproductive: to limit target to bare essentials and reproduce through rote learning.
DA: Deep	**Deep Motivation** (DM) is intrinsic: study to actualise interest in what is being learned; to develop competence in academic subjects.	**Deep Strategy** (DS) is meaningful: read widely, inter-relating with previous relevant knowledge.
AA: Achieving	**Achieving Motivation** (AM) is based on competition and ego-enhancement: to obtain highest grades, whether or not material is interesting.	**Achieving Strategy** (AS) is based on organising time and working space; to follow up suggestions; behave as a 'model' student.

Following Biggs (1987a) and Murray-Harvey (1994)

The structure of the approaches to learning was established using confirmatory factor analysis and the LISREL 8.30 computer program (Jöreskog & Sörbom 1999). The best-fitting model, of the many that were tested, was a baseline model whereby the six approaches to learning (latent variables) were allowed to correlate freely with each other and the observed (manifest) variables. The approaches to learning were investigated by a examination of pertinent research on the relationship between predictor variables, the student variables, the learning approaches and outcomes of learning.

The Surface Approach is utilitarian. The motivation is to gain maximum qualifications and strike a balance between working too hard and failing. The strategy that is employed is reproductive and often utilises rote learning. The Deep Approach is based on actualising what is learned by reading widely and relating new knowledge to previously gained information whereas in the Achieving Approach, the student's primary motivation is to gain the highest possible grades by being an ideal student, being punctual to class and using strategies such as extra reading and research that assist an individual student to attain his or her desired goals

All students are likely to manifest all three approaches to learning to some degree at some time in the process of studying and learning. Some students may even manifest approaches that are intermediate in nature that is at a point on a continuum between surface and deep approaches to learning (Leung et al., 2008). However, the primary concern of this investigation is to understand more about how CHC students learn, particularly in an educational environment that is physically, socially and culturally Western. Chan (1997, 1999) and Guan and Dodder (2001) found that some Chinese sojourners adopted the local culture particularly after longer stays overseas, but others resisted acculturation and preferred to preserve their ethnic culture and learning styles abroad. Other investigators have found similar results based on cultural distance and difference (Nisbett, 2003; Tafarodi et al., 2009). They preserved a form of 'Chinese separateness' by reinforcing their home cultural identity with their home language and forms of paid employment that perpetuated this identity. Some groups even formed enclaves of individuals from their home culture within the local population and only mixed with local people out of necessity. Therefore, for Chan (1999, p.294), learning was 'a question of style' and followed the Confucian adage that noted, "A learned man is very careful and timid in every word he says; but in action, he works swiftly and is not lazy." Other investigators have examined learning within students' home cultures (Watkins & Biggs, 1996, 2001). Still others have concluded that the Achieving Approach to learning is not as important to as the Deep and Surface Approaches (Biggs, Kember, & Leung, 2001; Kember, Biggs & Leung, 2004). However, for students in this study, the Achieving Approach to learning appears to be critically important to academic success.

Learning approaches – Definition and prior research

Kember and Gow (1990, 1991) found that students from Confucian cultures were often stereotyped as superficial learners. While this may have been an observable manifestation, it was found to be directly linked to the students' heavy workload, the demands of a surface assessment, the over-lecturing and a didactic teaching style that was adopted by the concerned staff, all of which caused the students to adopt reproductive approaches to learning in order to succeed academically. The pressure to succeed in any way possible appeared to result in a decline in intrinsic motivation which was manifested as a decline in the deep and achieving approaches to learning and a rise in the superficial approach to learning over the course of their academic study. Therefore, students' approaches to learning tasks appeared to be a function of the curriculum, the teaching environment and the students' need to achieve. Biggs (1987a) based his ideas on perceptions of learning in students from Western cultures. Kreber (2003), using a Canadian sample, found similar results and concluded that the workload and a facts-oriented assessment were the greatest contributors to adoption of a more superficial approach to learning particularly in younger students. However, deep learning tended to be self-directed and involved critical thinking and the evaluation of new ideas and concepts and proved to be the foundation for lifelong learning. Struyven et al. (2006), using a Belgian sample, found that the teaching and learning environment influenced students' preferred approaches to learning. Contrary to what had been anticipated, after pre-test and post-test measurements, first year education students appeared to preferred lecture-based rather than student-activated system of learning that pushed students towards a more superficial approach to learning.

Watkins, Biggs and Regmi (1991) distinguished between an approach to learning and a learning style which they found to be a relatively fixed and permanent method of undertaking learning with origins in a given learner's personality whereas an approach to learning was a sign of the interaction between the teaching context and the learner's motivation. They found that approach to learning taken by students reflected the students' confidence as the more self-assured a student was, the more highly organised were their learning strategies and the more likely they were to succeed. This was particularly evident in poorer families in Hong Kong where the students' English language level was initially lower. However, this was accompanied by a higher level of achievement motivation to succeed and enter Hong Kong University where the language of instruction was English. The educational success of the young was seen as a means of social and economic advancement and academic achievement for the whole family.

Chung and Chow (1999, 2004) used the SPQ to measure tertiary students' learning preferences and examine any changes that occurred in approaches to learning as a result of imbedded problem based learning (PBL) that had been implemented in an Asian higher education setting. They found their preference for the deep approach to learning increased from the pre- to post-PBL implementation which they interpreted as indicating that students valued the experience as independent thinkers with their own opinions on learning. In contrast, the surface approach decreased after the lecturer had provided clear instructions for all required course activities. The students also felt they had taken more responsibility for their own learning and were better able to decide what and how to learn. However, they still wanted the assurance from the lecturer that they had the correct answers. They felt that the experience of PBL over a 14 week period in one subject was insufficient to cause permanent changes in their learning behaviour and approach to learning, but became more interested in trying out a more involved learning style.

Ng and Renshaw (2002, 2003) correlated achievement goals with values that were assumed to mediate and affect the means of achievement. Results of the study showed that mastery goals were associated with motivations and strategies that were consistent with a deep approach to learning. This approach was related to positive learning outcomes. In contrast, performance goals were associated with motivations and strategies that tended to be superficial in nature and consistent with a surface approach to learning and yielded a lower level of achievement (Chan, 2002; Grant & Dweck 2001; Hau & Salili 1996; Lai & Biggs 1994; Salili 1996; Watkins, 2003).

Cano's (2005) research demonstrated that older female students tended to score higher on the deep and achieving approaches to learning than younger male students. However, he noted that these results may have been tempered by academic demands such as a dense curriculum and time limitations. Research also appeared to confirm the conclusions that (a) deep and achieving approaches to learning tended to be associated with academic success and (b) surface approaches were negatively linked to learning (Cano, 2005, Matthews et al., 2007; Watkins, 2001). Further, Cano (2005) and Schommer (1998) concluded that epistemological beliefs, students' beliefs about learning and knowing, and study orchestrations were linked to academic performance and therefore, approaches to learning often became more complex as students advanced in their academic pursuits. In addition, the relationship between epistemological beliefs and academic outcomes appeared to have been mediated by students' approaches to learning that can be envisaged as predictors of academic achievement. As well, the more able students were in constructing meaning from their learning, the higher their achievement was found to be. Another explanation of this observation might have been that students' conceptions of learning became deeper over time as their temperament changed with the method of assessment used and the influence of the study environment on learning (Biggs, 2001; Cano, 2005; Cano & Cardelle-Elawar, 2004; Rodrígues & Cano, 2006).

Haggis (2003, 2004) considered the perceptions and approaches to learning, beginning with an assessment of the work of Martön and Säljö's on surface and deep acquisition conceptions that were found to underlie the two basic approaches to learning and discernment of knowledge as linked to the outcomes of learning. Haggis (2003, p.100) emphasised the need to shift from a view of success based on the idea that university study was like 'an apprenticeship into new ways of thinking and expression for students…that took a number of years to develop, and needed to be explicitly modelled and explored.' Haggis (2004) also argued for the need for a wider range of approaches to thinking about learning in higher education as she conceived that learning should be understood as an individual and situated process based generalised models that often led to unexpected changes and ones that educators had not expected and therefore outcomes that were unpredictable.

Wilding and Andrews (2006) found that a higher GPA was related to greater use of the achieving approach to learning with a concomitant decrease in importance of wealth and status goals. They also found that older students with higher entrance scores on the British Advanced ('A' level) examinations tended to achieve better academic results overall regardless of the academic discipline.

Whitehead (1984) found that although home background partially influenced motivation and learning. students who sought financial reward and a high status job were pragmatically motivated and only interested in the pursuit of a qualification that would lead to those goals tended to use a surface approach to learning whereas students who preferred to discover knowledge for themselves were intrinsically motivated and tended to be those who went on to study at tertiary level were students who preferred a deeper approach to learning and did not require clearly defined tasks to maintain their motivation and interest in study as they were able to read about the topics and find things out for themselves.

Newell (1999, pp.286-289) found the transfer of knowledge from the West to the East to be problematic because the tacit knowledge underpinning Chinese or CHC learners was found to be different to Western educators articulation of knowledge. The variations were based on differences in value systems and resulted in contrasting views of knowledge. In the West a cognitive model appeared to underpin objectively defined concepts as analytical entities possessed by individuals whereas in Confucian cultures knowledge was socially constructed and based on experience. Further, in the West knowledge was transferred through texts and lectures, but in the East it was developed through social networks in the community and occupational groups and depended on strong interpersonal relationships (*guanxi*) and a concern for face. This was correlated with the perceived value of 'correct behaviour' and the hierarchical ordering of individuals in society for smooth and harmonious functioning based on conforming to standards of behaviour that were appropriate to an individual's social role in life. Therefore, in CHC cultures, learning was founded on the community model that stressed interaction rather than a one-way communication of knowledge from East to West. To attain this goal Newell (1999) advocated the development of learning communities that would promote dialogue and shared knowledge.

Gong (1989) noted that the Confucian elements of *Ren,* or the obliteration of individual desire, personal sacrifice, self-restraint, a lack of privacy and ideological control and the elements of *Li* were a social order based on the individual's work unit, the attainment of residence in a major city and the subordination of the lower classes to the elite were strong Confucian beliefs that influenced the government and society in cultures where Confucian traditions impacted strongly on life and learning.

Duff et al. (2004, p.1917) found that both the deep and the achieving approaches to learning correlated positively with academic achievement whereas surface learning was negatively correlated. Further, there was a strong relationship between learning approaches and the Big Five factors measuring personality characteristics. The results from their study showed the presage factors of age, gender, personality and prior educational achievement were significant determinants of an individual's orientation to learning. Of the Big Five factors, extraversion, conscientiousness and openness to experience were the most relevant in an educational setting and each was positively and significantly correlated, but neuroticism was negatively correlated to the results in Duff et al.'s (2004) research. The learning environment thus influenced student performance as did the nature of the assessment used. Therefore, the approaches to learning chosen were considered to be a learnt component of personality. McKenzie et al. (2004) also examined motivation, learning strategies and personality traits in a sample of first year Australian students. They found that previously high academic performance, the use of self-regulatory learning strategies and being introverted and agreeable were key indicators of academic success. Further, positive achievement motivation and conscientiousness were indirectly related to academic outcomes through the influence of these factors on the students' use of self-regulatory learning strategies in their sample of first year undergraduate students. Therefore, initial success in university studies necessitated both the skill to perform and the will to succeed.

Kember (1996, 2000, 2004) summarised some of the issues that have confounded researchers who study the Asian learner. He noted the commonly-held perception that Asian students relied exclusively on rote learning and preferred to be passively involved in the learning process seemed to be contradicted by their high level of achievement. What seemed to be a dichotomy was explained by an understanding that memorisation occurred in conjunction with the intention to understand. Leung et al. (2008) noted that approaches to learning were task-dependent and were often intermediate to what had previously been defined as either a surface or deep approach and combined memorisation with the intention to understand. This explanation helps to explain the paradox of the Chinese learner noted by Watkins and Biggs (1996). The students also appeared to prefer courses that provided good career preparation. These courses were a source of intrinsic rather than extrinsic motivation that led to a high level of achievement motivation which had a collective character and not the individual and competitive nature seen accompanying Western manifestations of high achievement motivation. In CHC students this revealed itself in an increased interest and involvement in the materials studied, high levels of out-of-class learning in groups and high learning outcomes in the form of a superior performance on group projects. Therefore, high levels of achievement motivation in CHC students was positive as it required the collective and supportive efforts of all group members for successful implementation that decreased individual competition and resulted in intrinsically-based group learning.

Tang (1996), studying tertiary science students in Hong Kong, also assessed a paired learning situation. She examined learning preference strategies by comparing two types of work undertaken by the same students. Learning was assessed using two different methods. For the first assessment, the students were asked to prepare for an examination on a specific topic. The second assessment task was based on a written assignment. Tang compared the learning strategies and preferences of the students before and after each of the assessment tasks. She found that students taking the examination used a form of deep memorisation as a strategy for achieving high grades in the examination. Their principal objective was to do well and they employed strategies that ensured that result. Memorisation was a form of learning that had previously been associated with surface learning only. The students' efforts were systematic and highly organised. They focussed on likely examination questions and sought cues from their lecturers as to what their expectations were. Therefore, an initially paradoxical situation existed wherein strategies previously associated with surface learning were now being used in a deep learning context.

Interviews were conducted before and after each type of assessment. It was expected that surface strategies would be the preferred method of preparation for the examination and that deep strategies would be employed for the written assignment. Tang (1996) found that the opposite had occurred. She explained her findings as follows: the research assignment was a new form of assessment and the majority of the students initially perceived that only a superficial or surface strategy was required. However, the examination was a form of assessment familiar to the students and they were well equipped to handle this form of assessment by using a strategy that involved deep memorisation. The outcome of the study provided Tang (1996) with the idea that the way in which students handled assessment depended on what they perceived the assessment task required. When faced with a new form of assessment, such as a written assignment, that involved research and the construction of an integrated argument, the students had few skills to undertake the new task. They perceived that deep strategies should be used but lacked experience in using these strategies. Furthermore, even students who were able to use deep strategies reverted to memory-based strategies in preparation for the examination.

Therefore, deep memorisation is an approach to learning that must be accepted as a part of the Asian learning environment where it can comfortably co-exist with both surface and deep learning strategies. However, deep memorisation may not be successfully transposed or fully utilised out of the specifically CHC learning context (Tang, 1996; Tang & Biggs, 1996). Investigators have shown through experimental research that rote memorisation only takes place as a learning strategy under the pressure of last minute study for an examination. Tang (1996) points out that rote or superficial memorisation is not an exclusively

Chinese approach; all students use this technique to learn details for an imminent examination. Therefore, it is a universal form of learning employed to meet an immediate need. Once the need has been met, the retention of information ceases to be important.

Overview of the HLM procedure

Hierarchical linear modelling (HLM) allows the investigator to model possible changes in student's approaches to learning by the examination of changes: a) within individual students over time as an estimation of the occasion of measurement in the students in the sample (intra-student change); b) between different students in the same study sample in two distinct ways (inter-student change): i) as a direct effect on the level of an approach to learning taken by students which may increase, decrease or remain the same; and ii) as an interaction effect between students, the approach to learning and time as measured by occasion. Further, HLM is a regression-based statistical method that permits the investigator to analyse multi-level data involving repeated measures that may show change over time both within and between students (Bryk & Raudenbush, 1992; Shin et al., 2004)

The outcome variables used in separate analyses in the multilevel modelling were the six approaches to learning listed above. At Level-1, the predictor variable was time, as determined by the five occasions of measurement. At Level-2, the student level characteristics that had been measured on the first occasion were accepted if they were significant at the 0.05 level. The statistically significant variables are included in Table 2.

Therefore three questions acted as a guide to the investigation of the data collected on changes in students' approaches to learning: (a) was there change in the learning approaches over time; (b) was the level of the measured approaches to learning the same or different between students (a direct effect); (c) if there was a difference between the approaches to learning of groups of students, what was the probable cause of the change that was an interaction effect between three factors: time, the student characteristic and the approach to learning scale?

The effects for the three approaches to learning

Figure 1 shows results for the student characteristics that are associated with the Surface Approach to learning, Figure 2 shows results for the student characteristics associated with the Deep Approach to learning and Figure 3 shows results for the student characteristics associated with the Achieving Approach to learning. Specific results are discussed in the section that follows.

Table 2. Predictor and interaction effects for learning

	Surface Motiv	Surface Strat	Deep Motiv	Deep Strat	Achiev Motiv	Achiev Strat
Direct effects; Adjusted occasion slope (Slope at β_1)	-0.087†	-0.070	0.073	0.164	0.053‡	0.127
Gender-sex of students: Female = 1, Male = 0					0.216†	
Granrel-grandparents religion: Different beliefs to grandparents = 1 Same beliefs as grandparents =0					-0.217	
Agegp-age group of student: Over 25 = 1, 25 years or under = 0					0.206	0.223
Mominflu-Most influential person in childhood: under five years Mother = 1, Another person=0		0.147				0.174‡
Arinoz –When student arrived in Australia: >1 year ago = 1, <1 year ago = 0		0.134				
PREST-Where student prefers to study: Home = 1, Not at home = 0			-0.166			
Prefer How the student prefers to study:Alone = 1, With others = 0			0.189			
Trip –Student lived overseas before coming to Australia: Trip = 1, No trip = 0				0.280		
Country -Country of birth: Less developed=1, Developed = 0					0.166	
UniSA - uni attended by student: UniSA or TAFE = 1, Another university =0					-0.189	
Parrel- Parents' religion: Different beliefs to parents = 1, Same beliefs =0				0.166		
Interaction effects						
Study- Student's major area of study: Business subjects=1, Other subjects=0	0.121	0.155				
PREST-Where student prefers to study: Home = 1, Not at home = 0	-0.113					
Parrel- Parents' religion : Different beliefs to parents = 1 Same beliefs =0	0.111					
Hours-Hours of study: 10+ hours each week =1, <10 hours each week=0			-0.093	-0.156		-0.267
Couvis-Countries visited before Australia: Beyond Asia = 1, Asia only = 0						0.113‡
Travel- Travel before coming to Australia: Travel = 1, No travel 0		-0.115				
Speak-Language spoken at home in Australia: English = 1, Home or other =0					-0.104	

‡These variables are significant at the ten per cent level †All significant slopes and effects

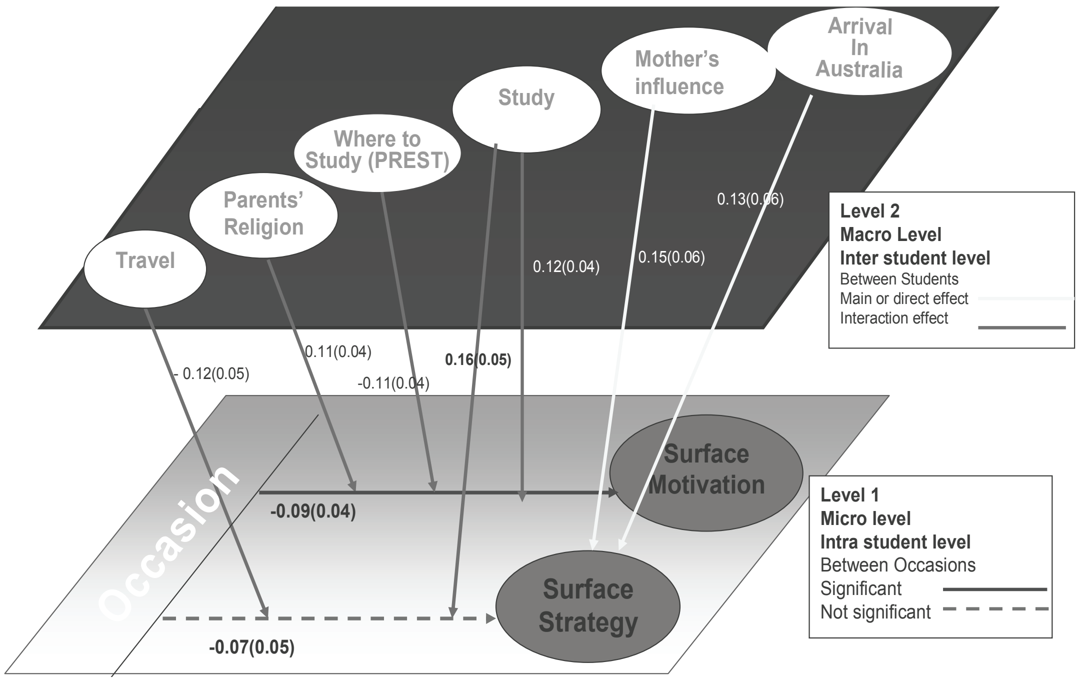


Figure 1. The Surface Approach to learning

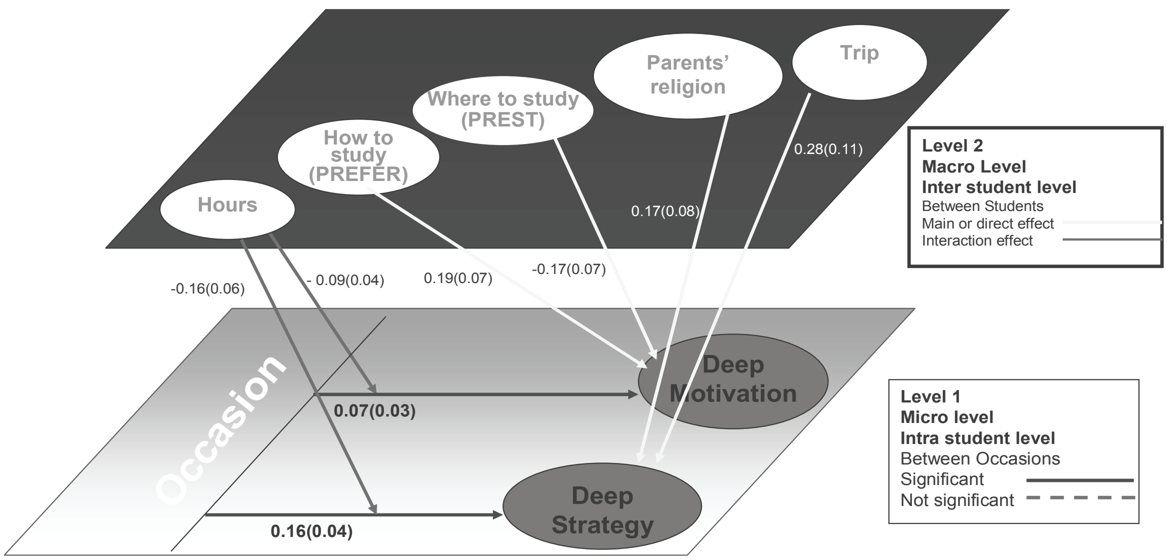


Figure 2. The Deep Approach to learning

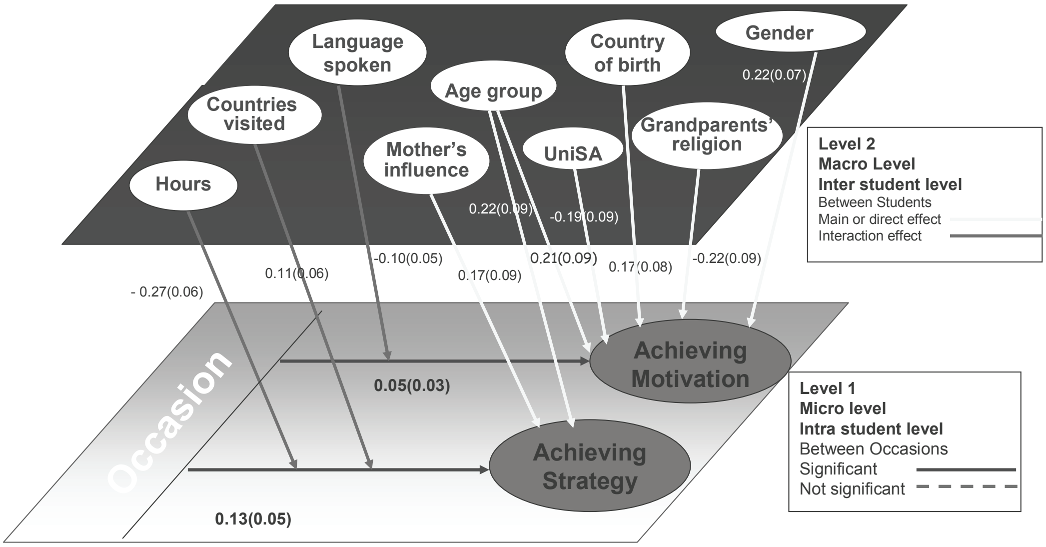


Figure 3. The Achieving Approach to learning

All the significant effects for the three approaches to learning are summarised in the three figures below. At Level-1, the significant Occasion effects are indicated as solid lines while non-significant effects are seen as dashed lines. At Level-2, an effect that influences an outcome variable directly is indicated by lines that pass from the effect to the outcome variable directly whereas an effect that influences a variable indirectly is shown by a line that passes from the effect to the outcome variable by way of the associated occasion variable. Therefore, interaction effects are moderated through the Level-1 variable Occasion. Table 2 shows the predictor and interaction effects for learning. It shows the significant effects and specifies the ways in which each of the variables has been coded. Greater detail on these data, the variance explained, the deviance and the methods of analysis utilised may be found in Matthews (2004).

The analyses undertaken showed that changes with respect to five of the six approaches to learning were statistically significant. Therefore, a negative occasion slope for Surface Motivation indicated that superficial learning decreased while the positive occasion slopes associated with the Deep and Achieving learning approaches indicated that problem-based learning approaches that were directed toward achievement increased over time. These results showed that although five of the six approaches to learning changed during the period in which measurements were made, not all groups of students changed their approaches to learning. Further, while there was no significant change for the Surface Strategy approach over time, there were significant direct and interaction effects associated with this approach to learning and the characteristics of particular groups of students. Therefore, although the study provided evidence that learning approaches do change measurably over a two year period, three important questions were raised: why do some groups of students change whereas some do not and still other groups remain the same?

Differences between groups of students shown by Level-2 direct effects

Age of students. Older students, those over the age of 25, are higher in Achieving Motivation and Achieving Strategy than students aged 25 years or less. Older students would appear to be more highly motivated to succeed in their studies than younger students. They seem to be stimulated to make more of an effort and study harder than younger students and to use strategies that assist them to achieve better results. Their motivations and strategies are guided by the need to achieve that comes from family, teachers and others in their home countries who want the skills these students have been sent to Australia to acquire. The high levels of Achieving Motivation and Achieving Strategy seem to occur because of a need to achieve so that they may be of greater use on their return to their country of birth. These results are consistent with research reported by Butcher (2002) and Ward and Kennedy (1993a, 1993b, 1999).

Gender of students. Women are significantly higher in Achieving Motivation than men in this study. Women may feel a greater need to attain and are, therefore, more highly motivated to achieve in Australia than men from Confucian cultures. Zhang Zhen (2001) comments that women have a stronger motivation to achieve than men do and that this effect is particularly strong in contemporary Asian women who live in urban environments. Further, if the women students are married, their own and their husbands' families may be encouraging them to reach higher levels of Achieving Motivation. In accordance with Confucian culture, the domestic or inside sphere of life is for women, whether rich or poor and world outside is for men (See 2006, pp. 28-29). Therefore, women should not pass beyond the inner chambers in either their thoughts or actions. Two Confucian ideals are thought to have guided women's lives in the past and still influence them today; the first relates to the Three Obediences: 'When a girl, obey your father, when a wife obey your husband, when a widow, obey your son. The second ideal relates to the four virtues that delineate women's behaviour, speech, carriage and occupation and states: "be chaste and yielding, calm and upright in attitude; be quiet and agreeable in words; be restrained and exquisite in movement; be perfect in handiwork and embroidery." If girls followed these principles, they would grow into virtuous women.

Thus it not unexpected that women from Confucian cultures are high in Achieving Motivation even when they are studying outside their country of birth.

Most influential person. Students who say their mothers have been the greatest influence in their pre-school years are higher in both Surface and Achieving Strategies. Mother's influence is marginally significant in its effect on Achieving Strategy. Mothers encourage higher levels of excellence and appear to demand higher academic outcomes than other, less interested persons. Therefore, it would seem that students who are influenced by their mothers' encouragement choose learning strategies that assist them to succeed at a higher level in their chosen field of study. The effect of increasing Achieving Strategy confirms the importance of mothers as people who encourage high levels of achievement. Yue and Ng (1999) have found similar results in their research. The coefficient of mother's influence indicates that this encouragement has a significant and positive effect on Surface Strategy. Ho (1994, 1996) has also found that mother's influence is important in the lives of students from Confucian cultures. Although the mother is not the head of the family *per se,* her influence is important particularly in the collectivist cultures that are hierarchical in the ordering of their members. As Confucius said, 'higher-ups govern; lower ranks obey." Therefore, mothers guide the young in the development of correct behaviour that involves learning to conform in accordance with one's present and future role in life (Gong, 1989). In Asian society, learning begins in the home with the mother, in particular, and continues in the school where the teacher is the expert authority and the students' role is to listen and learn, but they do not enter into discussions with these expert providers of definitive knowledge. In Asian cultures, dialogue and encouragement are key features of the community model of knowledge where 'loss of face' is avoided by the acceptance of a hierarchically ordered society based on interaction and the acceptance of older and wiser individuals as guides (Newell, 1999).

Grandparents' religious beliefs. Students who have different religious beliefs to their grandparents are lower in Achieving Motivation than students who have the same religious beliefs as their grandparents. In Asia many young people are brought up or strongly influenced by their grandparents. This is particularly true in countries where the extended family is of cultural and practical importance. Young persons may spend more time with their grandparents than with their parents. Therefore, grandparents often have an important role in the cultural and attitudinal development of the young so that if grandparents have strong religious beliefs, it is likely that these beliefs may have been passed on to their grandchildren (Yen Mah, 2000). In contrast, the results show that these students may have rejected their traditional religious beliefs or may have sought a new support network in Australia that a religious group is able to provide and one that has replaced the encouragement given by grandparents and members of the extended family in Asia. Therefore, if grandparents have positively influenced students' religious beliefs, this influence is likely to remain with them because it provides a measure of stability in their lives in a new academic setting and is related to the high level of Achieving Motivation noted in this study and in a recent article published by Butcher (2002).

Time of arrival in Australia. Students who have been in Australia for more than one year are higher in Surface Strategy whereas those who have been in Australia for less than one year are lower in this approach to learning. A superficial approach that uses rote memorisation may assist those students who have been in Australia for more time to achieve the qualifications they are seeking from their study in a Western learning environment.

Students' preferred place and manner of study. Students who study away from home and alone are higher in Deep Motivation than students who say they prefer to study at home or in a group. This outcome may reflect the conditions in which most students live. The places where students can afford to live are often noisy or crowded so that students who are serious about their academic work look for a quiet place to study away from the distractions found in the home environment. Further, studying on their own represents a change from what (Biggs, 1996a, 1996b and Tang & Biggs, 1996) have said about CHC students' preferences for studying in groups in their home environment in Asia. Therefore, finding a suitable place to study enables most students to ponder over what they are learning and may increase students' motivation to produce well-written assignments in English. This is related to Yee's (1989) assessment of East Asian universities role as transmitters of a recognised cultural heritage that is teacher-centred, where attending lectures and memorisation act as extrinsically motivating forces that determine advancement by success in examinations that are based on factual information. Yee (1989) notes that the problem-solving approach and the encouragement of critical or creative thinking are seen in the West as pathways to the development of intellectual independence, may act as an inhibitor to the uncritical acceptance of content presented by the lecturer. Thus CHC sojourner students acculturated overseas may succumb to socio-cultural pressures of the dominant way of thinking on their eventual return to their home environment.

Parents' religious beliefs. Students who have different religious beliefs to their parents are higher in Deep Strategy compared with students who have the same religious beliefs and practices as their parents. Butcher (2002) comments that a change in religious beliefs is a common phenomenon in Asian students who study in Western learning environments. These students read widely and relate what they are studying to previous knowledge. Holding different religious beliefs may be part of the strategy that helps these students in their studies.

Attendance at University of South Australia (UniSA) or Adelaide TAFE. Students who attend the University of South Australia or the Adelaide College of Technical and Further Education (TAFE) are much lower in Achieving Motivation. The information that may be deduced from these data is that students who attend either Adelaide or Flinders Universities may be more academically inclined and therefore, more motivated to achieve than University of South Australia students or those who come to Australia to study English at the tertiary level as TAFE students. These students may not be seeking the rewards that a higher level of achievement can provide, Yee (1989, pp.214-215) has noted that Asian students used different standards to judge overseas universities worldwide and those institutions of learning within their own societies. They almost universally ranked Harvard as number one, followed by Oxford and Cambridge, Stanford and MIT. Only Tokyo and Peking Universities attained universally high rankings from Asians and then only at the 3 and 1.5 per cent respectively. The criteria used for the assessment of overseas universities were research superiority, outstanding faculty and overall excellence whereas universities in East Asia generally derived prestige on the basis of how successful their graduates were in finding and securing the most coveted and well-paying careers in their own societies. Yee's (1989) research may thus explain the lower level of Achieving Motivation expressed by students attending the lesser ranked University of South Australia and Adelaide TAFE.

Development of the home country. Students who come from less developed countries are higher in Achieving Motivation than those students who come from more developed countries. This is understandable because students from countries less developed than Australia are generally sent to acquire skills needed to improve the educational, technological and agricultural level of development in their own countries. These students are likely to have been chosen because of their high motivation to achieve and enthusiasm to learn skills to take back to their own countries after study in Australia as well as their awareness of the particular needs of their home countries.

Cannon (1999) and Chan (1988) have found similar results in their own research.

Previous travel. Students who have travelled extensively before coming to Australia are considerably higher in Deep Strategy compared with those who have not had the opportunity to travel. This variable is coded 1 for students who have travelled for a period of a month or more and 0 for students who have not travelled at all or only travelled to the Asian region. As Ward and Kennedy (1993a, 1993b, 1994, 1999) and Ward et al. (2001) have commented, students who have lived overseas for longer than a holiday are markedly higher in deep approaches to learning than students who have travelled for short periods to a similar cultural environment or who have never travelled. Their previous travel would appear to have shown these students more of the world and may have resulted in students who are intellectually stronger compared with those who have not travelled. Previous travel may also have heightened their interest in what they are studying and thus encouraged them to use deeper strategies in their academic studies as noted in their approaches to learning.

Differences between groups of students shown by Level-2 interaction effects

Subject of study effects. Students who study business and commerce subjects show interaction effects with the Surface Motivation and Surface Strategy learning scales. The coefficient is positive and indicates that these students show a slight increase in scales associated with the Surface Approach to learning. This may occur because business and commerce subjects require more memorisation than deep thought whereas students who study other subjects show a decrease in this learning approach over time. This is shown by a negative value for time as measured by the occasion slope. The interaction effect shows that students who study other subjects to use a deeper learning strategy than students who learn by memorisation alone. Kember (2004) found that students' perceptions of workload were influenced by (a) the content of the materials studied, (b) the perceived difficulty of the material, (c) the types of assessment, and (d) the teacher-student and student-student relationships. Further, Kember (2004) confirmed that there was a positive correlation between heavy workload and the surface approach to learning as has been noted in the present study.

Students' preferred place of study effects. Students who prefer to study at home seem to show a greater decrease in Surface Motivation than students who have said they prefer to study away from home over their time of study in Australia. Students who choose to study at home would appear to be less motivated to use a superficial approach to learning than students whose prefer to study away from the home environment. These students may find that they are not as superficially motivated to learn as they were in the country of their birth. Moreover, it should be noted that this decrease in Surface Motivation is not accompanied by an increase in a deeper approach by this group of students. This effect may represent a decrease in superficial motivation caused by a need for students to change their approach to learning in response to the curriculum content and the mode of assessment in Australia.

Parents' religious beliefs effects. Students who have different religious beliefs to those of their parents show a slight increase in Surface Motivation over time. However, if students do not change their religious beliefs in Australia, there is a significant decrease in the value of the occasion slope as noted in Table 2. Butcher (2002), Chang (2000), Ward and Chang (1997) and Ward et al. (2001) comment that many students express the desire to follow different religious practices when they are studying in cultural environments dissimilar to those in their home countries. Even if students have changed their religious practices, there does not appear to be a large effect on the Surface Motivation that directs their study.

Student travel effects. Students who have travelled before coming to Australia to study show a greater decrease in Surface Strategy over time compared with students who have not travelled. This would appear to indicate that students who had travelled have decreased their use of a superficial approach to learning as a result of travel. These students may have been influenced by what they had seen and experienced during their travels which may have led to a decrease in Surface Strategy.

Hours of study effects. Hours of study is a variable that shows interaction effects with three approaches to learning. In the case of the Deep Motivation and Deep Strategy scales there is a significant, not a substantial change in students who study more than ten hours outside of classes each week over the occasions of measurement. The lack of sizeable change would seem to indicate that these students continue to use memorisation as a learning tool as they did in their Confucian cultural environment. However, there is a strong interaction effect with the Achieving Strategy learning scale over time. This observation may result from a change in approach to learning that is necessitated by a difference in the methods of assessment observed in Australia. In Asian countries education is examination-dependent (Biggs, 1996; Tang, 1996; Tang & Biggs, 1996) whereas Western education tends to rely more on written assignments and oral presentations as the primary methods of assessment. This effect also shows that students who spend ten hours or less each week on study show an increase in Achieving Strategy over the occasions of measurement. This effect is corroborated by Kember (2004) who found that (a) the nature of the subject, (b) the learning environment, and (c) the type of assessment determined the workload which influenced the approach to learning taken by students. Students spent longer hours on group project work particularly when this was the principal method of assessment. This form of assessment encouraged students to adopt a deep approach and was accompanied by a high level of achievement strategy as students committed themselves to longer hours of study without complaint. It would appear that workload and hours of study were strongly influenced by the curriculum content and the type of assessment which in turn affected perceptions of workload. Kember (2004) also confirmed that a high workload was generally perceived to reflect an examination-based curriculum that was reciprocally aligned with a surface approach to learning. Therefore, the development of a curriculum perceived by students to have an acceptable workload that also encouraged them to work long hours in order to achieve highly desirable learning outcomes necessitated (a) clearly defined objectives, (b) instruction that promoted understanding, (c) assessment that tested understanding, (d) an approach to teaching and learning that resulted in the active engagement of learners by the motivation of interest in the study materials, and (e) the promotion of a learning environment based on group discussion and supportive teacher-student relationships Choi (1997) also found that language issues, role expectations of both students and teacher based on differences in cultural norms affected both hours of study and preferred and places of study.

Students' preferred language effects. Students who speak English rather than their home language in their free time show a marginal, but significant decrease in Achieving Motivation whereas students who prefer to use their home or another language show an increase in Achieving Motivation. Speaking the home language may increase the students' confidence in their ability to study in a new social and cultural environment and have a positive influence on their motivation to achieve. It may also assist them to retain cultural ties with the country they expect to return to on the completion of their studies. Cannon (1999) has found this to be the case in his research with South Asian students. After a period of adaptation to the new language environment, even hesitant students are likely to show a stronger desire to learn and this may result in a further increase in Achieving Motivation.

Effects of counties visited on students' learning. Students who have travelled beyond the Asian region show a marginal, but positive interaction with the predictor variable Occasion and the outcome variable Achieving Strategy. Two factors are considered to have had an influence on the results encountered with this variable: (a) greater distances travelled and (b) travel for periods of a month or more. Travel of any kind gives individuals the opportunity to experience life in different living, learning, and cultural environments. This may explain why students who have travelled beyond Asia also show a marginally higher level of Achieving Strategy compared with students who have only visited Asian countries that are physically, socially and culturally similar to their home living and educational environment. Ward and Kennedy (1993a, 1993b, 1999) have confirmed this in their research with students who travelled to different cultures and countries to live and study. This effect is reflected in an increase in Achieving Strategy over the occasions of measurement in this study.

Summary

In summary, it may be noted that five of the six scales associated with learning show significant change over time. Surface Motivation generally decreases while the motivation and strategy scales associated with both the Deep and Achieving Approaches increase over the period in which measurements have been made. This information provides evidence that learning approaches change or are modified in some, but not all groups of CHC students in the Australian learning environment.

A key point in understanding why learning approaches changed was knowledge of the Australian teaching and learning environment. The Australian way of teaching appeared to be different to what students from Confucian cultures were accustomed to and it was this difference that seemed to be producing changes in some students' approaches to learning. This difference led to the question: why did some groups of students' approaches to learning change whereas some did not change and other groups' approaches to learning remained the same?

It would seem, therefore, that different teaching approaches and learning environments generated changes in motivations to learn that, in turn, required different learning strategies. Research by Watkins and Biggs (1996, 2001b) noted that Asian students previously considered surface learners often became deep learners if they encountered learning environments that encouraged and promoted this change. The results from similar longitudinal studies have been discussed in publications by Matthews et al., (2007); Renshaw and Volet (1995); Volet and Renshaw (1995, 1996); and Volet, Renshaw and Tietzel (1994).

Two issues that have been raised by these particular results need further consideration. The results may be improved by a refinement in the research design. Two variables: hours of study and time of arrival in Australia have proved somewhat problematic. These variables could be measured on several and not a single occasion. This would enable closer monitoring of any alteration in particular variables that are likely to change over time.

Conclusions

The observations noted were confirmed by the students in the study sample who generally showed a change to a deeper, more problem-based learning approach. The fact that the Surface Approach to learning generally decreased over the two years of the study at the same time as the Deep and Achieving Approaches increased indicated that some, but not all groups of students changed their approaches to learning.

Further, the motivation to achieve and the allied strategy that implements this motivation seem to be crucial factors to sojourner students from Confucian cultures. This knowledge necessitates a wider acceptance as well as a greater acknowledgement of overt and latent differences between individuals from different cultural groups. The appreciation of difference, as noted by Ang (2001) and Stepp (2008) is the first step toward the ultimate goal of social cohesion that is an acknowledgement of the co-existence of multiple cultures within a single nation-state. These findings and the concomitant appreciation of difference are of particular importance for individuals who are teaching and working with CHC students in Western learning environments.

Thanks to John, I have a greater appreciation of the value and importance of statistics in education and life.

References

Adams, R. J. and Khoo, S. T. (1993). *QUEST The interactive test analytic system.* Melbourne: Australian Council for Educational Research.

Ang, I. (2001). On not speaking Chinese: living between Asia and the West. London: Routledge.

Biggs, J. B. (1987a). *Student approaches to learning and studying. (Research monograph).* Melbourne: Australian Council for Educational Research.

Biggs, J. B. (1987b). *Study Process Questionnaire manual.* Melbourne: Australian Council for Educational Research.

Biggs, J. B. (1996a). Learning, schooling and socialisation, a Chinese solution to a Western problem. In S. Lau (Ed.), *Growing up the Chinese way* (pp. 147-167). Hong Kong: The Chinese University Press.

Biggs, J. B. (1996b). Misperceptions of the Confucian-heritage learning culture. In D. A. Watkins and J. B. Biggs (Eds.), *The Chinese learner: cultural, psychological and contextual influences* (pp. 45-67). Hong Kong: ACER and CERC.

Biggs, J. B. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education. 41,*221-238.

Biggs, J. B., Kember, D. and Leung, D. Y. P. (2001). The revised two factor Study Process Questionnaire, R-SPQ-2F. *British journal of educational psychology, 71(1),* 133-149.

Bryk, A. S., Raudenbush, S. W. and Congdon, R. T. (2000). *HLM5 for Windows.* Chicago: Scientific Software International.

Butcher, A. (2002). A grief observed: grief experiences of East Asian international students returning to their countries of origin. *Journal of Studies in International Education, 6(4),* 354-368.

Cannon, R. (1999). International education and a professional edge for Indonesian graduates: the third place? In D. Davis and A. Olsen (Eds.), *International education: the professional edge.* A set of research papers presented at the 13th Australian international education conference, Fremantle, 1999 (pp.15-36). Canberra: IDP Education Australia.

Cano, F. (2005). Epistemological beliefs and approaches to learning: Their change through secondary school and their influence on academic performance. *British Journal of Educational Psychology, 75*, 203-221.

Cano, F. and Cardelle-Elawar, M. (2004). An integrated analysis of secondary school students' conceptions and beliefs about learning. *European Journal of Psychology of Education, 19(2),* 167-187.

Chan, H. (1988). Chinese students, adaptation, life satisfaction and academic achievement. In P. Hanks and A. Perry (Eds.), *The Chinese in Australia*. Conference held 19 March 1988 (pp.25-36).

Chan, K. W. (2002, December). Students' epistemological beliefs and approaches to learning. Paper presented at the AARE Conference, Brisbane, Australia.

Chan, S. (1997). Migration, cultural identity and assimilation effects in entrepreneurship for the overseas Chinese in Britain. *Asia Pacific Business Review, 3(4)*, 211-222.

Chan, S. (1999). The Chinese learner-a question of style, *Education and Training, 41(6/7),* 294-304.

Chang, W. N. (2000). In search of the Chinese in all the wrong places. *Journal of Psychology in Chinese Societies, 1(1),* 125-142.

Choi, M. (1997). Korean students in Australian universities: Intercultural issues. *Higher Education Research and Development, 16(3),* 263-282.

Chung, J. C. C. and Chow, S. M. K. (1999). Imbedded PBL in an Asian context: Opportunities and challenges. In J. Marsh (Ed.) *Implementing problem-based learning: Proceedings of the First Asia Pacific Conference on Problem-Based Learning (pp.35-46)*, Hong Kong: Hong University Press.

Chung, J. C. C. and Chow, S. M. K. (2004). Promoting student learning through a student-centred problem-based learning curriculum. *Innovations in Education and Teaching International, 41(2),* 157-168.

Duff, A., Boyle, E., Dunleavy, K. and Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36,* 1907-1920.

Gong, W. (1989). The legacy of Confucian culture in Maoist China. *The Social Science Journal, 26(4),* 363-374.

Grant, H., and Dweck, C. S. (2001). Cross-cultural response to failure: considering outcome attributions with different goals. In F. Salili, C. Y. Chiu and Y. Y. Hong (Eds.), *Student motivation: The culture and context of learning. (Plenum series on human exceptionality)* (pp. 203- 219). New York: Plenum Publishers.

Guan, J. and Dodder, R. (2001). The impact of cross-cultural contact on value and identity, a comparative study of Chinese students in China and in the U. S. A. *Mankind Quarterly, 41(3),* 271-288.

Haggis, T. (2003). Constructing images of ourselves? A critical investigation into 'approaches to learning' in higher education. *British Educational Research Journal, 29(1)*, 89-104.

Haggis, T. (2004). Meaning, identity and 'motivation': expanding what matters in understanding learning in higher education? *Studies in Higher Education, 29(1)*, 335-352.

Hau, K-T. and Salili, F. (1996). Achievement goals and causal attributions. In S. Lau (Ed.), *Growing up the Chinese way* (pp. 121-145). Hong Kong: The Chinese University Press.

Ho, D. Y. F. (1994). Cognitive socialisation in Confucian heritage cultures. In P. M. Greenfield and R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 285-313). Hillsdale: L. Erlbaum Associates.

Ho, D. Y. F. (1996). Filial piety and its psychological consequences. In M. H. Bond (Ed.), *The handbook of Chinese psychology.* (pp. 155-165). Hong Kong: Oxford University Press.

Jöreskog, K. G. and Sörbom, D. (1999). *LISREL 8.30.* Chicago: Scientific Software International.

Kember, D. (1996).The intention to both memorise and understand: Another approach to learning? *Higher Education, 31(3),* 341-354.

Kember, D. (2000). Misperceptions about the learning approaches and study practices of Asian students. *Higher Education, 40,* 99-121.

Kember, D. (2004). Interpreting student workload and the factors which shape students' perceptions of their workload. *Studies in Higher Education, 29(2),*165-184.

Kember, D., Biggs, J., and Leung, D. Y. P. (2004). Examining the multidimensionality of approaches to learning through the development of a revised version of the Learning Process Questionnaire. *British Journal of Educational Psychology, 74,* 261-279.

Kember, D. and Gow, L. (1990). Cultural specificity of approaches to study. *British Journal of Educational Psychology, 60,* 356-363.

Kember, D. and Gow, L. (1991). A challenge to the anecdotal stereotype of the Asian student. *Studies in Higher Education, 16(2),* 117-128.

Kreber, C. (2003). The relationship between students' course preparation and their approaches to studying in undergraduate science courses: A Canadian experience. *Higher Education Research and Development, 22(1),* 57-75.

Lai, P., and Biggs, J. B. (1994). Who benefits from mastery learning? *Contemporary Educational Psychology, 19*(1)*,* 13-23.

Leung, D. Y. P., Ginns, P. and Kember, D. (2008). Examing the cultural specificity of approaches to learning in Hong Kong and Sydney. *Journal of Cross-cultural Psychology, 39(3),*251-266.

Martön, F., and Säljö, R. (1976a). On qualitative differences in learning−I, outcome and process. *British Journal of Educational Psychology, 46,* 4-11.

Martön, F., and Säljö, R. (1976b). On qualitative differences in learning−II, Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology, 46,* 115-127.

Matthews, B. (2004). Life values and approaches to learning: a study of university students from Confucian heritage cultures. Studies in comparative and international education series Adelaide: Flinders University Institute of International Education.

Matthews, B., Lietz, P., and Darmawan, I.G.N. (2007). Values and learning approaches of students at an international university. *Social Psychology of Education, 10*(2)*,* 247-275.

McKenzie, K., Gow, K. and Schweitzer, R (2004). Exploring first-year academic achievement through structural equation modelling. *Higher education Research and Development, 23(1),* 95-112.

Murray-Harvey, R. (1994). Learning styles and approaches to learning, distinguishing between concepts and instruments. *British Journal of Educational Psychology, 64,* 373-388.

Newell, S. (1999). The transfer of knowledge to China: building learning communities rather than translating Western textbooks? *Education and Training, 41(6/7),* 286-293.

Ng, C. H. and Renshaw, P. D. (2003). Motivation and school learning. In J. P. Keeves (Ed.), *Handbook of educational research in the Asia-Pacific region.* (pp. 495-510). Dordrecht: Kluwer.

Ng, C. H., and Renshaw, P. D. (2002). Self-schema, motivation and learning: a cross-cultural comparison. In D. M. McInernay and S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol. 2, pp. 55-87). Greenwich, CT: Information Age Publishing, Inc.

Nisbett, R. E. (2003). The geography of thought: How Asians and Westerners think differently…and why. New York: Free Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Renshaw, P. D. and Volet, S. E. (1995). South-east Asian students at Australian universities: a reappraisal of their tutorial participation and approaches to study. *Australian Educational Researcher, 22(2),* 85-106.

Rodríguez, L. and Cano, F. (2007). The learning approaches and epistemological beliefs of university students: a cross-sectional and longitudinal study. *Studies in Higher Education, 32(5)*, 647-667.

Salili, F. (1996). Achievement motivation: a cross-cultural comparison of British and Chinese students. *Educational Psychology 16*(8)*,* 271-281.

Schommer, M. (1998). The influence of age and education on epistemological beliefs. *Journal of Educational Psychology, 68,* 551-562.

See, L. (2006). *Snow flowers and the secret fan.* Edinburgh: Bloomsberry.

Shin, J., Espin, C. A.,Deno, S. L. And McConnell, S. (2004). Use of hierarchical linear modelling and curriculm-based measurement for assessing academic growth and instrumental factors for students with learning difficulties. *Asia Pacific Education Review, 5(2),* 136-148.

Stepp, L. S. (2008, February 29-March 6). Latinas who want to die. *Guardian Weekly, 178* (*11*)*,* 31.

Struyven, K., Dochy, F., Janssens, S. and Gielen, S. (2006). On the dynamics of students' approaches to learning: The effects of the teaching/learning environment. *Learning and Instruction, 16,* 279-294.

Tafarodi, R. W., Shaughnessy, S. C., Lee, W. W. S., Leung, D. Y. P., Ozaki, Y., Morio, H., and Yamaguchi, S. (2009). Disregard for outsiders: A cultural comparison, *Journal of Cross-Cultural Psychology, 40(4),* 567-583.

Tang, C. and Biggs, J. B. (1996). How Hong Kong students cope with assessment. In D. A. Watkins and J. B. Biggs (Eds.), *The Chinese learner: cultural, psychological and contextual influences* (pp. 159-182). Hong Kong: ACER and CERC.

Volet, S. and Renshaw, P. (1995). Cross-cultural differences in university students' goals and perceptions of study settings for achieving their goals. *Higher Education, 30*, 407-433.

Volet, S. and Renshaw, P. (1996). Chinese students at an Australian university, adaptability and continuity. In D. A. Watkins and J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 205-220). Hong Kong: ACER and CERC.

Volet, S., Renshaw, P. and Tietzel, K. (1994). A short-term longitudinal investigation of cross-cultural differences in study approaches using Biggs SPQ questionnaire. *British Journal of Educational Psychology, 64,* 301-318.

Ward, C. A., Bochner, S. and Furnham, A. (2001). *The psychology of culture shock (2nd edition).* London: Routledge.

Ward, C. and Chang, W. C. (1997). "Cultural fit": a new perspective on personality and sojourner adjustment. *International Journal of Intercultural Relations, 21(4),* 525-533.

Ward, C. and Kennedy, A. (1993a). Where is the "culture" in cross-cultural transition? *Journal of Cross-cultural Psychology, 24(2),* 221-249.

Ward, C. and Kennedy, A. (1993b). Psychological and socio-cultural adjustment during cross-cultural transitions: a comparison of secondary students overseas and at home. *International Journal of Psychology, 28,* 129-147.

Ward, C. and Kennedy, A. (1994). Acculturation strategies, psychological adjustment and sociocultural competence during cross-cultural transitions. *International Journal of Intercultural Relations, 18(3),* 329-343.

Ward, C. and Kennedy, A. (1999). The measurement of sociocultural adaptation. *International Journal of Intercultural Relations, 23(4),* 659-677.

Watkins, D. A. (2001). Correlates of approaches to learning: a cross-cultural meta-analysis. In R. J. Sternberg and L-F. Zhang (Eds.), *Perspectives on thinking, learning and cognitive styles. (The education psychology series)* (pp.165-195). Mahwah, NJ: Lawrence Erlbaum Associates.

Watkins, D. A. (2003) Student learning: a cross-cultural perspective. In J. P. Keeves (Ed.), *Handbook of educational research in the Asia-Pacific region* (pp. 441-462). Dordrecht: Kluwer.

Watkins, D. A. and Biggs, J. B. (Eds.) (1996). *The Chinese learner, cultural, psychological and contextual influences.* Hong Kong: ACER and CERC.

Watkins, D. A. and Biggs, J. B. (2001a). The paradox of the Chinese learner and beyond. In D. A. Watkins and J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp. 3-23). Hong Kong: CERC and ACER.

Watkins, D. A. and Biggs, J. B. (Eds.) (2001b). Teaching the Chinese learner: Psychological and pedagogical perspectives. Hong Kong: CERC and ACER.

Watkins, D. A., Biggs, J. B. and Regmi, M. (1991). Does confidence in the language of instruction influence a student's approach to learning. *Instructional Science, 20(4),* 331-339.

Whitehead, J. (1984). Motives for higher education: A study of intrinsic and extrinsic motivation in relation to academic attainment. *Cambridge Journal of Education, 14(2)*, 26-34.

Wildiing, J. and Andrews, B. (2006). Life goals, approaches to study and performance in an undergraduate cohort. *Journal of Educational Psychology, 76*, 171-182.

Yee, A. (1989). Cross-cultural perspectives on higher education in East Asia: psychological effects upon Asian students, *Journal of Multilingual and Multicultural Development, 10(3),* 213-232.

Yen Mah, A. (2000). *Watching the tree to catch a hare*. London: Harper Collins Publishers.

Yue, X. and Ng, S. H. (1999). Filial obligations and expectations in China: current views from young and old people in Beijing. *Asian Journal of Social Psychology, 2,* 215-226.

Zhang Zhen (2001). Mediating time: "rice bowl of youth" in fin de siècle urban China. In A. Appadurai (Ed.), *Globalisation* (pp. 131-154). Durham: Duke University Press.

20

Studying in higher education: students' approaches to learning, learning environment and learning outcomes

Pauline Goh Swee-Choo
Sultan Idris University of Education Malaysia

Introduction

This chapter presents a section of a study of students undertaking the twinning programs from the Private Higher Educational Institutions (PHEIs). Typical twinning program arrangements are '1+2' (one year in local PHEI and two years in overseas partner university), '2+1' or '2+2' year arrangements. In a '3+0' arrangement, students are allowed to complete the foreign degree entirely at local PHEIs. As an educator who personally experienced a PHEI twinning program environment, I realised that there was little literature in area of PHEI students undertaking overseas western curricula in Malaysia. I was thus motivated by a deep concern at the lack of attention given to the whole process of PHEI students' learning in light of the unprecedented growth in student numbers attending PHEIs since 1995 (Department of Private Education, 2003), the current call for educational quality and for Malaysia to be a 'centre of educational excellence'. Furthermore, the democratisation of education led to wider participation of students coming from different societal groups who had different English language abilities, needs, and levels of motivation. However, there had been few attempts to seek opinions that reflected the views of PHEI students, who had been seen as passive participants to becoming consumers, customers, and contributors in their own learning processes, and the improvement of PHEIs as a whole.

Context

Concerns about quality of student learning in tertiary education are not a new phenomenon. In recent years, there have been significant efforts by researchers and educators to address this issue. The expanding field of student learning research has produced many suggestions of what we should be doing to encourage quality learning. Although much of the research originated in the west, there has also been research carried out in recent years on the learning and learning processes of Asian students studying in western universities, with particular respect to student approaches to learning (SAL).

The concept of student approaches to learning was derived from an experiment by Martön and Säljö (1976). It examined students' experiences of a particular learning situation. They demonstrated that the way each student went about their learning was different, and so were their perceptions of the way they should handle their learning. This resulted in differences in learning outcomes.

Much of the student approaches to learning research emanated from Australia and Britain (Harris, 1997; Matthews, 2003; Smith, 2001). Findings showed that how students went about their learning and how they perceived their own learning contexts including good teaching, clear goals, appropriate assessment and workload appeared to be important variables that influenced achievement outcomes in tertiary education (Kember, Ng, Tse, Wong, & Pomfret, 1996; Lizzio, Wilson, & Simons, 2002; Ramsden, 1984, 1992). While the research into students' approaches to learning could be assumed to be relevant to students studying in western universities, there was less confidence in the pertinence of learning approaches to Asian students' in their home country (Kember & Gow, 1990). Research that attempted to repeat the investigations in students' home countries tended to be limited to looking at students undertaking their own institution's curriculum in their national language (Leung & Kember, 2003; Tan, 1990; Wan Ali, 2000; Watkins & Ismail, 1994). Therefore, while the research contributed to the growing knowledge of Asian students' approaches to learning, the perception of their learning context and learning outcomes, the extent to which Asian students were able to accommodate different approaches to learning within an imported Australian and British curriculum that occurred entirely in the English language, in their home country had not been examined. One such example was found in Malaysia where PHEI students came from different cultural traditions and different educational systems and practices. Matthews (2003) and Volet and Kee (1993) indicated that issues relating to students' learning in culturally different contexts had received very little attention.

In addition, Malaysian education stressed that students should undertake life-long learning (Zakaria, 2000; Abd Rashid, 2002). Students were encouraged to develop skills, which would enable them to do research and make learning decisions based on their needs, talents, and interests. The Ministry of Education indicated that graduates should be imbued with an aspiration for life-long learning where they learned to think, do, and create (Zakaria, 2000). The government of Malaysia wanted graduates to possess process expertise such as problem solving, decision-making, and creativity skills, and the ability to function as a team (Zakaria, 2000; Abd Rashid, 2002). Therefore, in addition to the traditional concerns of achievement outcomes of students in PHEIs, there seemed to be a great need to be informed about how student learning occurred at PHEIs and how their learning environments contributed to the accession of these skills. According to MacNair (1990) these skills were also known as competency skills and enabled students to apply what they have learnt effectively through the content and subject skills of higher education in their work environment. Lizzio, Wilson, and Simons (2002) contended that there was evidence of the impact that learning approaches and learning context had on the development of such process skills. It would seem of practical importance that the statement made by a former education minister that PHEIs students 'were being taught but not educated and were hence unable to think critically or to be analytical…' has great relevance (Indramalar, 1999, p.2).

The Study

The purpose of this study is to investigate the associations in PHEI students' English language competency, their perceptions of the learning context, approaches to learning, and related learning outcomes for students from different cultural backgrounds (Malay, Chinese, and Indian). Both quantitative and qualitative methods of analysis are used.

Student sample and selection

The population of students that the study seeks to investigate is composed of students undertaking the 3+0 twinning programs that are offered by either Australian or British universities. In particular, the target population is second and third year undergraduates taking degrees in business and business-related subjects, computer science, or engineering.

The selection of participatory PHEI students is based on their accessibility and willingness of the Principal or the President of the PHEIs to participate in the study. Willingness to participate is important as, in some instances, the lecturers in some of the participating PHEI programs may be required to administer the questionnaires. Eventually six colleges have agreed to participate.

Out of the 368 participants from six PHEIs, 166 are studying Engineering and Computer Science programs, while the other 202 are in business, commerce, accounting, finance, or management programs. The sample is made up of equal numbers of 184 males and 184 females of which 168 students are 21 years of age and younger. The ethnic divide of the total sample is 82 per cent Chinese, 10 per cent Indians, 5 per cent Malay, with the remaining coming from other indigenous races (consisting of Iban and Kadazan).

Quantitative measures

Gender differences, choice of academic discipline and age

Students are required to provide personal information such as gender, their choice of academic discipline and their age.

English language competency

A Perceived English Language Competency Questionnaire (PELCQ) has been developed to explore students' perceived competency in the use of the English language in various learning situations. It gathers data relating to students' perceived competency in the use of English in areas of writing, reading, understanding, discussion, and in conversation. The questions use a five-point Likert-type scale ranging from a value of 5 ('Very Good'), 4 ('Enough'), 3 ('Only Just Enough'), 2 ('Uncertain'), and 1 ('Definitely Not Enough'). Students' English language competency is assessed by the combined scores from their self-reporting of their competency in using the English language in the five skill areas. Students are divided into high, medium, or low competency groups on the basis of their total score on the five-item scale of the PELCQ. The low competency group represents 20 per cent of the total sample, while 40.8 per cent makes up the medium competency group, and 39.1 per cent of students are in the high competency group. The Cronbach alpha reliability coefficient has been used as an index of scale internal consistency. The scale reliability for the PELCQ scale is 0.83.

Perceptions of learning environment

To measure students' perceptions of their learning environments, the present study uses a modified version of the Course Experience Questionnaire (CEQ) by Wilson, Lizzio, and Ramsden (1997). The modified CEQ is a self-report questionnaire and has 13 items that comprises the following scales: Aspects of Good Teaching (six items); Aspects of Appropriate Assessment (four items); and Aspects of Clear Goals and Standards (three items). Students' responses are recorded on a five-point Likert-type scale of 1 ('Strongly disagree') to 5 ('Strongly agree'). Summing the scores on the appropriate items provides scores on the three scales, with a high score corresponding to a perception of a good learning environment. The value of Cronbach's alpha from the study's data for the overall reliability of the 13 items of the modified CEQ is 0.73, which is within the level of reliability of 0.70, suggested by Watkins and Mboya (1997).

Approach to learning

To measure students' learning approaches, the present study uses a modified version of the Revised Two-Factor Study Process Questionnaire (R-SPQ-2F) (Biggs, Kember, & Leung, 2001). It is also a self-report questionnaire designed to measure higher education students' approaches to learning and their preferred methods of study. The modified R-SPQ-2F consists of 20 items measuring the Deep Approach (DA) and Surface Approach (SA). The DA main scale uses Deep Motivation (DM) and Deep Strategy (DS) as subscales, while SA uses Surface Motivation (SM) and Surface Strategy (SS) as subscales. Each of the subscales (DM, DS, SM, and SS) contains five items. Each item within the subscales is rated on a five-point Likert-type scale: between 1 ('This item is never or only rarely true of me') and 5 ('This item is always or almost always true of me'). Subscale scores are calculated by summing up the scores on the relevant items. Subscale scores range from five to 25 with higher scores indicating those who make greater use of that approach, its motivation or strategy in learning. The Cronbach alpha reliability estimate for the 20 items in the questionnaire is 0.72, indicating an acceptable degree of overall internal consistency (Watkins & Mboya, 1997). More details of the factor analysis and Cronbach alpha values for each scale are reported fully in Goh (2006).

Academic attainment

The academic attainment records of students are obtained from the participating PHEI. Academic attainment consists of the end of session (course work and examination) results in each program of study. For all six PHEIs, the fail mark (F) is 49 per cent, pass mark (P) is 50-59 per cent and high distinction (HD) pass is 80-100 per cent. To divide the sample into contrasting performance groups, a decision has been made to standardise the results by subdividing them into numerical groups: 1 (0-49), 2 (50-59), 3(60-69), 4(70-79), and 5 (80-100). The decision has been made to provide a rating that can differentiate between relatively more or less successful students. Ratings are used to identify the way that achievement is operationalised within each of the PHEIs rather than to compare achievement scores across the six PHEIs.

Process skills

To measure students' development and acquisition of their process skills, six items of the Generic Skills Scale have been modified from Wilson, Lizzio, and Ramsden (1997). The skills measure PHEI students' development and acquisition of their process skills relevant to employability and lifelong learning. They include problem solving and analytical skills, teamwork, the ability to plan one's own work, written communication, and confidence in tackling new situations. Each of the six items requires the participants to indicate their agreement on a five-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency reliability (Cronbach alpha coefficient) for the six items of the Generic Skills Scale is 0.79.

Course satisfaction

Course satisfaction has been measured by a single item 'Overall I am satisfied with the quality of this course'. The item is rated on a 5-point Likert scale from 1 'Disagree', to 5 'Agree'.

All measures have been administered to students forty-five minutes before the end of a class or lecture. Students who volunteer complete the questionnaire before leaving the lecture hall or class room.

Students have been given the option to provide their contact details such as name, mobile and telephone numbers, and email addresses if they decide to volunteer for the qualitative part of this study. There is no time limit for the completion of the questionnaires; however, on average they take approximately 35 minutes to complete.

Multiple regression analysis has been used to examine associations. The results in Table 1 indicate that students' choice of academic discipline, high English language competency, and approaches to learning have been combined and show a medium association with academic attainment scores (R^2 = 19%, with an effect size of 0.23). For acquisition of generic skills, high English language competency, positive perceptions of aspects of good teaching, clear goals and standards, and approaches to learning combine to be related to a large amount of variance (R^2 = 39%, with an effect size of 0.63). In addition, the three environment scales combined have a large association with students' satisfaction with course scores (R^2 = 46%, with an effect size of 0.85). There are no significant associations between gender, age, and any of the learning outcomes, after taking into account the other predictors. Similarly, there is no significant association between perceptions of the learning environment and academic attainment. Deep approaches to learning are related positively to academic attainment and acquisition of generic skills while surface approaches have negative associations with these outcomes. The predictors combine to have a medium association with students' academic attainment and large associations with the acquisition of generic skills and satisfaction with the course they are undertaking.

Qualitative Methods

Qualitative data have been collected by means of semi-structured interviews from 52 participant students who have volunteered for this aspect of the study. The semi-structured interviews contain both 'warm up' and main questions to guide the course of discussion. 'Warm up' questions provide the opportunity to develop rapport with the interviewees. The main questions are designed with the research interests in mind and to obtain differing and complementary viewpoints from the students about their experiences of studying their chosen course at PHEIs, and to investigate if there is a link between how students approach their learning, the perceptions they have of their learning environments, and their learning-related outcomes. In addition, the questions allow the freedom to prompt, probe, and follow up responses for clarification and elaboration. Each interview took approximately 40 minutes.

Generally, students have been asked questions that deal with:

- the approaches they use in their learning;
- an exploration of their feelings and perceptions of their learning environment such as the teaching strategies used by their lecturers, influential person(s) who facilitate their learning or help in optimising their learning outcomes, and their overall satisfaction with the course and institution;
- elements they want to see improved in order to maximise their learning.

Table 1. Results of multiple regression analyses with student characteristics, perceptions of learning environment, and approach to learning as predictors of learning outcomes

	Academic Attainment			Acquisition of Generic Skills			Satisfaction with Course		
Variables	**b**	**β**	**t**	**b**	**β**	**t**	**b**	**β**	**t**
Individual Characteristics									
Gender (male = 1)	0.02	0.01	ns	0.32	0.04	ns	-0.04	-0.02	ns
Academic Discipline (science = 1)	0.71	0.36***	6.80	0.42	0.06	ns	-0.12	-0.10	ns
Age (less or equal to 21 = 1)	0.13	0.06	ns	-0.29	-0.04	ns	0.06	0.03	ns
English Language Competency	0.05	0.15***	3.00	0.15	0.13**	2.84	0.02	0.10	ns
Perceptions of Learning Environment									
Aspects of Good Teaching	-0.01	-0.03	ns	0.26	0.28***	5.99	0.10	0.48***	10.76
Aspects of Appropriate Assessment	-0.02	-0.04	ns	-0.03	-0.01	ns	0.04	0.10**	2.47
Aspects of Clear Goals and Standards	-0.02	-0.03	ns	0.58	0.29***	6.07	0.10	0.23***	4.96
Approach to Learning									
Deep Approach	0.02	0.16**	2.84	0.08	0.15***	3.16	0.01	0.06	ns
Surface Approach	-0.02	-0.10*	-1.97	-0.07	-0.11**	2.50	-0.01	-0.04	ns
Multiple *R*		0.44***			0.62***			0.68***	
R^2		0.19			0.39			0.46	
Effect Size		0.23[a]			0.63[b]			0.85[b]	

*p<0.05 **p<0.001 ***p<0.001 ns – non significant Effect size: [a] medium [b] large

Discussions of Findings

The findings from the study are based on a synthesis of both the quantitative and qualitative methods.

The findings indicate that certain factors which relate to students' chosen courses have encouraged them to adopt deeper level learning approaches and are important influences on their attitudes to study. The factors include the learning environment that students perceive gives them freedom of choice over their subjects. Care needs to be taken from students' comments about their enjoyment and interest in doing work that they perceive has real life relevance to them in their future professional environment and that is likely to be related to actual working practices. When perceived relevance and interest are low, students tend to adopt surface level approaches and indicate that they are studying the subject merely because it is required. McKeachie (1999) cautioned that students were intrinsically motivated in subjects that they chose rather than subjects that were compulsory for them to take. Similarly, Dart (1994) stated that students' perceptions of the lack of relevance of their subject matter discouraged them from moving towards deeper level learning approaches.

Students have been able to discern between teaching practices that encourage or are barriers towards their use of deeper learning approaches. Deeper level learning approaches are reflected in students' ability to (a) use analytical skills, (b) think critically, (c) share and apply ideas, and (d) cultivate a positive philosophy and interpretation of their world. Students are critical of teaching practices that are unconstructive, non-committal, apathetic, show favouritism, and that do not cater to different needs and student diversity. The findings suggest that appropriate teaching practices, supportive teacher attitudes, and the provision of positive learning environments are important factors for positive student learning.

Scouller (1998) identified the importance of assessment in influencing students' approaches to their learning. Even for those students who would like to demonstrate deep approaches to their learning, the very nature of assessment practices might have discouraged them from doing so. Moreover, the type, style, and number of assessments had also been shown to be associated with the approaches adopted by students. In the present study, the findings show an association between surface approaches to learning and assessments which students perceive to measure a lower knowledge-based level of intellectual processing. Also, the study has revealed that students have found dissatisfaction with learning when they perceive that the assessments have no relevance to their areas of learning, give a perception of being overwhelmed by the amount of assessment and the time frame given for completion. Moreover, external problems such as stress and anxiety caused by assessments are also related to their perceptions of overload.

PHEIs students often enter the twinning programs with the minimum English language requirement specified by the overseas partner universities. Nevertheless, the present findings suggest that studying in English is still a disadvantage for some students with lower language competency because these students perceive their learning environment less favourably, tend to adopt surface approaches to learning, and have poorer learning outcomes. Johnston (2001) observed that students who were less competent and less confident in English tended to deteriorate further in their English language skills over time. Gow, Kember, and Chow (1991) found that students who were less confident in the language were more likely to rely on rote learning without trying to understand what they were learning.

Conclusions

Although the quantitative data indicate the associations between English language competency, perceptions of learning environments, approaches to learning, and learning related outcomes of students from the twining programs, it is through the qualitative study that four differences in students' approaches to learning have been identified. They are:

- a surface approach is used where the intention is to memorise formulas, calculation methods, and theories to pass examinations;
- a surface approach is taken where the intention is a reaction to the unconstructive behaviour and negative attitudes evident in the teaching staff;
- a deep approach to learning is used where the intention is to understand meaning;
- a deep approach to learning is taken where the intention is a need to relate what has been learnt to reality and to future use.

Particularly revealing has been the discovery of the unexpected existence of a personality trait better known as *kiasu*-ism or the fear of losing out to their friends which is reflected in students (a) not questioning, (b) demonstrating passive memorisation, (c) being overly careful, and (d) showing their conformity within the learning environment. *Kiasu*-ism is often associated with Singaporeans. The interview data also show that the perceptions of being *kiasu* are prevalent in all the three ethnic groups' (Malays, Chinese and Indians) learning context and inhibit their full adoption of deep approaches to learning in Malaysia as well. More detailed definition and practical operationalisation of this personality trait and its relationship to the processes of learning as it pertains to Malaysian students is important and has the potential to enhance the teaching and learning environment.

This study has implications for both research and practice. From a research perspective, it has provided a detailed elaboration of the complex interaction between students' approaches to learning, their perceptions of the teaching and learning environment, and the ensuing quality of their learning outcomes. In addition, it has endorsed the need to establish a greater understanding of the cultural phenomenon of *kiasu*-ism which has been shown to be a link between students' approaches to learning and their learning environment.

In practical terms, it shows how important it is to understand from a student's perspective the way the learning context is perceived, what influences them to adopt a particular approach to learning, and what allows them to derive satisfaction from their learning.

Some suggestions that follow from the study are:

Overseas partner universities should ensure that they:

- Export a course curriculum that is relevant to students' work and lives.
- Understand the importance of guiding students towards adopting deeper approaches to learning.
- Understand the importance of creating and encouraging an impartial learning environment.
- Provide flexibility within each course for PHEI academics to structure and design assessments, or to re-structure some of the academic programs to suit local needs.
- Plan and design assessment activities that can help students assimilate more western educational concepts if that is desired.
- Understand that the influence of cultural elements (specifically manifestations of *kiasu* -ism) within Malaysian PHEI students' learning is important as it influences students' desire and willingness to use deeper approaches to learning.
- Accommodate students' actual English language competency.

In conclusion, I would like to briefly return to a comment made by a former Malaysian minister of education that PHEI students 'were being taught but not educated and were hence unable to think critically or to be analytical' (Indramalar. 1999, p.2). First, the current study has indicated that students who use deep approaches to learning tend to have better academic attainment, higher process skills, and are more satisfied with their learning. Possibly then, PHEI students who adopt deep approaches do think critically and are analytical. Students do desire to learn with deep meaning and understanding and they do desire to acquire the ability for analytical and creative learning because possessing these skills assists them in their aspirations to succeed in the real world. The present study also shows that students' learning environments play an important role in achieving their desired learning approaches. The psychological characteristics of *kiasu*-ism that students have reported, indicates an emphasis on submissiveness, passive memorisation, last minute cramming for examinations, and speculating on examination questions that may undermine the intentions of students to achieve deeper learning. Therefore, academics possess the ability to foster learning by constructing learning environments that are related to high level deep learning approaches. In addition, the awareness of the interrelationship of students' perceptions to the learning context, their approaches to learning, and the cultural processes of *kiasu*-ism need to be managed. At the very least, the present results should assist the former Minister and concerned academics to recognise that the relationships found in this study indicate that it is important for students to learn in a more effective way and every effort should be made to create a conducive learning environment that can support critical and analytical thinking.

Acknowledgements

I want to take a moment to write a special acknowledgement to you - Professor Keeves, to express my heartfelt thanks for the extra time you took to ensure that I was comfortable with the content of 'Structural Equation Modeling' and understand it well. It is people like you who make this world a better place. My heartfelt appreciation also goes to the late Emeritus Professor Kevin Marjoribanks (my thesis supervisor) because it was through him that I came to know you. Both of you stand head and shoulders above others in the facilitation of knowledge and certainly you both live your lives above and beyond 'the call'. Thank you for being outstanding people.

I would also like to gratefully thank Dr Bobbie Matthews for her helpful comments on earlier versions of this article.

References

Abd Rashid, A.R. (2002). Dasar inovasi pendidikan dalam konteks agenda wawasan 2020. In S.Hussin (Ed), *Inovasi Dasar Pendidikan Perspektif Sistem dan Organisasi* (pp.19-47). Kuala Lumpur: Penerbit Universiti Malaya.

Biggs, J.B., Kember, D., & Leung, D.Y.P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133-149.

Dart, B. (1994). A goal-mediational model of personal and environmental influences on tertiary students' learning strategy use. *Higher Education,* 28, 453-470.

Department of Private Education (2003). Statistik Institusi Pendidikan Swasta Retrieved July 30, 2003, from http://www.studymalaysia.com/jps/info/stats_negeri.htm

Goh, P.S.C (2006). Assessing the approaches to learning of twinning program students in Malaysia. *Malaysian Journal of Learning & Instruction*, 3, 93-115.

Gow, L., Kember, D., & Chow, R. (1991). The effects of English language ability on approaches to learning. RELC Journal, 22, 49-68.

Harris, R. (1997). Overseas students in the United Kingdom system. *Higher Education*, 29, 77-92.

Indramalar, S. (1999). Short cut or shortchanged? *Education Sunday Star*, 29 Aug., p. 2.

Johnston, C. (2001). Student perceptions of learning in first year in an economics and commerce faculty. *Higher Education Research and Development*, 20, 169-184.

Kember, D., & Gow, L. (1990). Cultural specificity of approaches to study. *British Journal of Educational Psychology*, 60, 356-363.

Kember, D., Ng, S., Tse, H., Wong, E.T.T., & Pomfret, M. (1996). An examination of the interrelationships between workload, study time, learning approaches and academic outcomes. *Studies in Higher Education*, 21, 347-358.

Leung, D.Y.P., & Kember, D. (2003). The relationship between approaches to learning and reflection upon practice. *Educational Psychology*, 23, 61-71.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. *Studies in Higher Education*, 27, 27-52.

MacNair, G. (1990). The British enterprise in higher education agreement. *Higher Education Management*, 2, 60-71.

Martön, F., & Säljö, R. (1976). On qualitative differences in learning – I outcome and process. *British Journal of Educational Psychology*, 46, 4-11.

Matthews, B.M. (2003). *Life Values and Approaches to Learning: A Study of University Students from Confucian Heritage Cultures*. PhD Thesis, The Flinders University of South Australia.

McKeachie, W.J. (1999). *Teaching Tips: Strategies, Research and Theory for College and University Teachers,* (10th edition). Boston, MA: Houghton, Mifflin.

Ramsden, P. (1984). The context of learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds), *The Experience of Learning* (pp.144-164). Edinburgh: Scottish Academic Press.

Ramsden, P. (1992). *Learning to Teach in Higher Education*. London: Routledge.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472.

Smith, S.N. (2001). Approaches to study of three Chinese national groups. *British Journal of Educational Psychology,* 71, 429-441.

Tan, M.C. (1990). Some factors influencing student performance in laboratory experimental work in physiology with implications for curriculum deliberations and instructional design. *Higher Education,* 19, 473-479.

Volet, S.E., & Kee, J.P.P. (1993). Studying in Singapore, studying in Australia – a student perspective. Occasional paper 1, Murdoch University Teaching Excellence Committee, Murdoch University, Perth.

Wan Ali, W.Z. (2000). *Memahami Pembelajaran*. Kuala Lumpur: Utusan Publications & Distributors Sdn Bhd.

Watkins, D., & Ismail, M. (1994). Is the Asian learner a rote learner? a Malaysian perspective. *Contemporary Educational Psychology*, 19, 483-488.

Watkins, D., & Mboya, M. (1997). Assessing the learning processes of black South African students. *The Journal of Psychology*, 131, 632-640.

Wilson, K.L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education*, 22, 33-53.

Zakaria, A. (2000). Educational development and reformation in the Malaysian education system: challenges in the new millennium. *Journal of Southeast Asian Education*, 1, 113-133.

21

The genesis of new ideas: Models, feeling and solutions

Carol Aldous
School of Education, Flinders University

Introduction

The roles played by feeling, intuition and imagination (together with other non-cognitive approaches to reasoning) in the solving of novel mathematics problems are both intriguing and ambiguous. This is because the identification, description and measurement of such non-cognitive activities are notoriously elusive. Nevertheless this chapter describes the journey of one such undertaking. Under the tutelage of John P. Keeves, a two pronged educational measurement approach is adopted. One path involves a small scale study of five expert problem solvers, the other a large scale study of 405 middle school students solving two novel mathematics problems. The first path entails the use of protocol analysis, and the development of a conceptual model of creative problem solving. The second involves testing this model using multi level structural equation modelling procedures, indicative of cutting edge educational measurement research. The use of methodologies from diverse fields reported in this chapter is testament to John's capacity to embrace the richness of educational measurement and to his ability to transform complex research problems into seemingly simple solutions. This chapter is for you John.

Background

The work presented here is a summary of findings that took more than three years to complete. Its contents are profound in the sense that they overturn conventional wisdom which suggests that feeling interferes with the ability of individuals to solve problems. Contrary to traditional viewpoints, this study finds that attending to feeling is an essential ingredient of successfully solving novel problems and, moreover, that in the absence of this feeling individuals are unlikely to solve novel problems at all. Indeed, no model of solutions could be found that was without the involvement of feeling. Novel problems were solved directly from a feeling but problems were unlikely to be solved successfully independent of feeling.

Feeling, in this study, is interpreted to mean a feeling of knowing, a feeling of connection, an intimation, or a feeling of cognition in the aesthetic sense of comprehending a new intellectual order (Ghiselin, 1963). Thus, the finding that 'no model of solutions was without the involvement of feeling' is something of a revelation in this study. This is all the more so when it is realised that, at the outset of this research, the likelihood of adequately modelling non-cognitive systems involved in novel mathematics problem solving was not at all certain. Hence the conclusion that feeling is not only involved, but is also likely to have a key role in successfully solving novel mathematics problems is important. It is a finding that, if verified with further research, has enormous implications for theory, for teaching and for future research.

The route to this conclusion has not been simple. It has involved engaging with a range of methodological tools from the qualitative to the quantitative and applying them with integrity and rigour. It has necessitated courage in the face of scepticism, persistence in the face of frustration and openness to receiving new ideas. It has involved embracing the paradoxes of creativity and the ambiguities of interpretation in the hope that with adequate model development some explanatory contribution to the fields of creative problem solving and mathematics learning and teaching could be made.

The discovery that 'no model of solutions was without the involvement of feeling' is indeed an example of stochastic modelling making an explanatory contribution to the fields of creative problem solving and mathematics learning and teaching. But this explanatory contribution is also likely to have application to other areas of life. For, if feeling is present in the successful solution of novel mathematics problems, it is also likely to be present at the solution of other novel problems of life. After all "Feelings" Damasio (1994, p.xvi) contends "form the base of what humans have described for millennia as the human soul or spirit".

Two questions lay at the heart of this research: 'From where do new ideas come?' and 'Is there any substance to the idea that individuals can feel their way to a solution in a novel mathematics problem solving event?' In seeking answers to these questions it was necessary to:

- identify and describe this so-called feeling (*Stage One*);
- find some way of measuring it (*Stages Two and Three*);

and by so doing

- modelling whether this so called feeling made any difference to solving novel problems successfully (*Stage Four*).

Five individuals participated in *Stage One* – qualitative path. Four hundred and five individuals participated in *Stages Two, Three and Four* – quantitative path. A summary of the procedural findings for each path and stage of the study is given below. It is followed by a discussion concerning the implications for theory, for teaching and for future research.

Overview of research

Purpose and plan

The purposes of this study were to identify and describe cognitive (thinking) and non-cognitive (feeling) elements used in solving novel mathematics problems and to investigate how such elements may interact. Although the outcome of such a goal was always uncertain, the construction of a conceptual framework of creative problem solving comprising the identified cognitive and non-cognitive elements, was hypothesised in order that the use of such elements by adolescents in solving novel mathematics problems could be examined. In this way any relationships to success in solving novel mathematics problems could be identified.

Building on the theory of Mooney (1963) and others (Rhodes, 1961; Richards, 1999; Tardif & Sternberg, 1989), the influence of antecedent factors, **person-**related factors, **environmental**-related factors and **product**-related factors on the use of cognitive and non-cognitive processes in problem solving needed to be considered. This ensured that the processes of creative problem solving could be examined in a rich contextual setting. Through the formation of a comprehensive model of creative problem solving, the collective interactions of each of the **person, environment**, **process** and **product** dimensions of creativity were to be assessed. In particular, the cognitive and non-cognitive elements of the **process** dimension comprised the microscopic aspect, while the **person**, **environment** and **product** dimensions comprised the macroscopic aspect of the comprehensive model.

Thus the microscopic aspect of creativity was embedded within the macroscopic one, permitting both cognitive and non-cognitive elements of the framework of creative problem solving to be examined in a comprehensive setting. The comprehensive model of creative problem solving thus formed could then be examined for applicability and generality to other novel problems.

Procedural findings

The study proceeded in four stages. The first stage involved identifying and describing cognitive (thinking) and non-cognitive (feeling) elements used in solving novel problems in particular instances of verbal protocols and developing a conceptual framework of creative problem solving. The second stage involved testing for these elements on a large scale through the formation of a self-report instrument capable of measuring the identified elements within the conceptual framework, using exploratory and confirmatory factor analysis. The third stage involved the formation of a comprehensive causal model of creative problem solving that not only incorporated the cognitive and non-cognitive **Process** elements of the conceptual framework, but also the macroscopic elements related to the **Person**, **Product**, and **Environment** dimensions of creativity. The fourth stage involved testing the structural relationships arising between the variables, both manifest and latent, within the comprehensive model of creative problem solving. This testing proceeded in two parts. The first part involved the use of Partial Least Squares (PLS) path analysis and the proposal of theory. The second part involved the use of a confirmatory structural equation modelling program with AMOS graphics (Arbuckle, 1999) employing maximum likelihood estimation and multi-group modelling techniques and the testing of theory. The procedural findings obtained at each stage of the study are summarised below.

Stage One – Qualitative path

As stated above, the first stage of this study involved the use of protocol analysis to identify and describe cognitive and non-cognitive elements used in solving novel mathematics problems. This step was essential since measuring cognitive and non-cognitive processes on a large scale was not possible until that which was to be measured could be identified. Thus, based on the information-processing model of cognition (Ericsson & Simon, 1993), concurrent and retrospective reports were collected from five experts while solving a novel mathematics problem. However, unlike traditional studies focusing largely on cognitive components, the non-cognitive elements of feeling, intuition, imagination and inspiration were also examined. Three themes were identified:

- the interaction between visual-spatial and analytical reasoning;
- the interaction between conscious and non-conscious reasoning; and
- the role of feeling in listening to the self.

These themes were used to construct a conceptual framework of creative problem solving containing both cognitive and non-cognitive elements. These elements are:

- the visual-spatial and linguistic circuits within the brain (as represented by approximate and exact arithmetic brain imaging evidence(Dehaene et al, 1999));
- conscious and non-conscious mental activity (as represented by rule-based and associative systems (Sloman, 1996) and by the rational and experiential systems (Epstein, 1998)); and
- the generation of feeling in listening to the self including that of intuition (as represented by neuro-science research linking rationality with feeling (Damasio, 1994; Immordino-Yang and Damasio 2007).

It was proposed that the interaction of each of these elements facilitated both cognitive and non-cognitive knowing and the manifestation of creativity. Central to the conceptual framework was the Intuitive function serving to evaluate and filter information generated in conscious (Self State One) and non-conscious activity (Self State Two), and in the linguisitic and visuo-spatial circuits. Feeling, it is proposed, was implicit to the mechanisms of interaction between each of the elements and was involved in the evaluation and communication of ideas at each stage of the creative process. The conceptual framework and proposed interactions are represented diagrammatically in Figure 1.

Stage Two – Quantitative path

The second stage of the study involved testing and verifying the cognitive and non-cognitive elements of the conceptual framework, identified in particular instances of protocol analysis, on a larger scale. This required the formation of an instrument that would tap the main elements within the conceptual framework of creative problem solving, which was achieved through the development of a self-report instrument termed the Systems of Reasoning Questionnaire (SRQ) (Aldous 2005; 2006).

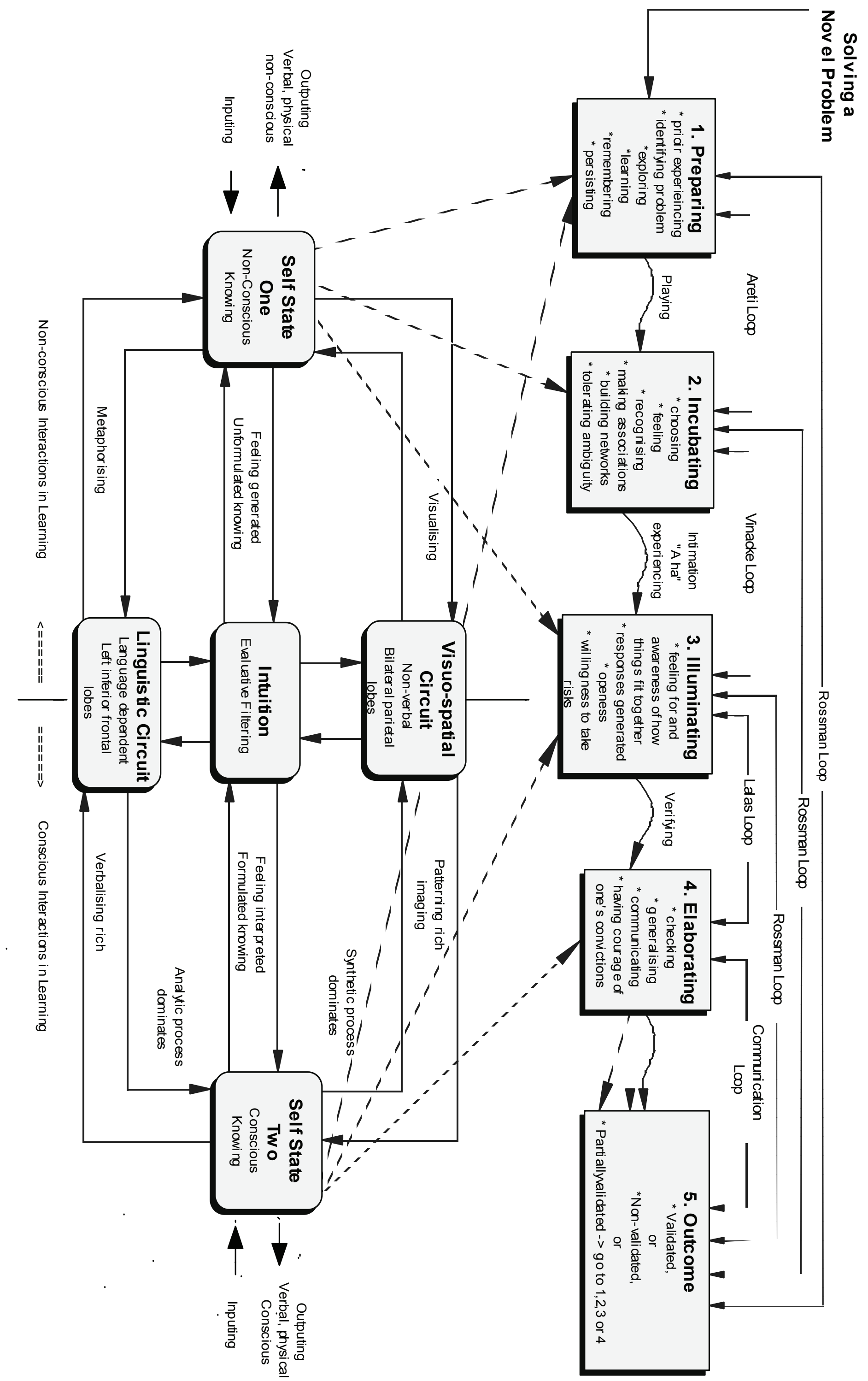


Figure 1. An emerging conceptual framework: Listening to the self in learning (dynamic interactions of flow)

The SRQ was grounded in Sloman's (1996) theory of two systems of reasoning and in Epstein's (1994) two patterns of mind. It was administered in both trial and main data collections to students participating in the Mathematics Challenge for Young Australians. The Mathematics Challenge for Young Australians is an annual program in novel mathematics problem solving for students in Grade 5 through to Grade 10. Participants in this program had three weeks in which to answer six novel problems. Respondents to the SRQ completed the questionnaire with respect to two of the six problems. These problems were the Cute Numbers and Birthday Cake problems. A copy of these two problems is given in Appendix 1.

Exploratory and confirmatory factor analysis, involving the testing of models with data collected through the SRQ, and triangulated in three structural equation modelling (SEM) programs using STREAMS (Structural Equation Modelling Made Simple) (Gustafsson and Stahl, 2000), revealed the presence of five distinct factors. These factors, formalised into five process scales, were the:

- **Strategic** approach to reasoning (**Strat**),
- **Free-flowing** approach to reasoning (**Free**),
- **Spatial-verbal** approach to reasoning (**Spat/Vb**),
- **Feeling** approach to reasoning (**Feel**), and
- **Systematic** approach to reasoning (**Syst**)

A system of *a priori* classification showed the scales for the Strategic and the Systematic approaches to reasoning to be predominately cognitive in nature, the Free-flowing and Feeling approaches to reasoning to be predominately non-cognitive in nature. The Spatial-verbal scale was found to be both cognitive and non-cognitive in nature depending on whether processing was simultaneous or successive and on whether verbal or non-verbal reasoning or both were involved (Aldous 2005; 2006).

The five approaches to reasoning were mapped to the cognitive and non-cognitive elements identified within the conceptual framework. This served the function of putting into operation the conceptual elements of the framework within the context of the study. In particular the Systematic factor was mapped to conscious activity (designated as Self State Two), the Free-flowing factor was mapped to non-conscious activity (designated as Self State One) and the Spatial-verbal factor was mapped to both the visuo-spatial and linguistic circuits according to their respective emphases. The Strategic factor was mapped to the elements of the stage model of creativity while the Feeling approach to reasoning was mapped to the centrally placed element of Intuition (refer to Figure 1).

Stage Three – Quantitative path

The third stage of the study involved the formation of a comprehensive model of creative problem solving which, along with the antecedent constructs, was found to incorporate all four dimensions of creativity including its macroscopic and microscopic aspects. This was done through the formation of a causal model in which the sequence of relationships arising between each of the latent variables within the model was specified. The logical and temporal order of variables within the comprehensive model was defined with due reference to theory. In particular, Carroll's (1963) model of school learning and Keeves' (1986) cycle of performance were used. Variables related to the **Antecedent** construct were identified first, followed by variables indicative of the **Abilities** construct, then variables representative of the **Experiences** construct (subset Opportunity to learn and Quality of instruction). Next came variables related to the **Perseverance** construct, followed by variables representative of the **Approaches to reasoning** construct, leading ultimately to the **Outcome** construct. In particular the **Outcome** construct was measured using the scores obtained on each of the Cute Numbers and Birthday Cake problems.

Embedded within this macroscopic causal model of creativity was a minor and detailed or microscopic one. The microscopic causal model comprised a specified sequence of process variables reflecting the intent of the creative problem solving process. This sequence comprised: **Strat → Free → Spat/Vb → Feel → Syst → Score**. Each of these variables was mapped to elements of the conceptual framework.

The comprehensive model of creative problem solving so formed, put into operation Mooney's conception of creativity as the interaction of its four dimensions (namely the **environment**, **person**, **process** and **product**) as well as Collins and Amabile's (1999, p.307) re-conceptualisation of this idea as the "creativity intersection". It also provided the desired specificity recommended by Lubart and Sternberg (1995) and others (Mumford 2003a; 2003b). The mathematical formulation of the comprehensive model enabled the creativity intersection to be mapped and influential factors intrinsic to successful novel mathematics problem solving to be identified.

Stage Four (Part One) – Quantitative path

The fourth stage of the study involved testing the structural relationships arising between variables within the comprehensive model of creative problem solving for both the Cute Numbers and Birthday Cake data. This was done initially through Partial Least Squares (PLS) path analysis since it made no assumptions about multivariate normality or the independence of variables (Sellin & Keeves, 1997) and allowed theoretical propositions to be examined. A diagram summarising the direct and indirect effects for each of the Cute Numbers and Birthday Cake path models is given in Figure 2.

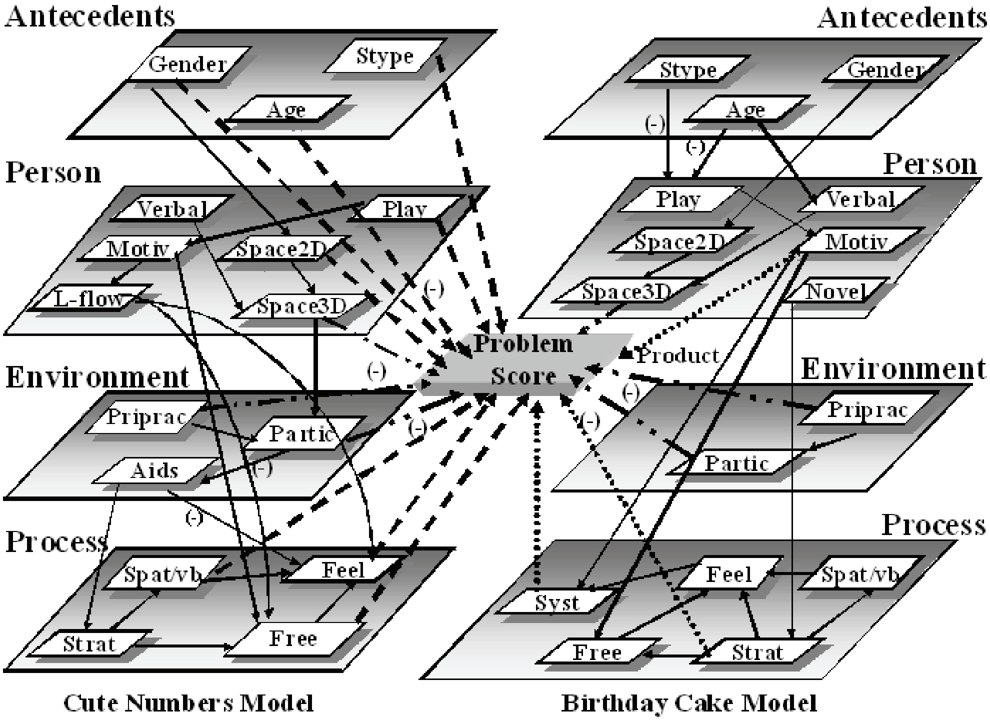


Figure 2. Comparison of direct and indirect effects for the Cute Numbers and Birthday Cake models

Variables representing the macroscopic aspect of the models are, given by the **Antecedents,** the **Person**, the **Environment** and the **Product** dimensions of creativity. Variables representing the microscopic aspect of the model are given by the **Process** dimension. Paths of direct influence common to both models are indicated by alternating dotted and dashed lines. Paths of direct influence unique to the Cute Numbers model are indicated by dashed lines. Paths of direct influence unique to the Birthday Cake model are indicated by dotted lines. Paths indicating indirect effects within each model are represented by continuous solid lines. However, for ease of reading not all indirect effects are shown in the diagram. Tables documenting the size of the direct path coefficients and the presence of indirect effects for both the Cute Numbers and Birthday Cake problems are given in Appendix 2.

Through analysis of paths, the conclusion was reached that different creative processes were likely to be at work within the Cute Numbers and Birthday Cake problems. Different spatial and verbal processing requirements were indicated for each problem and different approaches to reasoning were identified in each model. The Cute Numbers problem was characterised mostly by non-verbal processing and the explicit use of non-cognitive factors, namely the Free-flowing (**Free**) and Feeling (**Feel**) approaches to reasoning, while the Birthday Cake problem was characterised mostly by verbal processing and the dominant use of cognitive factors, namely the Systematic (**Syst**) and Strategic (**Strat**) approaches to reasoning.

However, while the Birthday Cake problem was dominated by the use of cognitive factors, the use of the non-cognitive factors of **Free** and **Feel** were also in evidence. In particular, an indirect effect of **Feel** on **Pscore** (Problem score)**,** operating through **Syst** was observed and an indirect effect of **Free** on **Pscore** operating through **Feel** and **Syst** was also present. Further, while the path from **Strat** to **Pscore** was negative, the path from **Strat** to **Pscore** passing through **Feel** was positive. Such indirect effects pointed to a significant role for **Feel** in the problem solving process.

Thus the Birthday Cake problem was characterised by direct effects of cognitive factors and by indirect effects of non-cognitive ones. However, the reverse was not the case for the Cute Numbers problem. Although the direct effects of non-cognitive factors characterised the Cute Numbers problem, the indirect effects of cognitive factors did not. Indeed, the factor for the Systematic approach to reasoning (**Syst**) played no role at all in the successful solution of the Cute Numbers problem. Moreover, an indirect effect of Strategic approaches to reasoning (**Strat**), operating through the non-cognitive factor **Free** was observed. This was also supported by an indirect effect of **Free** operating through **Feel**.

In addition, a negative direct effect for the Spatial-verbal approach to reasoning (**Spat/Vb**) was observed in the creativity intersection of the Cute Numbers problem. Nevertheless, a positive indirect effect of **Spat/Vb** operating through **Feel** was also present. This indicated that spatial-verbal processing, occurring without recourse to the Feeling approach to reasoning, was unsuccessful. Thus the central placement of **Feel** between the visuo-spatial and linguistic circuits within the conceptual framework of creative problem solving was once again indicated (refer to Figure 1).

Consistent with the Cute Numbers problem containing a greater visuo-spatial processing component than the Birthday Cake problem, both direct and indirect effects of three-dimensional and two-dimensional spatial ability within the **Abilities** construct were observed for the Cute Numbers problem, while only a direct effect of three-dimensional spatial ability and an indirect effect of two and three-dimensional spatial ability were found in the creativity intersection of the Birthday Cake problem.

Interestingly, a negative direct effect of **Gender,** as well as positive direct effects of **Age** and School type (**Stype**) from the Antecedent construct in the Cute Numbers creativity intersection, indicated that older girls from co-educational schools were likely to have performed better on the Cute Numbers problem than boys of similar standing. However, a positive indirect effect of **Gender** and **Age** as well as a negative indirect effect of **Stype** indicated that older boys from single sex schools were likely to have performed better in both the Cute Numbers and Birthday Cake problems than girls of similar standing from single sex settings.

Further, within the **Perseverance** construct, play with novel mathematics problems and games **(Play**) had both direct and indirect effects in the creativity intersection of the Cute Numbers problem, but only an indirect effect within the Birthday Cake problem. Meanwhile both a direct and indirect effect of intrinsic motivation (**Motiv**) within the creativity intersection of the Birthday Cake problem, but only an indirect effect of **Motiv** was observed in the Cute Numbers problem. This indicated that persistence and intrinsic motivation were more critical to success than play with novel mathematics games and puzzles in the Birthday Cake problem, but that **Play** was more critical to success in the Cute Numbers problem. Such differences reflected different creative processes at work within the two problems. An associative approach was implicated for the Cute Numbers problem, but a rule-based approach was implicated for the Birthday Cake problem. Such differences were also indicated by the fact that successful students in the Cute Numbers problem were more likely to experience flow as measured by the latent variable **L-flow** than successful students in the Birthday Cake problem.

Finally, it was found that students in both the Cute Numbers and Birthday Cake problems who had participated in previous Mathematics Challenge events performed better than those who had not. However, students with prior practice of Mathematics Challenge problems, but who had not previously participated in Mathematics Challenge events did not perform better than students having such previous Mathematics Challenge event experience.

Stage Four (Part Two) – Quantitative path

The findings of the Partial Least Squares (PLS) path analysis indicated that different creative processes were at work between the two problems. However, the means to test for model significance between groups using PLS was not available. Thus, confirmation of the propositions described above and generalisation to other types of novel problem were sought. Consequently maximum likelihood, multi-group modelling, using Analysis of Moments Structures (AMOS) (Arbuckle, 1999) software was employed to assess how similar or different the Cute Numbers and Birthday Cake path models were in their mechanisms of operation.

Employment of a confirmatory SEM technique required refining the comprehensive model of creative problem solving in order that the model could be specified within the confirmatory SEM framework. All process variables were retained within the model, although some variables related to the **Perseverance** construct and the **Experience** construct, were removed. Still others, such as the variables for two and three-dimensional spatial ability within the **Abilities** construct were combined to form the new latent variable **Space**. A complete list of latent variables within the refined comprehensive model of creative problem solving is given in Table 1. Nevertheless, each dimension of creativity was represented within the refined model.

Through the process of multi-group modelling, in which the refined model of creative problem solving was simultaneously tested for model fit in both the Cute Numbers and Birthday Cake data, an index of good model-to-data correspondence was obtained with very adequate fit statistics (RMSEA = 0.036). This enabled firm comparisons regarding the similarities and differences arising between the Cute Numbers and Birthday Cake data to be made.

Microscopic aspect

Despite the linear placement of latent variables within the causal model, a number of feed back loops were in evidence within the measurement or outer model of the comprehensive model.These loops were consistent with Shaw's (1989) cyclical representation of the creative process that necessitated both conscious and non-conscious processing. (For a summary of Shaw's proposed feedback loops please refer to the upper section of Figure 1).

Table 1. Latent variables included within PLSPATH and AMOS path models

Creativity Dimension	Construct	Latent Variable (LV)	PLSPATH	AMOS	AMOS revised LV
	Antecedents	Gender	●	●	
		Age	●	●	
		Stype (School type)	●	●	
Person	Abilities	Verbal (Verbal ability)	●		
		Space 2D (2 dimensional spatial ability)	●	▲	Space
		Space3D (3 dimensional spatial ability)	●	▲	Space
	Perseverance & Motivation	Play	●	●	
		Motiv (Intrinsic motivation)	●	●	
		Novel (Novelty)	●		
		L-flow (Flow)	●		
Environ-ment	Opportunity to learn	Priprac (Prior practice)	●	▲	Experience
		Partic (Participation)	●	▲	Experience
	Quality of instruction	Aids (Resources)	●		
		Help (Teacher help)	●		
Process	Approaches to reasoning	Strat (Strategic approach)	●	●	
		Free (Free-flowing approach)	●	●	
		Spat/Vb (Spatial-verbal approach)	●	●	
		Feel (Feelng approach)	●	●	
		Syst (Systematic approach)	●	●	
Product	Outcome	Pscore (Problem score)	●	▲	Score

● Indicates latent variable included in path model
▲ Indicates latent variable included in the model in a condensed form

The identified cycles involved moving between:

- preparation and incubation consistent with the Areti loop (This loop was identified in the Strategic approach to reasoning (**Strat**) in both the Cute Numbers and Birthday Cake data sets);
- incubation and illumination indicative of the Vinacke loop (This loop was identified in the Free-flowing approach to reasoning (**Free**) in both the Cute Numbers and Birthday Cake data sets);
- illumination and explication representative of the Lalas loop (This loop was identified in the Feeling approach to reasoning (**Feel**) in each data set with a notable distinction. Intimation led to the appearance of a sudden illumination in the Cute Numbers problem, but to a gradual illumination in the Birthday Cake problem); and
- explication and validation consistent with the Communication and Rossman loops (These loops were identified in the Systematic approach to reasoning (**Syst**) in each data set with a notable difference. The process of elaboration typified the Cute

Numbers problem while the process of verification typified the Birthday Cake problem).

Indication of different processing events arising in each of the Cute Numbers and Birthday Cake problems was also in evidence in the Spatial-verbal approach to reasoning (**Spat/Vb**). In particular, simultaneous spatial and verbal processing were indicated in the Cute Numbers problem while sequential spatial and verbal processing were indicated in the Birthday Cake problem. A number of points were raised. These included the ideas that:

- sudden illuminations may be associated with simultaneous processing while gradual illuminations may be associated with sequential processing;
- preconscious and conscious activity may occur simultaneously in sudden illuminations but sequentially in gradual illuminations (Thus the stages of illumination and explication could occur simultaneously in sudden illumination but sequentially in gradual illumination);
- the activities of working memory may be supplemented by processing occurring outside awareness such as that arising through defocused states of attention including incubation.

A diagram depicting the feedback loops associated with each element in the conceptual framework of creative problem solving is given in Figure 3. It should be recalled from Figure 1 that Self State One was aligned with non-conscious processing and Self State Two with conscious processing. The element connecting these two states of self is the Intuitive function, an evaluative filter involved in the generation and interpretation of feeling. The Intuitive function also mediates the interactions of the visuo-spatial and linguistic circuits.

The Areti loop is associated with cycling between Self State One and Self State Two. This cycling can occur through the visuo-spatial and or the linguistic circuits. It may also occur through the Intuitive function. The Vinacke loop is associated with cycling between Self State One and the Intuitive function. The Lalas loop is associated with cycling between the Intuitive function and Self State Two.

The Communication and Rossman loops are associated with cycling between Self State Two, the problem-solving outcome and every other phase of the creative problem solving process. The extent of recycling depends on the validation of the problem-solving outcome. A complete validation of the outcome may result in the individual exiting the creative problem solving process entirely, while a partial validation may result in the individual revisiting any stage or stages in the creative process, namely preparation, incubation, illumination or elaboration.

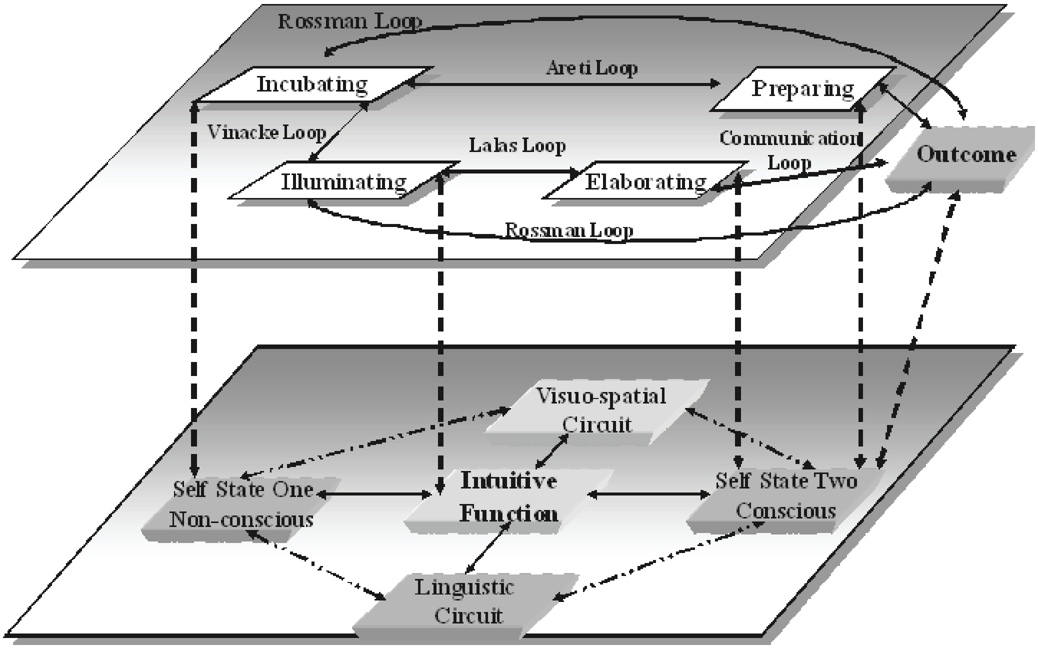


Figure 3. The conceptual framework and cycles of feedback

It may also be recalled from previous discussion that unless cycling between Self State One and Self State Two gives rise to a feeling, such as that interpreted by the Intuitive function, then a successful outcome is unlikely to occur. The same is true of cycling between the visuo-spatial and linguistic circuits. Should no intuition arise then the likelihood of finding a successful solution is low. Unsuccessful students take the paths indicated by the alternating dashed and dotted lines in the lower level of the diagram, while successful students draw on the paths indicated by solid lines depicted in the lower level of the diagram.

The importance of feeling in solving a novel problem was also evident in the structural model of creative problem solving. Two solution paths were present. One included the path; **Strat→ Free →Spat/Vb→ Feel → Syst →Score**, for the Birthday Cake problem. The other included the path; **Strat→ Free →Spat/Vb→ Feel → Score** for the Cute Numbers problem. This indicated that the Cute Numbers task relied more heavily on associative forms of reasoning while the Birthday Cake task relied more heavily on rule based forms of reasoning. However, while non-cognitive processing could operate independently or with minimum cognitive involvement in the Cute Numbers task (as represented by the path **Feel → Score**), the reverse was not the case in the Birthday Cake problem. Both cognitive and non-cognitive processing were needed. This is indicated by the dependence of **Syst** on **Feel** in the path **Feel → Syst →Score** for the Birthday Cake problem. Thus for creativity to be manifest, the likelihood was that attention to feeling was required.

It should also be pointed out that, in cases in which the operational path to **Score** from **Strat** (Strategic approach to reasoning) took place through the non-cognitive sequence of **Free→ Spat/Vb→ Feel** the cognitive processes involving **Strat** were also important to the successful solution of both problems. The path **Strat → Score** was negative in the Birthday Cake problem and not significant in the Cute Numbers problem. Thus, for the likelihood of strategic approaches being successful, the non-cognitive processes contained in **Free** and **Feel** were also required.

Contrary to expectations then, it was the non-cognitive feeling aspect of problem solving, rather than the cognitive rule based aspect, that set successful problem solving apart. Surprising as this may seem, this finding is entirely consistent with neuro-scientific research. Reason, neuro-scientist Damasio (1994) has proposed, is inseparably dependent on feeling.

A diagrammatic representation of the interactions arising between the cognitive and non-cognitive process elements within the microscopic aspect of the refined model of creative problem solving is given in Figure 4.

Macroscopic aspect

Figure 4 summarises the pattern of large to medium effects present within the comprehensive model of creative problem solving. For ease of viewing, these effects have been separated into macroscopic and microscopic aspects. Macroscopic variables are represented in the upper part of the diagram. Microscopic variables are represented in the lower part of the diagram. Paths that are significant in both the Cute Numbers and Birthday Cake models are drawn with a solid line. Paths that are significant only in the Cute Numbers problem are drawn with a dashed line. Paths that are significant only in the Birthday Cake problem are drawn with a dotted line. However an alternating dashed and dotted line has been used to indicate paths that are significantly different between models. A table displaying standardised and unstandardised estimates for each path in the structural equation model for Cute Numbers and Birthday Cake data is given in Appendix 3.

Of note are the five significantly different paths surrounding the variable **Space** (a measure of two and three-dimensional spatial ability), pointing to a difference in spatial processing between models. This is supported by the Spatial-Verbal factor being reflected in the measurement model by simultaneous synthesis in the Cute Numbers task and by sequential synthesis in the Birthday Cake task. Simultaneous synthesis, it may be recalled, has been aligned with visual pattern perception while sequential synthesis has been aligned with verbal mediation and ordered memory phenomena (Das et al., 1979). Further, a significant effect of **Space** on Feeling approaches to reasoning (**Feel**) is noted for the Cute numbers task but is absent in the Birthday Cake problem.

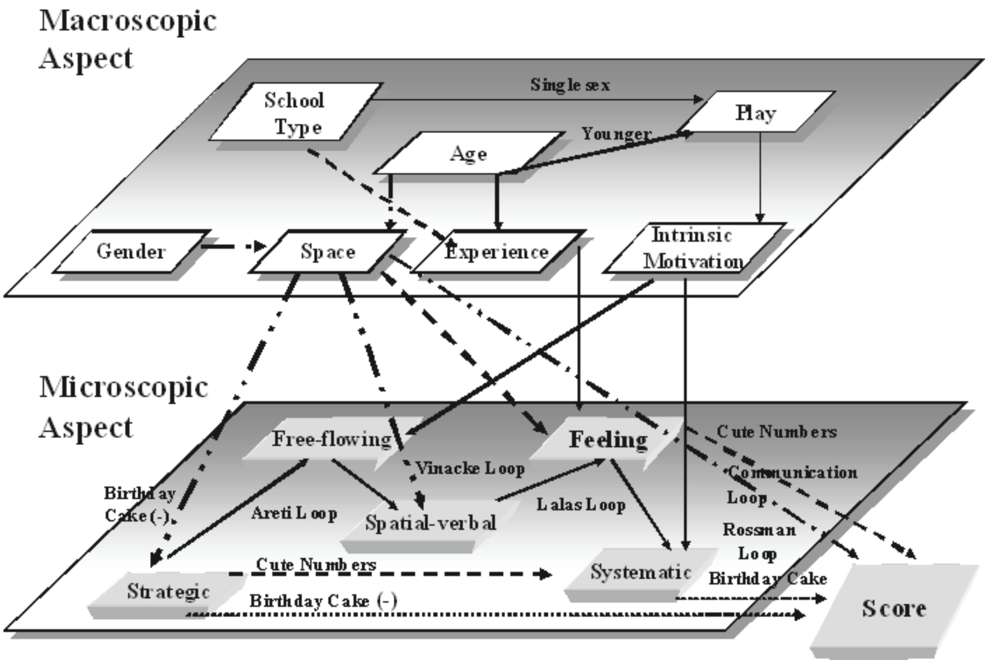


Figure 4. Pattern of large to medium effects in the comprehensive model of creative problem solving

Note: Variables in the macroscopic aspect are drawn in the upper part of the diagram Variables in the microscopic aspect are drawn in the lower part of the diagram. Paths that are significant in both problems are drawn with solid lines. Paths that are significant only in the Cute Numbers problem are drawn with dashed lines. Paths that are significant only in the Birthday Cake problem are drawn with dotted lines. Paths that are significantly different between models are indicated by alternating dashed and dotted lines.

However, these findings should not be interpreted to mean that visuo-spatial processing was absent from the Birthday Cake problem. Indeed, a large effect of two and three-dimensional spatial ability on the problem score was observed for both the Birthday Cake and Cute Numbers tasks. Nonetheless where simultaneous synthesis is accessible at all points in the process, sequential synthesis is accessible in only one direction (Das et al., 1979). It seems probable therefore that sequential synthesis may be more exacting on the problem solver in the Birthday Cake task.

Also of note is the medium to large effect of intrinsic motivation (**Motiv**) on both Free-flowing (**Free**) and Systematic (**Syst**) approaches to reasoning. This indicates that intrinsic motivation has a dual effect, acting not only at the conscious level, but also at the non-conscious level of awareness. Moreover, this situation is true no matter which problem is being solved.

It may also be noted that students with previous participation in the Mathematics Challenge event (**Experience**) make better use of Feeling approaches to reasoning (**Feel**). Thus increasing a person's knowledge base and raising that person's level of experience is likely to assist the production of creative new ideas. Knowledge and experience, however, do not constitute creativity. Recourse to non-cognitive resources such as that represented by **Free** and **Feel** is required.

Finally, it should be pointed out that the pattern of large to medium effects given in Figure 4 substantiates the proposed sequence of process variables put forward in the comprehensive causal model of creative problem solving. However, there is one notable addition. This involves the significant path from **Feel** to **Score,** which is in evidence in the Cute Numbers problem. Thus two alternate routes to **Score** are possible. One path leads from **Feel** to **Score** going through **Syst**. The other path leads direct from **Feel** to **Score**. Thus, while cognitive elements are dependent on non-cognitive ones in the derivation of a solution, a solution can largely arise from a non-cognitive state such as that represented by the Feeling approach to reasoning. Furthermore, a solution is unlikely to arise from a state independent of such a Feeling approach to reasoning.

Discussion

Implications for theory

In highlighting the need for rigorous empirical work to advance the scientific analysis of creativity (Runco, 2003), creativity researchers Runco and Chand (1995, p.243) made two salient points with respect to cognition and creativity. The first of these states that "… there is a need to understand processes which are not recognized unless one is specifically interested in creativity" and secondly, that some of these processes will be "inherently subjective, a fact which is often discarded by those hoping for a traditionally scientific analysis".

Through the use of "traditionally" and extended "scientific analysis" this study has not only identified cognitive processes relevant to the first point but also described non-cognitive ones relevant to the second. That such an outcome has occurred presents a challenge to theory inferred in the second point, that subjective processes are elusive to rigorous measurement. Exploratory and confirmatory factor analyses triangulated in three SEM programs in this study located five process factors relevant to creative problems solving. Two of these have been identified as non-cognitive, two as cognitive, while a third provides a unique measure of simultaneous or sequential synthesis interpretable in both cognitive and non-cognitive ways.

In particular, the non-cognitive factors of Free-flowing and Feeling approaches to reasoning have been identified with states of incubation and illumination consistent with the Vinacke loop, and with states of illumination and elaboration as are consistent with the Lalas loop respectively. The cognitive factors of Strategic approach to reasoning and the Systematic approach to reasoning have been identified with states of preparation leading into incubation consistent with the Areti loop and with elaboration leading into validation and communication consistent with the Rossman and Communication loops. The outcome of this study supports a cyclical view of the creative process in which progress towards a successful solution occurs in a recursive helical manner. This outcome is consistent with the structural view of creativity presented by Shaw (1989) and others (Lubart, 2001) and represents a potentially useful tool for translating theory into action. Use of this tool provides a way to examine the creative process in a range of contexts and settings and delivers a format by which other as yet unknown creative process cycles may be identified and described.

However, it is not sufficient to state that non-cognitive elements as well as cognitive elements of creative problem solving have been identified and described in this study. It must also be stated that such non-cognitive elements are also likely to be important for success. Feeling and Free-flowing approaches to reasoning were found directly and indirectly to influence success in creative problem solving. This presents a challenge to theory focusing solely on cognitive elements. The conclusion that solutions are unlikely to arise from a state independent of feeling shows feeling to be not merely external to the creative process, but also to be integrally involved. Contrary to traditional models of information processing, it was the non-cognitive processes, rather than the cognitive ones, which set successful novel problem solving apart. While creative solutions could be reached directly from non-cognitive states, no creative solutions were reached from cognitive states independent of non-cognitive ones. Thus, rather than placing feeling and intuition at the periphery of novel mathematics problem solving, the data presented here place feeling and intuition at the core of the process. This finding that cognitive processes are closely intertwined with non-cognitive ones, is entirely consistent with the neuroscientific theory in which reason is believed to be inseparably dependent on feeling (Damasio, 1994; Immordino-Yang, and Damasio, 2007).

The integrated nature of cognitive and non-cognitive processing observed in the current study supports the theory put forward by Epstein (1994) and others (Salk, 1985) that creativity results from the interaction between cognitive and non-cognitive synthesis. However, the finding that solutions could also be derived primarily from non-cognitive states, also gives credence to the view described by Sloman (1996) that creativity is largely a function of associative forms of reasoning. Nevertheless, what Sloman does not say in the context of two systems of reasoning, is that for creativity to be manifest, rule-based functions of reasoning are dependent on associative ones in order for a successful outcome to be reached.

In a review of creativity research Mumford (2003a; 2003b, p.147) called for more integrative models of creativity while "simultaneously conducting research evidencing greater specificity with regard to processes, tasks and domains…". Thus, through probabilistic causal modelling, this study sought to specify an integrated model of creative problem solving within the domain of novel mathematics problem solving that would permit elements of the **person**, **process, environment** and **product** dimensions of creativity to be considered. While not the sole model possible, the model examined in this study presents a framework upon which future theorists can build. Indeed, the way lies open to test other kinds of novel mathematics problems, to examine other kinds of environmental contexts, and to explore the influence of other variables relevant to the creative problem-solving milieu.

In studies on creativity and cognition Runco and co-workers (Runco and Chand, 1995; Runco, 1999a; 1999b) presented a "two tier model of creative thinking" in which the constructs of knowledge and motivation were placed on a second level while the process constructs of problem finding, ideation and evaluation were placed at a primary level. It is therefore interesting to note the large to medium effects of **Experience** and intrinsic motivation (**Motiv**) (presented at the second macroscopic level in Figure 4) on the process variables of creative problem solving (presented at the primary microscopic level in Figure 4) in this study. **Experience,** it may be noted, was a measure of past participation in

Mathematics Challenge events and prior practice with Challenge type problems. Such experience is consistent with the construct of knowledge. As with Runco's two tier model, **Experience** or knowledge in the current study was found to influence the ideation processes associated with the Feeling approaches to reasoning, while intrinsic motivation was found to influence both the ideation processes associated with the Free-flowing approaches to reasoning and the validation processes associated with Systematic approaches to reasoning.

Thus, in addition to validating the creative cycles hypothesised by Shaw (1989) and others (Lubart, 2001), the integrated model of creative problem solving examined here also supports the two tier representation of creative thinking advanced by Runco and co-workers (Runco & Chand, 1995; Runco & Dow, 1999). In particular, Runco theorised that knowledge and motivation were second tier variables because they contributed to, rather than controlled, the creative process, a view supported by the refined model of creative problem solving presented in this study. Thus the comprehensive model of creative problem solving presents implications for theory at both the microscopic and macroscopic levels. However, it remains for these implications to be translated into action through the development of sound practice.

Implications for practice

In Australia, student participation rates in senior school mathematics are declining (Brinkworth, 1999; Dekkers & De Laeter, 2001; Dekkers, DeLaeter & Malone, 1991) and student levels of achievement in mathematics at the lower secondary level have fallen over the past three decades (Afrassa & Keeves, 2001). Such outcomes are cause for concern, not only for Australia's future social and economic well being (Raison & Etheridge 2006) but also for individuals dealing with day to day numerical issues within the community.

Calls to incorporate more real world problem solving into the curriculum have increasingly been seen as the panacea for dealing with such issues. However, the findings of a TIMSS video study (Hollingsworth, Lokan & McCrae, 2003), for example, reveal students in Grade 8 Australian classrooms continuing to solving problems of low complexity with much repetition and little mathematical reasoning.

The reasons for this are too numerous to relate, but an issue of relevance to this research, may well be the lack of metacognitive activity by teachers and students, concerning the solving of novel mathematics problems. Real world problem solving entails creativity and the engagement of both its cognitive and non-cognitive aspects. It would seem inevitable, therefore that teaching novel mathematics problem solving, focusing solely on its cognitive aspects is misguided. Further, while current approaches to teaching and learning target some conscious forms of thinking, little account is taken of other possible approaches to reasoning or knowing and the non-conscious aspects of thinking.

Creativity, as this study has found, entails fostering the interactions between conscious and non-conscious aspects of reasoning and between visual-spatial and linguistic circuits of reasoning. It entails giving due attention to feeling that arises at the intersection of these interactions, and that heralds the arrival of a creative new idea. Feeling serves an evaluative function and is a source of direction in navigating a solution path through the problem solving space. Indeed, as this study has shown, attention to feeling is likely to be required for novel mathematics problem solving to be successful.

Psychological research in education circles has given rise to the view that ability in spatial reasoning is likely to be crucial for successful mathematics problem solving. However, while students with good two and three-dimensional spatial ability were successful in this study, this ability was not the sole defining feature. Successful students were also marked by attention to Feeling and Free-flowing approaches to reason. Given the finding in this study that successful outcomes were predicated on feeling, it is perhaps interesting to note the small to negligible effect in favour of girls for both the Cute Numbers and Birthday Cake problems in this study. This was despite the finding that boys outperformed the girls in tests of two and three-dimensional spatial ability. This highlights the need to draw student attention, particularly the attention of boys, to the role that feeling and intuition play in solving novel mathematics problems. Also of assistance would be the teaching of strategies to defocus attention and break a mental set in order to allow different conceptual connections to be made.

It should be pointed out that the feeling described in this study is likely to be indicative of some deep inner knowing arising from conscious and non-conscious processing, resulting from motivated attention to the novel problem. It may include an intimation or an inner feeling of connectedness in the aesthetic sense of comprehending a new intellectual order. This feeling may be indicated at every point in the creative process particularly those points bringing non-conscious information into conscious appraisal. Such a feeling should not be confused with feelings at another level of awareness typically referred to as emotion. Indeed, emotion can have both informational and processing influences on cognition in the sense that it can influence both what and how individuals think (Forgas, 2000).

Of relevance to this study, however, is the finding by Forgas (2000) that emotion is unlikely to influence cognitive processing when the individual is highly motivated. This indicates that intrinsic motivation, in the educational setting, is crucial in both initiating and maintaining interest in novel mathematics problem solving and in enabling the individual to sense the self when navigating a solution path through the problem space. This is even more so when it is realised that tasks requiring complex processing are more likely to be infused with affect (Forgas, 2000). While the infusion of positive affect into the processing strategies employed by the individual may not give cause for concern, the consequences of the infusion of negative affect remains in doubt (Kaufmann, 2003). Indeed it was for no small reason that Damasio (1994, p.247) made the statement "…educational systems might benefit from emphasizing unequivocal connections between current feelings and predicted future outcomes".

Implications for future research

The ability to recognise and respond to an inner sense or feeling, giving rise to a creative idea, this study has shown, is critical to successful novel mathematics problem solving. Poincaré is reported to have said that such feeling or intimation is in fact common to all mathematicians, but that its expression may arise to different degrees. It is surprising to learn, therefore, that research in this aspect of mathematics problem solving has drawn relatively little interest. It can only be speculated as to the outcomes that may have arisen should the process of intimation, in addition to that of illumination, have been formalised into the stage model of creative problem solving.

Indeed, the findings of this study surrounding the processes of feeling and non-conscious thinking warrant further investigation. In particular, the association of simultaneous processing with the appearance of a sudden illumination, but the association of sequential processing with a gradual illumination, is of particular interest. If research about creative problem solving is to take place for the benefit of future students then the processes surrounding the so-called 'moment of insight' require special attention. Such research also necessitates investigation into the role that states of defocused attention may have on the manifestation of feeling. This includes an analysis both of simultaneous and sequential, conscious and pre-conscious synthesis. This study has found feelings of intimation arising during oscillation between verbal and non-verbal reasoning and between conscious and non-conscious thinking. This indicates an important role for an oscillatory mechanism in the creative problem solving process, requiring further examination.

The link made in this study, that problems requiring spatial thinking were likely to involve simultaneous processing which, in turn, was associated with the appearance of a sudden illumination, requires further investigation. Similarly, the finding that problems involving verbal processing were associated with sequential synthesis and the appearance of a gradual illumination is also material for ongoing research. At the same time, research surrounding metacognitive instruction on distinguishing this kind of feeling from other types of affect must not be forgotten or overlooked. Research extending Poincaré's (Ghiselin, 1952) assertion that all mathematicians experience this aesthetic insight would be of assistance in this regard.

At the macroscopic level of analysis, further research is needed into expanding and testing the comprehensive model of creative problem solving developed in this study, to investigate the influence of a whole host of variables not begun to be examined in this study. These variables include more extensive testing of environmental factors, person related factors including aspects of the personality and product related factors as well as process related factors. Indeed, the potential is enormous and can only serve to enhance enlightenment as to the ongoing illusive nature of the creativity intersection. But before divining new paths for future creativity research both in the fields of neuroscience, education and psychology, it may be well to heed Damasio's advice and give due attention to the inner world of feeling in the hope that our sense of rationality may be increased.

Concluding comments

No curriculum for schools and universities is today complete without reference to the importance of problem solving and creativity in learning, yet the tasks involved in problem solving and being creative are not easily taught or learnt. Moreover, the processes involved are elusive and not clearly understood. This study has sought to unravel or tease out some relationships that occur in the minds of individuals and has shown the importance of feeling in successfully solving novel mathematics problems. It has provided empirical evidence of the feedback loops of creative problem solving theorised by Shaw (1989), added weight to Runco's 'Two Tier Model of Creative Thinking' (Runco & Shand, 1995; Runco, 1999a, 1999b) and successfully modelled, within the novel mathematics problem solving context, the interconnectedness of cognitive and non-cognitive systems of reasoning, so strongly advocated by Damasio (1994, 1999). It has also provided a structure for conceptualising creativity in such a way that the paradox of its multidimensional nature (first described by Mooney, 1963) and the specificity of its interactions (sought by Lubart & Sternberg, 1995) are reconciled in the mathematical formulation of the creativity intersection (theorised by Collins & Amabile, 1999). In so doing, it is hoped, that better understanding of these ideas will help advance both teaching and learning in the fields of problem solving and creativity.

Finally, where terms such as 'non-cognitive system of reasoning' may once have seemed confounding, new research (such as that presented here), new understanding and a redefinition in terms (such as that currently occurring in the arena of the new unconscious (Hassin, 2005), are likely to reveal more about the complex nature of processes at work within a novel problem-solving event, than previously realised. Such redefinition and understanding will likely highlight the unique role that feeling, free-flowing and intuitive processes play in cognition, including their importance in creatively solving novel problems, not merely within mathematics, but in other fields of endeavour. For in seeking to understand the nature of creativity and the origin of new ideas, one is seeking the "… key to education in its fullest sense and … the solution of mankind's most serious problems" (Guilford, 1967, p.13).

References

Cognitive And Non-Cognitive Systems Of Reasoning In The Solving Of Novel Mathematics Problems. Unpublished PhD Thesis The Flinders University Of South Australia, Adelaide.

Aldous, C. R. (2006). Attending To Feeling: Productive Benefit To Novel Mathematics Problem Solving. *International Education Journal*, 7(4), 410-422.

Afrassa, T. M., & Keeves, J. P. (2001). Change In Differences Between The Sexes In Mathematics Achievement In The Lower Secondary School Level In Australia Over Time. *International Education Journal,* 2(2), 96-108.

Arbuckle, J. L. (1999). *Amos 4.0.* Chicago: Smallwaters.

Brinkworth, P. (1999). An Exploration Of Gender Differences In Subject Choice And Beliefs Among Year 12 Students. Paper Presented At The Proceedings Of The Twenty-Second Annual Conference Of The Mathematics Education Research Group Of Australasia Incorporated:*Making The Difference*, Adelaide, South Australia.

Damasio, A. R. (1994). *Descartes' Error: Emotion Reason And The Human Brain.* London: Papermac.

Damasio, A. R. (1999). *The Feeling Of What Happens*. London: William Heinemann.

Carroll, J. B. (1963). A Model Of School Learning. *Teachers College Record*, 64, 723-733.

Collins, M. A., & Amabile, T. M. (1999). Motivation And Creativity. In R. J. Sternberg (Ed.), *Handbook Of Creativity* (Pp. 297-312). Cambridge Uk: Cambridge University Press

Das, J. P., Kirby, J. R., & Jarman, R. F. (1979). *Simultaneous And Successive Cognitive Processes*. New York: Academic Press.

Das, J. P. (2003). A Look At Intelligence As Cognitive Neuropsychological Processes:Is Luria Still Relevant? *Japanese Journal of Special Education*, 40(6), 631-647.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources Of Mathematical Thinking :Behavioural And Brain-Imaging Evidence. *Science Magazine*, 284, 970-977.

Dekkers, J., & De Laeter, J. R. (2001). Enrolement Trends In School Science Education In Australia. *International Journal Of Science Education, 23*(5), 487-500.

Dekkers, J., Delaeter, J. R., & Malone, J. A. (1991). *Upper Secondary School Science And Mathematics Enrolment Patterns In Australia. 1970-1988*. Perth: Western Australian Institute Of Technology.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports As Data Revised Edition*. London, England: The Massachusetts Institute Of Technology.

Epstein, S. (1994). Integration Of The Cognitive And The Psychodynamic Unconscious. *American Psychologist*, 49(8), 709-724.

Epstein, S. (1998). *Constructive Thinking The Key To Emotional Intelligence*. Westport, Connecticut: Praeger.

Forgas, J. P. (2000). Affect And Information Processing Strategies. In J. P. Forgas (Ed.), *Feeling And Thinking: The Role Of Affect In Social Cognition* (Pp. 253-280). Cambridge: Cambridge University Press

Ghiselin, B. (1963). The Creative Process And Its Relation To The Identification Of Creative Talent. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition And Development.* New York: John Wiley &Sons.

Guilford, J. P. (1967). Creativity: Yesterday, Today And Tomorrow. *The Journal of Creative Behavior,* 1(1), 3-14.

Gustafsson, J. E., & Stahl, P. A. (2000). *Streams User's Guide: Version 2.5 For Windows*. Molndal, Sweden: Multivariate Ware.

Hassin, R. R. (2005). Nonconscious Control And Implicit Working Memory. In R. R. Hassin & J. S. Uleman & J. A. Bargh (Eds.), *The New Unconscious* (Pp. 196-222). Oxford: Oxford University Press.

Hollingsworth, H., Lokan, J., & Mccrae, B. (2003). *Teaching Mathematics In Australia: Results From The Timss 1999 Video Study*. Camberwell, Victoria: Australian Council For Educational Research.

Immordino-Yang, M. H., & Damasio, A. (2007). We Feel, Therefore We Learn: The Relevance Of Affective And Social Neuroscience To Education. *Mind Brain And Education*, 1(1), 3-10.

Kaufmann, G. (2003). Expanding The Mood-Creativity Equation. Creativity Research Journal, 15(2&3), 131-135.

Keeves, J. P. (1986). The Performance Cycle. *International Journal Of Education Research,* 10, 143-158.

Lubart, T. (2001). Models Of The Creative Process:Past Present And Future. *Creativity Research Journal*, 13(3&4), 295-308.

Lubart, T., & Sternberg, R. J. (1995). An Investment Approach To Creativity: Theory And Data. In S. M. Smith & T. B. Ward & R. A. Finke (Eds.), *The Creative Cognition Approach* (Pp. 269-302). Cambridge, Ma: Mit Press.

Mooney, R. L. (1963). A Conceptual Model For Integrating Four Approaches To The Identification Of Creative Talent. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition And Development*. New York: John Wiley & Sons.

Mumford, M. D. (2003a). Where Have We Been, Where Are We Going? Taking Stock In Creativity Research. *Creativity Research Journal*, 15(2 & 3), 107-120.

Mumford, M. D. (2003b). Taking Stock In Taking Stock. *Creativity Research Journal*, 15(2&3), 147-151.

Raison, M., & Etheridge, M. (2006). *Macquarie University Science, Engineering And Technology Study*. Melbourne: Macquarie University.

Richards, R. (1999). Four Ps Of Creativity. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia Of Creativity* (Vol. 1, Pp. 733-742). San Diego: Academic Press

Runco, M. A. (2003). Commentary On Personal And Potentially Ambiguous Creativity: You Can't Understand The Butterfly Unless You Also Watch The Caterpillar. *Creativity Research Journal,* 15(2&3), 107-120.

Runco, M. A. (1999a). Critical Thinking. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia Of Creativity* (Vol. 1, Pp. 449-452). San Diego: Academic Press.

Runco, M. A. (1999b). Divergent Thinking. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia Of Creativity (*Vol. 1, Pp. 577-582). San Diego: Academic Press.

Runco, M. A., & Chand, I. (1995). Cognition And Creativity. *Educational Psychology. Review,* 7(3), 243-267

Runco, M. A., & Dow, G. (1999). Problem Finding. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia Of Creativity* (Vol. 2, Pp. 433-436). London: Academic Press

Salk, J. (1985). *The Anatomy Of Reality*. New York: Praeger.

Sellin, N., & Keeves, J. P. (1997). Path Analysis With Latent Variables. In J. P. Keeves (Ed.), *Educational Research, Methodology, And Measurement: An International Handbook* (Pp. 633-640). Oxford: Pergamon Press.

Shaw, M. P. (1989). The Eureka Process: A Structure For The Creative Experience In Science And Engineering. *Creativity Research Journal*, 2, 286-298.

Sloman, S. A. (1996). The Empirical Case For Two Systems Of Reasoning. *Psychological Bulletin*, 119(1), 3-22.

Sternberg, R. J., & O'hara, L. A. (1999). Creativity And Intelligence. In S. R. J.Sternberg (Ed.), *Handbook Of Creativity*. (Pp. 251-272). New York,.: Cambridge University Press.

Tardif, T. Z., & Sternberg, R. J. (1988). What Do We Know About Creativity? In R. J. Sternberg (Ed.), *The Nature Of Creativity: Contemporary Psychological Perspectives* (Pp. 429 - 440). New York: Cambridge University Press.

Appendix 1: Mathematics Challenge for Young Australians: Questions Selected for Response in the Main Study

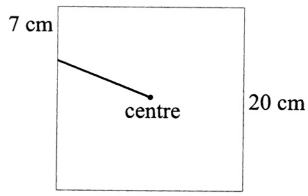
A: Junior and Intermediate levels

Question: Birthday Cake *(Students answered the SRQ at the completion of Part **a**.)*

Ravina has a square birthday cake. Its side length is 20 cm as shown. She wishes to cut the cake into five pieces, one for each of her four friends and one for herself. She wants to use straight vertical cuts to make five pieces of equal volume.

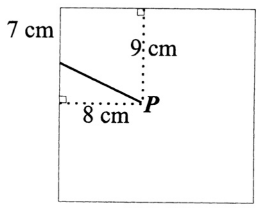
a. Suppose that Ravina makes the first cut from the cake's center to a point 7 cm from the top left corner as shown.

Where must she make the other four cuts if they all start from the cake's centre?



b. Suppose that Ravina makes the first cut from the point P, 8 cm and 9 cm from the nearest edges, to a point 7cm from the top left corner as shown.

Where must she make the other four cuts if they all start from P?



c. Liquorice of length 80 cm decorates the top edges of the cake. If Ravina gets more liquorice than any of her four friends, and the cake is cut in one of the ways described in **a** and **b**, what length of liquorice does she get?

Marking Scheme

a. The correct answers with reasoning - **1 mark**

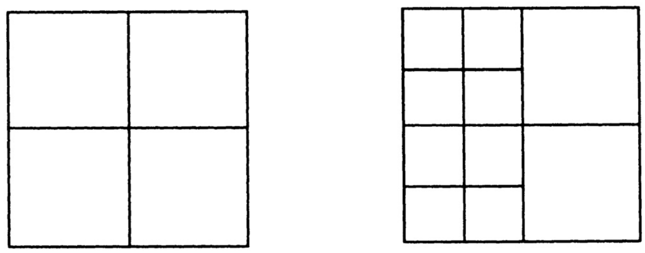
b. Appropriate method carried through - **1 mark**

The correct answers with reasoning - **1 mark**

c. The correct answer (18 5/9) with reasoning - **1 mark**

Question: Cute Numbers
(Students answered the SRQ at the completion of Part a.)

If a square can be cut into n squares of at most two different sizes, then n is called a cute number. For example, 4 and 10 are cute numbers.



a. Show that 6 is a cute number
b. Show that 2001 is a cute number
c. Show that every integer greater than 5 is a cute number

Marking Scheme

a. A correct answer - **1 mark**
b. A correct answer - **1 mark**
c. Any one correct infinite set (e.g., all even numbers > 6) – **1 mark**
The correct answer with reasoning – **1 mark**

Problems reproduced with permission from: Australian Mathematics Trust (2001) Mathematics Challenge for Young Australians: Teacher's reference book for primary, junior & intermediate maths challenge stages. Australian Mathematics Trust, University of Canberra, Canberra ACT

Appendix 2: Table showing direct path coefficients for the Cute Numbers PLS Path model

		Direct Path Coefficients for the Cute Numbers Model (Mean R-square = 0.23)																
Latent		**Verbal**	**Space2D**	**Space3D**	**Priprac**	**Partic**	**Aids**	**Help**	**Play**	**Motiv**	**Novel**	**L-flow**	**Strat**	**Free**	**Spat/Vb**	**Feel**	**Syst**	**Pscore**
variables		x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	y
Gender	x_1	*	0.34	?	*	?	-0.16	0.22	*	~	~	~	-0.22	~	-0.14	~	0.11	-0.15
Age	x_2	0.34	0.20	0.14	*	0.28	-0.20	~	-0.20	-0.10	~	~	~	~	~	0.17	~	0.20
Stype	x_3	*	*	-0.13	0.24	?	*	*	-0.24	~	~	~	~	~	-0.11	0.12	~	0.19
Verbal	x_4		*	0.20	*	?	*	*	*	*	*	*	*	*	?	?	-0.12	?
Space2D	x_5			0.28	*	*	0.16	?	*	-0.12	-0.15	-0.10	~	*	~	~	~	0.19
Space3D	x_6				*	0.11	~	*	*	*	~	*	*	*	0.10	0.20	-0.13	0.13
Priprac	x_7					0.20	~	~	*	*	-0.17	*	~	~	?	*	*	-0.10
Partic	x_8						-0.13	~	*	*	-0.10	*	~	~	*	?	~	0.13
Aids	x_9							0.31	*	*	*	?	0.13	*	?	-0.19	0.18	~
Help	x_{10}								*	*	0.12	*	0.14	*	?	?	0.10	*
Play	x_{11}									0.59	*	?	?	?	?	?	?	0.14
Motiv	x_{12}										*	0.61	?	0.26	0.13	?	0.21	?
Novel	x_{13}											*	0.18	?	-0.13	-0.10	~	*
L-flow	x_{14}												*	0.13	?	0.29	?	?
Strat	x_{15}													0.29	0.29	?	?	?
Free	x_{16}														0.33	0.11	0.22	0.18
Spat/Vb	x_{17}															0.31	0.33	-0.12
Feel	x_{18}																*	0.11
Syst	x_{19}																	*
R-square		0.11	0.17	0.23	0.06	0.15	0.10	0.13	0.09	0.39	0.09	0.39	0.25	0.27	0.37	0.39	0.49	0.23
Disturbance		0.94	0.90	0.87	0.97	0.92	0.95	0.93	0.95	0.77	0.95	0.77	0.86	0.85	0.79	0.78	0.71	0.87

Note: * indicates a nonsignificant path
? indicates the presence of a positive indirect effect
~ indicates the presence of a negative indirect effect, girls scored 0, boys 1, single sex schools 0, co-educational schools 1

Labels: **Gender** (sex), **Age**(age &grade), **Stype** (school type), **Verbal** (verbal ability), **Space2D** (2-dimensional spatial ability). **Space3D**(3-dimensional spatial ability), **Priprac**(prior practice), **Partic** (previous participation), **Aids** (use of resources), **Help** (discussion with teacher), **Play**(plays maths games and puzzles), **Motiv**(intrinsic motivation),**Novel** (rating of problem novelty), **L-flow**(flow), **Strat** (Strategic approachto reasoning), **Free** (Free-flowing approach to reasoning), **Spat/Vb** (Spatial-verbal approach to reasoning), **Feel** (Feeling approach to reasoning), **Syst** (Systematic approach to reasoning), **Pscore**(Problem score)

Table showing direct path coefficients for the Birthday Cake PLS Path model

		Direct Path Coefficients for the Birthday Cake Model (Mean R-square = 0.20)																
Latent		**Verbal**	**Space2D**	**Space3D**	**Priprac**	**Partic**	**Aids**	**Help**	**Play**	**Motiv**	**Novel**	**L-flow**	**Strat**	**Free**	**Spat/Vb**	**Feel**	**Syst**	**Pscore**
variables		x_{4}	x_{5}	x_{6}	x_{7}	x_{8}	x_{9}	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	y
Gender	x_{1}	*	0.38	?	*	?	-0.16	0.15	*	*	*	*	-0.20	~	~	~	~	?
Age	x_{2}	0.35	0.21	0.22	*	0.23	*	-0.11	-0.20	~	~	~	~	~	0.13	0.25	0.17	?
Stype	x_{3}	*	*	-0.13	0.23	?	*	*	-0.21	-0.11	~	~	?	~	~	~	~	?
Verbal	x_{4}		*	0.21	*	?	*	*	*	*	-0.15	*	~	~	~	~	~	?
Space2D	x_{5}			0.25	*	*	*	*	* *		*	*	~	~	~	~	~	?
Space3D	x_{6}				*	0.14	*	*	*	*	*	*	-0.21	-0.17	~	~	~	0.22
Priprac	x_{7}					0.20	*	*	*	*	*	*	*	*	-0.14	~	~	-0.11
Partic	x_{8}						*	*	*	*	*	*	*	*	*	*	*	0.17
Aids	x_{9}							0.18	*	*	*	*	0.18	*	?	?	0.13	*
Help	x_{10}								*	*	*	*	0.14	*	?	*	*	*
Play	x_{11}									0.55	?	?	?	?	0.13	?	?	?
Motiv	x_{12}										0.12	0.62	?	0.16	0.14	0.12	0.38	0.13
Novel	x_{13}											*	0.28	?	?	?	?	~
L-flow	x_{14}												*	0.18	?	?	?	*
Strat	x_{15}													0.21	0.32	0.12	?	-0.24
Free	x_{16}														0.26	0.11	?	?
Spat/Vb	x_{17}															0.30	0.21	?
Feel	x_{18}																0.13	?
Syst	x_{19}																	0.26
R-square		0.12	0.21	0.26	0.05	0.14	0.03	0.06	0.08	0.34	0.09	0.38	0.23	0.21	0.30	0.25	0.33	0.30
Disturbance		0.93	0.88	0.86	0.97	0.92	0.98	0.97	0.96	0.81	0.97	0.78	0.87	0.88	0.83	0.86	0.81	0.83

Note: * indicates a nonsignificant path
? indicates the presence of a positive indirect effect
~ indicates the presence of a negative indirect effect, girls scored 0, boys 1, single sex schools 0, co-educational schools 1

Labels: **Gender** (sex), **Age**(age &grade), **Stype** (school type), **Verbal** (verbal ability), **Space2D** (2-dimensional spatial ability). **Space3D**(3-dimensional spatial ability), **Priprac**(prior practice), **Partic** (previous participation), **Aids** (use of resources), **Help** (discussion with teacher), **Play**(plays maths games and puzzles), **Motiv**(intrinsic motivation),**Novel** (rating of problem novelty), **L-flow**(flow), **Strat** (Strategic approachto reasoning), **Free** (Free-flowing approach to reasoning), **Spat/Vb** (Spatial-verbal approach to reasoning), **Feel** (Feeling approach to reasoning), **Syst** (Systematic approach to reasoning), **Pscore**(Problem score)

Appendix 3: Table of AMOS inner path estimates for the Cute Numbers and Birthday Cake data

Standardised and unstandardised results are shown. Between-model comparisons are made by applying a t–test to unstandardised estimates.

The Structural Model

			Cute Numbers N= 387					Birthday Cake N= 360					
			Unstandardised				Standardised	Unstandardised				Standardised	t-value
Structural Path			Estimate	S.E.	C.R.	P	Estimate	Estimate	S.E.	C.R.	P	Estimate	Cutenos-Bcake
Space	<--	Age	6.45	1.041	6.2	0.00	0.407	0.706	0.103	6.8	0.00	0.487	5.49 *
Space	<--	Gender	12.51	2.102	6.0	0.00	0.392	1.167	0.204	5.7	0.00	0.401	5.37 *
Space	<--	Stype	-3.074	1.947	-1.6	0.11	-0.095	-0.559	0.179	-3.1	0.00	-0.190	-1.29
Play	<--	Space	0.006	0.004	1.6	0.11	0.130	0.049	0.046	1.1	0.29	0.099	-0.93
Play	<--	Stype	-0.401	0.084	-4.8	0.00	-0.268	-0.333	0.09	-3.7	0.00	-0.232	-0.55
Play	<--	Age	-0.200	0.048	-4.1	0.00	-0.274	-0.185	0.055	-3.4	0.00	-0.261	-0.21
Experience	<--	Age	0.128	0.022	5.7	0.00	0.596	0.140	0.024	5.9	0.00	0.513	-0.37
Experience	<--	Stype	0.144	0.039	3.7	0.00	0.327	0.079	0.044	1.8	0.07	0.142	1.11
Motiv	<--	Play	0.915	0.075	12.2	0.00	0.753	0.868	0.079	11.0	0.00	0.714	0.43
Motiv	<--	Gender	0.166	0.093	1.8	0.08	0.093	0.187	0.098	1.9	0.06	0.108	-0.16
Motiv	<--	Stype	-0.009	0.085	-0.1	0.91	-0.005	-0.183	0.089	-2.1	0.04	-0.104	1.41
Motiv	<--	Space	-0.014	0.004	-3.6	0.00	-0.242	-0.085	0.042	-2.0	0.05	-0.144	1.68
Strat	<--	Gender	-0.162	0.076	-2.1	0.03	-0.156	-0.126	0.081	-1.5	0.12	-0.117	-0.32
Strat	<--	Space	-0.006	0.003	-1.9	0.06	-0.173	-0.135	0.037	-3.7	0.00	-0.367	3.48 *
Free	<--	Strat	0.449	0.083	5.4	0.00	0.685	0.371	0.078	4.7	0.00	0.608	0.68
Free	<--	Motiv	0.182	0.033	5.5	0.00	0.477	0.178	0.037	4.8	0.00	0.470	0.08

* indicates t-values > 1.96 are considered significant at the five percent level

The Structural Model (continued)

Structural Path			Cute Numbers N= 387					Birthday Cake N= 360					
			Unstandardised				Standardised	Unstandardised				Standardised	t-value
			Estimate	S.E.	C.R.	P	Estimate	Estimate	S.E.	C.R.	P	Estimate	Cutenos-Bcake
Spat/Vb	<--	Free	1.002	0.196	5.103	0.00	0.889	1.018	0.228	4.469	0.00	0.825	-0.05
Spat/Vb	<--	Gender	-0.149	0.053	-2.817	0.01	-0.195	-0.165	0.065	-2.555	0.01	-0.204	0.19
Spat/Vb	<--	Space	0.005	0.002	2.242	0.03	0.205	0.105	0.032	3.318	0.00	0.379	-3.12 *
Feel	<--	Spat/Vb	0.643	0.137	4.694	0.00	0.716	0.518	0.121	4.285	0.00	0.543	0.68
Feel	<--	Experience	0.693	0.199	3.476	0.00	0.434	0.471	0.185	2.546	0.01	0.336	0.82
Feel	<--	Space	0.006	0.002	2.949	0.00	0.274	0.039	0.023	1.697	0.09	0.147	-1.43
Feel	<--	Motiv	0.067	0.030	2.231	0.03	0.173	0.032	0.035	0.929	0.35	0.072	0.76
Syst	<--	Feel	0.401	0.137	2.921	0.00	0.308	0.588	0.126	4.655	0.00	0.505	-1.00
Syst	<--	Motiv	0.194	0.040	4.814	0.00	0.387	0.250	0.039	6.414	0.00	0.481	-1.00
Syst	<--	Age	-0.090	0.032	-2.836	0.01	-0.201	0.049	0.028	1.749	0.08	0.110	-3.27 *
Syst	<--	Gender	0.088	0.049	1.794	0.07	0.098	-0.163	0.052	-3.159	0.00	-0.181	3.51 *
Syst	<--	Strat	0.296	0.076	3.870	0.00	0.343	0.002	0.056	0.041	0.97	0.003	3.11 *
Score	<--	Syst	-0.033	0.236	-0.141	0.89	-0.012	1.407	0.295	4.771	0.00	0.430	-3.81 *
Score	<--	Feel	0.933	0.363	2.570	0.01	0.263	-0.153	0.369	-0.414	0.68	-0.040	2.10 *
Score	<--	Gender	-0.559	0.166	-3.362	0.00	-0.228	-0.096	0.180	-0.534	0.59	-0.033	-1.89
Score	<--	Space	0.027	0.008	3.371	0.00	0.352	0.3	0.095	3.168	0.00	0.297	-2.86 *
Score	<--	Strat	-0.217	0.179	-1.209	0.23	-0.092	-0.713	0.199	-3.584	0.00	-0.260	1.85

* indicates t-values > 1.96 which are considered significant at the five percent level

22

The use of partial least squares (PLS) path modeling analysis in a cross-cultural investigation into the relationships of values, classroom cultural climate, and learning approach

Gregory Arief D. Liem
Faculty of Education and Social Work, University of Sydney

> As always, all stories end somewhere. This particular one ended on the Isle of Spetses in Greece where I received the Harry and Pola Triandis Doctoral Thesis Award. Looking back at the many hours and sleepless nights I had gone through in putting the thesis together, I realised that I could not have done this Sisyphean task without the selfless help from John Keeves in Australia for his valuable help in the data analysis. (Liem, 2007, p. 35)

Introduction

The purpose of this chapter is to report a cross-cultural investigation into the relationships of values, the cultural climate of the classroom, approaches to learning, and perceived academic performance among a representative sample of secondary school students in Australia, Singapore, the Philippines, and Indonesia. The chapter especially aims to demonstrate the use of partial least squares (PLS) path modeling analysis that was recommended by Professor John Keeves in September 2004 and whose assistance in the actual process of the analysis was sincerely and enthusiastically rendered during the course of my doctoral candidature at the National University of Singapore. I begin the introduction of this chapter with a review of the Student Approaches to Learning paradigm, which underpins the research reported here, and end this section by articulating the purposes of the study and a research question. This is followed by the description of the method, with special foci given to the description of PLS path modeling analysis, the presentation of the results, and a concluding remark.

Student approaches to learning theory

The Student Approaches to Learning (SAL) theory (Biggs, 1993; Entwistle & Ramsden, 1983) is a learning paradigm that takes seriously the influence of contextual factors (sociocultural and educational contexts) on motivation and learning. This paradigm is in agreement with cultural and indigenous viewpoints which maintains that culture and mind are inseparable (Cole, 1990), or as explicitly stated by Säljö (1991, p. 184), "… human experiences are inescapably cultural in nature and learning takes place within cultural boundaries". The SAL perspective is supported by situated cognition theorists (e.g., Brown, Collin & Duguid, 1989) who have even gone further by stressing the importance of cultural beliefs, values, assumptions, and social expectations in shaping what are considered culturally determined cognitive processes, attitudes, and behaviours.

The emergence of the SAL theory can be traced back as early as 1976 in the work of Martön & Saljö. In their experiment, a group of Swedish tertiary students were asked to read an academic article and to explain what they had learned and how they had achieved that learning. Two groups of students were identified. The first learned by memorising words or sentences, and thus simply reproduced what the author said. The second, however, tried to understand the message and focused on what the author meant. The way the first group learned was designated a *surface approach to learning*, whereas the second was called a *deep approach to learning*.

Biggs (1987) in Australia and Entwistle and Ramsden (1983) in the United Kingdom pioneered the quantitative research approach to the SAL theory and linked the results to the findings discovered by the researchers using a qualitative approach. Biggs and Entwistle and Ramsden have gone beyond describing "why" and "how" of learning phenomena by measuring the *degree* of students' agreement with the approaches to learning using inventories administered to students. These inventories are the *Learning Process Questionnaire* or LPQ (Biggs, 1987) for use with secondary school students, and the *Study Process Questionnaire* or SPQ (Biggs, 1987) and the *Approaches to*

Studying Inventory or ASI (Entwistle & Ramsden, 1983) for use with tertiary students.

Biggs (1987, 1993) has postulated that the process of student learning is based on motive and strategy components. Learning motive refers to why students undertake learning. For example, a student may see learning as a means of obtaining high grades. Another may learn in order to be able to pass an examination, whereas others are motivated to find out more about a particular topic due to personal interest. A particular learning motive would determine what learning strategy is adopted by a student in handling a learning task. The combination of a learning motive and its corresponding learning strategy is what has been designated as *approach to learning* (Biggs, 1987, 1993), the term that was first coined by Martön and Saljö (1976).

In his earlier studies, Biggs (1973) conducted a series of factor analyses of secondary and tertiary samples' responses to an earlier version of his inventory, the Study Behaviour Questionnaire – SBQ (Biggs, 1970). The factorial solutions of the SBQ had consistently demonstrated three dimensions of study processes designated as *surface, deep* and *achieving.* Using the ASI, Entwistle and Ramsden (1983) also identified very similar dimensions to those found by Biggs (1973). As a result, three types of approaches to learning are postulated in the SAL theory (see Table 1), each of which has its own motive and strategy component.

Table 1. Three types of approaches to learning in the SAL theory (Biggs, 1987)

Type	Description
Surface Approach	
Surface Motive	Motivation is utilitarian; the main aim is to gain qualification or avoid failure at minimum allowable standard – extrinsic motive.
Surface Strategy	Learning strategy used is aimed to reproduce bare essential; often using rote learning.
Deep Approach	
Deep Motive	Motivation is interest in and curiosity about the Subject being studies and its related areas – intrinsic motive.
Deep Strategy	Learning strategy used is to understand what is to be learnt through interrelating ideas and reading widely.
Achieving Approach	
Achieving Motive	Motivation is to obtain the highest possible grades as expressed in competition with other students.
Achieving Strategy	Learning strategy used is aimed to achieve high marks by being an "ideal role model" student, eg being punctual, doing extra readings, or whatever else that is instructed by the teacher.

3P model of classroom learning

In line with the emphasis in the SAL theory on the role of contextual factors on student motivation and learning, Biggs (1987, 1993) adapted Dunkin and Biddle's (1974) "Presage-Process-Product" model from the context of classroom teaching to explain student learning in the classroom setting. The so-called 3P model of classroom learning, shown in Figure 1, delineates the relationships involving students' characteristics, teaching context, students' learning processes, and learning outcomes. The ways students approach their learning, namely surface, deep, and achieving, constitute the *process* component, whereas conceptions of learning are part of the *student-presage* component of the model.

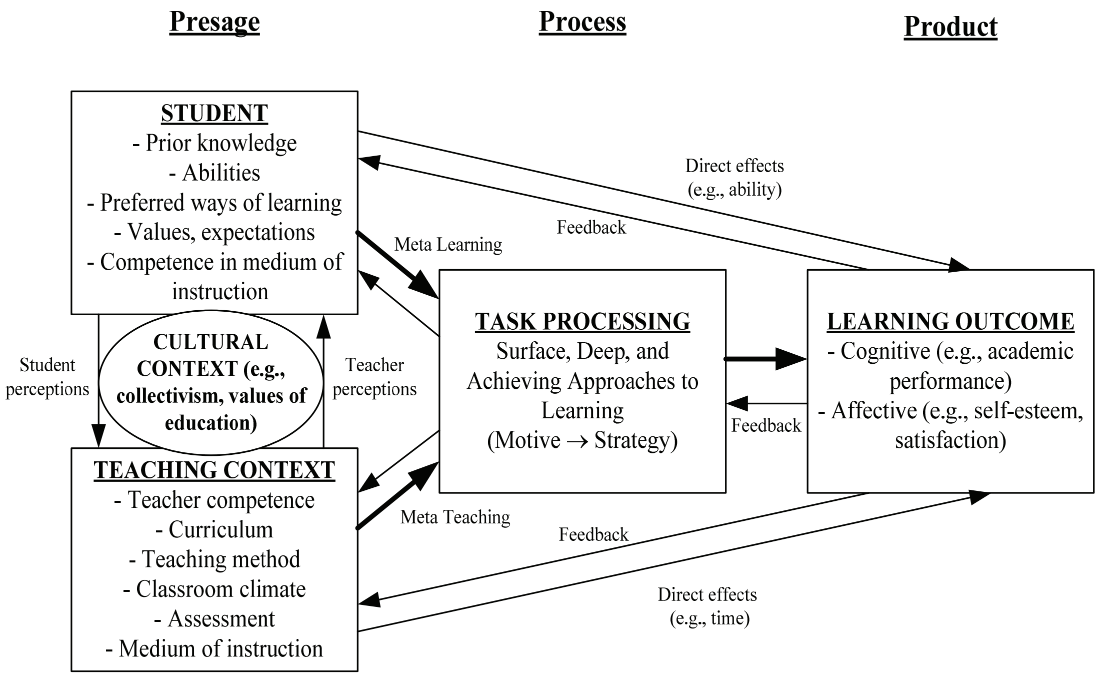


Figure 1. Culturally-Modified 3P Model of Classroom Learning (adapted from Biggs, 1996)

As shown in Figure 1, the 3P model of classroom learning consists of three major components, namely the presage, the process, and the product. The "Presage" component exists prior to learning and consists of two aspects: (a) student-presage factors that are relatively stable learning-related characteristics. This domain includes students' prior knowledge and experience relevant to the task, students' adherence to values and expectations concerning achievement, abilities, and approaches to learning as a *predisposition* to engage in academic activities (learning motive, learning strategy, and conception of learning), and competence in the medium of instruction (Biggs, 1987, 1993); (b) the teaching-presage factor that refers to the context and superstructure set by the teacher and the institution. Factors personal to the teacher would include teaching competency and teaching style, conceptions of teaching and learning, and the classroom climate established. Institutional factors would include the course structure, curriculum content, methods of teaching and types of assessment (Biggs, 1993).

The "Process" component refers to the way students actually handle, engage, or approach a learning task, which is determined by their perceptions of the teaching context, their motives and predisposition, and their decisions for immediate action. Their approach may be of low or high cognitive level. The low-level cognitive process, or surface approach, results in correspondingly low-level outcomes; whereas the high-level cognitive process, or deep approach, results in high-level outcomes (Biggs, 1987, 1993). Finally, the "Product" of learning may correspondingly be of a low or high cognitive level. Low-level outcomes emphasise quantitative recall, while high-level outcomes focus on quality, in addition to being correct, well structured, abstract-conceptualizing, relevant, and even elegant (Biggs, 1996). Biggs (1993) categorised the product of learning into two categories. The first is cognitive outcomes, that is, how much is learnt and how well it is learned, which is commonly reflected in grades awarded by the school (institutional outcomes). The second is affective outcome that has to do with how students feel about their learning, such as self-concept.

The model describes a linear progression (shown by thick arrows in Figure 1) from presage to process to product, in which each component *interacts* with all other components (indicated by double arrows pointing from and to each component in the model) forming an integrated system operating in a state of equilibrium (Biggs, 1987, 1993). As such, the system generates predictions of the degree to which each component is able to create a new equilibrium and hence a new system, or the changed component will be absorbed and the system reverts to the *status quo*. A steady state is the norm; when a system is not in balance, something is amiss and change may be expected (Biggs, 1987, 1993).

The equilibrium principle of the model must be highlighted because it explains and predicts how students approach their learning. That is, the ways students go about their learning will derive from the teaching/learning environment in which students are placed. For example, when the classroom or learning context is conducive to deep approaches, the students will tend to approach their learning deeply and in turn produce complex learning outcomes. In contrast, when the learning context is more restrictive, and students perceive that surface approaches will be sufficient, they tend to adapt to it and learning outcomes are poorly structured. In both cases, the 3P system is in equilibrium.

Studies have shown that the 3P model of classroom learning (Biggs, 1987, 1993) was not fully applicable for explaining learning phenomena in Confucian-oriented educational contexts such as those in China, Hong Kong, Korea, Singapore, and Taiwan (see Watkins & Biggs 1996, 2001 for a fuller discussion). Therefore, Biggs (1996) re-conceptualised the 3P model by taking cultural factors into account. In the model, Biggs (1996) postulated that the influences of cultural factors, individual differences, and teaching context are so closely entwined that it may not be possible to separate these components. In other words, classroom processes are complex and operate at many different levels, from that of individual learners and teachers, to the class, the school, the school system, the particular society, the wider culture, and the global level.

The present study

The aim of the study reported in this chapter is to examine the relationship between students' adherence to *values* (as a representation of variables in the student-presage component of the 3P model) and their adoption of *approaches to learning* (the process component). To include the teaching-presage and product components of the 3P model (Biggs, 1987, 1993, 1996), I also measured *classroom cultural climate* and *academic performance* - both are as perceived by students. To address the effect of *cultural context* emphasised in the culturally-modified 3P model (Biggs, 1996), this study drew samples from different cultural backgrounds. This objective was to obtain a more homogenous group of students who had been enculturated, socialised, and taught in the same sociocultural and educational contexts.

Two countries, namely Singapore and Australia, were selected. Singapore represents a broadly *Confucian-oriented* society, whereas Australia represents a *Western, non-Confucian-oriented* tradition. This design allowed a cross-cultural comparison in the relationships between the 3P-model components. In addition to the Singaporean and Australian samples, this study also included samples from two *Asian, non-Confucian-oriented* countries, namely Indonesia and the Philippines. The interest in understanding Indonesian and Filipino students' approaches to learning was especially driven by the fact that

Indonesian and Filipino students have consistently demonstrated poorer academic performances than Singaporean and Australian students (e.g., International Association for the Evaluation of Educational Achievement [IEA], 2004). This is an intriguing phenomenon that seeks for explanations.

Based on the conceptualisation above, this chapter reports a study that addresses the following research question: What are the cross-cultural similarities and differences in the relationships between the adherence to values, classroom cultural climate, approaches to learning, and the perceived academic performance across Singaporean, Indonesian, Filipino, and Australian students? Figure 2 shows a hypothesised model to be tested in the present study by using partial least squares path modeling analysis (cf. the 3P Model of Classroom learning by Biggs, 1996, shown in Figure 1).

Unlike Biggs's (1996) 3P model, inter-construct/variable relationships in the hypothesised model are depicted by single-headed arrows, respectively. It must be remembered that, when adapting the 3P model of classroom teaching (Dunkin & Biddle, 1974) to explain student learning in the classroom context, Biggs (1987, 1993) perhaps did not consider testing the model statistically. In fact, inter-variable relationships in Biggs's (1987) earlier 3P model were represented by single arrows, each of which pointed from the more left components to the more right components. So, in the present study, students' adherence to values is depicted as a predictor of their attitudes toward classroom interpersonal relations (as conceptualised in the cultural factors of the classroom learning environment), their approaches to learning, and their perceived academic performance. The cultural factors of the classroom learning environment, in turn, are posited to predict the adoption of approaches to learning and the perceived academic performance. Lastly, students' approaches to learning become a proximal predictor of their perceived academic performance. The summary model in Figure 2 is in a way a simplified representation of the 3P model (cf. Figure 1). This is so because each of the student-presage, teaching-presage, process, and product components of the 3P model (Biggs, 1996) is only represented by a single construct, namely students' adherence to values, the cultural climate of the classroom learning environment, approaches to learning, and the perceived or self-rated academic performance, respectively.

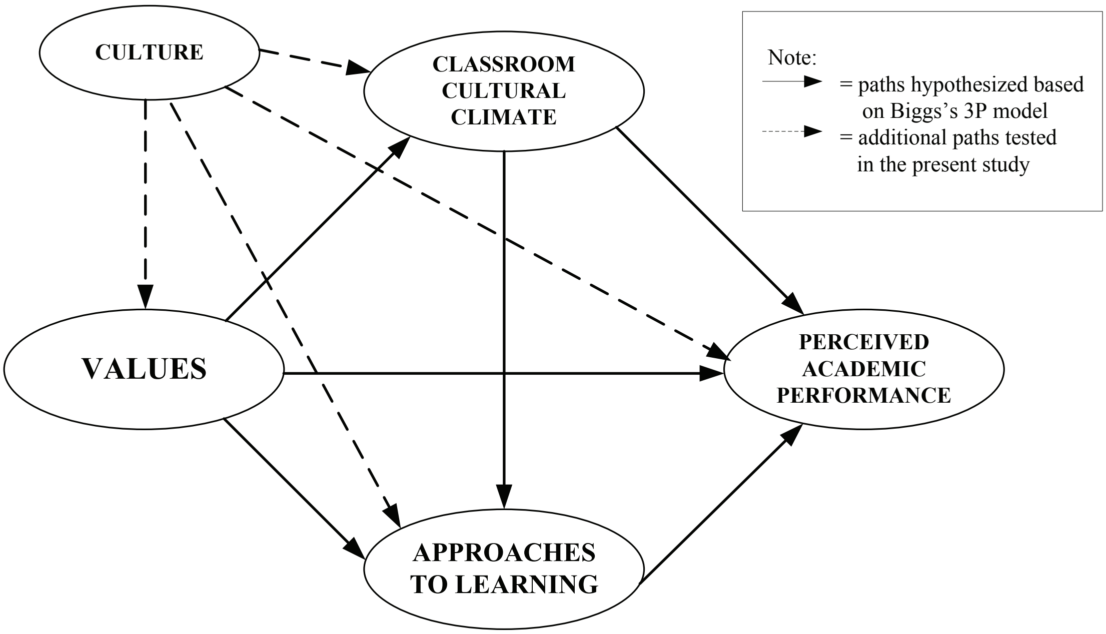


Figure 2. A model summarising the relations between values, cultural factors of the classroom learning environment, approaches to learning, and perceived academic performance

Method

Participants

Four broadly-matched samples of Year 10 students were drawn from co-educational public schools in Sydney, Singapore, Manila, and Jakarta. Each cultural sample comprised 230 students, with boys and girls equally represented (n=115). The average age was comparable across samples at $\underline{+}16$ years. Participants included in this study were born or had lived in their respective countries for at least 10 years, and they were all citizens of the country where they lived.

Measures

A set of questionnaires consisting of two parts was administered to participants in the four countries. In the first part, demographic information was gathered, including name (optional), gender, age, ethnic background, nationality, religion, mother tongue, socio-economic status (SES), and father's and mother's educational backgrounds. The second part comprised pre-existing self-report instruments to measure values, classroom cultural climate, and approaches to learning, respectively. Each of these instruments is described below.

Values. The Portrait Values Questionnaire (PVQ; Schwartz, 2005) was used to measure different orientations of basic human values postulated by Schwartz (2005). The PVQ comprises 40 items measuring 5 broad value orientations: Conservation (13 items), Self-Transcendence (10 items), Openness to Change (7 items), Self-Enhancement (7 items), and Hedonism (3 items). Each of the 40 items consisted of a two-sentence short verbal portrayal of a person's goals, aspirations, ambitions, wishes, or desires that refer implicitly to the importance of a single value type. For example, *"Being very successful is important to him. He likes to impress other people."* is an item used to measure a Self-Enhancement value. To respond to each portrait, participants were asked, *"How much like you is this person?"* Six response choices, ranging from *very much like me* (scored 6) to *not like me at all* (scored 1) were provided.

Cultural classroom climate. The Cultural Learning Environment Questionnaire (CLEQ; Waldrip & Fisher, 2000) was developed to examine culturally-sensitive factors in the classroom, which specifically measure a diverse range of student perceptions, tendencies, and attitudes towards their classmates and teachers. The seven 5-item CLEQ subscales are: Competition (e.g., *"I like to compete against the other students*"); Collaboration (e.g., *"I like to work with other students"*); Teacher Authority (e.g., *"It is OK for me to disagree with teachers"*); Deference (e.g., *"I try to say what I think the teacher wants rather than give my own opinion"*); Modeling (e.g., *"I like to have teachers tell me how to work in class"*); Gender Equity (e.g., *"I think that both females and males make excellent teachers"*); and Congruence (e.g., *"What I learn at home helps me to do things at school"*). To respond to the CLEQ items, participants were provided with a 5-point Likert scale ranging from *disagree* (scored 1) to *agree* (scored 5). In a preliminary analysis using principle component analysis, deference and modeling items were found to intermix and split into two components that can be differentiated based on the "social reference of the interaction" explicitely referred to in the items (i.e., either the teacher or the classmates). I then labeled them *teacher conformity* (4 items) and *classmate conformity* (4 items) (see Liem, 2006 for details).

Approaches to learning. The *Learning Process Questionnaire* or LPQ (Biggs, 1987) is a 36-item, self-report questionnaire used to assess student approaches to learning. The LPQ consists of six 6-item subscales: Surface Motive (e.g., *"I want to study hard only for subjects* *that I think will help me to get a job, not for those that are more interesting"*), Surface Strategy (e.g., *"I tend to study only what the teacher says, no more"*), Deep Motive (e.g., *"I find that my school work can give me a good feeling inside"*), Deep Strategy (e.g., "*While I am learning things in school, I try to think of how useful they would be in real life"*), Achieving Motive (e.g., *"I try to obtain high marks in all my subjects because I like to beat the other kids"*), and Achieving Strategy (e.g., *"I have my way of keeping my books, notes and other class things so that I can find them easily"*). Participants were asked to report whether the item was true or false in their respective case, using a 5-point scale: 1 (*this item is never or rarely true of me*), 2 (*this item is sometimes true of me*), 3 (*this item is true of me about half the time*), 4 (*this item is frequently true of me*), and 5 (*the statement is always or almost always true of me*).

Cronbach's alpha internal consistency coefficients were computed for each cultural sample for each of the PVQ, CLEQ, and LPQ subscales mentioned above. The results were comparable across cultual groups (see Liem, 2006).

Procedure

The measures were administered together with a survey on demographic information (e.g., age, gender, ethnic background, religion, and so on) in a classroom setting. It was emphasised that the participants' responses would be confidential and would not affect their school grades. Participants were also told that there were no right or wrong answers to any of the questions and that honest responding was of great importance in the study.

The English version of the measures was administered to the Singapore, Filipino, and Australian samples. Since Bahasa Indonesia (BI) is the native language for the Indonesians, a BI version of the PVQ and the CLEQ was developed. The development of each of the BI measures was carried out through a translation and back-translation procedure (Van de Vijver & Poortinga, 2005) with close and iterative consultation with the original author of the PVQ (S. H. Schwartz, personal communication, December 2001-February 2002), the CLEQ (B. Waldrip, personal communication, December 2001), and one of the main learning approach researchers who had frequently used the LPQ in his cross-cultural studies (D. Watkins, personal communication, January-February 2002). The very first validation study of the BI measures was carried out with a sample of 482 adolescents prior to the present study (see Liem, 2006 for details).

Partial Least Squares (PLS) Path Analysis Modeling

Partial Least Squares (PLS) Path Modelling Analysis (Jöreskog & Wold, 1982; Noonan & Wold, 1983, 1997; Wold, 1982) was employed to examine the *predictive relationships* specified in the hypothesised model (see Figure 2). This analysis was performed using PLSPATH version 3.01 (Sellin, 1990). PLS path analysis, or also known as path analysis with latent variables (Sellin & Keeves, 1997), is an approach to structural equation modelling developed by Wold (1982) and subsequently advanced by Lohmoller (1989), to bridge researchers' conceptual ideas about the world and empirical observations. Wold (1980) believed that behavioural and social sciences are based on "soft" theories (unlike "hard" theories in engineering and mathematical sciences), and accordingly are developed through soft empirical observations. Strict assumptions made on theory, measurement scales, sample size, and residual distribution, according to Wold (1980), will only restrain the development of behavioural and social sciences. As a result, PLS Path Analysis "soft modelling" was created. However, note that "… it is not the concepts nor the models nor the estimation techniques which are soft, only the distributional assumptions" (Lohmoller, 1989, p. 64; see Jöreskog & Wold, 1982, Noonan & Wold, 1983; Wold, 1982 for fuller details of the development of PLS Path Modelling including its basic PLS algorithm).

Unlike its Maximum Likelihood (ML) or Generalised Least Squares (GLS) counterparts, which are covariance based estimation methods commonly used as a tool for theory testing and development, the PLS path modelling analysis is a variance based approach used to examine prediction in an empirical model. This was clearly stated by Jöreskog and Wold (1982, p.270):

> ML is theory-oriented, and emphasizes the transition from exploratory to confirmatory analysis. PLS is primarily intended for *causal-predictive* analysis in situations of high complexity but low theoretical information (italic added).

Unlike the ML estimation model-building technique commonly performed using LISREL (Jöreskog & Sorböm, 1989), EQS (Bentler, 1992), or AMOS (Arbuckle & Wothke, 1999), the PLS Path Analysis builds on the use of least squares regression analysis to predict and explain the effects of a set of predictors/predicting variables on one or more criterion/predicted variable (see Chin & Newsted, 1999; Fornell & Bookstein, 1982; Lohmoller, 1989; Noonan & Wold, 1983; Wold, 1989 for the comparison between PLS and ML estimation modelling). So, the emphasis of this approach is on the predictive power of inter-variable relationships, i.e., the maximisation of the amount of variance of the criteria (or DVs) explained by the predictors (IVs) and the minimization of the residuals on all the variables, latent and manifest, caused by measurement errors (Chin & Newsted, 1999; Falk & Miller, 1992; Noonan & Wold, 1983, 1997; Sellin & Keeves, 1997; Wold, 1982). The preference in the PLS Path Analysis for minimising the residuals, especially on the manifest variables, is based on the assumption that the theory is softer (less precise, less developed) than the empirical observations (Falk & Miller, 1992).

The PLS Path Analysis is a more appropriate model-building tool when the conditions and circumstances for ML estimation are not met. The statisticians dealing with a model-building technique (e.g., Chin & Newsted, 1999; Falk and Miller, 1992; Fornell & Bookstein, 1982; Lohmoller, 1989; Noonan & Wold, 1983; Wold, 1982) have classified these conditions into four categories: (1) theoretical conditions, i.e., (1a) when hypotheses are derived from an abstract, conceptual, or higher-level theory in which salient and/or relevant variables are not known, (1b) when the relationships between theoretical constructs and their manifest variables are ambiguous, (1c) when the relationships between constructs are hypothetical; (2) measurement conditions, i.e., (2a) when the manifest variables are represented by different levels of measurements (nominal/categorical, ordinal, interval), (2b) when the manifest variables indicate some degree of unreliability, (2c) when the residuals on manifest and latent variables are correlated (a condition that is also known as heteroscedasticity); (3) distributional condition, i.e., when the response distribution is non-normal or unknown; and (4) practical conditions, i.e., (4a) when a cross-sectional, survey, secondary data, or quasi-experimental research design are used, (4b) when a large number of manifest and latent variables are modelled, and lastly (4c) when too many or too few cases are available.

Note that, given the abovementioned conditions, the concept of *causation* must be abandoned in the PLS Path Analysis and be replaced by the concept of *predictability* (Falk & Miller, 1992; Jöreskog & Wold, 1982; Lohmoller, 1989). That is, while in a causal relationship antecedents are guaranteed to have an effect on a DV, the concept of predictability suggests that a set of predictors have only a limited degree of effect on certain variables. Noonan (1982) called this a *quasi-causal* relationship or weak causality.

The use of the PLS estimation technique in cross-cultural educational research has been primarily done by Noonan (1976, 1978, 1982, 1989; Noonan & Wold, 1983, 1985). In these studies, the PLS path modelling analysis was used to approximate the predictive power of sets of predictors (e.g., student's backgrounds, home and school environment variables, teacher training, teaching methods) on the student learning outcomes, such as verbal ability, science achievement, student attitude toward education and science (e.g., Noonan, 1976, 1978, 1982, 1989; Noonan & Wold, 1983, 1985). The application of the PLS path modelling in psychology is still scarce (but see Schneider, Borkowski, Kurtz, & Kerwin, 1985).

Model specification. It has become common practice to distinguish *theoretical* from *observational* variables in structural equation modelling. The former are called *latent* variables (LVs) and graphically represented by *circles* or *ellipses*, whereas the latter are called *observed* or *manifest* variables (MVs) and graphically represented by *squares*. MVs are directly measured, whereas LVs are the underlying theoretical constructs constituted by a combination of a set of MVs often inferred with the help of factor analysis.

The predictive relationships among LVs are commonly indicated by *single-headed arrows*. In PLS Path Analysis, both latent-manifest and latent-latent variables relationships will be computed and taken into account to yield the coefficient values of the parameters. More specifically, the relationships among latent variables are termed *inner model* or *latent variable path model*, whereas the relationships between latent and manifest variables is specified as *outer model* (Noonan & Wold, 1983, 1997; Wold, 1982) or a *measurement model,* because it consists of the links between observations and theoretical constructs (Jöreskog & Sörbom, 1989). The arrows between a LV and its associated MVs in a measurement model are also known as a *block* (Falk & Miller, 1992; Noonan & Wold, 1983, 1997). As noted by Falk and Miller (1992), it is important to acknowledge that because we are dealing with soft theories, the arrows between LVs in an inner model tend to be conjectural and not necessarily firmly held relationships. These arrows represent relationships of interest and may be thought of as paths of influences. With an increased theoretical knowledge, they may become predictive or (of??) functional relationships (Falk & Miller, 1992).

Model estimation. The PLS parameter estimation proceeds in two steps (Chin & Newsted, 1999; Lohmoller, 1989; Sellin, 1986; Wold, 1982). The first step involves the iterative estimation of LVs as linear composites of their associated MVs. This denotes the estimation of the outer models. There are two optional modes of weight estimation for researchers to choose in order to obtain the parameter values in the outer models: (1) *outward* mode and (2) *inward* mode (Noonan & Wold, 1997; Sellin & Keeves, 1997; Wold, 1982). In an inwardly-directed block, the arrows are pointed from the MVs to the LV, whereas in an outwardly-directed block the arrows are pointed from the

LV to its MVs. It is important that all arrows run in the same direction within a single block (cf. Rogowski, 1989).

Sellin and Keeves (1997) pointed out that the distinction between the outward and inward blocks corresponds to the differentiation between *reflective* and *formative* indicators made by Hauser (1973). Outward or reflective indicators are assumed to *reflect* the corresponding latent construct. They illustrated that, for example, a set of attitude items are assumed to reflect an underlying attitudinal dimension, which is usually extracted using principal component analysis (Noonan & Wold, 1983, 1997). Inward or formative indicators, on the other hand, are assumed to *form* or *produce* a latent construct. For example, a set of teacher behaviours is assumed to form a specific teaching style, or in the present case, the seven *cultural factors* are assumed to form the *cultural climate* of the classroom learning environment. The estimation of outward blocks is based on an iterative sequence of simple ordinary least squares (OLS) regressions where the respective MVs are considered as dependent variables. The estimation of inward blocks, on the other hand, is based on multiple OLS regressions where the MVs are used as independent variables. Fornell and Bookstein (1982) and Sellin and Keeves (1997) have advised researchers to use the inward mode for the LVs serving as predictors, and outward mode for the LVs serving as criteria or predicted variables, in order to increase the predictive power of the inner model.

The second step involved in the PLS Path Analysis is the estimation of inner model coefficients and loadings which are estimated by means of standard least squares methods (Noonan & Wold, 1983, 1997; Wold, 1982). Thus, these inner model coefficients are estimated using standard path analytical procedures (Sellin, 1986; Sellin & Keeves, 1997; Wold, 1982). That is, the inner model coefficients are obtained by the OLS regression applied to each inner model equation separately.

In addition to predictor specification applied to inner and outer model equations, a fundamental principle of the PLS modelling is the assumption that all information between observed variables is exclusively conveyed by the latent constructs (Sellin & Keeves, 1997). This has two implications: (1) that PLS models do not involve any direct relationships among MVs; and (2) the outer residuals of one block are assumed to be uncorrelated with the outer residuals of all other blocks (Sellin, 1986; Sellin & Keeves, 1997).

Model trimming. As in most structural equation modelling techniques, a parsimony principle is important to follow, that is, to include in a model only variables and paths of the effects of one variable on another that contribute in a significant amount to prediction and the proportion of variance (Falk & Miller, 1992). Since the PLS Path Analysis avoids the assumptions that observations follow a specific distributional pattern, there is nothing to prevent the application of the distributional assumption, and therefore allowing the assessment of Standard Errors by Tukey's Jackknife (Chin & Newsted, 1999; Wold & Noonan, 1983). Thus, the significance of the regression coefficients in the model is determined by the critical value in the t-test. Statistically, a regression coefficient should be at least *twice* (or precisely 1.96) its corresponding standard error to be determined as a significant parameter, which means the t critical value ≥ 1.96 for $p < .05$ (Falk & Miller, 1992).

PLS Path Analysis, however, does not place great reliance on tests of statistical significance, since the computation and use of standard errors in a regression analysis necessarily assumes both a simple random sample and a multivariate normal distribution for the observed and the latent variables (Sellin & Keeves, 1997). The common practice in the PLS Path Analysis is to eliminate variables if a regression weight between an MV and its associated LV is less than 0.l in an outer model, and if a path coefficient between a predictor and a criteria in an inner model is less than 0.1, which means this predictor is likely to contribute less than one per cent (1%) to the explanation of the criteria. This practice was not followed here because the purpose of this study was to compare the four cultural models when an identical set of variables was used.

Results

The analysis conducted aimed to explore cross-cultural similarities and differences in the relationships between values, classroom cultural climate, approaches to learning, and perceived academic performance. The PLS path modelling analysis was performed to examine the predictive powers of the "predictor-criterion" relationships in the hypothesised model (see Figure 2). These relationships are: (1) values and the classroom cultural climate, (2) values and approaches to learning, (3) classroom cultural climate and approaches to learning, (4) values and perceived academic performance, (5) classroom cultural climate and perceived academic performance, and (6) approaches to learning and perceived academic performance.

The use of the PLS Path Analysis was thought to be appropriate for the purpose of this analysis for the following reasons: (1) a cross-cultural comparison of the predictive power in the hypothesised model is one of the main interests, (2) the ratio between the number of variables and each culture-specific sample size is large, (3) several variables were found to violate the assumption of normal distribution (e.g., gender equity and some PVQ items, as reported in Liem, 2006), and (4) the possibility of examining the effect of "culture" as a categorical data on the three constructs of interest as well as the perceived academic performance.

Model specification and estimation

The specification of model to be tested was the first step in this analysis. The three constructs examined in the present study, namely, values, classroom cultural climate, and approaches to learning, served as latent variables (LVs) in the model. To simplify the model, the LV for values was formed by five manifest variables (MVs) representing the five broad value types, i.e., Self-Enhancement, Openness-to-Change, Self-Transcendence, Conservation and Hedonism (John P. Keeves, personal communication, September 28, 2004). The LV for classroom cultural climate was formed by the seven cultural factors of the classroom learning environment identified earlier using principal component analysis (i.e., competition, congruence, teacher authority, gender equity, collaboration, teacher conformity, and classmate conformity). The LV for approaches to learning was formed by the three types of learning motive and learning strategy originated from the SAL theory (i.e., surface motive, deep motive, achieving motive, surface strategy, deep strategy, and achieving strategy).

The inwardly directed or formative mode, as opposed to the outwardly directed or reflective mode, was chosen because each set of MVs mentioned above was assumed as the explanatory components forming the LVs representing the constructs of values, the classroom cultural climate, and the approaches to learning, respectively. The use of the inward mode estimation for exogenous LVs would also increase the predictive power of the predictive relations in the inner model (Sellin, 1986).

Although the perceived academic performance was directly measured by only one item, it was still regarded as an LV (and was therefore graphically represented by an ellipse) for it was part of the inner model specified earlier (see Falk & Miller, 1992 for a detailed explanation). However, since the value of parameter between the MV and the LV that has only one MV is one or unity, irrespective of whether such a single MV block is specified as inward or outward (Falk & Miller, 1992; Sellin, 1986), for the clarity of the pictorial models shown in this analysis, the MV representing the perceived academic performance directly observed or measured was not drawn in the models.

Figures 3-6 show the four PLS models for the Singapore, Indonesian, Filipino, and Australian samples, respectively. It must be noted that, since the focus of this part of the analysis was on the predictive relationships in the inner model (specified by bold lines), the relationships between MVs and their respective LVs were not discussed. Furthermore, as pointed out by Sellin and Keeves (1997), PLS models do not involve any direct relationships among MVs because all information between observed variables (i.e., MVs) is exclusively conveyed by the latent constructs.

Singapore model. Figure 3 shows the PLS model for the Singaporean sample. As shown in the inner model, the direct effect parameter or path coefficient from values pointing to the classroom cultural climate was significant (β = 0.53, SE = 0.04). Twenty-eight per cent of the variance in the classroom cultural climate was fully explained by values because the latter was the only predictor for the former in the model.

The Singaporean students' adoption of approaches to learning was significantly predicted by values (β = 0.42, SE = 0.06) and the classroom cultural climate (β = 0.39, SE = 0.06) as shown as direct paths in the model. Values (26%) and the classroom cultural climate (24%) accounted for 50 per cent of the variance in the Singaporean sample's approaches to learning[1]. Apart from the direct effect, the relationship between values and the approaches to learning was mediated moderately by the classroom cultural climate (β = 0.21). This made the total effect of values on the approaches to learning even stronger (β = 0.63)[2].

The direct effects of values (β = 0.16, SE = 0.10, ns), the classroom cultural climate (β = -0.07, SE = 0.09, *ns*), and the approaches to learning (β = 0.15, SE = 0.10, *ns*) on the perceived academic performance were considerably small. The indirect effect of values on the perceived academic performance through the cultural climate and the approaches to learning was also small (β = 0.06). Likewise, the indirect effect of the classroom cultural climate on the perceived academic performance mediated by the approaches to learning was also

[1] The variance accounted for explained by a single predictor can be calculated by multiplying the path coefficient (β) between two Latent Variables (LVs) with the correlation coefficient between the corresponding LVs, and taking the absolute value (Falk & Miller, 1992). Note that this calculation applies only for the direct effect. See Table 4.51 for the correlation coefficients for all possible relationships between two LVs in the model across the four cultural groups.

[2] Note that the notion of statistical significant, as indicated by the Jackknife Standard Error [SE], does not apply to indirect and total path coefficients or effects (Noonan, 1982).

low (β = 0.06). As a result, only seven per cent of the variance in the perceived academic performance was accounted for by the three predicting constructs, namely values (3%), the classroom cultural climate (1%), and the approaches to learning (3%). This caused a high Residual Standard Deviation or R (0.96) in the perceived academic performance in the Singapore model[3]. The Residual SD can be interpreted as the path coefficient of some hypothetical residual explanatory variable which would explain all variance not already explained by the variables/constructs in the model (Noonan, 1982). From the model, it can be seen that these hypothetical residual variables have greater direct effect than the three constructs in the model in predicting students' perceived academic performance.

The effects of values, the classroom cultural climate, and the approaches to learning on students' perceived academic performance will not be elaborated further for the Indonesian, Filipino, and Australian models because of their small and nonsignificant effects, and because they are not the primary interest in the study. Table 2 presents the direct, indirect, and total effects, as well as the variance explained by a predictor (or a set of predictors) in the various linear relationships in the model. It should also be noted that since the cross-cultural comparison of the magnitudes of coefficients (predictive power) was of interest in this analysis, nonsignificant paths were not eliminated. As pointed out by Sellin and Keeves (1997), PLS path analysis does not place great reliance on tests of statistical significance, since the computation and use of standard errors in regression analysis necessarily assumes both a simple random sample and a multivariate normal distribution for the observed and the latent variables.

Indonesian model. Figure 4 shows the PLS model for the Indonesian sample. In the inner model, values was found to show a direct effect on the classroom cultural climate (β = 0.56, SE = 0.05). Thirty-two per cent of the variance in the classroom cultural climate was explained by values. The adoption of approaches to learning was significantly predicted by values (β = 0.48, SE = 0.05) and the classroom cultural climate (β = 0.39, SE = 0.06) as shown as direct paths in the model. Values (33%) and the classroom cultural climate (26%) accounted for 59 per cent of the variance in the approaches to learning.

---

[3] The Residual Standard Deviation can be calculated by the following formula $\sqrt{1 - R^2}$ (Noonan, 1982)

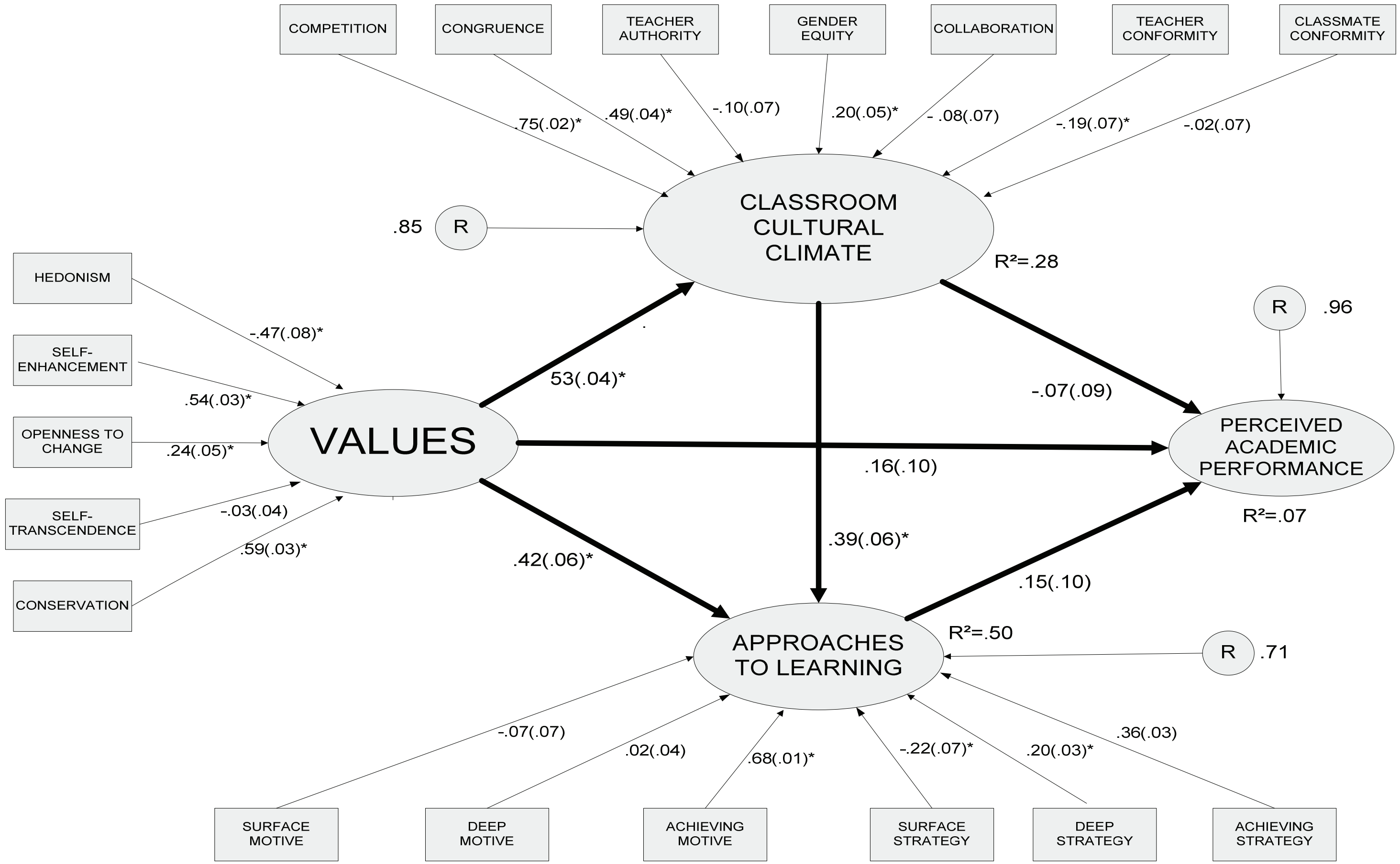


Figure 3. PLS path model of values, classroom cultural climate, approaches to learning, and perceived academic performance for the Singaporean sample (n=230).

(Note: the values in the parentheses are Jackknife Standard Errors (SE); * = a significant path)

Table 2. Cross-cultural comparison of the predictive powers and the percentage of variance in the model

No	Relationships in the Model	Effect	Singapore β	Singapore r	Indonesia β	Indonesia r	Philippines β	Philippines r	Australia β	Australia r
	PREDICTOR CRITERION									
1	Values – Classroom Cultural Climate	direct	.53	.53	.56	.56	.58	.58	.57	.58
		indirect	-	-	-	-	-	-	-	-
		total	.53	-	.56	-	.58	-	.57	-
2	Values – Approaches to Learning	direct	.42	.63	.48	.70	.44	.70	.14	.52
		indirect	.21	-	.22	-	.26	-	.38	-
		total	.63	-	.70	-	.70	-	.52	-
3	Classroom Cultural Climate – Approach to Learning	direct	.39	.61	.39	.66	.44	.70	.66	.74
		indirect	-	-	-	-	-	-	-	-
		total	.39	-	.39	-	.44	-	.66	-
4	Values – Perceived Academic Performance	direct	.16	.21	.18	.21	-.04	.26	.05	.18
		indirect	.06	-	.04	-	.30	-	.13	-
		total	.22	-	.22	-	.26	-	.18	-
5	Classromm Cultural Climate – Perceived Acad Perf	direct	-.07	.11	-.10	.08	.11	.32	.09	.22
		indirect	.06	-	.05	-	.15	-	.09	-
		total	-.01	-	-.05	-	.26	-	.18	-
6	Approaches to Learning – Perceived Acad Perf	direct	.15	.20	.13	.18	.33	.39	.14	.24
		indirect	-	-	-	-	-	-	-	-
		total	.15	-	.13	-	.33	-	.14	-
	PERCENTAGE OF VARIANCE		Singapore		Indonesia		Philippines		Australia	
1	Variance in Classroom Cultural Climate explained									
	By – *Values*		28%		32%		33%		33%	
2	Variance in Approaches to Learning explained by									
	- *Values*		26%		33%		31%		8%	
	- *Classroom Cultural Climate*		24%		26%		31%		48%	
	Total variance explained		50%		59%		62%		56%	
3	Variance in the Perceived Acad Perf explained by									
	- *Values*		3%		4%		1%		1%	
	- *Classroom Cultural Climate*		1%		1%		4%		2%	
	- *App to Learning*		3%		2%		13%		3%	
	Total variance explained		7%		7%		18%		6%	

Note: β = the standardised beta coefficient; r = correlation between two Latent Variables in the model

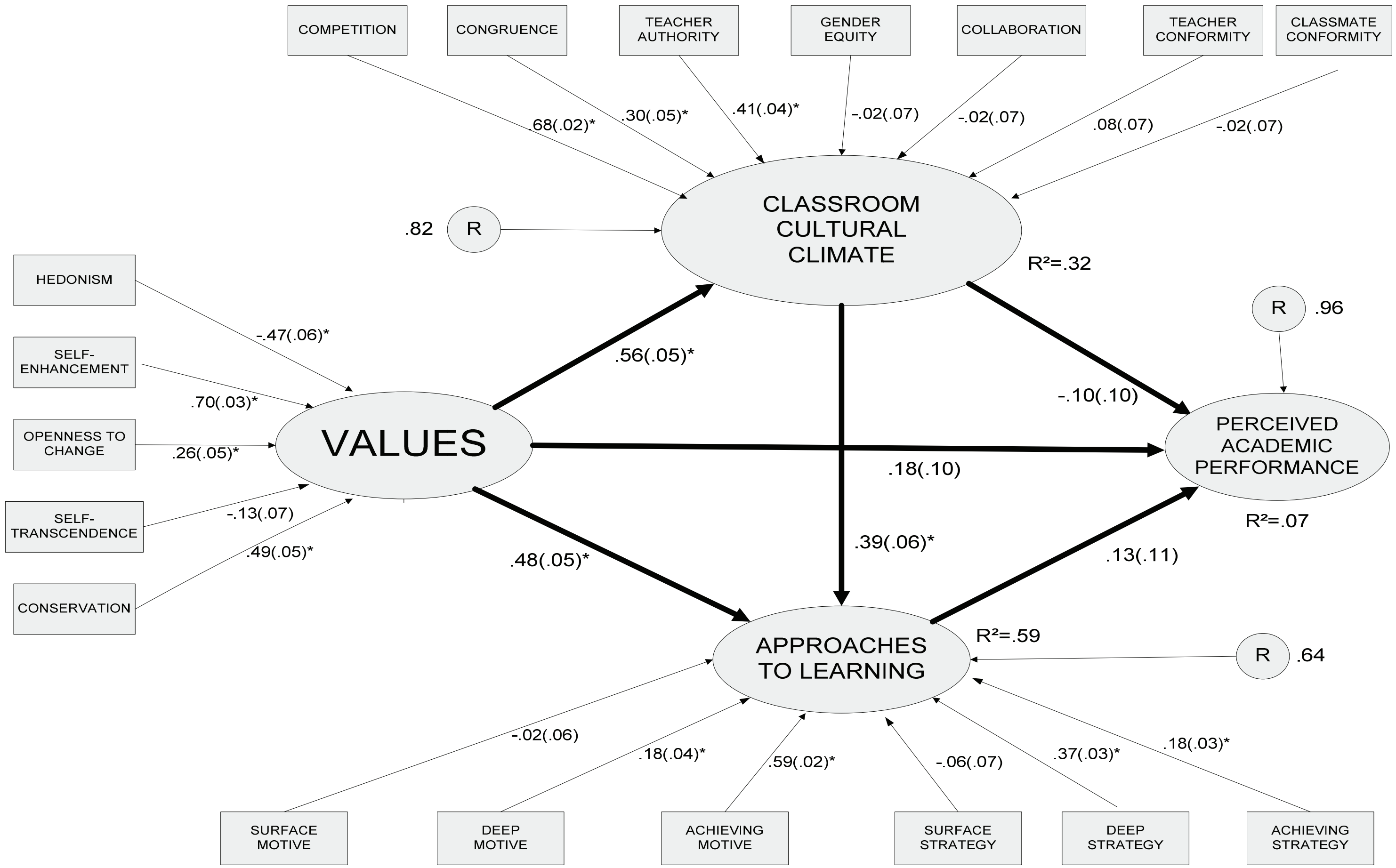


Figure 4. PLS path model of values, classroom cultural climate, approaches to learning, and perceived academic performance for the Indonesian sample (n=230)
(Note: the values in the parentheses are Jackknife Standard Errors (SE); * = a significant path)

Apart from the direct effect, the relationship between values and the approaches to learning was mediated moderately by the cultural climate of the classroom learning environment (β = 0.22). This made the total effect of values on the approaches to learning larger (β = 0.70).

Filipino model. Figure 5 shows the PLS model for the Filipino sample. In the inner model, values was found to be a significant predictor for the classroom cultural climate (β = 0.58, SE = 0.04). Thirty-three per cent of the variance in the classroom cultural climate was accounted for by values. The adoption of the approaches to learning was significantly predicted by values (β = 0.44, SE = 0.05) and the classroom cultural climate (β = 0.44, SE = 0.05) as shown as direct paths in the model. Each of them accounted for 31 per cent of the variance in the approaches to learning. Apart from the direct effect, the relationship between values and the approaches to learning was mediated by the classroom cultural climate (β = 0.26). This made the total effect of values on the approaches to learning stronger (β = 0.70).

In the Filipino model, the approaches to learning was found to be a moderate and significant predictor for students' perceived academic performance (β = 0.33, SE = 0.10). As a consequence, (1) the indirect effect of values on the perceived academic performance through the classroom cultural climate and the approaches to learning was found to be moderate as well (β = 0.30), and (2) the indirect effect of the classroom cultural climate on the perceived academic performance mediated by the approaches to learning was slightly higher (β = 0.15) than its direct effect (β = 0.11). Eighteen per cent of the variance in students' perceived academic performance was accounted for by the three predictors in the model, namely values (1%), the cultural climate (4%) and the approaches to learning (13%).

Australian model. Figure 6 shows the PLS model for the Australian sample. In the inner model, values was found to be a significant predictor for the classroom cultural climate (β = 0.57, SE = 0.04). Thirty-three per cent of the variance in the classroom cultural climate was accounted for by values. The adoption of the approaches to learning was significantly predicted by the classroom cultural climate (β = 0.65, SE = 0.04) and, to a lower extent, values (β = 0.14, SE = 0.05). Values (8%) and the classroom cultural climate (48%) explained 56 per cent of the variance in the approaches to learning. Although the direct relationship between values and the approaches to learning was low (β = 0.14), the indirect effect between the two mediated by the classroom cultural climate was found to be higher (β = 0.38).

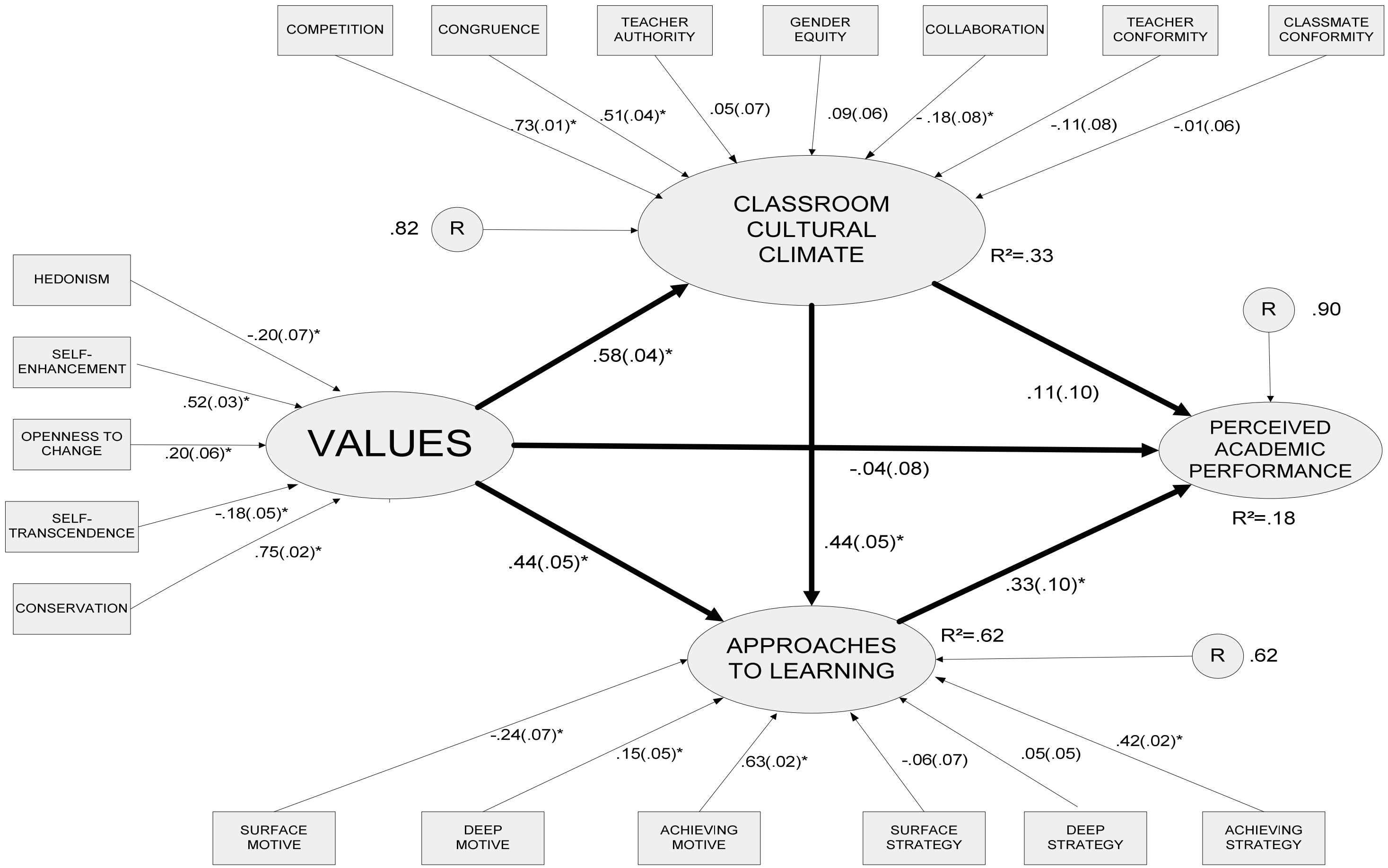


Figure 5. PLS path model of values, classroom cultural climate, approaches to learning, and perceived academic performance for the Filipino sample (n=230)
(Note: the values in the parentheses are Jackknife Standard Errors (SE); * = a significant path)

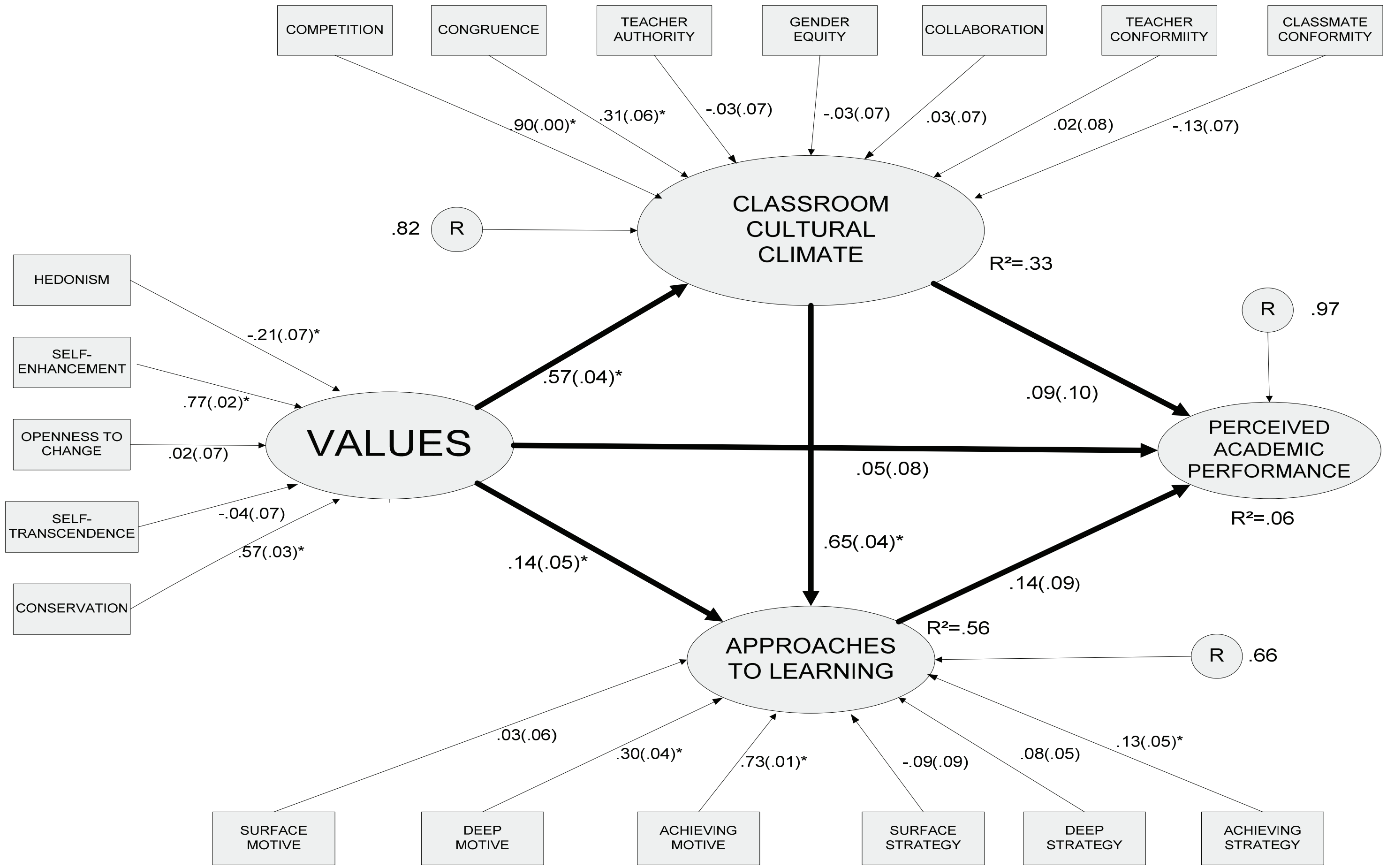


Figure 6. PLS path model of values, classroom cultural climate, approaches to learning, and perceived academic performance for the Australian sample (n=230)
(Note: the values in the parentheses are Jackknife Standard Errors (SE); * = a significant path)

This made the total effect of values on the approaches to learning moderate ($\beta = 0.52$).

Cross-cultural comparison of the predictive relationships in the model

To examine cross-cultural similarities and differences in the predictive relationships in the (inner) model, Table 2 was developed. The evaluation of the four culture-specific models (Figures 3-6) indicated two similarities and four differences in the predictive relationships specified in the model.

Note that, since the cross-cultural comparisons made here were based on the standardised path coefficients, a statistical caution must be taken in the interpretation of the similarities and differences. Standardised coefficients are known to be sensitive to the amount of variation prevailing in the sample because these coefficients refer to units equal to one standard deviation (SD) in the predictor and criterion variables (Noonan, 1978, 1982; Pedhazur, 1997). Assuming equal variation in the criterion variable and equal underlying predictor-criterion predictive relationships in the four cultural groups, the standardised regression coefficient would be higher when the SD in the predictor variable was higher. So, the comparison made here was based on the assumption that the distributions of scores for each variable across the four cultural groups were similar. In making a cross-cultural comparison, I shall distinguish between universal relationships which would refer to the relationships found in all four cultural groups, and culture-specific relationships, which would denote the relationships shown in only one or two or three cultural groups.

Values and classroom cultural climate. The first universal relationship observed across the four cultural groups was the presence of a significant and positive direct effect of values on the classroom cultural climate. The path coefficients between the two constructs were comparable across the four cultural groups (Singapore $\beta = 0.53$; Indonesia $\beta = 0.56$; Philippines $\beta = 0.58$; and Australia $\beta = 0.57$). As a result, the proportions of variance in the classroom cultural climate accounted for by values were also cross-culturally similar (Singapore 28%; Indonesia 32%; Philippines 33%; and Australia 33%).

Approaches to learning and perceived academic Performance. The second universal predictive relationship was observed between the approaches to learning and the perceived academic performance. In all cultural groups, the direct effect of approaches to learning on the perceived academic performance was positive. Although the path coefficient in the Philippines model was apparently higher (β = 0.33, SE = 0.10) than those of the other three models (Singapore β = 0.15, SE = 0.10; Indonesia β = 0.13, SE = 0.11; and Australia β = 0.14, SE = 0.09), a series of pair-wise t-test comparisons indicated that the four coefficients were not significantly different from one another. To illustrate, a comparison between the largest path coefficient (Philippines β = 0.33, SE = 0.10) and the smallest path coefficient (Indonesia β = 0.13, SE = 0.10) showed no significant difference between the two (t = 1.43, two-tailed, *ns*).

Classroom cultural climate and approaches to learning. The first difference was found on the direct effect between the classroom cultural climate and the approaches to learning. Two culture-specific patterns were observed and they distinguished the three Southeast Asian cultural groups from the Australian sample. The direct effect between the two constructs was moderate in the Singaporean (β = 0.39, SE = 0.06), Indonesian (β = 0.39, SE = 0.06), and Filipino (β = 0.44, SE = 0.05) modela. This effect was higher in the Australian model (β = 0.66, SE = 0.04). A series of pair-wise t-test comparisons showed that the Australian path coefficient was significantly higher than the one in the Singapore (t = 3.74, $p < 0.01$; two-tailed), Indonesian (t = 3.74, $p < 0.01$; two-tailed) and Filipino (t = 3.44, $p < 0.01$, two-tailed) models. However, the path coefficients in the Singapore, Indonesian, and Filipino models were not significantly different from one another. As a result, while the classroom cultural climate accounted for 24 per cent, 26 per cent , and 31 per cent of the variance in the approaches to learning in the Singapore, Indonesian, and Filipino models, respectively, it explained relatively higher, 49 per cent, in the Australian model.

Values and approaches to learning. The second difference was observed for the direct effect of values on the approaches to learning. The path coefficients between the two were moderate in the Singaporean (β = 0.42, SE = 0.06), Indonesian (β = 0.48, SE = 0.05), and Filipino (β = 0.44, SE = 0.05) models. In the Australian model, this coefficient was lower (β = 0.14, SE = 0.05). A series of pair-wise t-test comparison showed that there were significant differences between the path coefficient in the Australian model and those in the Singaporean (t = 3.59, $p < 0.01$, two-tailed), Indonesian (t = 4.79, $p < 0.01$, two-tailed), and Filipino (t = 4.23, $p < 0.01$, two-tailed) models. However, the path coefficients in the Singapore, Indonesian, and Filipino models were not significantly different from one another. As a consequence, while values accounted for 26 per cent, 33 per cent, and 31 per cent of the variance in the approaches to learning in the Singapore, Indonesian, and Filipino models, respectively, it explained only 8 per cent in the Australian model. Although the classroom cultural climate explained 48 per cent of the variance in the approaches to learning in the Australian model, the total variance in the approaches to learning explained by values and the classroom cultural climate became similar across the four cultural groups (Singapore 50% Indonesia 59%; the Philippines 62%; and Australia 56%).

Values and perceived academic performance. The third difference was found for the direct effect of values on the perceived academic performance. The path coefficient between the two constructs was positive in the Singapore (β = 0.16, SE = 0.10, *ns*), Indonesian (β = 0.18, SE = 0.10, *ns*), and Australian (β = 0.05, SE = 0.08, *ns*) models. This coefficient was negative in the Filipino model (β = -0.04, SE = 0.08, *ns*). However, since the indirect effect between values and the perceived academic performance mediated by the classroom cultural climate and the approaches to learning was relatively higher in the Filipino model (β = 0.30) than in the Singapore (β = 0.06), Indonesian (β = 0.04), and Australian (β = 0.13) models, the total effect of values on the perceived academic performance became comparable (Singapore β = 0.22; Indonesia β = 0.22; the Philippines β = 0.26; and Australia β = 0.18).

Classroom cultural climate and perceived academic performance. The fourth difference was observed for the direct effect of the classroom cultural climate on the perceived academic performance. Two culture-specific patterns were found and they distinguished the Singapore and Indonesian models on the one hand, and the Filipino and Australian models on the other. The path coefficient between the two constructs was negative in the Singapore (β = -0.07, SE = 0.09, *ns*) and Indonesian (β = -0.10, SE = 0.10, *ns*) models. However, this coefficient was found to be positive in the FIlipino (β = 0.11, SE = .10, *ns*) and Australian (β = 0.09, SE = 0.10, *ns*) models.

Pancultural model

A pancultural model, which was derived from the combined data set (n=920), was developed to examine the effect of "culture" on (1) values, (2) the classroom cultural climate, (3) the approaches to learning, and (4) the perceived academic performance. This aim was summarised in the following question: How much variance in (1) values, (2) the classroom cultural climate, (3) the approaches to learning, and (4) the perceived academic performance is explained by culture?

As asserted by Sellin and Keeves (1997), PLS path modelling has an advantage for cross-national comparative studies in that the technique enables researchers to use variables to represent countries in the equations. That is, the researchers are able to investigate *the effect of culture* on the constructs of interest.

To avoid a misinterpretation, it is important to explain what *the effect of culture* means. In a true experimental study, participants are randomly assigned into different groups; participants in a particular group will typically be given a different treatment from those in other groups. By doing so, between-group differences found after the experiment can be explained by or attributed to the treatments given (Campbell & Stanley, 1963). Van de Vijver and Leung (1997) asserted that a cross-cultural study can be regarded as adopting a *quasi-experimental* research design. It is called *quasi* because is not possible to randomly "allocate" participants into different cultural groups. It is called *experimental* because, in a cross-cultural study, the researcher would typically make a comparison of the constructs of interest across cultural samples, and the between-cultural group differences found in the study are then attributed to the "treatments" given to participants by cultures. The treatment is given through a process called *cultural transmission*, which can be in the forms of enculturation, socialisation, and education (Berry et al., 2002).

The participants in this study were selected because they had experienced a similar process of enculturation, socialisation, and education in their respective cultural milieus (in this case, Singapore, Indonesia, the Philippines, and Australia). So, although there were a few participants from Indonesia and Australia who were racially different from the racial majority in the Australian and Indonesian samples, respectively, this was counterbalanced or compensated by selecting those who were born and raised in the country (like the Indonesian sample) and who had lived and studied in the country for at least 10 years (like the Australian sample). The selections of participants in the Singaporean and Filipino samples were easier because participants in each of these samples were all from the same ethnic background, and importantly, born and raised in the country. So, the selection of the participants in each cultural group was done with an assumption that they had received a set of cultural treatments (i.e., socialisation, enculturation, and education) from their respective cultures for a significant period of time. The effects of these cultural treatments on the constructs in the hypothesized model (see Figure 2) were what I called the effect of culture.

In this analysis, each participant was dummy-coded according to the culture he or she was from. The Singaporean participants were coded 1, the Indonesian participants 2, and the Filipino participants 3. The Australian participants were coded 0 (zero) and served as a reference point for the dummy variables. These dummy codes were entered into PLS regression equations as categorical variables (Falk & Miller, 1992; Sellin & Keeves, 1997) serving as explanatory Manifest Variables for a Latent Variable called "culture".

Figure 7 shows the pancultural model depicting predictive relationships between values, the classroom cultural climate, the approaches to learning, and the perceived academic performance, all of which served as LVs, and also the effect of "culture" on these LVs. Table 3 presents the summary of the direct, indirect, and total effects, as well as the percentages of variance in the pancultural model.

As carried out earlier with each of the four culture-specific models, the predictive relationships among the LVs in the model were first evaluated before the effect of culture on each LV was examined. In the inner model, the three main predictive relationships must be noted. First, the direct effect of values on the classroom cultural climate was positive and significant ($\beta = 0.55$, SE = 0.02). Second, the adoption of approaches to learning was predicted significantly by values ($\beta = 0.39$, SE = 0.03) and the classroom cultural climate ($\beta = 0.44$, SE = 0.03). Third, approaches to learning ($\beta = 0.14$, SE = 0.05) was a weak predictor for the perceived academic performance. These findings are consistent with those found in the four culture-specific models, and the magnitudes of direct effects in the pan-cultural model were all within the range of the direct effects found earlier in the four culture-specific models.

Effect of culture. The path coefficients between 'culture' and the LVs in the model were evaluated. 'Culture' showed a direct positive effect on values ($\beta = 0.41$, SE = 0.03), and the correlation coefficient between the two was $r = 0.41$. As a result, 17 per cent of variance in the values construct was accounted for by the cultural membership (see Table 3).

In contrast, the direct effect of 'culture' on the classroom cultural climate was rather low ($\beta = 0.14$, SE = 0.03). As a consequence, 'culture' explained only 5 per cent of the total variance in the classroom cultural climate. This suggested that the relationship between 'culture' and the cultural climate was largely mediated by values. As presented in Table 3, the indirect effect of 'culture' on the classroom cultural climate was higher ($\beta = 0.23$) than its direct effect ($\beta = 0.14$).

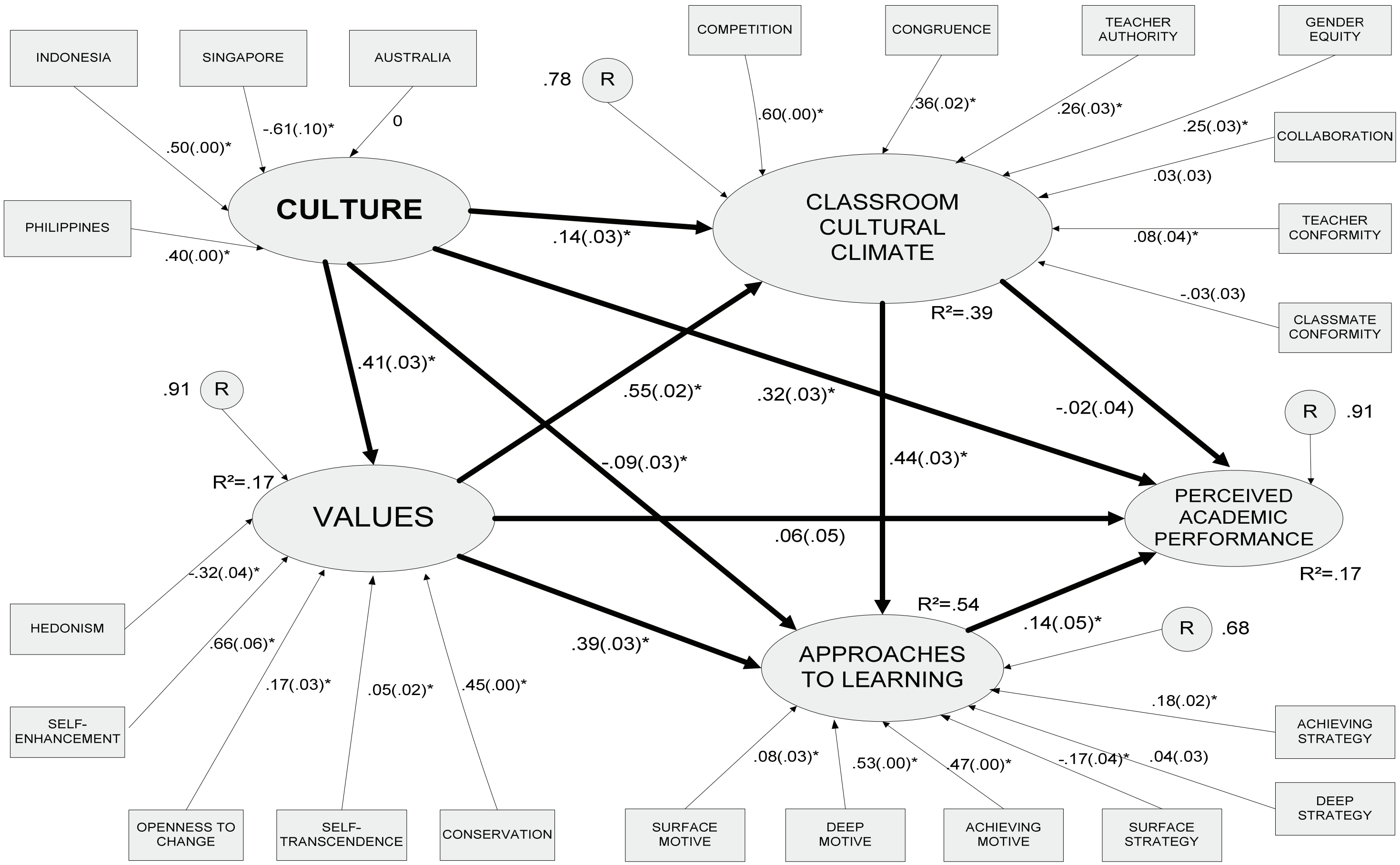


Figure 7. PLS path model of culture, values, classroom cultural climate, approaches to learning, and perceived academic performance: A pan-cultural analysis (n=920)

(Note: the values in the parentheses are Jackknife Standard Errors (SE); * = a significant path)

Table 3. Summary of the direct, indirect, and total effects and the percentage of variance in the pan-cultural model

No.	Predictor - Criterion Variables Relationship	Effect			LV-LV
		Direct	Indirect	Total	Correlation
	PREDICTOR - CRITERION				
1.	Culture - Values	.41	-	.41	.41
2.	Culture - Classroom Cultural Climate	.14	.23	.37	.37
3.	Culture - Approaches to Learning	-.09	.32	.23	.23
4.	Culture - Perceived Academic Performance	.32	.05	.37	.37
5.	Values - Classroom Cultural Climate	.55	-	.55	.61
6.	Values - Approaches to Learning	.39	.24	.63	.62
7.	Values - Perceived Academic Performance	.06	.07	.13	.26
8.	Classroom Cultural Climate - Approaches to Learning	.44	-	.44	.65
9.	Classroom Cultural Climate - Perceived Academic Performance	-.02	.06	.04	.22
10.	Approaches to Learning - Perceived Academic Performance	.14	-	.14	.23
	PERCENTAGE OF VARIANCE				
	Variance in Values explained by:				
1.	- *Culture*		17%		
2.	Variance in the Classroom Cultural Climate explained by:				
	- *Culture*		5%		
	- *Values*		34%		
	Total variance explained:		39%		
3.	Variance in the Approaches to Learning explained by:				
	- *Culture*		2%		
	- *Values*		24%		
	- *Classroom Cultural Climate*		28%		
	Total variance explained:		54%		
4.	Variance in the Perceived Academic Performance explained by:				
	- *Culture*		12%		
	- *Values*		2%		
	- *Classroom Cultural Climate*		0%		
	- *Approaches to Learning*		3%		
	Total variance explained:		17%		

'Culture' was observed to have a negative direct effect on the approaches to learning (β = -0.09, SE = 0.03). However, the correlation coefficient between the two was found to be positive although rather low (r = 0.23, see Table 3). The negative sign marking this path coefficient was possibly due to a *suppressor effect* (Falk & Miller, 1992; Keeves, 1997), which occurs when the path coefficient and the correlation coefficient between two latent constructs do not have the same sign. Falk and Miller (1992) offered three alternative reasons for the occurrence of suppressor effects. First, there are two or more variables that contain the same information (multicollinearity) and are therefore redundant. Second, the fact that the original relationship between the two variables is so close to zero that the difference in the signs simply reflects random variation around zero. The third is due to real suppression. Falk and Miller (1992) also suggest that the difference between redundancy and real suppression can be detected by eliminating an arrow elsewhere in the model. Should this lead to a change in the sign of the path coefficient between two latent constructs and not decrease the R^2 for the predicted variable, then the arrow is redundant and should be eliminated.

The next step carried out was therefore the elimination of a nonsignificant path coefficient in the model considered to be redundant. Since the path coefficient between the classroom cultural climate and the perceived academic performance was so small and nonsignificant (β = -0.02, SE = 0.02), this path was deleted and the model was re-run. The elimination of this path did not change much the magnitude of the other path coefficients and the R^2s in the modified model. In fact, the path coefficient from 'culture' to the approaches to learning remained the same (β = -0.09). Next, the path coefficient from values to the perceived academic performance was also eliminated because it was nonsignificant. Again, the magnitude of the other path coefficient and the R^2s in the model was still unchanged. Since the rest of the path coefficients in the model were significant, the deletion of other paths was not possible because it would lead to a *specification error*, i.e., where all relevant and statistically significant variables are not specified in the model (Falk & Miller, 1992). It was thus concluded that the negative sign marking the path coefficient from 'culture' to the approaches to learning was due to the real suppressor effect. When the real suppresser effect takes place, the correct sign interpretation is the one given by the correlation coefficient (Falk & Miller, 1992). Since the impact of the deletion of the direct paths pointed from values and the classroom cultural climate to the perceived academic performance, these paths were retained in the final model. After all, these nonsignificant paths were also informative for the evaluation of the hypothesised model. So, the final pancultural model is the one shown in Figure 7.

While we can now infer that the direct effect of 'culture' on the approaches to learning was low (β = 0.09), its indirect effect (where values and the cultural climate of the classroom learning environment served as mediators) was moderate (β = 0.32). This was, of course, due to the rather high direct effects between both mediators and the perceived academic performance as elucidated earlier. On the whole, 54 per cent of variance in the approaches to learning was explained by values (24%), the classroom cultural climate (28%), and the ‘culture’ (2%). Lastly, ‘culture’ also showed a significant direct effect on the perceived academic performance ($\beta = 0.32$, SE = 0.03). As a result, ‘culture’ accounted for 12 per cent of the variance in the perceived academic performance.

Concluding remarks

This chapter has presented a part of my doctoral study that was conducted between 2001 and 2005 at the National University of Singapore. The use of partial least squares (PLS) path analysis modeling in this study was made possible by the generous favour from Professor John Keeves, who was introduced to me by Dr Bobbie Matthews, the principal editor of this special volume, when I happened to be visiting Adelaide, South Australia, in September 2004 for a friend’s wedding in Port Elliot, an hour’s driving distance from Adelaide. I found Professor Keeves as somebody who is generous in sharing his expertise and knowledge, particularly to students who come from the Asian region, and humble, given his remarkably outstanding achievement and the work he had done – including his 1997’s edited volume, *Educational Research, Methodology and Measurement: An International Handbook,* which I have been using as a reference again and again. Professor Keeves is undoubtedly an exemplary role model for young scholars and researchers, or for that matter, for all scholars and researchers who have been fortunate to know him, including myself. A period of five years has gone by fast since I had my three-day one-to-one workshop on PLS path analysis modeling with Professor Keeves in September 2004, but memories of him will always endure in my mind. Thank you for being who you are to all of us. Happy 85th birthday, John.

References

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 user’s guide.* Chicago, IL: SPSS Pub.

Bentler, P. M. (1992). *EQS structural equation program manual.* Los Angeles: BMDP Statistical Software.

Biggs, J. B. (1970). Faculty patterns in study behaviour. *Australian Journal of Psychology, 22*, 161-174.

Biggs, J. B. (1973). Study behaviour and performance in objective and essay formats. *Australian Journal of Education, 17,* 157-167.

Biggs, J. B. (1987). *Student approaches to learning and studying*. Melbourne: Australian Council for Educational Research.

Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*, 3-19.

Biggs, J. B. (1996). Western misperceptions of the Confucian-heritage learning culture. In D. A. Watkins, & J. B. Biggs (Eds.), *The Chinese learners: cultural, psychological and contextual influence* (pp. 45-67). Melbourne: ACER.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Research, 18*, 32-42.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modelling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307-341). London: Sage.

Cole, M. (1990). Cultural psychology: A once and future discipline? In J. J. Berman (Ed.), *Nebraska symposium on motivation.1989: Cross-cultural perspective* (pp. 279-336). Lincoln: University of Nebraska Press.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart, & Winston.

Entwistle, N., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

Falk, R., & Miller, N. B. (1992). *A primer for soft modelling*. Akron, OH: University of Akron Press.

Fornell, C., & Bookstein, F. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*, 440-452.

International Association for the Evaluation of Educational Achievement (IEA) (2004). *Trends in international mathematics and science study 2003*. Retrieved December 16, 2004, from www.timss.org

Joreskog, K. G., & Sorbom, D. (1989). LISREL 7: A guide to the program and applications. Chicago: SPSS.

Joreskog, K. G., & Wold, H. (1982). System under indirect observation: Causality, structure, prediction (Part 2). Amsterdam: North Holland.

Liem, A. D. (2006). *The influences of sociocultural and educational contexts on approaches to learning*. Unpublished doctoral dissertation. National University of Singapore.

Liem, A. D. (2007). Triandis Award: Appreciating differences, celebrating similarities – An experience of learning in culturally diverse contexts. *Cross-Cultural Psychology Bulletin, 40*(1-2), 31-35.

Lohmoller, J. B. (1989). Latent variable path modelling with partial least squares. New York: Springer-Verlag.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning -1: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Noonan, R. (1976). School resources, social class, and student achievement: A comparative study of school resources allocation and the social distribution of mathematic achievement in ten countries. *IEA Monograph Series, No. 5*. Stockholm: Almqvist & Wiksell International.

Noonan, R. (1978). An empirical study of two countries: Chile and India. In T. Husén, L. Saha & R. Noonan (Eds.), Teacher training and student achievement in less developed countries. *Staff Working Paper, No. 310.* Washington D. C.: World Bank.

Noonan, R. (1982). School environments and school outcomes: An empirical comparative study using the IEA data. In M. Niessen & J. Peschar (Eds.), *Comparative research on education: Overview, strategy and applications in Eastern and Western Europe* (pp. 169-202). New York: Pergamon Press.

Noonan, R. (1989). Evaluation of school system using partial least squares (PLS): An application in the analysis of open social system. In H. Wold (Ed.), *Theoretical empiricism: A general rationale for scientific model-building* (pp. 31-62). New York: Paragon House.

Noonan, R., & Wold, H. (1983). Evaluating school system using partial least squares. *Evaluation in Education, 7*, 219-364.

Noonan, R., & Wold, H. (1985). Partial least squares. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopaedia of education* (Vol. 7, pp. 3767-3775). Oxford: Pergamon Press.

Noonan, R., Wold, H. (1997). Partial least squares path analysis. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (2nd ed.) (pp. 710-716). Oxford: Pergamon Press.

Rogowski, J. (1989). Estimation of RE-models, and some remarks on model specification in econometrics. In H. Wold (Ed.), *Theoretical empiricism: A general rationale for scientific model-building* (pp. 75-77). New York: Paragon House.

Saljo, R. (1991). Introduction: Culture and learning. *Learning and Instruction, 1*, 179-185.

Schneider, W., Borkowski, J. G., Kurtz, B. E., & Kerwin, K. (1985). Metamemory and motivation: A comparison of strategy use and performance in German and American children. *Journal of Cross-Cultural Psychology, 17*, 315-335.

Schwartz, S. H. (2005). Robustness and fruitfulness of a theory of universals in individual human values. In A. Tamayo, & J. B. Porto (Eds.), *Valores e comportamento nas organizações* [Values and behavior in organisations] (pp. 56-95). Petrópolis, Brazil: Vozes.

Sellin, N. (1986). Partial least squares analysis. *International Journal of Educational Research, 10,* 189-200.

Sellin, N. (1990). *PLSPATH version 3.01: Program manual.* Hamburg, West Germany.

Sellin, N., & Keeves, J. P. (1997). Path analysis with latent variables. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (2nd ed.) (pp. 633-640). Oxford: Pergamon Press.

Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational tests and psychological tests for cross-cultural assessment.* Mawhaw, NJ: Erlbaum.

Waldrip, B. G., & Fisher, D. L. (2000). The development and validation of a Learning Environment Questionnaire using both quantitative and qualitative methods. *Journal of Classroom Interaction, 35*, 25-37.

Wold, H. (1980). Soft modelling: Intermediate between traditional model building and data analysis. *Mathematical Statistics, 6*, 333-346.

Wold, H. (1982). Soft modelling: The basic design and some extensions. In K. G. Joreskog & H. Wold (Ed.), System under indirect observations: *Causality, structure, prediction* (Part 2, pp. 1-54). Amsterdam: North Holland.

Wold, H. (1989). Introduction. In H. Wold (Ed.), *Theoretical empiricism: A general rationale for scientific model-building* (pp. VII-XL). New York: Paragon House.

23

A need to research ‘reflection’

T. Gibbons
University of South Australia

While reason and reasoning attracts much analysis and research, reflection does not. Yet the word and its synonyms appear frequently in research documents, books and pronouncements. Reflection is linked to reasoning yet is not simply reasoning. Reflection is linked to the core beliefs that a person holds and it is to these beliefs that a person turns when reflecting. This raises the question of how we develop core beliefs and the question of what should be researched.

The word ‘reflection’ and its synonyms are scattered like confetti throughout the pages of curriculum documents, research reports, and political statements. The word and its synonyms are used with abandon by advisors, researchers and department heads in all manner of enterprises, not least of which is education. However, one is hard-pressed to find the word in any of the indices of books on the social sciences or books on education. One is hard-pressed to find in any of the documents that use the word, some account of what it means to reflect, which, in passing, is strange. Reason and the development of the ability to reason has been studied extensively in educational research and the results of that research have influenced those who devise curricula, train teachers and administer education. These same people sprinkle ‘reflection’ and its synonyms liberally throughout their writings and their speech. By not saying what they mean by the word are they supposing that it is synonymous with reason and reasoning? If they are supposing so, I think it a mistake. And, if I am correct, then there is the real possibility there is a serious flaw that lies embedded in much educational research, curriculum documents and administrative pronouncements. The purpose of this essay is to examine some aspects of what it is to reflect in the hope that some ideas of use to those who use the word may emerge.[1]

Consider the phrase ‘rational reflection’. The implication is that reflection is different from reason. McDowell considers the case of a rational wolf, that is, a wolf who:

> …would be able to let his mind roam over possibilities of behaviour other than what comes naturally to wolves. (McDowell, 1995, p.152).

However, later, in describing the rational wolf, faced with the question of what he should do, McDowell’s wolf becomes the ‘reflective wolf’ (McDowell, 1995, p.153). Following this, reason is described as that which:

> …enables and even obliges us to step back from it [our nature] (McDowell, 1995, p.154).

There is here the notion of both reason and reflection and they have been confused. They are both concerned with thought but this is not to say that there is no distinction between them. Reflection cannot take place without reason but reasoning can take place without reflection. The crux of the distinction is that in reflection we step back from our thought and review and evaluate. That is, reflection involves thinking about thinking or, as the behavioural sciences would have it, meta-cognition. However, I am wary of accepting meta-cognition as a synonym for reflection. Meta-cognition is a word that operates within a particular type of inquiry and can be expected to have meanings associated with that inquiry which are not associated with the ordinary language use of reflection. Meta-cognition has occupied the attention of those in the field of educational psychology in recent years. I shall confine myself to ordinary language usage.

The distinction between reasoning and reflecting is evident in the English language. We may say that we dispute, disagree with or reject a line of reasoning, but it sounds odd to say the same of reflecting. We may say that we should reflect on something but to say the same of reasoning is odd. We reflect ‘on’ but reason ‘about’. We ‘check’ reasoning but not reflection. Reasoning can be valid or invalid but not reflection, though it might be irrational. While making a distinction, none of this divorces reasoning from reflection.

---

[1] The substance of the argument on reflection appears in Gibbons T. (2009) *Reflection, Science and the Virtues*. Rotterdam, Sense Publishers where various aspects of it are developed further. The material appears with the permission of the publishers.

With reflection, in the first instance, we return to previous thinking or thoughts; we revisit them. This is a necessary precursor to giving attention to that past and reflecting on it. If I reflect on the past I am recalling incidents that may or may not involve myself. I may make a judgement that a wrong move was made or a wrong direction taken. On the other hand I may pinpoint a decision in the past with which I am pleased because it produced a successful outcome. I may, however, merely recall the past and, as it were, run it before my mind without making judgements but simply noting that it had occurred and observing the links in the chain of a personal narrative. This is recall which does not necessarily involve reflection. Is it possible to reflect on the future? The immediate response might be, surely not, for the very root of the word 'reflection' is connected with the past. It may be said that I may reflect that certain possibilities may or may not arise in the future. I may reflect that this decision now will close or open a door in the future. But reflect on the future?

On this account reflection is the giving of attention towards past thinking. To reflect is a transitive verb. There must be something to reflect upon, we do not simply reflect. And this can be further broken down into the giving of attention to:

- thinking and this may be in the distant past or that just gone;
- the content of that thinking; and
- the evaluation of that thinking.

In all of these the reflection may be concerned with reasons and reasoning.

But to confine the analysis of reflection to the past is perhaps a mistake. What is the point of reflection? We step back and consider and concentrate on the past in order to do a number of possible things, among them:

- to correct previous thinking or plans; and
- to revisit our plans so that we may plan further.

The first presupposes that something has gone awry and we need to rethink the present and the possible future. The second presupposes that we have completed or nearly completed previous plans and need to consider the next move. The introduction of the notion of planning seems at odds with the assertion that reflection is concerned with the past. Any planning incorporates a future dimension. It also incorporates the ability to imagine the future. Reflection about past or current plans necessarily incorporates a time factor.

The first step in planning may be an attempt to establish the possibilities and evaluate them. For instance, when confronted with a situation the like of which we have not met before, we typically step back from the situation and make a number of possible moves. We may try to determine whether or not masterly inactivity is called for or if we must act. Both masterly inactivity, as the phrase implies, and action, demand an assessment of the results – what happens if I do nothing? What happens if I do this? We speculate. We are in the position of McDowell's reflective wolf. If we decide to do something, then foresight and imagination are involved and planning. But foresight and imagination are also implied in masterly inactivity. We may attempt to find a suitable comparison between something from our past or the past of others for the situation which confronts us. This is a situation akin to finding an analogous theory in science to use in tackling a new problem. This is a reflective activity. Reflection is necessarily connected with the past and may be connected with the future. Planning necessarily involves reflection and is thus involved with both past and future. The element which takes reflection into the future is imagination. To give attention to the future we must imagine it. Once we have imagined it, then it is possible to reflect upon it. Though in saying this there is the appearance of a step by step process and it is a mistake to think it so. Reflection is both retrospective and prospective.

In order to reflect we must concentrate, that is, we must single out and consider certain aspects of what we have thought or planned or done and we do this by focusing on those aspects to the exclusion of others which may be extraneous and irrelevant. Faced with what ought to be done we have to concentrate on the problem if we are to solve it. Simple attention will not do.

Consider some examples in order to unpack further the concept of reflection.

The years 1845-1849 were the years of what the Irish call 'The Great Hunger' and the English call 'the Irish Famine'. The population of Ireland in 1841 was thought to be 8,175,124. The expectation for 1851 was 9,018,799 given the normal rate of increase at the time. The census of 1851 placed the population at 6,552,385. It is estimated that 1,500,000 emigated to England, North America, Australia and New Zealand between 1845-49. This leaves a shortfall of approximately 1,000,000 in the anticipated figure for 1851. The million are the deaths from the famine. The deaths were, in the main, due to starvation and typhus caused by the conditions of the famine (Kee, 1982).

The Great Hunger was caused by the failure of the potato crop upon which the Irish poor depended for their existence. They existed by hiring out their labour in return for a small plot on which they could grow potatoes. In many respects, for the poor, it was a cashless economy. The failure was caused by *Phytophthora Infestans*, a fungal disease which struck every year from 1845 to 1849. The disease did not effect other crops and, in fact, the years 1845-49 saw a rich harvest of grain and excellent conditions for farming livestock. The Irish poor lacked the money to have access to this food. Even in normal times a quarter of the Irish population, approximately 2,000,000, lived between potato harvests in conditions of semi-starvation.

The Irish Parliament having been abolished by England in 1800, Ireland was ruled by the English government in London. The political ideology at the time of the Irish Famine was known as 'political economy'.

> Certain almost unshakeable, sincerely held economic beliefs were to underlie all governmental policy towards the Famine. And the greatest of these was that principle of political economy which maintained that you should interfere to the absolute minimum with the market forces of supply and demand because if you *did* so interfere, you endangered the natural flow by which supplies could reach the market. (Kee, 1982, p.82)

Implementing this principle resulted in a continued large export of food from Ireland to the continent of Europe during the famine years and the control of any relief for fear of affecting the market. The causes of the Great Hunger and the deaths of so many are clearly a complex and difficult subject to explore. If, for the sake of argument, we cease exploration at this point and consider just the above, then, what might be said?

There are causal links between the deaths and a number of factors. These links can be analysed biologically, politically, culturally and economically. Reasons can be given for the links. What happens if we stand back from the causal network that has been constructed and evaluate it? It might be said that one sense of reflection occurs when we stand back and examine the validity of the causal links. That is, we look at the evidence for causation and evaluate that evidence. We might, as a result, give weight to some causes over others. We might weigh the strength and weakness of the overall case and this might point the way to further avenues of investigation. However, it seems to me that it is a mistake to call this reflection. It is more properly called reasoning.

Reflection is the thinking in which reasoning is involved but which goes beyond reasoning. We can stand back and evaluate the acts and omissions of the actors in the historical drama. For instance, we might say that the ideology of 'political economy' was morally bankrupt and for the English government to act on the principles of 'political economy' in the way that they did was immoral. We might contrast the government's actions with the Christian virtues of charity and benevolence towards the poor and their needs, which the English government publicly professed. In this we are engaged in making judgements based on criteria which are embedded in our ontological and epistemological beliefs. This I would consider merits the term 'reflection'. To reflect is to ponder, to mull over. It is to go beyond the construction of an argument and the checking of that argument. It is to step back and evaluate, and, because the Great Hunger was concerned with what happened to and was done to human beings, the evaluation is fundamentally a moral evaluation. In making such an evaluation we measure what happened against what we think ought to have happened. This is to place the matter before criteria which are fundamental to our conception of the world in which we exist. Reflection goes beyond but includes reasoning.

Consider a different example. Crystals take regular forms. They have a certain symmetry. Fluorite is an octahedron. The snowflake crystal has a six-fold symmetry. Iceland spar is rhomboid. In chemistry the student may learn to recognise the various shapes and link them to their chemical composition. Crystals may be grown in classroom experiments. A wealth of detail may be acquired, arranged, analysed and catalogued. It is possible to view the analysis and the catalogue arrangements to see whether or not they have any validity. However to reflect on the appearance of crystals is to confront questions about reality. Why is it that the crystals have flat planes; why is that the fluorite crystal is an octahedron? Pursuing these questions may lead to the thought that the atoms of a crystal are governed by the properties of three dimensional space. Repeated symmetry is only possible in (?) certain ways. Crystals are natural kinds and the atoms of the crystal are examples of fundamental natural forces. Reflection leads us to considerations about the nature of the world in which we exist.

The production of the spectrum from white light has been known for millennia. For some time it was thought that the spectrum was produced by the modification of white light. White light which passed through the thinner part of a prism was modified less than that which passed through the thicker part. If we reflect on that explanation then it would appear that another prism should produce further modification. It does not. Another explanation was needed. Newton provided it. The white light is separated into constituent parts, not modified. The reflection lies in asking why it should be that white light produces a spectrum under certain conditions. Science is the attempt to provide a description and explanation of the underlying structure of nature. This is a description of a reflective activity and reflection in science takes place whenever we ask questions that bear upon the underlying structure of nature, whenever we ask the ontological questions.

The Greeks produced buildings in which columns were connected by and supported a straight beam. Their buildings are rectilinear. The Romans invented the arch. Their arches were semicircular. The physical advantages of the Roman arch over the Greek beam can be explained by some elementary physics. The results of the advantage can be seen all over Europe. At some point around 1200 A.D. an oval arch appeared in Europe. This had advantages over the Roman arch which again could be calculated. The advantages resulted in Gothic architecture. It is not known how humanity came to progress from one to the other. Was the idea sparked by the humble hen's egg which is Roman at one end and Gothic at the other and whose strength lies on the plane between the ends? This is to reflect on the historical narrative and to speculate as to its causes.

Students may be asked to solve problems in physics which involve calculations – inclined plane problems for example. To solve the problem they will have to recall how to analyse the forces involved and the formulae to apply. Having completed the calculations, being good students, they will check their work. None of this involves reflection. Reflection, in my view, enters the arena when students place the problem within its history, which, of course, implies that if you wish to teach for reflective students then you must teach the history of science. Reason and reflection are connected. Certainly we can and do reason and reflect on particular issues but I would argue that the appearance of reflection signals the appearance of something wider and more important than the issue of the solution of this particular problem that confronts us here and now and the checking of that solution. Reflection always brings with it the activity of placing the thinking in the context of the narrative and the context of the nature of reality. This is the case in science.

In order to reflect we must place the thought or the plan or the deed in context. All thoughts, plans and deeds have a history, a narrative which leads to their existence and without which we cannot understand fully the thought, plan or deed. This we need to bring into focus, to concentrate on, if we are to reflect with the possibility of success. Consider $T = 2\pi\sqrt{l/g}$. There is no need to itemise the meanings of the symbols to some people. They recognise it as the equation describing the motion of the simple pendulum. Some of those, a smaller number than the original set, will recognise the equation as arising within a limited system. That is, they will recognise that there are specific limits set with regard to the arc of swing, friction, elasticity and so on. A smaller number still will recognise it as part of a revolution in physical science. At what point might it be said that a person reflects on the equation? More than merely recalling the connection of the equation with the simple pendulum is required. More than recalling that if the equation is to be derived from a simple pendulum then limits must be imposed on the system. This is simply a matter of mathematical practice. I would suggest that reflection is the proper description for that which takes place when, with recall as a base, the equation and the physical phenomenon start to be seen in the revolutionary context of which they form a coherent part. It is at this point that we talk of 'mulling over', 'pondering', 'appreciating'. This is not to suggest that recall is always a necessary base for reflection.

Plans and deeds are necessarily teleological. They are formed or performed with a purpose. Thoughts are too, though the connection is not so clear. However, the fact that we have in the language the phrase 'idle thoughts' seems to indicate that the norm is thought which is not idle but formed towards a purpose. The above points all find their expression in the language. Reflection is essentially a teleological enterprise and a very human enterprise. It is this capacity which, in my view, is a far better possibility as that which distinguishes us from the other animals. On reflection, reason will not do in this regard!

Following MacIntyre (1999), I would argue that the capacity which may be distinctively human is expressed in at least four things:

- the ability to distance ourselves from our beliefs, decisions and actions both in the past and the present;
- the ability to evaluate our beliefs, decisions and actions;
- the ability to imagine and attempt to choose, which presupposes evaluation, our future; and
- the ability to imagine, which presupposes evaluation, our past.

Taking these ideas a step further I would argue that the four things above express the capacity to reflect. If we are to teach for reflective students then it is these four things that are pre-eminent. If we are to research how best to produce a reflective student then these are the four areas of research. The danger is that such research and such teaching will fail to see that the four elements are inter-dependent.

The word 'speculate' illustrates these elements. The Macquarie Dictionary says that to speculate is:

> 1. to engage in thought or reflection, or meditate (oft. fol. by *on, upon* or a clause). 2. to indulge in conjectural thought. (Deldridge, 1987)

Speculation in the above sense often begins with the query – what if? What if the earth moved round the sun; what if I were to ride on a beam of light looking back at a clock; what if the continents are floating and can move? Not all 'what if' questions lead to speculation. What if I switch off the electricity connection to this computer? It will shut down. The speculative queries are those which lead a person to question the basis on which he or she views reality. The speculative questions demand the exercise of the imagination. In this sense it becomes clear that speculation is the life blood of science. Indeed, it becomes clear that speculation is the life blood of all attempts to advance our knowledge in whatever field or discipline.

The concept of 'regret' provides further illustration of reflection. We may regret the past, the present or the future. In regretting we always admit that things should have been and should be managed differently while at the same time admitting that this might not be possible. There is always evaluation and that evaluation is against what we conceive as the way the world ought to be. In regretting the past we recall what has happened, we imagine what might have been different and we evaluate what has occurred. In regretting the present we see how it has come to pass and how it may proceed. In regretting the future there is an element of helplessness in the face of what we see impending. To regret is a reflective process which requires reasoning, imagination and evaluation.

Reflection is a capacity which cannot be exercised in a vacuum. A human child goes:

> …beyond the reasoning characteristic of dolphins when they become able to reflect on and to pass judgment on the reasons by which they have hitherto been guided. This transition is one that dolphins have not made, so far as we know, but we can learn a good deal from them – and from chimpanzees and members of various other species – about the preconditions for making it. (MacIntyre, 1999, p.57)

The acquisition of the language of reflection is a pre-condition of rationality (MacIntyre, 1999, p.54). Rationality is only intelligible in the context of the ability to reflect and consider whether or not this or that reason is appropriate or relevant. And this rationality would not be possible without the language that enabled it. Rationality and reflection are intimately associated.

The principle of political economy held by the English government in 1846-50 can be classed as part of the core of beliefs of that government against which proposed actions were measured. The principle defined what was to count as reasonable, as evidence and as sound judgement in a particular area. What does it mean to say that the principle of political economy was part of the core beliefs of the government? The core beliefs define identity and, for the person or the agency, reality. The consequence is that an attack on the core beliefs of an agency, such as a government, is an attack on their existence and to be repelled with vigour.

The core of beliefs of an individual, define for that individual, what are to count as reasons, as evidence and as sound argument. They may also be said to start to define the individual. The flat-earther looks at the world in one way, the spheroid earther another. Jane Austen looks at the world very differently from Graham Greene. The identity of an individual is bound up in his or her core beliefs, that is, those beliefs which define how that individual sees the world and acts in it. They are the ontological and epistemological dicta of that individual. An attack on the core beliefs is an attack on the identity of an(?) individual and is defended with vehemence. The defence typically includes the erection of defensive ideas and if and when these defensive ideas become more and more ad hoc the individual can find himself or herself under great psychological strain. Lakatos (1978b) has expressed such ideas in the philosophy of science and I would argue that his work has a wider scope than science. It expresses something about human nature and about reflection. When Austen and Greene engage in reflection, they do so against the way in which they see the world. Their books are a reflection of and on that world. The reflection takes place in the context of their core beliefs. Rationality is only intelligible in the context of the ability to reflect and rationality is defined by this core. The core is defended from attack by the erection of auxiliary ideas which counter the opposing arguments.

Consequently, we may say that there is a capacity to reflect which, for its exercise and development, requires a core of belief which is the basis of the ontological and epistemological beliefs of an individual. As young animals we have certain basic desires and needs. We have a variety of instincts, though perhaps not many compared with other animals, which help us to satisfy those needs. We also appear to have been hard-wired in such a way that we are able to discriminate the world of our experience on the basis of natural kinds (Markman and Gelman, 1986; 1987; Kornblith, 1993). From this beginning we develop the ability to find a reason for action in the satisfaction of our needs and desires and it seems evident that we need something like this beginning in order to develop so. We can do this rather than that because this is more productive of the satisfaction of our desires. On an evolutionary basis it would appear that we are performing as the higher animals perform. We start to differ, as far as we know, from the animals when we start to develop the capacity to stand back from our needs and desires and judge whether or not they are worthwhile. We reflect rather than simply reason. The initial notion of worthwhileness is with reference to what it is worthwhile for me to do, what it is best for me to desire. And this is assessed against my small world and my place and the place of others in that small world. To put it another way, assessed against a developing core. I take account of others not merely myself. With development this world enlarges and the measure of what counts as worthwhile and rational enlarges and becomes more complex encompassing reflection on the past and consideration of the future. This is a process of education. The development of a reflective human being is not the development of, on the one hand, a capacity, and, on the other hand, a core of ontological and epistemological beliefs. They are interdependent. The aim of both informal and formal education should be to develop both. The capacity to reflect is developed through encouraging the child to imagine, to speculate, to ponder, to evaluate. The core is developed through the formal and informal processes by which children acquire their view of the world and their place in it. It is a mistake to think that the core is composed of empirical beliefs and moral principles. The human being is not that simple. Religion, myth and beliefs which are fervently held but for which there is no evidence may all be present. The 'Dreaming' of Australian Aboriginals plays a very real part in the core of those peoples. The Icelandic sagas still form part of the core of that culture (Smiley, 2000).

> What do you believe is true even though you cannot prove it? (Brockman, 2005)

And among the published responses of 120 scientists and thinkers is that from the physicist, Paul Davies, who believes that the universe holds other life forms though he cannot prove it:

> …I believe we are not alone because life seems to be a fundamental, and not merely an incidental, property of nature. It is built into the great cosmic scheme at the deepest level, and therefore likely to be pervasive. I make this sweeping claim because life has produced mind, and through mind, beings who do not merely observe the universe, but have come to understand it through science, mathematics and reasoning. This is hardly an insignificant embellishment on the cosmic drama, but a stunning and unexpected bonus. Somehow life is able to link up with the basic workings of the cosmos, resonating with the hidden mathematical order that makes it tick. And that's a quirk too far for me. (Brockman, 2005)

And in that reply it is possible to see elements of Davies' core.

Wertheim has argued that physics has always had a connection with religion. Discussing Einstein's view of the connection between physics and God, she says:

> One of his [Einstein's] most famous aphorisms is the oft-quoted "God does not play dice," which pithily summed up his objections to quantum mechanics. But there is another English translation of the same remark (originally uttered in German) that captures more fully the thrust of Einstein's thinking – "God casts the die, not the dice." Here was Einstein's philosophy of nature in a nutshell: The universe is divine creation, and it is the task of the physicist to discover the mathematical die from which it was cast. It is this essentially Pythagorean dream that inspires Stephen Hawking and others today. (Wertheim, 1997, p.185-86)

The core will be subject to attack either from others or from an individual's own observations of reality. The response to attack is defence. This is not to say that there can be no change in the core of a culture or a person, nor is it to say that change is not worthwhile. Clearly this can happen and does. The development from child to adult is witness to that. How then are we proceed so that the child has the best possible chance of turning into a reflective human being? This is the major question facing all researchers into education, all curriculum designers, all teachers, all parents and the culture in which the child exists. While dealing with their particular part of the jigsaw puzzle they must keep in mind the aim of their research, design or teaching. This is an absolute necessity too often honoured in the breach.

Consider one final example. The four cardinal virtues of the Middle Ages were prudence, justice, temperance and fortitude. Traditionally they have been considered to be the necessary possession of a flourishing human being. All these virtues have opposites. The opposite of prudence is rashness. Rashness is exhibited when a person, without forethought as to the possible consequences, acts. Forethought requires that we attend to the past, the present and the future and use our imagination to foretell the likely outcomes of our action. The actor who neglects to do this may or may not come to grief but, even if he or she is successful, we are entitled to claim the act was rash.

> Prudence presupposes uncertainty, risk, chance, and the unknown…Prudence is not a science; rather, it replaces science where science is lacking. One deliberates only when one has choice to make, in other words, when no proof is possible or adequate… (Comte-Sponville, 2003, p.32)

Prudence and rashness are possible in the type of situation presupposed. How do we develop one and limit the other? We start early with sayings such as – 'look before you leap'. All cultures have similar sayings. The problem then is to develop this beginning and in this the development of knowledge and imagination play a crucial part, for to be prudent is to be reflective. One cannot acquire the virtue of prudence without the ability to reflect and the ability to reflect depends upon the four things I enumerated earlier.

The narrative of a human life is called upon when we reflect on whether or not this should have been done or believed or whether or not we should do this or that, believe this or that. Reflection here involves who the human being is and his or her stance to the world. Reflection in any practice, whether it be science, fishing or painting, brings into play, for the reflecting human being, that human being's connection with and view of the world of his or her existence and the attempt to make and retain a coherent, consistent account of that world. It is to seek harmony. And this, John Keeves has always done.

References

Brockman, J. (2005). Edge. www.edge.org

Comte-Sponville, A. (2003). *A Short Treatise on the Great Virtues.* London: Vintage.

Deldridge, A. (Ed.). (1987). *The Macquarie Dictionary 2nd Revision.* Macquarie University NSW: The Macquarie Library.

Gibbons T. (2009) *Reflection, Science and the Virtues*. Rotterdam: Sense Publishers.

Kee, R. (1982). *Ireland: A History.* London: Sphere Books.

Kornblith, H. (1993). Inductive Inference and Its Natural Ground: an essay in naturalistic epistemology. Cambridge, MA: MIT Press.

Lakatos, I. (1978). Falsification and the Methodology of Scientific Research Programs. In J. Worrall, & G. Currie (Eds.), *The Methodology of Scientific Research Programs: Philosophical Papers Vol. 1.* Cambridge: Cambridge University Press.

MacIntyre, A. (1999). *Dependent Rational Animals.* Illinois: Open Court.

Markman, E. & Gelman, S. (1986). Categories and Induction in Young Children. *Cognition*, *23*, 183-208.

Markman, E. & Gelman, S. (1987). Young Children's Inductions from Natural Kinds: The Role of Categories and Appearances. *Children Development*, *58*, 1532-41.

McDowell, J. (1995). Two Sorts of Naturalism. In R. Hursthouse, G. Lawrence, & W. Quinn (Eds.), *Virtues and Reason: Philippa Foot and Moral Theory: Essays in Honour of Philippa Foot.* Oxford: Clarendon.

Smiley J. (2000). *The Sagas of the Icelanders: The Preface.* London: Viking Penguin.

Wertheim, M. (1997). *Pythagoras' Trousers.* London: Fourth Estate.

As the Wizard of Oz once said, "A heart is not judged by how much you love; but by how much you are loved by others." This book is testament to John Keeves, a wizard himself, and his generous heart, fine intellect and courageous approach to life.

Dr Katherine Dix
School of Education, Flinders University

This book brings together papers by 28 of John's former students and colleagues from around the globe, papers especially written as a tribute to their mentor and friend. That the papers encompass such breadth in content and research methodologies, demonstrates John's expertise as a scholar across a wide field and as an authority with an international reputation. The many chapters in this book will prove particularly useful to postgraduate students in education about to embark on the exciting adventure of research. Several chapters provide superb templates of the process of conducting educational research. Other chapters provide apt models for reporting and disseminating the results of research. Collectively, the various chapters point to the achievements of a unique individual - John P. Keeves.

Emeritus Professor Jonathan Anderson
School of Education, Flinders University

Shannon Research Press
Adelaide, South Australia
ISBN: 978-1-920736-39-2