The Seeker

Sivakumar Alagumalai
Murray Thompson
James Anthony Gibbons
Andrew Dutney

FLINDERS UNIVERSITY INSTITUTE OF INTERNATIONAL EDUCATION
RESEARCH COLLECTION
NUMBER 16

STUDIES IN COMPARATIVE AND INTERNATIONAL EDUCATION
Number 16

THE SEEKER

The wise man is he who knows
the extent of his own ignorance,
is he who knows how little he knows.

EDITED BY

SIVAKUMAR ALAGUMALAI
MURRAY THOMPSON
JAMES ANTHONY GIBBONS
ANDREW DUTNEY

FLINDERS UNIVERSITY
INSTITUTE OF INTERNATIONAL EDUCATION

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

Designed by Katherine L. Dix
Published by Shannon Research Press, South Australia
ISBN: 1-920736-11-5

Preface

We are designed to learn and grow by observing, doing and adjusting. Education provides the catalyst for such a growth, and for some the capacity to empower fellow seekers. Education here refers to the learning within, and for educators, the ability to deliver instructions to optimise learning. For an even smaller group, education includes the competence to undertake research to refine further the learning and instructional processes, which makes an additional contribution to children, youths and adults in society.

This book is dedicated to acknowledge and honour the work Prof John P Keeves. A seeker of knowledge, John is exemplary in highlighting the nexus between instruction, learning and research. John's diversity of learning experiences and contributions to students, colleagues and the broader community are highlighted through the broad range of articles in the book.

He constantly argues that learning is everyone's right and is a two-way process. Through learning, one interacts with the environment, activating both within thoughts and communications with the broader world. In the passion for learning, lies both the content and context of knowledge, which forms the basis for innovating instruction. John through his practice and beliefs highlights the importance of understanding the nature of knowledge and the nature of learning. Each challenge provides the cue for another learning experience, and an open-minded approach has allowed John to transcend both geographical and temporal barriers to make an even wider contribution to education.

Instruction (teaching) is part of the joys of learning. He constantly provides supervision and support to all those who seek his assistance, and is dynamic in his methods of instruction. Portals, interactive websites, online adaptive systems and simulations have been given further thrust through John. In the courses a number of us have undertaken, John constantly flags the need to 'anticipate the next generation of emerging technologies'. In a seminal lecture on the use of technologies in the teaching of statistics, John traced his learning from logbooks to handheld computing devices to calculators to interactive simulations. He has been exemplary in infusing technology with education, and highlights that learning is fundamental to instruction and not just the delivery of information alone.

Research is John's passion and life! Research is not about adopting a particular school of thought, and guarding it dogmatically through myopic one-off experiences. John, through his countless discourses and publications, highlights the unbiased, practical issues associated with conducting educational research and associated methodological pathways. He reiterates that researchers need to understand the relationship between epistemology, methodology, practice and reporting. He is not afraid to examine both innovative and sometimes controversial approaches to educational research.

The research articles in this book illustrate the need for understanding the 'design and structure' of social science research. John through his work and publication

highlights the need for a 'logic of research', and an understanding of research design is paramount to effective research. In his research experiences from his postgraduate work, through ACER, IEA and in a number of countries lies the basis for his belief in the learning-instruction-research nexus. The articles in the book are only snap-shots of his life and passion. We had the privilege to be able to learn, work and imbibe from his experiences. John, thanks for being a lighthouse, and for invigorating the passion for learning within us. We also thank you for providing us the base for connecting with the 'path you have come through'. We wish you a special Happy 80th Birthday.

Sivakumar Alagumalai
Murray Thompson
James Anthony Gibbons
Andrew Dutney

September 2004

Contents

Part 1

From School To University

Murray Thompson
Senior Secondary College, Inc.

There is no doubt that the contribution of John Philip Keeves to education and educational research throughout the world has been enormous. But what were the beginnings, what were the influences that led to this extraordinary and life-long career that transcends life-long learning into life-long mastery?

This section of this tribute to the life and work of John Keeves explores the influences in his early life, his family, his early schooling, and the beginnings of a career as a teacher at Prince Alfred College in Adelaide. We begin with some perspectives of John Keeves from his family, written by his nephew and namesake, John Keeves, with contributions from his sister (John Senior's niece), Wendy Keech (nee Keeves). This chapter explores some of the family and cultural influences that must have shaped his life and work and we begin to see insights into John's love of the outdoors, of sport and of course learning.

There follows a contribution from his old school-mate and life-long friend Professor Ren Potts. John finished school at Prince Alfred College in 1942 graduating from a class of extraordinary students, of whom Ren Potts is an outstanding example. Ren explores some of the experiences they shared and the emerging character of John Keeves.

Alan Dennis is a much loved former teacher from Prince Alfred College who taught John and was his Scoutmaster at the school. Some comments about John from a recent interview with Alan are included that make clear that John developed his sense of duty and his capacity for work at an early age.

In a series of interviews for the History of Prince Alfred College, noted Adelaide historian Ron Gibbs interviewed John Keeves about his experiences at school and a paraphrase of this material appears as Chapter Four. There is no doubt that John

continues to see Prince Alfred College as an important part of his life, both as a student, a young teacher and an enthusiastic old scholar.

David Prest, retired Principal of Wesley College in Melbourne, has made three wonderful contributions to this work. He begins with John's early teaching career at Prince Alfred College, when David himself was a student at the school. Some time later David returned to the school and worked closely with John as a colleague, thereby beginning an enduring friendship. This continued during John's time in Melbourne, when David recognized his talents and insights into educational issues by having John serve on the College Council.

Many of us have enjoyed John's generous hospitality at his much loved beach house at Port Willunga on the mid-south coast in South Australia and David, who with his wife Jean, has a house nearby, gives us some insights into John as part of the Port Willunga community.

Together, these contributions help us to understand some of the influences that have helped the shape the life of John Keeves.

1

Observations from a Family Perspective

John Storrie Keeves

***Dedications**: John's nephew and namesake, with contributions from his sister, Wendy Keech, reflects on the family influences that set John on the path of his extraordinary career. John Storrie Keeves is a prominent solicitor and barrister in Adelaide.*

Introduction

The purpose of this paper is to provide a slightly different perspective of John Philip Keeves – from the point of view of his family, both in the sense of his family history, and in the sense of our family experiences.

We are sure that John Philip Keeves - or Uncle John or even "UJ" as we call him – would want more than anecdotal evidence, would want some rigorous statistical analysis before we suggested even any tentative conclusions from the information presented in this paper. So it is largely left to the reader to draw his or her own conclusions on the influence of John's family history or circumstances have had on the man, his life and his work.

What's in a name?

The surname "Keeves" may be associated with the dictionary definition of "keeve" - an Old English word meaning a copper pot used for brewing or washing ore. It may well have been derived from the French "cuivre" for copper.

One quaint suggestion is that the name was, like many English surnames, a gift from the French at the time of William the Conqueror's Doomsday Book - a nickname - either for a redhead or a person who resembled a vat of … ale.

However, research by a distant family member suggests a more complex story. It seems that the surname only appeared - in various spellings - Quive, Queeve, Keev, Keevs, etc. - after the massacre of St Bartholomew (1572). The explanation offered is that the Anglicised name derives from the name of Huguenot refugees escaping from France at and after this time, some of whom went to England, bringing, among other things the silk weaving trade. The Hugenots were known to be an industrious and enlightened group.

The most promising start to the family chain was apparently a reference to one Lorezia Quivia “a widow of the French churche” in the year 1582.

There is a suggestion that “Keeves”, if derived from the French, could perhaps be related to the French sentry’s challenge of “qui vie” - or the Latin “quivis” - “whoever” or "whatever you please", both interesting possible “tags” for a Huguenot refugee.

William James Keeves

John's great-grandfather, William James Keeves, came to Australia in 1876, with his family including two sons, William Gill Keeves (John's grandfather) and John Gill Keeves. William James Keeves had originally settled in Fort William in Canada, before returning to England and later emigrating to South Australia.

William James Keeves’ family in London operated a tinsmith business that at one time had around 150 workers and boasted an illustrated price list of 688 pages. The business was later described as “tin and iron plate workers, hardware merchants and exporters”. The firm had been started under the name “John Keeves” in 1848.

William James Keeves left the firm in 1852 at around the time of his marriage to Emma Gill, before migrating first to Canada and subsequently to Australia. He established his own tinsmithing business in Rundle Street, Adelaide and lived across the Adelaide Parklands in Finnis Street, North Adelaide. The business was located in the East End of Rundle Street, not very far from the campus of the University of Adelaide.

In those days, Adelaide was a British colony, but the only Australian colony not to receive convicts. The settlers came by choice, and Adelaide became known as the "City of Churches" due their number and variety, evidence at the time of religious toleration and diversity.

The colony was premised on political and social freedoms and enjoyed the promise of representative government from the beginning – the reality occurring when the colony had 50,000 settlers.

South Australia, while conservative in some ways, has been very liberal in others, being, for example, one of the first places in the world to provide votes for women in 1894 but the most liberal – all women had the right to vote. The University of Adelaide admitted women as students from 1876 and legislation was passed to allow women to take degrees – apparently against the wishes of the British Government.

William Gill Keeves

William James Keeves died in 1890. William Gill Keeves and John Gill Keeves then operated the business before selling it to a rival firm, Simpsons, perhaps an early example of an anti-competitive merger. Simpsons became well known in Australia as a whitegoods manufacturer in later years.

Following the sale of the business, William Gill Keeves and John Gill Keeves became share brokers, sharing a seat on the Adelaide Stock Exchange from 1894, that was held in John Gill Keeves' name. It seems that William Gill Keeves at least lost some funds on mining speculation in the mining boom and crash of the 1890s and this apparently affected his son Ernest's attitude to investing in mining shares – he wouldn't touch them.

Ernest William Keeves

John's father, Ernest William Keeves was born in 1890.

John's grandfather, William Gill Keeves died at age 42 in 1895 when his son Ernest William Keeves was aged 5. John Gill Keeves retained the seat on the Stock Exchange and Ernest and his family were not well to do. Ernest was required to leave school early at the age of 11 years, to help support his mother, Elizabeth, who was from the Gill family and who had married her first cousin, William, after the death of her first husband (Snow).

His father's early death, and his enforced early departure from school certainly affected Ernest; he resolved that his children would have the opportunity to receive the education that he had missed, and each of his three children had the opportunity in the 1940s to attend university, and later study overseas.

His life was twice interrupted by war service, first in the Field Ambulance in Egypt and France in World War I – of which he, like many other soldiers, never spoke, and then in the Australian Army Medical Corps in World War II at the Woodside Army base, in the Adelaide Hills, then Springbank in the southern suburbs of Adelaide, which became the Daws Road Repatriation Hospital (now known as the Repatriation General Hospital).

In an obituary, Ernest was recognised as the "Patron Saint" of the Daws Road Repatriation Hospital. During the Second World War and afterwards Ernest looked after the soldiers at the Hospital as Amenities Officer. When he was discharged from the Australian Army after the end of the Second World War he immediately joined the Australian Red Cross and continued his work with the Red Cross at the Hospital until retiring in 1959. Thereafter, he continued as a volunteer worker at the Hospital until his death in 1972 aged 81. He was known as someone who was not concerned by "red tape", but would press on regardless to do something he thought ought to be done. Ernest was a well-known figure in ANZAC day marches (to commemorate the landings in Gallipoli in World War One and remember those lost in war), as a member of the Army Medical Corps and marched each year until his death.

Ernest was, however, fortunate that he was able to pursue this vocation. Originally employed by Harris Scarfe, then as now a well-known Adelaide retailer (much more recently known for financial difficulties, but now recovered and still operating under that name). He left the employ of Harris Scarfe in 1934 during the Great Depression that affected Adelaide as it affected much of the rest of the world.

He then became involved in a venture that imported china from Europe. However, the Second World War intervened and the business failed. During the early war years, Ernest was lucky enough to run into his former commanding officer, Dr F. N. Le Messurier, in the street. Dr Le Messurier gave him the opportunity to serve once again, and he joined the Army Medical Corps initially to assist with the establishment of a hospital facility at Woodside.

During this period in the early 1940s Ernest's uncle John Gill Keeves passed away and Ernest was fortunate to inherit some assets, including the seat on the Stock Exchange held by John Gill Keeves. The seat on the Stock Exchange was subsequently sold, neither of Ernest's sons (John or younger brother Andrew) being interested in stockbroking. There is a family story that the seat was sold to the famous cricketer Sir Donald Bradman (who at that time was setting up as a stockbroker on his own account) but Stock Exchange records do not support this tale.

Ernest was also fortunate to be able acquire during the war years a relatively large (6,000 m^2) property in suburban Leabrook, east of the City of Adelaide, enabling him to pursue his love of gardening. He was lucky enough to have funds available, and was able to acquire the property at a keen price at that time during the war when there were not many buyers. The family had previous lived in Kingswood, south of the City and then Medindie, just north of the City.

Ernest was, apparently, somewhat disappointed that neither of John nor his brother Andrew were not more interested in the garden, rather than sporting or academic pursuits.

The family home was in Statenborough Street, not far from the site of the original and famous Coopers Brewery, and it is said that Ernest would from time-to-time visit former soldier friends and, despite his wife's tee-totalling disposition, have occasionally imbibed a glass or two.

Jonathon Job Nicholls

John's other grandfather was Jonathon Job Nicholls, a Methodist Minister, born in Gunnislake, Cornwall, England in 1855. J.J Nicholls' mother (Mary Job) was said to have come from an old Methodist family. J.J. Nicholls' father died when the family was young – leaving nine boys and two girls. J.J. Nicholls arrived in South Australia in 1883 and subsequently brought his wife (nee Celia Philips Hodge) out from Gunnislake, Devon in England before being ordained in 1888.

He worked as a minister at many places in country South Australia including Yorketown, Wilmington, Yankalilla, Naracoorte, Riverton, Two Wells and Clarendon. He was, it is said, a free-thinker, pacifist, a supporter of the Federation of the Australian States (not necessarily overwhelmingly supported in South Australia) and argued against many social ills.

J.J. Nicholls' son, Jonathon (John's uncle) became a school teacher and later High School Headmaster at Gawler and Woodville High Schools.

J.J. Nicholls died in 1914 from tuberculosis after a long illness.

Ella Annie Nicholls (Keeves)

J.J. Nicholls' daughter Ella Annie Nicholls was John's mother ("Granny K." as we knew her).

We recall her as a somewhat stern and forbidding grandmother, who kept a jar of ginger nut biscuits on the kitchen mantelpiece and a crate of the famous South Australian "Woodroofes" lemonade in the cellar, the latter being alleged to be "good for the blood".

Ella was a kindergarten teacher, having been an early student at the Adelaide Kindergarten Training College under Lillian De Lissa, graduating in 1910. Ella started her own kindergarten school in Rose Park, after having assisted at Methodist

Ladies College following completion of her schooling, and having worked at a kindergarten at Hamley Bridge in the South Australian country. Her students at Rose Park were said to include a young (later Sir) Robert Helpmann, later to become one of Australia's most famous ballet dancers.

Ella remained involved in the kindergarten movement for many years. She was a member of the Education Committee of the Kindergarten Union (later the Pre-School Council) from 1931 to 1969.

Their daughter, John's sister Margaret (Trembath) followed in her mother's footsteps and worked with the Kindergarten Union for many years, including as Director of the Kindergarten Union Teachers Centre, after achieving BA and Dip Ed from the University of Adelaide. Margaret, like John, earned distinction as a Fellow of the Australian College of Education and leader in her field.

John's brother Andrew did not follow his older siblings into the field of education, but chose forestry, achieving distinction as a Fellow of the Institute of Foresters, the recipient of the Norman W Jolly Medal, and a Public Service Medal. However, even Andrew taught, including a forestry course in Mount Gambier in the South East of South Australia, where he was based for many years.

Ella's friends were often connected with the kindergarten movement and often quite radical, including the South Australian artist Dorrit Black. Ella herself was radical for her time, a feminist, socialist and pacifist.

Ella and Ernest met through the Walkerville Methodist Church, Ella's father having been involved with the parish during the building of the Church that survives in Smith Street, Walkerville. J.J. Nicholls laid the foundation stone on 13th April 1912. Ernest lived in Gilberton (adjacent to Walkerville) and played cricket and Australian rules football for Walkerville at the Walkerville Oval a few hundred metres down Smith Street from the Church.

John Philip Keeves

John's early life would have been much influenced by his mother's serious illness – tuberculosis - that occurred in the late 1920s during her pregnancy with John's brother Andrew and afterwards during Andrew's infancy. The fact that her father had died a few years before as a result of this condition is notable.

John was sent to Wilderness School in Medindie at age 3½ due to his mother's illness - an early form of childcare. At that time the family lived in Nottage Terrace, Medindie and so Wilderness School would have been geographically convenient.

At around 5 years of age John went to the Queens School in Barton Terrace, North Adelaide. The Queens School no longer exists.

He did not enjoy the climate of the school and afterwards went to live with his cousins and Ella's sister and her husband – the Bartholomaeus family on Pine Creek Station, near Broken Hill for a time. This was a much happier period for John, and he often returned to Pine Creek for school holidays in later years.

John started at Prince Alfred College in 1934. No doubt elsewhere is in this book much will be said about the influence of PAC on John and his influence on the school. There is no doubt that his relationship with the school has been one of the constant themes of his life.

It is worth recording what John has said of the influence of Norman Mitchell who joined the PAC preparatory school as Headmaster around the time John started, from

Trinity Grammar School in Melbourne. According to John, Mitchell transformed the school into an exciting place – children writing poetry, painting, drawing, acting in plays, recitals, and attending symphony concerts. As much as this seems uncontroversial for a leading boys school today, this was very radical for its time, and had an enormous positive effect on the boys – including, or especially, on John Philip Keeves.

It may also be recorded elsewhere in this book, but John was the first captain of the PAC preparatory school, and held that post for an unusual second year – due to a reorganisation transferring a year group from the senior school to the preparatory school, John spent a second year in the senior year at the preparatory school.

John has recalled that Saturdays were work (presumably school work) and sport and Sundays were church and Sunday school. John was a keen sportsman, playing Australian rules football for PAC and the University of Adelaide.

John chose science at the University of Adelaide, and did Honours in Physics before returning to Prince Alfred College in 1947 as a teacher. John has acknowledged that when he first started teaching he was very "green" and has since felt sorry for the students in his early teaching days.

He later went to teach at Radley College in England before returning to PAC and later becoming boarding house master. In those days, as now, young people from the country inhabited Australian boarding schools.

Perhaps surprisingly in contrast to his academic focus, John was and is a keen outdoorsman. He was heavily involved in scouting activities as a student and became the Scoutmaster at PAC, leading many expeditions, including a number in the Flinders Ranges in the north of South Australia.

Why Education?

Why did John choose education as a career?

He has said that his mother was a teacher, his uncle was a teacher, and he enjoyed his time at school, so it seemed an easy choice. Also, he had not enjoyed his experimental work in Honours Physics at the University of Adelaide, and the headmaster of PAC, J.F. Ward, had asked if he was interested in teaching. So teaching seemed a natural step and John thought he could teach for a couple of years then go overseas (and he indeed did so).

John's career at PAC is no doubt dealt with in detail elsewhere, but he is remembered fondly by his students – no least the boarders from his years as Housemaster, and the Scouts from his many years leading expeditions as Scout Master. He has been lucky enough in recent years to attend many class reunions of his former students, who have been delighted to see him.

But the turning point in his career came. He left PAC to join ACER – the Australian Council for Educational Research – and his interest in maths and science testing was triggered. This was to become a key to his later work. Studies in Education at the University of Melbourne followed, then a research appointment at ANU and a PhD, and later appointment as Director of ACER, and a second PhD from Stockholm University.

After his retirement from ACER, he continued his work at Melbourne University, and later (and still) at Flinders University. He has recently indicated that he is to return to

his alma mater to teach at the University of Adelaide. Hopefully John will not cause too much confusion by being the second John Keeves to teach at the University of Adelaide - although part-time undergraduate teaching in Corporate Law in the Adelaide Law School by one of us pales in comparison to John's teaching record. However, it has been amusing, from time to time to have received e-mails from both overseas and in Australia addressed to **john.keeves@flinders.edu.au** that had gone inadvertently to **john.keeves@adelaide.edu.au.**

One of the particularly interesting aspects of his work has been his travel to far-away places to undertake research and reviews - including seeing schools in desert provinces of India and for street kids in South America. To hear of his experiences in these places, where the educational opportunities we take for granted would be a dream, is at once sobering and uplifting.

John has had, to date, a remarkable career. He continues to work, to write, to teach and to supervise research students - all this at the age of 80 years.

We speculate that he has long believed in the importance of education - that education can make and has made a difference to many lives, created opportunities, given freedom. We think he believes in mathematics and science education as being of primary importance - a belief in the scientific method, scientific - statistical - rigour and scientific progress. And we think he believes that researching and measuring (and hence improving) mathematics and science education is thus the best way to make a difference and improve the society that he lives in. Or perhaps he just believes in doing his work to the best of his considerable ability. Or both.

"Uncle John"

John has been a kind and generous uncle to his nieces and nephews, the sons and daughters of his brother Andrew – and to our children, John's grand nieces and nephews, numbering 14 in total. He was also generous in many ways to Andrew, to his sister Margaret and no doubt to his parents as well.

He sometimes seems somewhat other-worldly, and largely focussed on his work, although having time for his family. On one occasion he was able to mix work, travel and family to appear at a family wedding in Whistler, Canada in 1992 and to visit Boston, Mass, in 2003 to join family for a terrific 4th of July.

In his later years he has amazed us with his continued hectic pace of life, work and travel, and impressed us with his summer exercise regime at Port Willunga – walking and swimming – almost daily without fail. We speculate that his continued keenness of mind is supported by his continued mental and physical activity.

Expeditions to John’s beach house at Port Willunga - his summer home for many years - are a part of our family life and memories. Port Willunga has one of the best beaches in South Australia (some would say the best in Australia). John's house is on top of the cliffs overlooking the sea, with a view down to "Gull Rock", the goal of John's daily constitutional of several kilometres, followed by a swim in the summer months.

We think of Port Willunga especially for the annual family get-together, where his nieces and nephews and great nieces and nephews gather for a barbeque, followed by an afternoon on the beach.

He has always been an intelligent and articulate – thoughtful and considerate – companion and conversationalist, with great insight. But while thoughtful, he is also

quick to laugh, to see humour, and not overly serious or dour. We cannot recall him ever uttering a rude or unkind word – he is a gentleman and a gentle man.

Conclusion

So we suspect that the John Philip Keeves that we family members know is the very same man that is known, admired and respected by his friends and colleagues in Australia and overseas. There is no doubt but that he deserves the recognition that this book gives to him on the occasion of his 80th birthday.

2

Student Days at PAC

Ren Potts

***Dedications**: The class of 1942 from Prince Alfred College was by any measure an extraordinary group of young men. At a time when the country was at war and there were serious shortages both of people and resources, this group of young men emerged. Within the group were outstanding academics, such as John Keeves himself, a pioneer heart surgeon, Dr John Waddy, to mention just a couple. One of these students, Renfrey Bernard Potts became a Rhodes Scholar and went on to take the Chair in Applied Mathematics at Adelaide University. Ren has provided an insight into what he affectionately calls the "Pug Days".*

Introduction

I entered PAC in 1935 in the Prep School. My father 'Gympie' was a master at the school, which was his life and passion; my elder brother was in the Big School; and I had already regularly attended Intercoll sport fixtures.

I find it hard to call John Keeves John, because he was and is still called 'Pug' by his school mates. He had entered PAC a year before me in 1934. We were for eight years in the same class, from 1935 in Senior B to 1942, our second year in V1u.

Since we left school our paths have crossed and recrossed. We both enrolled at the University of Adelaide in 1943, John for the BSc course and I in the war emergency BScEng course. After two years, when the emergency course was discontinued, I switched to the BSc course and John and I both graduated with BSc degrees in 1945. John continued with Honours in Physics while I took Honours in Mathematics.

We were together again in Oxford where I was studying for my DPhil in Mathematical Physics and John was teaching at nearby Radley College. Ralph Parsons, another of our class at PAC, was also studying at Oxford, and he, John, I and

my fiancee Barbara Kidman together invested in a 1935 Hillman Minx. John was best man at our wedding in Oxford in 1950. Barbara and I toured Europe in the Hillman for our honeymoon, and on our return to Calais, John and Ralph took over the car for their tour of Europe. With this arrangement we saved the cost of a second channel crossing for the car.

John and I both established our academic careers in Australia, John's in education, with national and international fame in mathematical and science education, and my own in applied mathematics.

It is a privilege to have been invited to contribute to the celebration of John's 80th birthday. I have been asked to comment on the impact of PAC on John as a student, and I do so under three headings, Leadership, Scholarship and Friendship.

Leadership

PAC provided John the opportunity to develop and enhance his signal leadership abilities.

John was a year older than most in our class, and in the Prep School, bigger and taller, as photos of school teams illustrate. In 1936, when we were in the top Prep Class Senior A, the new Prep Headmaster Norman Mitchell transformed the Prep School. Recognizing John's potential as a leader, he appointed him as the Prep School Captain. John took the responsibilities of the position with conviction and benefited greatly from his close interaction with Mitchell. In 1937, the third year form from the Big School was moved to the Prep and John was appointed Prep School Captain for a second year. John excelled in the classroom and on the sports fields and as Captain had the strong support of his class mates.

In the Big School, John was selected for service on a variety of committees including the Library committee, the SCM committee, the Games committee, President of the Debating Society, culminating in his appointment as a School Prefect in 1941-2. It was his service to the school Scout Troop which greatly fostered his leadership skills. He was a member of the Troop from 1936 to 1942, became a King Scout in 1940, and Troop Leader in 1942. He shared much of the control of the activities of the Troop with the Scoutmaster Alan Dennis. John helped establish high standards for the Troop and was firm in his insistence that these standards be met.

An Old Red, Cecil Plush, owned fruit blocks at Winkie on the River Murray and each year he would employ mainly PAC boys for the apricot season. It seemed natural that John should take responsibility for the organisation of a contingent of class mates to spend three or more weeks picking and drying Plush's apricots, sleeping in an orange packing shed and preparing meals. With his scouting experience, John successfully introduced rosters for cooking and washing up, and controlled the expenditure on food to nine shillings (a day's pay) per person per week with, I must admit, menus that mainly featured apricots. I well remember one meal: main course, apricot pie, with a crust which burned black and had to be thrown out; dessert, stewed apricots, barely distinguishable from the main course. John was cheerfully accepted as in charge – except perhaps by a couple of Saints boys who did not take so kindly to toeing John's PAC line.

There is no doubt that John's leadership skills were nurtured by the opportunities PAC provided.

Scholarship

The second major impact of PAC on John as a student was the school's emphasis for scholarship at the highest level. Unashamedly this was measured by performance in the public examinations, the Intermediate in the fifth forms, the Leaving in the sixth forms and the climax – the Leaving Honours in the upper sixth form. In the Prep School and throughout the Big School, this scholarship was underpinned by extensive homework, regular tests, and term exams with numerous handsomely bound books as prizes awarded on the annual Speech Night. The competition among the top students was extreme, stimulating, but almost cut-throat in the Leaving Honours where just twelve bursaries were granted giving free entrance to the University. For many students a bursary was the only way to get to the University. The resulting intense State-wide competition forced students to repeat the Leaving Honours year once or even twice.

In this blatant pursuit of top exam results, gradually the humanities and foreign languages were dropped by most students in favour of Maths I and II, Physics and Chemistry, with English the competing outsider and often a stumbling block. The challenge to the teachers was not only to deliver top results but also to stimulate interest. Two outstanding teachers, Spenny Williams in Maths and Ray Smith in Physics and Chemistry, were to achieve these two goals and they were to greatly influence John.

Spenny Williams was an animated slave-driver, orderly, conscientious in the extreme, who kindled our excitement and enthusiasm for his subject. He had his idiosyncracies; for example, he carried with him bundles of students' papers wrapped in newspaper, shaved while revising a table of formulae placed by the mirror to save wasting precious time, and had choice sayings such as "Grasp the opportunity of a lifetime at the lifetime of the opportunity". His occasional 100 minute Saturday morning lessons were gems, and he excited the Leaving students with the promise that in Leaving Honours their eyes would be opened – to the Calculus. We could hardly wait! What an impact on John! He responded by coming top of the State in Leaving Maths 1. John's subsequent internationally famed work on mathematics education rests squarely on the shoulders of Spenny Williams.

Ray Smith was a top teacher of Physics and Chemistry. His Leaving Honours Physics notes were sought after by students from other schools and he had written definitive texts on school chemistry. His teaching gave opportunity for individual expression in laboratory work, and the projects he set in Leaving Honours for the repeating students were challenging and exciting. John, and another class mate Bob Carter, shared a home chemistry set, and I would often join them out of school hours to experiment and try to satisfy our curiosities. With Ray Smith we were well prepared for the Leaving Honours exam and for University courses. Again John thrived on the teaching and it proved a firm foundation for his degree in Physics and his work on science education.

And what was the culmination of the impact of PAC scholarship on our class? In the Leaving Honours exam in 1942, no less than seven of the twelve available bursaries were won by PAC students – John and I among them. An eighth PAC student, John Waddy, had come top of the exam but had already won a bursary the year before. It was a record breaking result. What an impact!

Friendship

For us as students, PAC was a friendly school. As John's involvement spread from classroom activities to sports, to extra-curricula groups and especially to scouts, he established friendships with a wide spectrum of students. These friendships have stood the test of time, facilitated by the strength of the Old Collegians' Association. Our class mates have by age qualified for attendance at the annual Old Old Boys' Assembly and the Vintage Reds' lunch, and thanks to Don Davies, John and I meet for lunch every two months with Don and Scott Dolling, Reg French, Jim Gurner, Ross Reid and John Waddy.

Conclusion

As John reaches the magic age of 80, all his colleagues will share with him his satisfaction and pleasure in his continued active participation in science and mathematical education. John is still a leader, still a scholar, still a friend, and I am sure he will agree that he owes much to his student days at PAC.

3

Prince Alfred College 1934-1977

Murray Thompson & Alan Dennis

***Dedications**: One of John's teachers at Prince Alfred College was Mr Alan Dennis. Alan taught Latin, French and English over his 43 year career at Prince Alfred College, which began in 1934. He was also for a time the Scoutmaster of the school Scout Troop, and John was the troop leader. Alan is still fit and healthy and remembers John as an excellent student and later as a colleague. What follow are a few comments that Alan gave in an interview in January 2004.*

Reflections

"He was always willing to have a go at anything and to take on the responsibility of anything you had a chance to give him. My first memory of him was a memory of a Prep School Concert where he was figuring as the Judge in Gilbert and Sullivan's "Trial by Jury". He was the judge. At that stage, I didn't really know him because he was in the Prep. School. I was teaching in the upper school, but for some reason it has stuck in my mind. When John got into the big school, he joined the scouts and was eventually, I think, Troop Leader."
"… he was a natural leader. He put his best into everything."
"… he had a very strong sense of obligation and duty."
"Well, it's always hard being a teacher, but teaching that lot was very rewarding."

"… he always was (strong) - well built, healthy. There were never any doubts when you left things to him … We used to hike into Mylor and possibly Bridgewater. For that you had to go down the hill and cross over Cox Creek. He was tall, he was reliable and forever cheerful."

"He has a tremendous capacity for work hasn't he?"

4

John's Reflections of PAC and beyond

Ron Gibbs & Murray Thompson

***Dedications**: Insight into John's own feelings about his experiences as a student at Prince Alfred College. This material has been taken from an interview recorded by Ron Gibbs, historian for Prince Alfred College in which John gave his impressions of Prince Alfred College and its influence on him.*

The beginnings

John Keeves came from a home that he describes as basically Methodist. His grandfather was a Methodist minister and like many of his contemporaries there was not an overabundance of money. John Keeves commenced his first steps along the path of learning when he entered Wilderness School at Medindie in Adelaide. He remained there for two or three years and this was followed by two years at Queen's School in North Adelaide. However, as soon as he was old enough to travel through Adelaide by tram, he began what was to become a life-long association with Prince Alfred College.

Prince Alfred College, Adelaide

Prince Alfred College is a long established independent boys' school operating within the Uniting Church. It was established in 1869 as part of the Methodist Church and in 1934, when John Keeves enrolled it had established a strong reputation as an academic and sporting school with a proud tradition. Traditionally, Prince Alfred College has drawn its students from those not necessarily particularly wealthy. The school tended to work with the sons of merchants and an emerging professional class

rather than those who had inherited wealth. For these families, education was an opportunity to enter the professions.

> "So the school took boys who came mainly from Church homes, who were I guess, honest, hardworking, generally liberal, generally had some sense of service to the community through a professional life."

It was always expected that John would become a student at Prince Alfred College. His uncle had attended the school and his aunt Winnie Nicholls had taught in the primary school, traditionally known as the Preparatory School. He began at Prince Alfred College on the twelfth of February 1934, entering at year 4 level, known at the time as Senior C. He had happy years there in 1934 and 1935. It was a good and friendly school and he enjoyed his involvement in Australian Rules Football and Cricket as well as the challenges of the classroom.

An important inspiring teacher - Norman Mitchell

However, it was in 1936 that a new Head of the Preparatory School was appointed, Norman Mitchell. He transformed the school. New emphasis was given to music, speech training, plays and poetry writing. John became inspired with the new opportunities he was being offered and was particularly taken with learning how to draw, especially how to draw maps.

An important development in the school at the time was the upgrading of the Preparatory School library. In 1937, John Keeves was appointed the first Captain of the Preparatory School, a position that has remained at the school since that time and in 2003, he was asked to open the newly developed Preparatory School Library complex in recognition of his role as the first Captain of the Preparatory School. Under the leadership of Norman Mitchell, the Preparatory School grew, taking in more students and teachers. The development of the Preparatory School was important part in the redevelopment of the educational policies and philosophies of Prince Alfred College, undertaken at the time of the depression, bringing to the school the best of progressive education that was available in Australia. John himself upon reflection speaks very highly of Norman Mitchell and there is no doubt that he was a very important influence in his development. It is perhaps a pity that it was not until many years later that John came to realize the influence that Norman Mitchell had had on his life. Pivotal in coming to this realisation was a discussion with W. H. Frederick, who had been Headmaster of Wesley College, Melbourne and went on to become Professor of Education at the University of Melbourne. Frederick was a man of great standing in educational circles and his comments about Norman Mitchell caused John to reassess his estimation of him.

> "It was only when a man like W. H. Frederick praised him so clearly and so strongly, not in a situation where it was solicited or required. It was a spontaneous praise for him, that I realized if a man as liberal and as strong an educationist as Frederick praises somebody, it's not given lightly."

Other influences at school

A number of other "masters" greatly influenced John during his time in the Preparatory School of Prince Alfred College. In Year 7 he was taught by Maynard Close who became the Head of the Preparatory School and whom John described as a wonderful man and wonderful fun. "And I just worshipped Maynard Close."

Sport and involvement in other activities were important to John, including the Student Christian Movement, debating, but especially important for him was his involvement in the Scouting movement. He joined the Scouts in 1936.

"There was the Scout Troop that gave a sense of adventure, an opportunity for adventure activities, an alternative to sport." John spoke very highly of his two Scout Masters of the time, Alan Dennis and Maynard Close.

> "If you can pick out the two best of the junior staff members, Maynard Close and Alan Dennis, and assign to them job of running the Scout Troop, then you guarantee success."

Throughout John's time as a student at PAC, the Headmaster was Mr J. F. Ward. By all accounts he was a hard and demanding man, not often given to humour and often feared by his students and staff. However, John who knew Fred Ward both as a student and as a teacher, speaks of him as one of the great Headmasters. It must be remembered that these were very difficult times. Ward came to the school just before the depression and managed to lead the school through the trials of the depression and the ensuing war years, building a very talented and highly qualified staff and supporting them with a very strong discipline structure. A very significant number of the staff held Masters degrees, rare in those days. The school achieved outstanding results in the Public Examinations.

During this period the school had outstanding academic results, due in no small part to the leadership of J. F. Ward.

> "He was a liberal radical man in many respects. He argued, I can remember on one occasion, that he hoped no boy would leave the school without recognizing that on every issue there were two sides to the question. He saw the complexities. He argued for a liberal approach. He was committed to hard work, to ideals of service to the community. He ran the school, the administration of the school, single-handed. He steered it through the depression. He worked extremely long hours; that in his study till ten o'clock every night checking the detail himself. He had an eye for detail. He held the detail. He carried the detail. He ran the school in a way that the fees were held back. He ran it economically because he believed that the opportunity for some parents to send their sons to the school would be denied if he allowed the fees to rise. It was a deliberate policy to make the school accessible to the Methodist community of the State who sought to send their sons to the school."

One of the most influential of the teachers for John was the mathematics master "Spenny" Williams. Williams was a meticulous and methodical teacher who encouraged his students to solve problems in a systematic and thorough manner. He was instrumental in building the strength of the Prince Alfred College Old Collegians Association, currently the largest such association in the Southern Hemisphere. "Spenny" Williams has also been credited with the reshaping of the sporting program of the school. It was sad, therefore, for John, that when J. F. Ward retired and was replaced as Headmaster by John Dunning, that "Spenny" Williams resigned because of the lack of discipline that he perceived in the school.

The Leaving Honours Class of 1942 was by any measure an extraordinary group of young men. Prominent names like the pioneer heart surgeon John Waddy, the Professor of Applied Mathematics at Adelaide University, Ren Potts, the general surgeon Lehonde Hoare, the President of the Methodist Conference in South Australia, Rev Don Howland and my own father Rex Thompson, an analytical chemist who became a Churchill Fellow for his exploits in violin making, were all part of this outstanding group. The academic ferment, combined with outstanding

teachers such as “Spenny” Williams and Ray Smith must have been inspiring and one can imagine that this experience must have been influential in John’s ultimate choice of career.

In 1942, Professor Gartrell, Professor of Mining and Metallurgy at the University of Adelaide visited the school and impressed upon the students the importance of pursuing their studies and in particular following a career in the sciences which, if the war continued would be vital.

The University of Adelaide

In 1943, John Keeves enrolled at Adelaide University studying Science, and found himself soon working with uranium salts doing spectrographic analysis. In retrospect, John surmised that there was at that stage an awareness of the development of the atomic bomb and completed a First Class Honours degree in Physics in 1946. With a strong background in science, one might have expected that John would pursue a career in this direction but he felt that he was not particularly good at practical and investigatory work in Physics. It was during 1946 that John informally met J. F. Ward who said, “John, when the time’s ready, come and have a yarn.” Thus began the teaching career of John Philip Keeves.

5

Teaching Days at PAC 1947-49, 52-56, 58-61

David Prest

***Dedications**: It was when David Prest was a student at Prince Alfred College and John Keeves was a young teacher, that they first met, although John never actually taught David. It was a few years later, when David was appointed to the staff of Prince Alfred College that they began a friendship that has endured over many years. It was no wonder then that when David was the Principal of Wesley College in Melbourne, that he sought to appoint John Keeves as a member of the school's council. We are delighted that David has made three important contributions to this work.*

Recollections

John Keeves' association with Prince Alfred College (PAC) has been long and distinguished. Student from 1934-42, teacher during the middle part of the century and generous financial supporter, he is constantly aware of and interested in the overall condition of the school. He is a member of the PAC Hall of Fame, was invited to open a new building in 2003 and three Prep School awards are named in his honour. Curiously, and not to the school's credit, he was never a member of its governing Council. Wesley College Melbourne did not let this opportunity slip; he was a crucial member of the Wesley Council for about 10 years.

In the early seventies John started coming to dinner at 7.00pm on Monday nights at our home – and still does whenever we are at Port Willunga or he is in Melbourne. He has a sixth sense regarding our presence at the beach and unfailingly appears at 7.00pm on the first Monday after our arrival. There is a bang on the door at 7.00pm, not five to or five past, we watch the ABC news, we then eat and talk and he leaves at

9.30pm. These days he is always early to bed and early to rise; I have no idea whether he is wealthy but he is certainly healthy and wise. Remarkably so in both respects.

During Monday nights the discussion invariably turns, at least for some of the time, to PAC. As has been noted elsewhere John has an extraordinary memory and can recall great detail about the past and, indeed, about recent events. Not all of us are good at that. There is no doubt that the reason he became a teacher centres on the experiences he gained and the personalities and characters of teachers he observed at his old school. Maynard Close, in Prep School days, and "Spenny" Williams and Ken Smith in the Senior School were probably the most influential. Ken Smith was quite pro-active in this respect; he gave books such as *The Young Schoolmaster* to boys whom he thought should enter the profession.

After 4 years at Adelaide University John began his teaching career at PAC in 1947. He was 22, full of energy, enthusiasm, bright ideas and idealism. As such schools are inclined to do, much to their benefit but not necessarily to the benefit of the teacher, John was given a very high work load. In his first year he assisted with the Scouts, coached football and athletics, lived in the Boarding House where there was always much to attend to, and taught full-time including classes at Leaving Honours, now Year 12. And in a fit of misplaced generosity he allowed himself to be appointed Secretary of the Old Collegians Football Club. I have never heard him complain about his lot. This hectic school program continued during 1948 and 1949, with the added responsibility of being in charge of Scouts but without the madness of the Secretaryship of the Football Club.

But further education called. It was obviously essential for John to learn more about education and he did this in style. Brian Hone, who was Headmaster of Melbourne Grammar School, and an old boy and former master at Prince Alfred College, arranged a two-year teaching appointment at Radley College, a leading English Public School. Subsequently, the Warden of Radley College facilitated through K.A.H. Murray, the Rector of Lincoln College, Oxford for John to study for a diploma of education in Oxford, which John completed with distinction. This must have been a very stimulating time for John; he had had a sniff of schooling at PAC and now there was before him a chance to widen horizons and think about some of the general and international issues. Again one can imagine the continued expansion of John's views on educational matters.

Surprisingly perhaps, John returned to PAC in 1952 for his second period of service at his old school. In my opinion this period, 1952 to 1956, was the most significant of his time at PAC.

His duties were similar except that he was in charge of the Boarding House as well as the Scouts. He was ready for a further break at the end of 1956 and again went to England, this time to work with Encyclopaedia Britannica in London. Jean and I were in England at this time and he visited us at Southampton and later at Birmingham.

Refreshed he returned for his third term at PAC, 1958-61. This time he took a more measured approach, although he did include the Science Journal Committee in his extracurricular duties, but he sensibly left the very demanding PAC Boarding House behind him. In those days, indeed for decades before, the PAC Boarding House had been far too large for proper pastoral care and the Master-in-charge of the House faced a daunting task.

Eventually, of course, PAC did break the House down into more sensible components, much to the benefit of all concerned. And to explain or excuse the earlier arrangements for boarders it is perhaps worth remembering that the three great

Methodist institutions, Newington Sydney, Wesley Melbourne and Prince Alfred Adelaide were opened in arithmetic progression, 1863, 66, 69 and all had a similar design, at least in principle, with boarders found in the centre of the original school buildings.

At this point it might be worth asking why John left PAC at the end of 1956 to spend a year in London with Encyclopaedia Britannica. If he had spent a year in another Public School or pursued his studies at Oxford or elsewhere then the question would not arise. But why leave schools? I have asked a number of former students of PAC this question and received various answers. John Rowland, a Boarding House prefect in 1956, reports that John had a "huge" load during his term as Master-in-charge of the House. His observation was that John always went to bed late and got up early, that he appeared to make no time for rest, and that the school expected him to carry his load. John Rowland suspects that John was simply exhausted after five years in residence and that he was due for a year off. He saw him spending a great deal of time and effort running the Scouts, coaching football and athletics, playing in "scratch" matches (he was quite an athlete) and teaching a demanding load of mostly senior classes.

Under John the Scouts had taken a great leap forward, especially in terms of challenging outdoor activities, and in the second term vacation in 1956 a "party of 36 … led by Mr J Keeves and Geoff Bridge", went for "an eleven-day hike in the Northern Flinders Ranges". The October 1956 edition of the *Chronicle* reports that "Dicky Barrows fell ill and we had to get him transported out of the ranges to Hawker. He was then rushed to Adelaide for an operation. John's part in this was so heroic, according to Geoff Bridge, that he "deserved a medal". And former Scout Robin Oaten, whose father was the Barrows family doctor, reports that Dicky "just survived". Then a few days later we are told that while "hiking across the (Wilpena) pound for a windmill, whose exact position was unknown, one of the party became lost and half the party spent about three hours looking for him".

Was John a good Scout Leader? All the evidence I can find suggests that he was much better than good, that Scouts during his time at the helm were better than beforehand and subsequently and that, above all, he was safety conscious. He led boys to be adventurous, creative, and to show initiative but never lost sight of the fact that safety came first. He must not be judged according to present day criteria and present day training such as teachers with B.Ed degrees in Outdoor Education receive. Such professionals may be surprised, for example, that 36 Scouts were accompanied by only 2 adults for an eleven-day hike in the Flinders, but changes in safety rules and expectations, in every field of endeavour, have been enormous in the last fifty years.

The case of Dicky Barrows, different but nevertheless hair-raising enough, tells us more about John's care for boys – and about his sheer physical strength and stamina. Several former Scouts remember the incident which is far more dramatic than the bland report in the *Chronicle***.** On learning that Dicky was ill it was clear to John that he must walk out of the Aroona Valley to a Station and, hopefully, return with a vehicle. After some hours he returned in a four-wheel drive accompanied by a rouseabout from the Station. However on arrival at the PAC camp it was discovered that the vehicle had run out of petrol. And Dicky was getting sicker by the minute – he was seriously ill. So John walked out again. Another car, Geoff Bridge remembers it as a utility, was obtained and back to the camp went John. It was now quite dark. Dicky Barrows was loaded into the utility and off they went to Hawker hospital and thence, by plane, to Adelaide, where an operation saved his life.

Andrew Duguid, who left PAC at the end of 1948, and who was a prominent member of the PAC Scouts, emphatically declares that John was "very conscious of dangers on excursions". That year a group of boys, unknown to John, planned a walk along the South Coast of the Fleurieu Peninsula. John found out about the proposal and immediately went to see Andrew's father, the famous Dr Charles Duguid, to warn him not to let the group go without an adult for the terrain, in parts, was very steep. The good doctor did not take heed, there was an accident (a fall and a badly cut leg) and the boys were left to manage as best they could. It is worth mentioning that the roads at the bottom end of the Peninsula at that time were very rough indeed and in the Flinders were simply tracks.

John was conscientious, careful, caring and safety conscious as well as an adventurous and challenging leader. Camps and hikes took place in various locations – Cox's Creek, Mylor, Cradle Mountain – Lake St Claire, Mt Pleasant, Angaston, Yankalilla, Myponga, Curramulka, Mount Lofty, Gumeracha, Torrens Island (Sea Scouts), Yundi, the Northern and Southern Flinders, Kuitpo Forest, Port Vincent (Sea Scouts), the Grampians and doubtless other locations. Jamborees near Sydney and Melbourne were enthusiastically attended and international friendships made. The work at the school was greatly enhanced in 1955 with the "acquisition of a Scout Headquarters … an eight room cottage at the back of the School. Hazeldene, as the house was named, is already changing the scope of our activities". (*Chronicle*, October 1955). John was also instrumental in introducing Sea Scouts into the school. This important addition took place in 1947, the prime mover behind the scenes being Brian Hobbs, a weekly boarder from Victor Harbor and a very able student indeed. According to Brian the "profile of the Scouts lifted" when John arrived on the scene. Later on the Sea Scouts were to acquire a property at Goolwa and to sail in their own yachts on the Port River and St Vincent Gulf as well as on the Murray. In the forties there were four school activities, during school hours, and all boys in the three senior year levels were obliged to choose from Army Cadets, Air Training Corps, Scouts and the Ambulance group. According to Robert Johnson (1949) Headmaster Fred Ward rated these pursuits in the above order and the Scouts, until John arrived, were not seen by the boys as a particularly attractive alternative. In due course John's colleagues on the staff voted with their feet and sons of Bill Leak, Doug Provis and Kyle Waters, for example, were seen wearing the Scout uniform on Wednesdays.

Besides all the work, including invaluable map reading which, on one hike reported by Robert Johnson got the boys out of trouble, the Scouts had plenty of fun. And John joined in. His Canary Wagon is commented on without fail. This was an Austin A30 Wagon, not a big car by any means, and was obviously used a great deal for Scouting purposes. He still had it in 1972 when we arrived in Melbourne! On one occasion the Canary Wagon was parked in the street outside Hazeldene and another car crashed into it. According to Richard Blake (1960) John barely raised his eyebrows as he asked the boys to nominate a colour for its repaint. And Bill Riceman (1962) describes how, when the Wagon was parked in a lane near Hazeldene, the Scouts simply picked it up and left it at right angles to its former position, thereby blocking the lane and leaving the wagon apparently immoveable. And then there was the snake. Robin Oaten (Head of House 1956) tells about a snake, killed by the Scouts while marking at Dean Range for the Cadets (infra dig?), that was left wrapped around the steering wheel of the Canary Wagon. That night the Head of House was asked by Housemaster Keeves to report to his study in order to discuss a House problem. When Oaten eventually went to bed the snake was in it. The final appearance of the reptile was in the dining hall on John's hot breakfast plate covered

by an aluminium lid. The maid lifted the lid. All the boys were watching. She screamed. Mayhem.

Bill Riceman confirms John's good sense of humour and willingness to mix with the boys. Geoff Bridge, who left in 1951, and who also later helped with the Scouts, recalls John swimming naked with the boys on camp and, more than that, squatting near John alongside the toilet trench. Geoff Bridge and Bill Riceman were not the only former students to assist John during their post PAC years. Old Boys could see the excellent progress that the Troops were making and were keen to be part of a growing and dynamic group at their school. Norm Howard became inspired. He was not a former student (Saints I suspect) and made himself and his ocean-going yacht available to the Sea Scouts. It is not surprising then to read in the **Chronicle** that "there are far more boys wanting to join the 1st PAC Troop than we can possibly take and it is with considerable regret that we must refuse admission to so many who wish to join" (April 1954). And again in April 1955: "once more we regret that we are unable to take some who wish to join. The Seniors (the 3rds) and the Sea Scouts (the 2nds) are now working on the policy of accepting only entrants who have already had Scouting experience and who have passed their second-class tests".

John also formed a Scouts Parents and Friends Committee who raised, by dint of hard work and enthusiasm for the work being done by John, "large sums of money" (*Chronicle* April 1953). They also worked hard on various projects – tree removal, laying a new floor, treating the woodwork in Hazeldene for white ants and repairing the "shack" at Mylor. As reported in the October 1953 *Chronicle* the work at Mylor was substantial. "the tank and gutters have been repaired, the floor replaced, all outside wood-work painted, a large dead tree felled, a new camp fire circle built, a new driveway made, and …" No ordinary working bee this; John was clearly able to muster considerable forces for the long time benefit of PAC boys. The boys themselves also assisted in fund-raising for the Troops. Each year "Bobs for Jobs" and other student based exercises supplemented the outstanding efforts of the Parents and Friends. All this fund-raising and self-help may support Robert Johnson's contention that Fred Ward favoured the Cadets. No other extra-curricular activity, as far as I know, was expected to be self-funding, so to speak, although the acquisition of Hazeldene was partly funded by the Council. But the ongoing expenses of a vigorous Scouting movement at the School were considerable and the fact that so much was done is a great tribute to John. And the Sea Scouts must not be forgotten in this respect. "In the junior section of the Squadron six boats were raced and skippered throughout the summer by members of the Troop" (*Chronicle* October 1961).

Most teachers take great pride in the professional work they do, preparing and presenting lessons, marking homework and tests, assisting students on a 1:1 or small group basis, encouraging the able to read outside the curriculum, making recommendations to libraries for inclusion of materials related to their subjects, ensuring practical classes, excursions and tours are stimulating, and providing pastoral care for their charges. John did all that, extremely well, but in my estimation his work with the Scouts is the thing that historians should remember. R.M.Gibbs in his *A History of Prince Alfred College* (Griffin, 1984) unfortunately gives scant attention to Scouting at PAC and John receives the briefest recognition: "The school troop owed much to the interest of staff members such as (my underlining) John Keeves". And this is John's only mention in the 400 page volume. One of the interesting points made by Gibbs, however, is that in the seventies the school purchased a Field Centre at Scott's Creek, an inlet of the River Murray. By that time many independent schools in Australia had purchased or leased Outdoor Education camp sites and the professional training of Outdoor Education teachers was well

underway. Schools were able, therefore, to provide some of the experiences inherent in Scouting to whole Year groups and, in most schools, do this on a compulsory basis. Scouting therefore waned or, perhaps, became a redundant activity at PAC at the end of the seventies.

In concluding this section, and especially as it is not spelt out by Gibbs, let me repeat what former Scouts said of their leader.

"He had energy, commitment and enthusiasm".

"The profile (of Scouting) lifted".

"Pug (his nickname) got out more. There was far more hiking and camping".

"The Scouts buzzed along – it was the heyday of the movement".

"Pug was a good bloke, a perfectionist".

"He was hard working, well organised, got things done".

"Snored when asleep".

"After he left hikes were shortened".

"Inspired loyalty".

"Respected him and enjoyed him".

"Occasionally lost his temper".

"Funny mannerisms, we used to mimic him – behind his back".

"Conducted prayers well".

"Flinders hikes were the highlight".

"Gentle, kind. Incredibly tolerant of silly behaviour".

"Got upset by some things on camp".

"Conducted himself with resolve, politeness, wisdom and determination".

"Thought like a young person".

"Remarkable".

"Initiative given to the boys".

"Gentle – some took advantage of him".

The Housemaster in any boarding school is in loco parentis. It is a highly responsible and demanding position. And at PAC in John's time, and before, it was undervalued by Headmaster and Council. From 1952 to 1956, for five years, John was in charge of well over a hundred boarders aged between 12 and 18. (The traditional boarding house, based on the English Public School model, had about 60 or 70 boarders and its own separate building. Boarders, at the end of the school day "went home" to such a house). Typically John was assisted by one full-time resident teacher at the school, the legendary, colourful and eccentric Gordon Connell, together with a sprinkling of University or Teachers' College students. The Headmaster John Dunning was not far away but it was John Keeves who carried the load. Viewed from Dequetteville Terrace the old building had boarders at top floor level in each wing and also in the central block. Senior and Middle students occupied the South and North wings respectively and Juniors were crammed in to the central section – opened in 1869. This, in retrospect, was a firetrap but mercifully there was never a fire. From any area it was easy to escape, easy to find somewhere to smoke and easy to play the fool, or even occasionally, the bully. The day began with cold showers, roll call, breakfast and then the school day. Then, after an early evening meal supervised night school or "prep" was held, interrupted only by prayers in the assembly hall and concluded by supper in the dining hall. Then it was off to bed and lights out. Weekends were more relaxed and perhaps more difficult to manage and about twice a term (there were

three terms in those days) there was an exeat weekend to the great relief of the Housemaster. But organising and legitimising all exeat arrangements were not necessarily easy. Every Sunday morning the boarders attended the Kent Town Methodist Church and Sunday evening was taken up with a short service in the assembly hall. As John was coaching sport and looking after the Scouts I suspect that the only time he had to himself was Sunday afternoon. He probably went home to Statenborough Street and slept.

Discipline in the House was somehow managed and the very great majority of boarders loved every minute of it – except perhaps the initiation activities (they were hardly ceremonies) which didn't last long and didn't involve physical harm at all. Just a tradition one might say. Discipline was also a traditional matter and much was in the hands of the House Prefects who received no training or advice in such matters. It was all handed down, year after year.

The light in the dormitory goes on. Standing in the doorway is a prefect. "Who was talking after lights out?" "Come on, who?" Out of bed would creep one or two who were promptly asked to bend over, whacked on the bottom with a slipper or, worse, sandshoe, and sent back to bed. It was all a sort of a game, painful sometimes, but no one ever complained. And of course, the House Prefect always held the upper hand; repeated offenders were referred to the Housemaster where there was a discussion rather than a whack, but sometimes both. Punishment for minor offences was usually regarded by boys as some sort of minor achievement. Something to boast about. "How many did you get?" was not an uncommon question.

The regularity of boarding house life and the great deal of fun had by boarders contributed to strong esprit de corps and, for those not going home to the farm, to impressive results in public examinations. Former boarders of the Keeves era hold prominent positions in the professions and business as well as leadership positions in agriculture and local government. John, with his strongly self-disciplined life, would have set an excellent example. And doubtless with senior science and mathematics students he would have been a source of knowledge as well as inspiration. Boarders, as in some other schools, were known as the backbone of the school.

So what did the boys think of their Housemaster? Brian Hobbs, who left in 1947 and who therefore had one year in the House when John was Assistant Housemaster, remembers him as strict but fair, having no tantrums of temper and being very supportive. "We could talk to him". Brian also saw him as a very steady person, a man with a stable background. David Oaten boarded during the key Keeves years, 52-56, and remembers John coaching the PAC 3rd Eighteen, a team which consisted of boarders only. Included in the team was Tony Clarkson, now a distinguished renal physician and one time prominent SANFL footballer. According to David Oaten this was all John's doing. He (David) also recalls asking John whether, at the conclusion of a boarders' party he should take his partner to the bus. "Take her all the way home" was the Housemaster's reply. In those days the boarders' picnics were held on Saturdays during the warm weather at one of the beaches. It was not surprising that numbers of boys used acute sunburn as an excuse for not attending Kent Town Church on the Sunday – until, that is, an edict came down from the Housemaster that sunburn would no longer be an acceptable excuse. Brian Hurn (1956) who returned to the land and became a distinguished contributor to local government in South Australia (and who is also a member of the PAC Hall of Fame) says that John was well-respected, a totally fair disciplinarian and that he had an "educational approach" to punishment. He further describes John as calm, reserved, formal, a conservative and a good citizen – rather like Headmaster John Dunning. Other boarders "liked

him", described him as "really good", a "great educator", a "wonderful influence", and all, it seems, respected him.

During his time in the PAC boarding house John, as part of his duties, conducted evening prayers. (He had similar responsibilities when in charge of Scouts in the Flinders Ranges and elsewhere). On Sundays John accompanied the boarders to morning services in Kent Town Methodist Church. This involvement in religious matters was part of a tradition stretching back a long way, and not, for most boarders I suspect, one which was remembered with much enthusiasm. It is tempting to guess at the level of John's involvement. Were these daily prayers and weekly church services simply part of his duties and seen by him as such? To attempt to answer this question it is necessary to look at John's background.

At PAC as a student he was greatly influenced by "Spenny" Williams, a redoubtable figure who was instrumental in keeping PAC students at or near the top of State honours lists in mathematics during the thirties and forties. John still talks about him. He also talks about Ken Smith, KWA Smith known to the boys as "KWAS". Ken and "Spenny" were opposite ends of the pedagogical pole in their approach to teaching mathematics. "Spenny" was a methodical, painstaking, systematic, energetic and inspiring teacher whose methods for problems were clearly defined. He worked from an axiomatic base in most areas and built logically, step by step, establishing conclusions. His processes or methods always seemed to work. He was a great influence on large numbers of PAC boys, including this writer. "KWAS" on the other hand was intuitive and creative; he saw the elegance of a quick solution to a problem and applauded different approaches. He encouraged the bright with mathematical puzzles on the periphery or outside the syllabus. And he too inspired his students and gently hinted to some that teaching might be a great profession.

John clearly came down on the side of "Spenny". For him logical reasoning was the key to development, to academic success and, as it turned out, to his eventual outstanding career at the Australian Council of Educational Research (ACER), universities and elsewhere. This commitment to logic, to observation and measurement, together with his phenomenal memory and determination have much to do with his extraordinary career in international education.

But back to the Christian faith. During preparation of this essay I have discussed this matter with a number of others and answers point to the same conclusion. John has a very inquiring mind and as a member of a flourishing SCM (Student Christian Movement) at Adelaide University he would have loved the discussion, debates, reading circles and so on that were part and parcel of SCM. This would have been enhanced, no doubt, by stimulating debate at Radley and even more so at Lincoln College Oxford. Seriously inquire. Yes. Commitment to a faith, other than a faith in logical inquiry, I doubt it. He would not, of course, be an atheist for it too requires a leap in faith, but an agnostic, one who keeps on inquiring and is open to logical persuasion, but not one who in a fit of wild emotion is likely to be, as it were, converted. And I have no doubt that John sees in Christ a person whose philosophy, whose thoughtful way of life, was exemplary, a model for countless millions. Former students and colleagues use such phrases as "not an atheist", "an inquirer", "not a man of Faith", "not a risk taker", "a tolerant man", "looked at religion in the same way as he looked at other fields of endeavour", "probably religious", "probably believes in a God", "had Christian virtues", "no supernatural belief".

So there it is. Perhaps when he reads this John might let us know the hidden truth. Perhaps.

And now to John the teacher. He began teaching formally in 1947 and is still hard at it in 2004. He first went to school in 1929 and is still in full-time attendance in 2004. A couple of years ago I rang him at Flinders at midday. "Sorry, David, can't speak now; ring me at home at 8.30 tonight". John still teaches with enthusiasm and clarity as only someone passionate about his subject could at the age of 80. It is a pity that Tony Clarkson, a teacher himself from time to time, could not have seen that in his teenage years for it was certainly there. And like so many teachers John is also a life-long learner and was ever thus. It is this that probably makes him such an interesting person to know.

I was a student during John's first three years at PAC, 1947-49 and a teacher there during his last period of service 1958-61. However he did not teach me and as teachers we were all so busy that it is difficult to give an opinion about his performances in the classrooms and laboratories. But I got the impression that he was well-respected by students and by his colleagues and that his contribution to the life of the school was considerable. He was certainly friendly and helpful to a younger physics and maths teacher freshly returned from England in 1958. In those days there was no induction of new or young teachers. Furthermore there were no staff meetings as we know them today and no departmental meetings of any kind. The Headmaster occasionally got us together and issued a few instructions and those in charge of mathematics and physics left us to our own devices. Under these circumstances, the avalanche of work and, put positively, the professional freedom which we all had, it wasn't easy to form an opinion of a colleague – as a teacher.

Former teacher David Mattingley knew John for longer than I and his view of him as a professional is similarly restricted. David saw him as thoughtful, idealistic and friendly. John lent David the Yellow Canary Wagon when he needed transport for a few days in 1955 – and locked the petrol cap so that he couldn't refill it. A nice touch. David also saw John as single purposed, committed, with a phenomenal memory and now, in 2004, as the Grand Old Man of PAC. The only negative David could muster was that in discussion John listened intently enough to the points of view of colleagues and then stuck to his own. School Chaplain Kyle Waters remembers John as conscientious, perhaps over-conscientious, effective, concerned about the development of the whole person and interested in the art of teaching and the philosophy and psychology of learning. Geoff Woollard found him independent, approachable and one who had great pride in his teaching.

To get alternative and perhaps better views of John the teacher at PAC it is profitable to hear what his former students have to report. Like all beginning teachers John had his difficulties with some classes in his early years. At that time the classes were streamed so that teaching Year 10D was a very different experience to teaching Year 10A, especially if your subject was physics or mathematics. And the lower streams were often handballed to younger teachers who, to be fair, were also given the odd A stream class. Some of the less academic respondents refer to giving John a "hard time" in the classroom. One, who prefers to remain anonymous, simply says that "we gave him hell". A "D" stream, a common curriculum for all streams, physics, a new teacher – not a good combination of circumstances.

The majority however are glowing in their praise. He was "a good teacher who knew his stuff". He "didn't appear to be a new teacher". He was "confident, not afraid of the students". "Conscientious, organised, precise". "Suffered the less academic well". "Went far beyond the text book". "Never abused the automatic authority which a teacher had in those days". "Gentle, kind, tolerant". "Curious, thoughtful". "Loved problem solving". And from John Willoughby, a distinguished neurologist and

Professor at Flinders University: “He taught me physics. A lovely person. I like him more and more. We have lunch occasionally and the discussion is never lightweight”.

There are a couple of questions that remain about John the teacher. Why did he keep returning to PAC rather than seek fresh fields elsewhere? Why did he wait so long (he was 37 years of age when he finally left secondary school teaching) before entering educational research which became his major contribution to Australian education and which saw him become an international figure? We may never know the answers to either of these questions but can make some educated guesses. The three terms at PAC, stretching from 1947 to 1961 and representing 12 years of service to the school, constitute 20% of John’s working life – so far! In my opinion he kept returning to his old school simply because he loved it and wished to return something in recognition of the great impact it had had on him in his student days. I do not believe that John’s three terms at PAC were due to him being conservative or unadventurous; his distinguished work with the PAC Scouts and his post 1961 career show that to be far from the case. Perhaps there were family reasons that brought him back to Adelaide. The second question is more interesting. Not many professionals wait until they are 37 years of age before changing direction. Some have suggested that he may have applied unsuccessfully for headships or that he saw himself as a Mr Chips until he actually saw some at PAC and that either of those factors may have contributed. I think not.

From the commencement of 1957, without the hectic boarding house life, John had time to think about his future. He had all the qualities of an academic, of a researcher, but not a higher degree. First-class honours in physics at Adelaide was an indication of more to follow but it alone perhaps made selection panels falter. The perceptive Director of ACER did not. It is significant that John was offered a position at ACER early in 1962. In the *School Notes* of the *Chronicle* (April 1962) we read that “Early in January Mr J.P. Keeves was appointed to the staff of the Australian Council of Educational Research. He had given excellent service in the classroom, with the scouts, on the playing fields, and earlier in the boarding house”. John would not have left his old school at such an inconvenient and awkward time for the Headmaster unless he was very, very keen to start a new career.

The following people have been helpful in the preparation of this essay and I thank them all sincerely. Richard Blake, Geoff Bridge, Don Burns, Sandy Clark, Tony Clarkson, David Cornish, Andrew Duguid, Margaret Giddings, Bob Hale, Geoff Hall, Brian Hobbs, Brian Hurn, Robert Johnson, David Mattingley, David Oaten, Robin Oaten, Jean Prest, Elizabeth Puddy, Bill Riceman, John Rowland, Kyle Waters, David Wibberley, John Willoughby and Geoff Woollard.

I also warmly thank Kathleen Logan (Trinity College, Melbourne University) who did the word processing.

6

Wesley College Council

David Prest

***Dedications**: David Prest, retired Principal, Wesley College, Melbourne, Old Scholar of Prince Alfred College and teacher at Prince Alfred College.*

Recognition and Support

During his time as Principal of Wesley College in Melbourne, David Prest was quick to recognise in John Keeves qualities that made him ideal in the important role as a member of the College Council.

> "In the '60's and early '70's there were a number of educators/academics on the Wesley College Council. These people sat comfortably with Dr Tom Coates, the then Headmaster, and the non-academics on the Council. They had close ties to the Methodist Church, later the Uniting Church, and Melbourne University.
>
> Later the Council consisted mainly of people of a more pragmatic style – 'business men' (very predominantly men) – such as engineers, bankers, lawyers, accountants. Neither the Council, through its Selection Committee, nor the Principal, David Prest, felt the need for educators on the Council. The Council clearly regarded the field of education as the domain of the Principal. The Principal strenuously avoided interfering in appointments to Council.
>
> David Prest did however suggest the appointment of his former colleague, John Keeves to Council in late 1977. This was readily accepted by the Selection Committee, and the Council. Unlike most others on the Council, John had little to say, his contribution being by way of thoughtful and thought-provoking papers, erudite comments on educational matters, and absolute pearls of wisdom on subjects such as enrolment trends, and professional education of teachers.

> One or two more vociferous 'educators' on the Council were not well received by the majority of councillors, but this was not so with John. Councillors valued his wise input and his kindly manner ensured he was well liked and well respected by his peers". (Doug Oldfield)
>
> "As a member of the Council of Wesley College, Melbourne, John Keeves was both a unique resource and a luxury.
>
> His knowledge of global developments in education was particularly helpful at a time when Wesley was visualising, and then determining, its place in Australian education over the next forty to fifty years. Further, with this analytical mind and his concise presentation, his contributions to Council strategic planning were always sharply relevant and keenly heard.
>
> Beyond that, John Keeves fulfilled another, yet equally valuable role. On matters educational, he held the respect both of the Council and the Principal. Thus, he was able to provide personal support to the Principal – without compromising the Principal's authority and, at the same time, defusing any perceived need for the Council to intrude into the educational affairs of the day". (Geoff Stephenson)
>
> "When he spoke everyone listened". (Warrick Mitchell)
>
> "He would quote from time to time from the results of educational research, sometimes his own". (John Gellie)
>
> "Keeves acted as a good foil against the economic rationalists to the extent that he was able to provide an assertive and irrefutable face for the Principal's ideas. He was really the only Councillor who could ask educational questions of the Principal". (John Hall)
>
> "Rarely spoke and when he did, Prest would lean forward and prick up his ears". (John Hicks)

John Keeves served on the Wesley Council from early 1978 until the end of 1986, an extraordinary period in the history of that school.

In 1977, Wesley was a boys' school of 1605 students, 927 at Prahran and 678 at the recently (1966) established Glen Waverley campus. The school had a boarding house at Prahran of about 100 boarders which was run on lines similar to those which John had employed at Prince Alfred College in the fifties. The Prahran campus catered for students from Year 7 to Year 12 and Glen Waverley from Year 4 to Year 8. Ten years later, in 1987, there were no boarders, both schools were coeducational throughout, Glen Waverley ran classes from Beginners to Year 10 and Prahran from Year 5 to Year 12. The enrolment totalled 2262 and each campus had its own Head, reporting, as it were, to the Principal. (Ten years after that both schools operated from Beginners to Year 12 and a third campus at Elsternwick had students from Beginners to Year 10. The total enrolment was 3427).

John contributed generously and significantly during his nine years of service, years which saw change across the board – in curriculum, educational structure, administration, coeducation and vision. It was customary at that time for new Councillors to accept membership of Council and one of the Standing Committees. John was offered, and accepted, Property Committee membership but the records show that he never attended a meeting! He was, however, quickly snapped up by the ad hoc Education Committee, convened by the Principal, and became a regular attender of it and Council. (The Principal had his own Senior Staff Committee; the Education Committee provided academic or professional opinion, from people outside primary and secondary schools, and also acted sometimes as a sort of Upper House).

A couple of key decisions had been made in 1977 before John's arrival; the College would become coeducational beginning with B-3 in 1978 and, adding one Year incrementally, B-12 in 1987. Secondly, the College would be split into Preparatory (B-4), Junior (5-7), Middle (8-10), Senior (11-12) schools, each with its own territory, coordinator, curriculum and, as far as possible, its own staff.

At the Education Committee meetings in 1979 a number of topics were discussed including the Whitfield curriculum model, open entry and curriculum diversification, the territory for the Senior Schools and the role of the Committee itself. John was an ardent supporter of the Principal and explained, at a Council meeting in October 1979, that the Education Committee was in fact an advisory committee and that Council members, constitutionally, were not free to step into the Principal's domain except in that capacity. In particular it was confirmed by Council that

> the principal alone be referred to when criticism is received (by Councillors) relating to education.

John was a strong supporter of the Senior School and

> felt that the key question is whether 11 and 12 should be built around the library, which should occupy a central position in the School. (Education November 79)

Then at Council in December 1979 John presented a paper on **Wesley College – An Equal Opportunity School** which raised much discussion e.g. on the provision of remedial and gifted programs and the educational challenges which students from migrant (mainly Greek) families, where English was not spoken at home, presented.

In 1980 discussion at the Education Committee included the functioning of the mini schools (as they became known), senior positions for women on staff, women Council members, the role of the Principal as Chief Executive Officer and the relationship of the College with the recently formed Uniting Church. In all these matters John made useful contributions and at the Education Committee meeting in November 1980 he presented a paper on **The history of the relationship of the President (of Council) to the Church.**

1980 was the last year of boarding at Wesley, for the Council, when taking the coeducational decision in 1977, determined that the total number of boys should not be decreased. Later, and John was involved in this decision, it was decided that the boy:girl ratio would not fall outside the bounds of 1:2 and 2:1 (this still applies in 2004). Space was therefore needed and the boarding house at Prahran had plenty, although, to tell the truth, boarding was not being done well in the seventies. At Glen Waverley the College owned acres of vacant land on the other side of the road and there was a vigorous debate about selling. John had considerable input. He discussed future educational needs in terms of international students, the burgeoning demographics, the possibility of schools taking on a Year 13 in the future, additional space for games, and pleaded for research and caution before a sale.

It was clear that as Wesley College was to become rather large, curriculum content should continue to be discussed in the early eighties. How many and which languages should be taught and at which levels? John suggested (Education June 1981) that the introduction of further Asian languages "might be a possibility at Senior College level only". This was a clever idea for some universities were about to offer first year language courses which did not demand a traditional detailed background of the language in question.

John was also asked to comment on a NCIS (National Council of Independent Schools) paper on **The Relationship with the Education Department**, he joined in the debate on the re-introduction of boarding (coeducational) at Wesley, provided information on PE programs in South Australian Departmental schools when a staff paper on **Principles and Practice of Sport at Wesley College** was up for discussion and won the debate on the proposed sale of excess land at Glen Waverley.

> It was recommended that Council does not part with any of this land, but, in order to finance future projects at the school, non-contiguous properties owned by the school should be sold. (Education Committee July 1982).

Sadly, perhaps, Council reversed this some years later. The economic rationalists had their way.

John was very good at giving the Council broad views of the educational scene and at peering into the future. His statistical information and research were greatly appreciated and it was against this background that decisions were sometimes taken. By mid 1983 it seemed highly likely that Commonwealth per capita and building grants would be reduced for high fee schools such as Wesley. This was confirmed by an undated circular from Minister for Education and Youth Affairs Susan Ryan: per capita grants for schools such as Wesley "for 1984 will be reduced by 25%". This announcement produced a kerfuffle up and down the country and public meetings of parents were common. The Wesley answer to the threat, at a Council Executive meeting in May, was to set up a Task Force of the Principal and three members of Council including, of course, Dr Keeves. The group was asked to "decide (!) how the school will handle the next decade". A considerable amount of work ensued with John heavily involved and some senior members of staff also.

The Educational Goals of the School were reviewed and John presented to Council (October 1983) a paper on **Education in 1984 and Beyond**. This so impressed Council members that it was distributed to all staff and parents. Then at the Executive meeting in February 1984 the following note and accompanying paper was received from John.

> Doug Oldfield (President) asked to be kept informed of enrolment projections in Victoria. Please find enclosed the latest figures that I have seen. These figures are sufficient reason for the harsh policies of the Victorian Schools Board and the Commonwealth Schools Commission.
>
> Yours sincerely
>
> J P Keeves
>
> Director

The figures showed a predicted decline in Government school enrolments to 1992 and a dramatic increase in Catholic and non-Catholic Independent Schools. At the Council meeting later in February John presented his paper again and warned that "it was likely that considerable changes could occur in the next 12 months". Then in May 1984 at a Council meeting John continued to expand members' views beyond the school gates. His paper this time included statistics on

(i). Student-teacher ratios Government / Non-Government Schools 1952-1984.

(ii). Real Values of Taxation Concessions for Current Expenditure on Education 1948-1981 in 1979-1980 prices.

(iii). Real Private Price of Private Schooling 1948-1981 and percentage of Students enrolled in Government Schools.

Then came John's participation in the formulation of the 1984 Wesley Parent Questionnaire. He must have wondered sometimes whether he was employed at ACER or Wesley. The purpose of the questionnaire was to provide information for the Council to help in making proper decisions about the school's future and to put appropriate submissions to governments. Significantly the questionnaire was designed at the same time as Senator Ryan's 1984 paper, in which it was made clear that Wesley would receive less funding, in real terms, from 1985.

From then on the Council and its Education Committee were less demanding on John although he continued to have a good attendance record until his resignation in December 1986. The matter of a coeducational boarding house continued to raise its head and John was a participant at the special Council meeting at the Wesley Chum Creek Camp near Healesville to discuss the matter. John's boarding experience at Prince Alfred College was useful in this respect and he was instrumental in the final decision – coeducational boarding would be established at Glen Waverley. The nuts and bolts of this were referred to the Education Committee but in 1989 the terrible Prahran fire put the project on hold. It still is.

The completion of the mini schools received continued attention from the Education Committee before John departed from the scene. At the March 1986 meeting the construction of a Senior College at Glen Waverley was recommended and also a Preparatory School near the Prahran campus. (The possibility of a Preparatory School on the crowded Prahran campus was out of the question). Dr Keeves moved "that we see strong educational arguments for the establishment of co-ed classes B-4 near the Prahran campus". Later at an Education meeting (June 1986) the new Head of the Glen Waverley campus, Barbara Lynch, gave her views on aspects of the proposed Senior College at Glen Waverley and "Dr Keeves mentioned general technical and technological education taking place in Italy and the United Kingdom".

The matter did not fall on deaf ears. Barbara Lynch gave another paper (Education August 1986) on **Mathematics, Education and Girls**.

> Dr Keeves found the paper an excellent summary of the area and endorsed its contents. He believed that attitudes and aspirations were important and that these were developed in the home; therefore we have a message to get across to parents as well as to staff.

On the lighter side John recommended (Education March 1986), after much pressure from the Parents' Association, that

> we do not recommend that a raffle be held … given the possible reactions in the Wesley community we think it advisable not to approve raffles.

At the Council meeting in December 1986 John's resignation was accepted. He had made a great contribution to Wesley.

> The Chairman spoke of Dr Keeves contribution and service to the Council and his resignation was accepted with regret … Dr Keeves thanked the Council and wished the Council, Principal and Staff every success in the implementation of the developments that are in hand and those that are projected in the future.

The extraordinary thing with him is that at 80 years of age he is still capable of making a significant contribution to any school Council.

7

Port Willunga

David Prest

***Dedications**: David Prest, retired Principal, Wesley College, Melbourne, Old Scholar of Prince Alfred College and teacher at Prince Alfred College.*

Beyond Work

Just about anyone who knows John Keeves has enjoyed his hospitality at his beach house overlooking the cliffs at Port Willunga. David Prest tells of these experiences.

In the summer of 1960, or thereabouts, John invited Jean and me to stay for a weekend at Port Willunga. He had, a year or so before our visit, purchased a property in William Street, one street back from the Esplanade, and in the favoured Old Section. The house was modest: small, hot, somewhat airless and a haven for flies and mosquitoes. It was constructed of corrugated iron. But it was a start and John loved it. Perhaps ten years later he acquired the present solid brick property along the Esplanade which he improved significantly some years later with the expertise of architect Ian Berriman. Old PAC of course. In fact John must have been attracted to Port Willunga partly because there were a number of PAC friends, from student days, with houses, or shacks as they are sometimes known down there, close at hand. Crompton, Schafer, Provis, Bloomfield are some names that come to mind and more were to follow.

John would not have made a quick decision before choosing Port Willunga – or indeed in making up his mind about any matter of substance. I can imagine him driving, in the Canary Wagon, to many venues up and down the coast, and perhaps in the hills, before settling on Port Willunga. The District was settled very early, being first visited by Colonel Light and a party from the **Buffalo** in 1837, and soon became a thriving wheat growing area. Two jetties were erected in order to transport wheat and other produce by ship; the first of these was out of order by the early sixties and the second, a small remnant of which remains, was of no use by the end of the

century. Storms had taken their toll. In fact, as cited in G.H. Manning: **The Tragic Shore**, by 1889 "Port Willunga, indeed, ranks among the 'has been' settlements of the colony". John and others would have had a good sense of that when he purchased his William Street shack in the late fifties and doubtless would have liked to have kept it that way.

So what does John do at Port Willunga? Most of the time, at least in the waking hours, he does what he has always done and always will do – he works. To this end Ian Berriman did him proud. He designed a study at first floor level with a view of the gulf. There John can be seen seven days of the week hard at it. Except in the hot summer afternoons when the blistering sun is too much he retreats to his downstairs office on the cooler eastern side at the back of the house. But it's not all work at Port Willunga for John. He is a regular walker, at about noon, to and from Gull Rock which is at the northern end of the Port bay, a round trip of 3 or 4 kilometres. Then, if the air temperature is 24ºC or greater he goes for a swim and perhaps these days after lunch has a sleep. He has made his place a wonderful source of rest and recreation and has some beautiful works of art, particularly paintings and wall hangings. And, of course, there are hundreds, perhaps thousands, of books in crowded but orderly shelves.

The house has become a great place for his extended family, brother Andrew (recently deceased) and sister Margaret who sadly died when she was comparatively young. And Andrew's children and their children have greatly appreciated John's warm hospitality. Although it has to be said that he is a terrible cook.

One of the unexpected and fortuitous consequences of living at Port Willunga is that it is now, since the construction of the freeway to the Onkaparinga River, an easy drive to Flinders University. Flinders has been a wonderful institution for John in his "retirement" years and it has given him very great pleasure indeed. From time to time he entertains staff and students, particularly international students, at Port Willunga and they are joyous occasions indeed. Others seem to do the cooking and from time to time some of the Prests have been able to accept John's invitation to attend. Local residents get a vicarious pleasure at the sight of numerous international students suddenly invading their beach and smile at the thought of Dr Keeves entertaining again. It is clear that many at Flinders regard him warmly as colleague and friend.

John never tires of Port Willunga. He loves the beach, the cliffs, the water, the people, the history, the view and the chance to escape. He has walked thousands of kilometres to and from Gull Rock, swum his regulation hundred strokes thousands of times and written or edited millions of words in his study overlooking the beautiful Gulf St Vincent. One day I'll persuade him to come fishing.

8

Teacher and Scout Leader

John Willoughby

***Dedications**: John Willoughby, professor of neurology at Flinders Medical Centre and a former student and scout at Prince Alfred College*

Leadership and Care

I knew John Keeves as a teacher and a leader of a scout group at a secondary school.

In his class-room role, John Keeves took things seriously. Lessons seemed to have been prepared well and the content was treated with genuine respect, naturally enough, because it was physics and chemistry that he taught. The ideas were, of-course, more complex during our senior years, a fact that John Keeves appeared to address by, on the one hand, increasing his deep personal interest in the subject matter and, on the other, by having us use the new knowledge in tackling larger numbers of problems - it was very solid work for us. I do not recall class-room discipline being an issue, though a student's memory on such a matter can always be questioned. However, my recall is that John Keeves dealt with students in a most even-handed manner, always with courtesy, sometimes quite self-effacingly, and invariably with a long view. Any view longer term than one lesson probably seemed long to us as students. In retrospect, I think his view was broad and not only on the long-term teacher-student relationship over the time the student might be at the school but also on the attitude students might have in their future years to whatever subject matter was under discussion. He provided an excellent model for how to conduct oneself during a strong discussion – even being so much senior to us, he remained open to hearing our ideas and he responded without an authoritarian stamp.

As a supervisor of youths while they developed skills in out-door activities and in bush living, John Keeves' patience was the quality that he most strongly expressed. The activities varied, taking the form of weekly half-day activities as well as week or longer bush walks and camps. Looking back now, it is hard to conceive how anyone

could maintain an even-handed approach in the face of pranks, pointless complaints and the general hyperactivity that flourishes around male youths. (I do not claim I was not guilty of these.) If as adults we are challenged by being in such a role, a half-day's activity might be seen to have an end, however, seven to ten days would seem for ever. By what means did he cope? In retrospect, I think it was through his ability to communicate even-handedly to individuals, to patiently explain reasons why limitations had to be imposed and to point out the consequences of actions – and he coped because this strategy worked. More often, though, he was able to support requests for freedom and allow students an opportunity to undertake their plans, as much as that may sometimes have been anxiety provoking. As a leader, organiser and responsible person on camps and treks, John Keeves was a marvellous senior companion. Being knowledgeable, he could provide good direction on answers about natural phenomena and being wise, he handled the fine judgements necessary in dealing with personal behaviours. We students regarded the trekking experiences with John Keeves as high moments in our early lives and we still do.

Together with a colleague from school days, I enjoy a yearly meal with John Keeves at a quiet restaurant. He is definably the same individual of 45 years ago. I would expect my impressions of John Keeves to strike a chord in his current students: courtesy, curiosity, deep interest, wisdom and more hard work!

Part 2

Contributions and Collaborations beyond Australia

Sivakumar Alagumalai
Flinders University

The contributions in this section highlight a very small 'sample' of John's international collaboration. The authors share their passion for educational research, measurement and assessment, an interest that John demonstrates through his beliefs and practices. The four 'research gems' in this section highlight the broader notion of inquiry, and flag John's collaboration with other eminent researchers across the world.

Posthethwaite discusses the criteria for sound educational surveys. As highlighted in preface, John reiterates the need to understand the relationship between epistemology, methodology, practices and reporting when undertaking research. Posthethwaite's work in wanting excellence in undertaking educational surveys sits well with John's motivation. Howie and Plomp are concerned with the effects of language proficiency on students' performance in mathematics. This research method mirror the rigour with which John strives in his research studies, and ensures in his students and colleagues.

Black examines the influence of formative assessment. Black indicates that John is rigorous in working with evidence, its assembly and subsequent interpretation. Black highlights the importance of these 'virtues' when dealing with assessment. Kok Aun is concerned with the diversity of research on teaching, an area which John constantly spends his time examining the related issues. Kok Aun highlights the studies undertaken across cultures and traditions, and raises concerns that needs address.

The four articles provide an insight to the depth and breadth of John's interests and work. John's keen sense of measurement and tracking change over time highlight his ongoing learning about perfecting both the tools and techniques. The need for refined educational research and interest in assessment issues indicate the breadth of knowledge John (has acquired) is acquiring. Collaboration with eminent authors of these articles further supports John's interest and mastery of the nexus between learning, instruction and research.

9

Ten Questions by which to Judge the Soundness of Educational Achievement Surveys

T. Neville Posthethwaite
University of Hamburg, Germany

Dedications: *I am writing the preface to my contribution to your Festschrift in the form of a letter to you. My memory may have played tricks on me about exact years but you will remember the events to which I refer. I first met you in 1965 when you attended your first IEA meeting. You had been sent to the meeting by Dr. Radford, the then Director of the Australian Council for Educational Research (ACER). Sir Fred Schonnel was living in Australia at the time and it was he who had mentioned to Radford about this new venture called IEA. At about the same time, there was man called Stephen King-Hall who lived in the house next to where I lived later in St. Albans (but I did not know him) whose wife produced a newsletter for children and you had read an article that she had written about IEA and you drew this to Radford's attention. The upshot was that Australia participated in the first international mathematics study (FIMS) and you conducted the research in Australia.*

In 1971, you came to Stockholm to work on the data analysis of the first three subjects in the six-subject study (in all there were six subjects: Science, Reading Comprehension, Literature, French as a Foreign Language, English as a Foreign Language, and Civic Education). At the forefront of multivariate analyses in those days was multiple regression analysis. You and I were in a large room with blackboards on three walls, the fourth wall having windows. We had variables as rows and countries as columns. The cells were beta coefficients. Each country had been analysed separately but with all of the variables in

the initial regression runs. The reason for the very large tables is that we were hunting for variables that had strong or very strong association with educational achievement in all countries. In short, was it possible to generalise about the effects of variables in different subject matters in education? Arriving at these tables had been a great deal of work. You and I worked intensively on the study for three months. Those who know you will realise that this meant twelve hours per day for six days per week. You often came to our home on the Sunday for Sunday lunch and of course we 'talked shop'. But, we also learned of your school days, both as a pupil and later as a Physics teacher at Prince Alfred College in Adelaide, South Australia. We learned of your days at Radley College, again as a Physics teacher and later at Oxford. It was in 1962 that you moved from Adelaide to the Australian Council for Educational Research (ACER). You even regaled our children with stories of cricket and this usually included some mention of the Chappell brothers. More importantly, you taught them the words to 'Waltzing Mathilda' as we walked along the coast of the Swedish archipelago. They loved that. I shall never forget that time.

In 1967 you had taken leave from the ACER and gone to Canberra to conduct a study from which you used the data for your doctoral thesis. The data set was a good one and has been used on many occasions to try out new multivariate approaches. Partial Least Squares (PLS) and more recently Hierarchical Linear Modelling (HLM) have been tried out with this data set. In 1972 you returned to the ACER and remained there, lastly as Director, until you retired in 1984.

In 1972, our paths then diverged for a few years while I moved to the International Institute for Educational Planning (IIEP) in Paris and from there to the University of Hamburg. In 1978 I became the Chair of IEA. I became the Chair of IEA and had to build it up by increasing the number of projects, the number of participating countries, and to get some money into the bank. One of the projects we decided to start was a second science study. I asked you to be in charge of it. At that time, you were Director of ACER and did not have much time. But you accepted. Money was very tight. You visited us in Hamburg a couple of times and met several of my students and data processing team. Above all you learned to know both Norbert Sellin and Andreas Schleicher. You learned Partial Least Squares. In 1986 I managed to raise some money for the analyses of the Second Science Study data but on condition that the money was spent in Sweden. The money was from the Bank of Sweden Tercentenary Fund and from the Swedish Ministry of Education. I asked you if you would be prepared to go to Sweden to do the work when you retired from the ACER directorship. You agreed, even though it meant that you would miss the Australian summer in Willunga. Somehow, some people at ACER gained the impression that this money could have been used in Australia and that I had deliberately stopped it going to Australia. This is just to set the record straight. The money had to be spent in Sweden and I was very grateful to you that you were prepared to help IEA. The Second Science study took place when it was difficult to obtain funding for international studies of educational achievement. In the end the whole study was run on a shoe string and there were many strings attached to the US money

that became available. It was due to you that the study was brought to a reasonable conclusion. Thank you. Indeed, if the Second Science study had not been brought to a reasonable conclusion there might not have been any IEA today.

Another common venture was the International Encyclopedia of Education. The first edition of ten volumes was published in 1985 and the second edition of twelve volumes in 1994. Again in the years preceding publication you agreed to be the Section Editor of the methodology articles. In both cases this resulted in your famous 'International Handbooks of Educational Research, Methodology, and Measurement'. The first one, as you know, was a best seller for the publishing company. Thank you, too, for helping on that venture.

Finally, I should like to pay tribute to your great ability to help students with their theses and ensure that they learn. You have helped many people across the last four decades. They are from several continents and I am sure that they are all grateful for what they learned from you.

There are two areas of research where I have developed a special interest and both are connected with you. The first concerns the relationship between research results and the dissemination of them. In 1968 you wrote up the results of the first national survey of educational achievement in Australia. The book was called 'Variation in Mathematics Education in Australia: some interstate differences in the organisation, courses of instruction, provision for and outcomes of mathematics education in Australia'. The data had been collected in 1964 and the book was to have been published in November 1968. But the book was shredded because one of the findings did not support the initiatives taken by the Department of Education in New South Wales. This was my first experience of a report being suppressed because of the findings being presented in it. It was important to note that the finding was a research finding and not an opinion of the researcher involved. I am sure that you still feel a bitter taste in your mouth when you recall the experience.

We have both experienced other examples of research reports not appearing at all or not appearing on time. In the United States a mathematical sociologist friend of mine conducted a large study of private and public schools. The publication was delayed on the grounds that the results should really be replicated first. Ken Ross and I tested Grade 6 teachers in Indonesia and compared their scores with those of the pupils. It was deemed better not to publish the results. From about 1980 onwards there seemed to be a relative calm about research reports for about two decades. But, in 2003, I suddenly experienced two occasions when results were suppressed. The first concerned a national study of student achievement and its determinants in Grade 5 in Vietnam. I was heavily involved in the study. By mid-2004, the Ministry had still not given permission for the publication of the volumes of results. This was the first time that Vietnam had had a study of this kind and I suspect that the policy-makers in the Ministry were not used to the idea of using research results as a basis for the decision-making. In 2003, I was also involved in an African cross-national study where a Minister of Education of one of the

participating countries declared that the research in his country was flawed and therefore no comparisons should be made among provinces in his country and between his country and other countries. The research was not flawed but it was politically not in the Minister's interest to have comparisons made and hence his remarks. His successor immediately rescinded his ban. It is still my hope that ministries of education and researchers will work together to conduct research and use the results for improving their system of education. I now have a great interest in how ministries of education deal with the research results at their disposal.

The second area of research in which I am interested and which I connect with you is that of maintaining high standards in the conduct of research. We have often complained to each other about some unfortunate trends that have appeared in some studies where we believe that standards of research have been falling. Others may think that it is just two old fuddy-duddies complaining to each other. I am convinced that there are very large differences in the quality of research studies, both national and international. The best are very good and the worst are so full of technical errors that they should be consigned to the waste-paper basket. Indeed, where the research is of poor quality it is dangerous to disseminate the results. Readers of all reports should read critically and judge the quality of the research and be prepared to consign some reports to the waste-paper basket.

In what follows I have tried to write down a set of questions that readers of research reports of surveys of educational achievement might like to pose before deciding about the technical quality of the work. I have deliberately kept the discussion non-technical. You may not agree with all of the points. I would have loved to have your opinion and comments before publishing but, alas, in this context it has not been possible. But, I do hope that the researchers in national centres of research as well as the international researchers involved in studies such as those run by the International Association for the Evaluation of Educational Achievement (IEA), the Programme for International Student Assessment (PISA), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Programme d'analyse des systèmes éducatifs de la Confemen (PASEC) (Confemen is the conference of the Francophone ministers of education), and the Latin American Laboratory for the Assessment of Quality in Education (LLECE) will pay attention to what I have written.

John, the last four decades have been interesting and we have experienced and learned a lot. I hope that you continue to be able to live and work in South Australia for many years to come.

Ten Points to Ensure that Cross-National Studies are of Good Quality

There are several cross-national studies in existence at the beginning of the 21st century. They are of different quality. Some have very high standards of quality and others are of very questionable quality such that the results cannot be trusted. Whereas it is the prime responsibility of the researchers themselves to ensure that the quality of their research is good, it is also for each Ministry of Education to ensure that the quality of all aspects of the research is high by not allowing poor researchers to have tax payers' money with which to conduct poor research. In 1992 IEA published 'The IEA Technical Handbook' (Keeves 1992). In 1994, Andreas Schleicher wrote up 'Standards for the design and operations in IEA Studies' (Schleicher, 1994). The International Academy of Education has published a book that includes many aspects of research that must be heeded (Beaton et al., 1999). At the same time IEA produced its 'Technical Standards for IEA Studies' (Martin, Rust, and Adams, 1999). It is in such publications that many technical details will be found. The aim of this short article is to highlight ten points that readers should look for in any study in order to be able to judge the technical soundness of the research.

Some points have been marked with an asterisk (*). This denotes that these aspects of the research are particularly critical and that if the researchers have failed to do their work well in one or more of these aspects, then the results of the study are not to be trusted. It is up to readers to demand from researchers that they describe what they have done in detail and accurately.

Have the aims of the study been stated explicitly?

What were the aims of a study? Have they been clearly stated? Has the relationship of the aims to the policy and theory-oriented issues been described? Then, have the aims been operationalised into research questions? It is always troublesome to read research reports where it is unclear from the beginning which research questions the researchers were attempting to answer. Indeed, one sometimes forms the opinion that the researchers themselves were not too sure what they were trying to do. Has evidence presented in documents or reports to show that the research questions that had been developed addressed important policy and theory-oriented issues in the country or countries concerned? (If this is not the case, then there is a danger that the research issues were the favourite topics of the researchers rather than those of the practitioners.) Is there evidence to show that the design of the study was specifically developed to allow the policy and theory-oriented issues to be answered? In some studies, great effort has been invested in the identification of the policy issues common to many systems of education. The research questions are developed to answer the policy questions and then 'blank' or 'dummy' tables are developed in order to show how the answers will be reported. If this has been done, then the researchers will have reported it in their research report. Sometimes the researchers write of a conceptual model being developed. It is to the reader to ensure that the conceptual model has resulted in specific research questions that can be answered by examining the data.

In international studies, it is sometimes stated that the interests of different systems of education are too different to be able to have a set of research questions to guide the study. In my experience, all countries are interested in levels of provision and attainment (of inputs, processes and outcomes) and also in the equity of these levels

across administrative units such as regions or provinces within a country as well as among schools.

The questions to pose are:

- Have the aims of the study been stated clearly and are they relevant?
- Have the research questions been developed with care?

Was the defined target population appropriate (and comparable)?

If, say, the desired target population was all pupils in second grade, the reader must ask if this grade was appropriate for the kind of questions posed about the system of education.

Where comparisons were made across countries, was like being compared with like? For example, if students in a specific grade group were being compared for their achievement, were all of the students in the grade included in the target population or were some students excluded? It is usual to have some students 'excluded' either because they are small in number (and it would be exorbitantly expensive to collect data from them - for example, in very isolated areas) or because they are in special education schools (for example, for students with visual or hearing impairments). These students are normally referred to as the 'excluded' population. It is normal to keep these 'excluded' down to less than five per cent of all students in the 'desired' target population. The 'defined' population is the 'desired' population minus the 'excluded' population. What is **not** acceptable is to have two per cent excluded in some countries and 14 per cent in others. Were the different extents of school and student level exclusions and the likely impact of these exclusions on comparisons of means and distributions across countries reported? What should make the reader extremely suspicious is when no excluded students are reported. The researcher who knows what he/she is doing will always report the extent of the excluded population with the reasons for such exclusions. If information has not been reported on this matter, then it is likely that no attention was paid to it and the reader therefore has no idea what is being compared with what. This is a sign of a bad study.

The same argument applies when age groups are being compared. One argument for using age groups rather than grade groups is to discover the achievement of the students born between certain dates (for example, a calendar year). This approach seeks to examine how systems of education have coped with the education of an age cohort. Where systems have high rates of grade repeating it is possible to have students of, say, age 13 or 14 spread across several grades. Some systems will argue that the tests are too difficult for those students who are three grades behind the others and these should therefore be 'excluded'. In this case, either the tests do not have enough 'bottom' to them (in which case it can be argued that the tests are not appropriate for all of the students), or the students should be awarded zero or chance scores. One way of dealing with this problem is again to apply the rule that not more than five per cent of the students should be 'excluded'.

Some of the questions to be posed are:

- Were the 'excluded' population and the ensuing 'defined' population described?
- Was the excluded population less than five per cent of the desired population?
- Were the target populations really comparable?

Was the sampling well conducted?

The main object of sampling is to ensure that each student in the 'defined' target population has a specified, non-zero chance of entering the sample. Was this done? Since there is usually a shortfall for various reasons between the designed sample and the actual sample, it is common to calculate and use sampling weights to correct for any disproportionality among sampling strata. Any study that does not report how this was done is suspect. The explanation will always be there, be it in a footnote or technical chapter or report. The more there are differences among schools the higher the number of schools that have to be in the sample. The statistic used for describing the difference among schools is rho. Has this been mentioned?

If it is anticipated that a sector of the system or special group of students should be studied in depth this will require more students for that group than would normally be the case and this will have implications for the total sample size. There should also be a table in which the planned sample figures (for schools and students) and achieved sample figures (for schools and students) are presented. The response rate (proportion of schools responding multiplied by the proportion of students responding) should be greater than 0.85 (see also Point 7).

Furthermore, the population estimates derived from the samples should have a sampling error that is acceptable with respect to the policy decisions that are based on the results. Since the mid-1960s, many of the major international studies have adopted the standard of having sample designs that have the same or better sampling precision as a simple random sample of 400 students for educational outcome measures. This level of sampling precision provides sampling errors for results on test items (percentage correct) of no more than 2.5 per cent for one standard error and no more than 5 per cent for two standard errors. This means, for example, that for a population estimate of 50 per cent then one can be sure, 19 times out of 20, that the true value of the 50 per cent lies between 45 and 55 per cent. Since in nearly all countries the sample is a two-stage sample (first a sample of schools and then students within schools) it is important that the standard error is calculated to take this into account. Many make the mistake of using SPSS that produced a standard error that assumes that the sample was a one stage simple random sample. This is an incorrect standard error because it has not taken into account the two-stage nature of the sample and will produce smaller standard errors than what the real case is. Where differences between means were reported (say for gender or urban-rural) then differences would be found that were not really significant. A characteristic of a good study is that the correct standard error is calculated and the researchers report what they did.

The question for the reader then is 'Was the sampling conducted in such a way as to yield standard errors of sampling that were acceptable for the purposes of the study?' It is usually the case that researchers who are knowledgeable in the area of sampling will have provided a detailed description of the steps of sampling and the correct sampling errors. If this information has not been provided, then there is a distinct possibility that the samples are suspect. It is also usual for the standard errors of sampling to be presented in the tables of results. If they are not there, then the reader should be wary.

The reader should be wary because if the achieved sample is too small (large difference between the planned sample and achieved sample) or the excluded population is greater than five per cent or the correct rho was not known and so on, then the calculated means and variances for any one variable may be very wrong.

It can sometimes be the case that the sampling has been good and that significant differences have been found. But with a large sample it is usually the case that significant differences are found. The question then arises as to whether these differences are educationally meaningful. Where a significant difference can be only worth one item on a test, then it is not educationally meaningful to report it. So, although the significant differences must be correctly calculated the interpretation must be exercised with caution.

The questions to be posed about sampling are:

- Was the confidence limit for the sampling mentioned?
- Was the rho used for sampling mentioned?
- What was the response rate (schools x students) greater than 0.85?
- Were sampling weights calculated and used?
- Were sampling errors calculated and reported for every estimate?
- Was care taken in the reporting about the difference between a statistically significant difference and an educationally meaningful difference?

Were the tests well constructed and pre-tested?

It is clear that the tests must be seen to be appropriate for measuring the subject-matter being tested. If they are not shown to be appropriate and valid and reliable then the reader has every reason to be suspicious. This applies whether the test is a national or international one.

In most cases tests are meant to measure what the students should have learned by a particular point in the school system. Occasionally, they are meant to measure what the students will need when they enter society. Whichever the case may be, it is important to prove that the tests fulfil their function.

First, it is normal to have a fairly detailed description of what is meant by reading or mathematics (or whatever the subject-matter is) at the point in question in the school system. If this is missing from the research report (even as an appendix), then there are reasons to doubt the enterprise. Second, it is normal to have a test blueprint or assessment framework. This can take various forms but is usually a grid with content on the vertical axis and cognitive behaviours on the horizontal axis. Each cell in the blueprint represents an educational objective. Again, it is normal that the blueprint is shown in the report.

Where the study aims at measuring what the students have learned to date, the test instruments must cover the intended curriculum of the country or participating countries. This normally involves a two-stage process: first a content analysis of the curricula (via curriculum guides, textbooks, examinations, and what teachers say they teach) in the various countries; second, on the basis of the first step, the production of a national or international blueprint for the test(s). While many of the curricular objectives will be common across countries, some objectives will be common to only a subset of countries. Finally, the subject-matter is often broken down into domains. In reading this is often: narrative prose, expository prose, and documents. These domains must be specifically described.

In some cases, the aim of the study will focus on other outcomes such as whether the pupils can read well enough to ‘cope in society’ or ‘to progress to the next grade’.

Table 1. Skill Level

Skill level	Reading	Mathematics
Level 1	*Pre-reading*: Matches words and pictures involving concrete concepts and everyday objects, and follows short simple written instructions	*Pre-numeracy*: Applies single step addition or subtraction operations. Recognizes simple shapes. Matches numbers and pictures. Counts in whole numbers
Level 2	*Emergent reading*: Matches words and pictures involving prepositions and abstract concepts; uses cuing systems (by sounding out, using simple sentence structure, and familiar words) to interpret phrases by reading on	*Emergent numeracy*: Applies a two-step addition or subtraction operation involving carrying, checking (through very basic estimation), or conversion of pictures to numbers. Estimates the length of familiar objects. Recognizes common two-dimensional shapes
Level 3	*Basic reading*: Interprets meaning (by matching words and phrases, completing a sentence, or matching adjacent words) in a short and simple text by reading on or reading back	*Basic numeracy*: Translates verbal information (presented in a sentence, simple graph or table using one arithmetic operation in several repeated steps. Translates graphical information into fractions. Interprets place value of whole numbers up to thousands. Interprets simple common everyday units of measurement
Level 4	*Reading for meaning*: Reads on or reads back in order to link and interpret information located in various parts of the text	*Beginning numeracy*: Translates verbal or graphic information into simple arithmetic problems. Uses multiple different arithmetic operations (in the correct order) on whole numbers, fractions, and/or decimals
Level 5	*Interpretive reading*: Reads on and reads back in order to combine and interpret information from various parts of the text in association with external information (based on recalled factual knowledge) that completes and contextualizes meaning	*Competent numeracy*: Translates verbal, graphic, or tabular information into an arithmetic form in order to solve a given problem. Solves multiple-operation problems (using the correct order of arithmetic operations) involving everyday units of measurement and/or whole and mixed numbers. Converts basic measurement units from one level of measurement to another (for example metres to centimetres)
Level 6	*Inferential reading*: Reads on and reads back through longer (narrative, document or expository) in order to combine information from various parts of the text so as to infer the writer's purpose	*Mathematically skilled*: Solves multiple-operation problems (using the correct order of arithmetic operations) involving fractions, ratios, and decimals. Translates verbal and graphic representation information into symbolic, algebraic, and equation form in order to solve a given mathematical problem. Checks and estimates answers using external knowledge (not provided within the problem)
Level 7	*Analytical reading*: Locates information in longer (narrative, document or expository) texts by reading on and reading back in order to combine information from various parts of the text so as to infer the writer's personal beliefs (value systems, prejudices, and/or biases)	*Problem solving*: Extracts and converts (for example, with respect to measurement units) information from tables, charts, visual and symbolic presentations in order to identify, and then solves multi-step problems
Level 8	*Critical reading*: Locates information in a longer (narrative, document or expository) texts by reading on and reading back in order to combine information from various parts of the text so as to infer and evaluate what the writer has assumed about both the topic and the characteristics of the reader – such as age, knowledge, and personal beliefs (value systems, prejudices, and/or biases)	*Abstract problem solving*: Identifies the nature of an unstated mathematical problem embedded within verbal or graphic information, and then translate this into symbolic, algebraic, or equation form in order to solve the problem

In these cases exercises must first be undertaken in each country to have panels define what is required for these types of outcomes. This is a laborious process but must be conducted in a convincing way.

In yet other cases, it is normal to have a hierarchical set of skills or competencies that is typical of the grade or age group being tested. Each level is described by what the students can do. An example for Grade 6 from the SACMEQ study has been presented in Table 1. There were eight levels but this covered both Grade 6 students and their teachers.

In this case, it is important to write items that fit each skill level.

In general, there is much less variation among countries in subjects such as reading and foreign languages than in subjects such as mathematics, history and social studies. There must, however, be agreement on the international blueprint and this must cover the bulk of the curricula in all countries if it is the intention of the study to focus on the common contents of national curricula.

Test items must be written to cover all cells having objectives in the blueprint. The item formats must be agreed and justified. The items must be trial-tested and analysed. Where multiple-choice items are used then the distractors must be plausible not only in terms of content but also in their diagnostic and distracting power. Constructed response questions requiring students to construct answers should be pre-tested to ensure that they will yield a range of responses that can be reliably scored. Where scaling is being used there must be agreement on the substantive meaning of the scale in terms of student performance on specified tasks at specified points of the scale. There must be agreement on the appropriateness of the items and the tests must be shown to be reliable. Where there is an attempt to measure change over time, say from the last survey to the current one, then there must be sufficient common items between the two points in time to allow change to be reliably measured. Finally, items should be tested for item bias in each and every country. The psychometric properties of the test items should be similar over a sufficiently large number of countries. Where overlapping tests have to be used it must be shown at the trial stage that the common items used to allow calibration onto the same scale fulfil their purpose.

In some instances, ‘hands-on’ performance assessment tasks may be deemed necessary to cover the full range of objectives in a subject area. The design of such tasks should take into account the (usually) limited amount of time available for testing, the need to make use of equipment which is simple and available in multiple copies and not beyond the resources of participating countries, and the need to yield responses that can be graded reliably across countries. Where the rotation of subtests has been undertaken there must be proof that this was well conducted. There must, for example, be common items in the subtests such that all items can be brought onto one scale.

Finally, there must be evidence that the tests are valid. The kinds of validity tests will have been reported if the researchers have conducted them. It is for the reader to determine whether these are convincing or not. Certainly, if the researchers have not reported them, then they surely have not undertaken them. In this case the reader has no idea about the validity of the tests and the reader should be suspicious. When undertaking validity checks internationally, it is usual for the researchers to have asked the different participating countries to identify the items in the test that are certainly part of their curriculum. The researchers then calculate a national curriculum score as well as a total test score (all items in the test whether in the curriculum or

not). All nations are then allocated a series of scores: the total score and then a score based on the curriculum for country A, then country B, and so on. It has been shown in various international studies that the rank order of countries does not change significantly according to which score is used. This is an indication that the test is good in the sense that it measures the outcome variable in a way that satisfies each country.

The questions to be posed for test construction are:

- Was the subject-matter for the test well and convincingly described?
- Were the domains in each subject-matter well defined?
- Were the processes used to analyse the existing curriculum or to identify the skills needed by society convincing?
- Was the item-writing process convincing?
- Were the items tried out and analysed?
- How was the scaling organised?
- Were the validity checks convincing?
- Were the test reliabilities high enough?

Were the questionnaires and attitude scales well constructed and pre-tested?

Many believe that it is easier to construct questionnaires and attitude scales than to construct tests. They are mistaken. There is a whole technology that can be used to help with test construction. This exists to a much lesser extent for questionnaire construction. The secret for questionnaire construction (and attitude scale construction) is pilot, pilot, and pilot. If no piloting occurred then it is most likely that the measures were no good.

The **questionnaire** instruments must include questions to cover all of the indicators needed to answer the research questions raised at the onset of the study. Several of the indicators will be what are normally called ‘derived variables’ – that are constructed from the information obtained from one or more questions. Some will be simple ratio variables while others will be factors consisting of several variables. In nearly all cases there will be a scale for an individual question or derived variable. The questions must be written in a simple language easily understood by all of the students (able and less able) who have to answer them. All questions must then be trial-tested and analyses undertaken to ensure that the questions are providing accurate and reliable information for the indicators and ‘derived variables’. The lists of derived variables and how they have been formed are normally listed in an appendix in the report together with information on their reliabilities.

The **attitude** instruments, sometimes a part of the questionnaires, measure selected attitudinal dimensions. The dimensions must be described. Attitude items are normally collected through special small studies from the target population members. They too are trial-tested and analyses undertaken. Very often about three times as many items are needed for trial testing as for the final attitude scale measure. The final scale must be shown to be reliable and valid for the purposes for which it is intended. In the description of the construction of the attitude scales, it is important to see how the researchers arrived at the number of options for answers were arrived at as well as which items were inverted in order to avoid students replying to all in the same way.

The questions to be posed are:

- Was the process described to ensure that questions were written to cover all of the research questions for the study?
- Were the attitude statements collected from the population for which they were intended?
- Were the questionnaires and attitude instruments subject to several sets of piloting?
- Were the derived variables described?
- Where required, was the scaling of the measures described?

In cross-national studies involving the translation from a central language to others, were verifications of the translations carried out?

It is clear that all items should be translated and then checked, through a thorough verification process, to ensure that the linguistic difficulty of the item is about the same in all languages? There are elaborate procedures for doing this and the researchers will certainly have described them if they did them. The verification procedure is also quite expensive. If the verification process has not been undertaken, then the reader has no idea about the comparability of the test and questionnaire and attitude items.

The main question is:

- Was a thorough verification undertaken of the translation?

Was the data collection well-conducted?

The data collection stage for any study is crucial. The object of the data collection is to test all respondents selected in the sample and to have them complete every question in the questionnaires and all test items that they are able to answer. Normally, a manual is written for the persons in charge of the data collection at the national level in each country. This manual is required so as to ensure that the data collection procedures proceed in a manner that will provide valid data under conditions that are uniform at each data collection site.

The National Centre manual (sometimes called a National Research Co-ordinator (NRC) or National Project Manager (NPM) manual) should cover every possible detail that must be taken into account when conducting the data collection. This involves ‘school forms’ and ‘student forms’ to ensure that the correct schools are selected, the correct students are tested (and not others), and the correct teachers are selected (where questionnaires or tests are being administered to teachers). A second data collection manual is usually prepared for the data collectors and details everything to be done within each selected school. A third test administration manual spells out (a) what each test administrator has to do and say during the actual testing sessions, (b) the procedures and timing for the administration of the instruments, and (c) how to parcel up the instruments and return them to a central point. There should be very few, if any, missing schools and very few missing students in the data collection. Again, the authors of the reports should report the percentage of missing schools and missing students. It is often said that not more than 10 per cent of schools should be missing from the sample and not more than 20 per cent of the students.

However, since there are no completely valid procedures for dealing with missing data, these figures should be taken as absolute maximum levels.

In some studies, insufficient care is taken to ensure that there are as few non-completed questions as possible. It is essential that the research centre ensures that the tests and/or questionnaires are collected by someone who checks for this in the school before the instruments leave the school. In this way it is possible to spot questions in the questionnaires that have not been answered and have them completed before the instrument leaves the school.

In large scale studies, it is also often the case that quality control of the actual testing is carried out. Specially trained test administrators are sent to randomly selected schools to observe the testing to ensure that it was well conducted. Checks are made that the correct students were tested, for the seating in the testing room to ensure that students did not cheat, and so on.

As a result of the data collection, the response rate, as mentioned in Section 3 on sampling earlier, should be at least 85 percent (school response rate x student response rate).

The questions to be posed are:

- Were the manuals described in the research report?
- Were the tracking forms (school forms and student forms) described?
- Were the resultant response rates (without replacement schools) high enough?
- Were there only very few missing data?
- Was a quality control on the testing carried out?

Were the data recording, data cleaning, test scoring, and sample weighting well conducted?

The data are usually recorded on computers at the National Center. Typically the researchers provide the data entry software to be used. The researchers often undertake a double entry of ten per cent of the instruments for verification purposes. Good data entry software provides a number of initial checks on the data that can be corrected immediately during the data entry process. Then all sorts of further checks are conducted in both national and international studies. Where there are several countries in the study, a common set of cleaning rules should be used. It is very difficult to have comparability if each nation has used a different set of cleaning rules. There are always extra errors in data entry no matter how good the data entry program. By undertaking consistency checks it is possible to identify questions on the questionnaires where an error occurred on the part of the respondent. These problems are reported back to national centers; they then contact the schools for elucidation and then send the correct data back to the international data processing center. The necessary changes are then made. This cleaning process can take a long time especially when there are many countries in the study. However it should be mentioned that a data set from one country, where some carelessness is evident in the data collection and/or data entry, can take an inordinate time to clean.

It is important for the reader to be made aware of those variables where there were so many missing data that they could not be used in the analyses. If there are many variables with more than 20 per cent missing data then the reader should beware. Furthermore, it is important to see how the problem of missing data was tackled. There are several ways of dealing with missing data and one of them is to impute the values. Whichever approach was used, the researchers will have reported it in their

report. If no mention is made of how the researchers dealt with missing data, then the reader should beware.

Where constructed response items have been used in the test, those items will have to be scored and a test scale computed. Again, it is important that the scoring procedures have been reported, typically in an appendix or in a separate technical report.

Finally, in order to account for different probabilities of selection (due to shortfall in the data collection, disproportionate selection across strata, inaccurate sampling frames, missing data, etc.) sampling weights, have to be calculated. Since there is nearly always shortfall in survey studies, then sampling weights are needed. Either there will be a description in the report about how the weights were calculated or (but only in exceptional circumstances) there will be a description and justification of why sampling weights were not required. If there is no description of how sampling weights were calculated, it is very likely that they were not used and therefore the estimate of the means and variances of variables will be wrong.

The questions to be posed are:

- Was a data entry program used that included consistency checks?
- Were further checks carried out?
- Were there many variables with more than 20 percent missing data?
- Were sampling weights calculated and used?

Were the data analyses well conducted?

In all reports there are usually some univariate analyses and some multivariate analyses. Although the analyses must fit the research questions, most issues are sufficiently complex to warrant more than only univariate analyses.

Some of the analyses will be simple and others complex. It is normal for a study to have a set of dummy tables produced at the onset of the study. These dummy tables cover the research questions posed. The analyses will have been undertaken to complete the tables. If the reader is not experienced in data analysis, it is usually wise to have experts advise him or her on the appropriateness of the analyses for the questions posed.

Some examples of inappropriate analyses often encountered in poor studies might be of use to readers. It can happen that the researchers report a mean score for a cell on a test blueprint (an aspect of achievement) despite the fact that the number of items per cell was insufficient to derive a separate scale. It can also happen that the researchers comment on a zero-order correlation without considering and testing if it would be non-significant if other variables such as the socio-economic status of the students or school location (rural/urban) were controlled for. Another example would be undertaking multivariate analyses between schools using 100 variables when there are only 150 schools (normally in a case like this one would need at least six times more schools than variables).

Where new constructs (or factors) and scales have been produced during the data analyses, it is important that they have been described in the report.

There are also errors in interpretation that occur in poor studies. Sometimes the authors of the research reports exhibit a lack of caution when they forget that correlations do not necessarily signify causation. Or when they forget that the responses to perceptual questions do not necessarily depict what actually is the case (for example, teachers' perceptions of the goals of the school collected through a teacher questionnaire).

As mentioned earlier, it is important that each estimate be accompanied by a standard error of sampling. There are now good programs for the calculation of standard errors and it is reasonable to expect that every estimate should be accompanied in the tables and figures by the standard error of sampling. If a report does not include these, then the reader should be suspicious that the researchers know what they are doing.

Some of the questions to be posed are:

- In reporting test scores (either total scores or subscores) were sufficient items used to create the score? If not, then the reader should be suspicious that the researchers knew what they were doing.
- Were the appropriate variables taken into account when examining relationships between variables?
- Have the standard errors of sampling been reported for every estimate in the report?

Were the reports well written?

The reports should be clearly written and deal with each of the policy issues in turn. The source of the data under discussion should always be clear, as should arguments concerning the interpretation of the analyses. It should be made clear that in some studies the major univariate results are reported first and then major clusters of research questions are reported in separate reports.

It is important for the researchers to obtain feedback on their reports before the reports are finalised. In part, this is from other researchers but also from the intended users of the results in the report as well as from concerned persons such as school heads and teachers. Where the main user of the results will be the ministries of education it helps a great deal if the researchers have discussed their recommendations with those responsible in the ministries of education before publication. Again, if the researchers have done this they will describe the process. It is also useful to ministries of education if the researchers cluster their results not only by theme but also by cost (low, medium, high) and by length of implementation (short, middle term and long).

Conclusion

Finally, it is normal for the researchers to make the data set(s) available as an archive so that others can analyses the data themselves in order to check the veracity of the statements that the researchers made and also to explore the data to answer other questions that might be asked of the data. It is very important that the archives are made available very soon after the publication (or even at the same time) and in format that is user friendly.

References

Beaton A E, Postlethwaite T N, Ross K N, Spearritt D, Wolf R M (1999) *The benefits and limitations of international educational achievement studies*. Paris. International Academy of Education and International Institute for Educational Planning.

Keeves J P (Ed.) (1988) *Educational research, methodology, and measurement: an international handbook.* 1st edition. Oxford. Pergamon.

Keeves J P (Ed.) (1992) *The IEA Technical Handbook*. The Hague. IEA.

Keeves J P (Ed.) (1997) *Educational research, methodology, and measurement: an international handbook.* 2nd edition. Oxford. Pergamon.

Martin M, Rust K, and Adams R. (1999) *Technical Standards for IEA Studies*. Amsterdam. IEA.

Schleicher A (1994) *Standards for the Design and Operations in IEA Studies*. The Hague. IEA.

10

Exploring the Effects of Language Proficiency upon Secondary Students' Performance In Mathematics in a Developing Context

Sarah J Howie & Tjeerd Plomp
Centre for Evaluation and Assessment, University of Pretoria & IEA, The Netherlands

***Dedications**: John showed both incredible patience, kindness, not to mention his expertise, in directing Sarah to the appropriate methods for secondary analysis of the TIMSS data. By alerting her to the techniques of PLS and multilevel analysis (and demonstrating both packages) in response to her desire to explore the data, he provided the light at the beginning of what had seemed a considerably long and dark tunnel. He was also waiting at the end of the tunnel in the form of being an examiner for the doctoral dissertation on an international panel in front of which that dissertation was defended. He was also kind enough to recommend to Kluwer publishers that the dissertation be published. Thus, a day spent with John in 1999 at Flinders changed her life forever and sparked a passion for secondary analysis and further research based on system-level data. This has culminated in 2003 in Sarah setting up a Centre for Evaluation and Assessment at the University of Pretoria in South Africa, of which one of its primary focuses is large-scale international and national evaluations and a second one is introducing students to a variety of analytical techniques in quantitative analysis.*

For Tjeerd, John Keeves was a friend in IEA from the first day they met at the combined General Assembly and SIMS project meeting in 1978 in Tokyo. He was always willing to give advice on technical issues, which was especially important during the early years of my

chairmanship of the IEA. Furthermore, I could always call on John when I need his specialist help to strengthen the dissertations of my PhD students.

This article demonstrates how the PLS and multi-level techniques were applied to the data which was similar to that John himself had analysed extensively and the benefits of doing this for a developing context.

***Abstract**: South Africa participated in the Third International Mathematics and Science Study (TIMSS) in 1995 and again in 1999 (TIMSS'99) and in both studies the performance was extremely low compared to the other countries in the studies. In both studies more than 70% of the pupils wrote the achievement tests in their second or third language. A national option, an English test, was included together with the TIMSS'99 mathematics and science tests in an attempt to ascertain the level of the pupils' language proficiency. Furthermore, additional questions pertaining to the pupils and their teachers' exposure and usage of English both within and outside of school were also included in the background questionnaire.*

In this research project that included more than 8 000 pupils in 200 schools, all the items pertaining to English proficiency and language usage and their relationship to mathematics achievement were explored. Partial Least Square analysis was used to explore the relative contribution of these factors to pupils' achievement together with other background variables from the student, teacher and principal questionnaires, resulting in the presentation of a school-level, a classroom-level model, a student-level model and a combined class and school-level model. Multilevel analysis was employed whereby a 2-level model (school and class-level and the student-level) was analysed in order to investigate the main factors explaining achievement of South African pupils in mathematics.

The study revealed that the pupils' proficiency of English was a strong predictor of their success in mathematics. A number of other background variables on student and class-level were found to be significant. However, home language and class size were amongst those that were not found to have significant effect on achievement, whilst the effect of socio-economic status had a lesser effect once certain class-level factors were taken into consideration.

Introduction

This chapter reports on research that was conducted by including a national option in IEA's TIMSS'99. The research was a secondary analysis of the performance of the South African pupils in the Third International Mathematics and Science Study 1999 (TIMSS'99). Pupils wrote tests in mathematics and science, but in South Africa pupils also had to write an English test, which was included as a national option. The South African pupils' performance in mathematics was significantly below that of all other participating countries including other developing countries such as Morocco, Tunisia, Chile, Indonesia and the Philippines. This paper presents the final results of a

three-year research project (Howie, 2002) and includes the final Partial Least Squares analysis and multilevel analysis of the effect of language and other contextual factors on mathematics at student and school level.

Objectives of the Study

The research reported here concentrates on the final outcomes of the exploration of the performance of the South African pupils in mathematics and most especially the relationship between mathematics achievement and pupils' proficiency in English. Its relevance stems from the fact that English is spoken as a first language by less than ten per cent of the population and is the language of business and government. It is also one of two languages usually used at South African schools although it is not the most widely spoken language at home. However, the issue around the language policy for teaching and learning has become a sensitive and controversial topic in South Africa just as it is in many other post-colonial countries.

The factors relating to the pupils performance in mathematics and English language proficiency were explored in relation to the background information that was also collected from the pupils, teachers and principals of the schools included in the study.

The research questions addressed in this paper are:

1. What is the effect of language on South African pupils' performance in mathematics?
2. How does the effect of language relate to other background variables collected at school and student level?

The extent of the pupils' aptitude in English and mathematics was analysed. Every South African pupil in TIMSS'99 completed a standardised written skills and language usage test (HSRC, 1990). The performance of the pupils in this test indicates a measure of language proficiency in English. As English language proficiency in South Africa is correlated with other important background variables due to the nature of the country's political past other background variables were explored and included in a multilevel model developed to address the second research question.

Literature Review

A number of reports and articles have been written on the status of mathematics (and science) education in South Africa (Arnott & Kubeka, 1997; Kahn, 1993; Taylor & Vinjevold, 1999; amongst others). A number of factors have been reported pertaining to the poor performance of pupils in the matriculation examinations and in general (Adler, 1998; Arnott & Kubeka, 1997; Kahn, 1993; Monyana, 1996; Setati & Adler, 2000; Setati, Adler, Reid & Bapoo, 2001; Taylor & Vinjevold, 1999). These include: inadequate subject knowledge of teachers, inadequate communication ability of pupils and teachers in the language of instruction, lack of instructional materials, difficulties experienced by teachers to manage activities in classrooms, the lack of professional leadership, pressure to complete examination driven syllabi, heavy teaching loads, overcrowded classrooms, poor communication between policy-makers and practitioners, as well as lack of support due to a shortage of professional staff in the ministries of education. Most of these were reported on the basis of classroom observations and discussions with teachers and other stakeholders. Only Moyana (1996) collected data for the purpose of analysing the factors that had an effect on mathematics performance and utilised inferential statistics. In the Third

International Mathematics and Science Study (TIMSS) conducted in 1995, the performance of the South African pupils was significantly below that of all the other 40 participants in the study (Howie, 1997; Howie & Hughes, 1998). TIMSS provided South Africa with the first national representative overview of how South African pupils were performing in mathematics (and science). However, there were significant language and communication problems with South African pupils learning mathematics in a second language. Pupils in all three Grades (7, 8, 12) showed a lack of understanding of both mathematics questions, and an inability to communicate their answers in instances where they did understand the questions. Pupils performed particularly badly in questions requiring a written answer (Howie, 1997; Howie & Hughes, 1998).

Internationally, research addressing factors related to achievement in mathematics were found using data from, for example Belgium (van den Broek & van Damme, 2001), and Eastern Europe (Vari, 1997), but most were found in the United States (Sojourner and Kushner, 1997; Teddlie & Reynolds, 2000, amongst others). No studies were found either nationally or internationally that attempt to link English Language proficiency, or more general the language of instruction, to mathematics achievement at secondary level using such a comprehensive dataset with data on pupil, class and school levels.

Studies regarding the effects of language on mathematics achievement appear to indicate the importance of language in achievement generally, including mathematics (Berry, 1995; Clarkson, 1991; Tartre & Fennema, 1995; Young, 1997). There seems to be sufficient evidence internationally and some evidence locally to warrant the assessment of language and its relationship to mathematics on a large scale in South Africa.

Clearly there are many factors that have been found on school, class and student levels to have positive and negative effects on mathematics achievement (see Howie, 2002, for a comprehensive discussion). Here summary conclusions are presented whereby only references to African research and/or pertaining to less developed countries are included.

The factors on **student level** discussed in the literature included: socio-economic status (Afrassa, 1998; Cherian, 1992; Howie & Pietersen, 2001, amongst others), books in the home, parental education, parents occupation (Afrassa, 1998; Eshetu, 1988, Gennet, 1991), parental relationships (Eshetu, 1988), parental press, parent's self-concept, pupils' attitudes to mathematics, family size (Behutiye & Wagner, 1995), jobs in the home (Deresse, Wagner & Alemaychu, 1990; Daniel, 1995)), pupils' aspirations, peer group attitudes, pupils' self concept (Howie & Wedepohl, 1997), self expectations, pupils' anxiety (Maqsud & Klalique, 1991), enjoyment of mathematics, attitudes towards maths (Sayers, 1994), , reading ability, gender (Afrassa, 1998), age, attitudes towards teachers (Georgewill, 1990), time spent on homework (Afrassa, 1998). Of these most were investigated in this study as well. The exceptions are parents' occupations, parental relationships, parents' self-concept, pupil anxiety, cognitive ability, reading ability and attitudes to teachers. This is not because they are not important, but rather due to the limitations of the dataset that was used.

On **classroom level**, factors found in the literature were the learning environment, teacher's characteristics (including gender) (Moyana, 1996), teacher's personality, streaming, computers, teachers' competence (Georgewill, 1990; Taylor & Vinjevold, 1999), teacher's confidence, education background, teacher's qualifications (Arnott & Kubeka, 1997; Mpofana, 1989), teachers' methods, class size (Cohn & Ressmiller, 1987), time on task, disruptions in class, calculators, content coverage, and

assessment. Of these factors, most were explored in the study, the only exceptions being teacher's personality and content coverage. The former not having been included in the questionnaires and the latter due to data problems where the data could not be recovered.

Finally on **school-level** a number of factors have been investigated in previous studies. These included textbooks, teacher quality, time on task, leadership, organisation, management (Riddell, 1997), decision-making, within-school hierarchy, communication, school size, professional development (Cohn & Ressmiller, 1987), location, commitment, and the controlled environment. In this study, only textbooks, time on task, leadership, decision-making, school size and location were explored, as the other factors were not included in the data collection.

Theoretical Framework

In order to address the objectives of the study, a conceptual model was developed for this research. This was to allow for the exploration of contextual factors within different levels having an effect on pupils' achievement in mathematics within the context of South Africa. A summary of the model description is given here whilst a complete description can be found in Howie (2002).

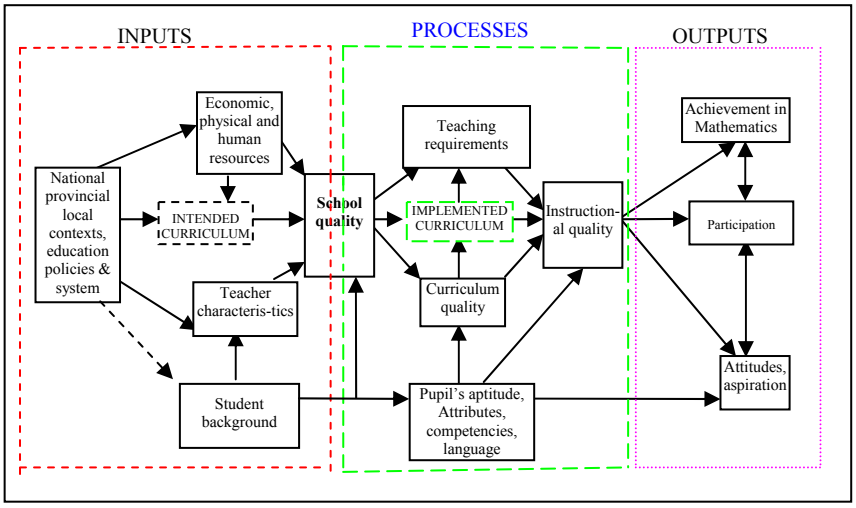


Figure 1 Factors related to mathematics achievement

(Howie, 2002, adapted from Shavelson, McDonnell & Oakes, 1987)

The model was informed largely by Shavelson, McDonnell and Oakes, 1987, as well as other literature discussed in the previous section. It is also underpinned by the IEA's curriculum framework, i.e: the intended, implemented and attained curricula. The model presents the education system in terms of inputs (including contexts), processes and outputs. The inputs are the policy-related contexts on a national, provincial and local level from which the intended curriculum (in the meaning of what should be taught in schools and learned by the pupils) is also designed and developed. They also include the antecedents: the economic, physical and human

resources supplied to different levels of the system; the characteristics of the teachers and the background of the pupils. Inputs into the system affect all the processes of education, which may also be seen as the practice in education. Different processes (relating to what is taught and how it is taught) take place within the districts, schools, and inside the classrooms in terms of the implemented curriculum (in the meaning of what is actually being taught in the classrooms), teaching (in the meaning of the context and conditions under which teachers work) and instruction. The outputs, also seen as the outcomes, eventuate in terms of the achievement of learners in specific subjects such as mathematics; participation in class and school activities, and finally learners' attitudes towards subjects and schooling and aspirations for the future. It is expected that, due to the dynamics of the processes included in the model, there will also be indirect benefits and outcomes, such as improved learner participation partly due to improved curriculum quality.

The model serves as an important theoretical and conceptual basis for the analysis of the TIMSS'99 data. As the data were collected on a number of education levels, namely, school, classroom and learner level, the model serves as a guide to explore the causal links for the learners' achievement.

Research Design

The study was exploratory, focusing on the secondary analysis of the TIMSS'99 data related to mathematics achievement and a summary of the methods as well as the findings are discussed in this paper. The data were explored to investigate the reasons for the pupils' performance and to examine the inter-relationships between achievement and the background variables revealed by pupils, teachers and the school principal. In particular, the exploratory part of the study was to determine the factors that influence mathematics achievement and performance of South African pupils and to ascertain the effect of South African pupils' language and communication skills on their achievement in mathematics.

Sample

The TIMSS requirements stipulated that a minimum of 150 schools be tested and that a minimum of one class (preferably one whole class) per school be tested. The South African initial sample was expanded to 225 to accommodate the inter-provincial analysis required. The nationally representative sample was drawn and stratified by province, school sector and medium of instruction. A two-stage stratified cluster sample of 225 schools was randomly selected and stratified according to province, type of education (government or private) and medium of instruction (English and Afrikaans). Tests and questionnaires were administered (by the Human Sciences Research Council based in Pretoria, South Africa) to more than 9000 pupils. Questionnaires were also administered to 200 school principals and 400 teachers of mathematics and science at Grade 8 level. After the data cleaning a representative sample of 194 schools and 8,146 pupils was used in the data analysis.

Instruments

In addition to the TIMSS'99 instruments, namely eight test booklets containing mathematics and science achievement tests, pupil questionnaire, mathematics teacher questionnaire, science teacher questionnaire, and school questionnaire, an English language proficiency test was included specifically for South African pupils. This instrument had previously been validated by the Human Sciences Research Council

and standardised for Grade 8 Second Language pupils in South African schools (HSRC, 1990). At the time of the TIMSS'99 study, this test was the only standardised South African second language test at the Grade 8 level that could be found. Questions were also included in the TIMSS'99 pupils' and teachers' questionnaires, to ascertain the extent and level to which the pupils are exposed to English. They included pupils' home language, ethnic group, the language spoken predominantly by the pupils in the mathematics class, the language used by the mathematics teacher in class, media languages pupils are exposed to and the language of their reading materials. In this research, data from the test booklets, pupil questionnaires, mathematics teacher questionnaires, school questionnaire and the national option were analysed.

Data Analysis

After examining frequencies, building constructs and reviewing correlation matrices, Partial Least Square analysis and multilevel modelling were applied. Given that there is a number of variables reported to influence pupils achievement as well as the vast number of variables in the database, and that some of these were intricately inter-related, Partial Least Square analysis (PLS) (Sellin, 1989) was used initially to analyse those student-level and classroom-level factors that influenced pupils' achievement in mathematics. This type of analysis allows one to estimate or predict both the direct and indirect effects of a set of independent variables on a dependent variable (with each path taking into account the effects of all the other variables).

Due to the fact that data were collected on three levels – student-level, class-level and school level, multilevel modelling (Institute of Education, 2000) was applied. In this study, multilevel modelling was used to distinguish between the variance in mathematics achievement uniquely explained by student-level factors as opposed to the variance uniquely explained by the classroom and school-level factors and to investigate the individual effects of variables inserted in the model once the multilevel structure of the data is taken into account. As only one class per school was sampled, only two levels could be analysed and this was due to the original TIMSS'99 design where class and school are considered to form one level.

Results

Results of the Mathematics Tests

Overall, South African pupils achieved 275 points out of 800 (standard error, 6.8) in the mathematics test, whilst the international average was 487. This result is significantly below the mean scores of all other participating countries, including the two other African countries of Morocco and Tunisia as well as that of other developing or newly developed countries such as Malaysia, the Philippines, Indonesia and Chile.

Pupils who spoke the language of the test more frequently attained higher scores on the mathematics test. When comparing those pupils that almost always or always speak the language of the test to those that never speak the language of the test, the former achieve scores that are more than 140 points higher than the latter. In South Africa, pupils who spoke either English or Afrikaans at home achieved higher scores (about 100 points above the national average) than those who did not and. What is illuminating is that children that spoke other languages at home (for example, Greek, Portuguese or Tamil) and therefore also learned in a second language, scored only 20

points on average less than first language speakers However, children speaking African languages at home attained 100 points less than the other group of second language speakers (Howie, 2001). No group of pupils came close to attaining the international average for mathematics

From a comparative analysis of South African pupils' results with other countries in TIMSS'99 (see Howie, 2001), some interesting observations were made. More than 70 per cent of pupils from South Africa, Indonesia, Morocco, Philippines and Singapore did not always speak the language of the test at home. Nonetheless, the mean achievement scores vary considerably across this group of countries and there are also some interesting trends in the data. Pupils in Malaysia generally did considerably better in mathematics than those from Indonesia. Nonetheless, there is a similar trend in both countries where pupils who never speak the language of the test at home, namely 9 per cent in Indonesia and 10 per cent in Malaysia, still appeared to outperform those who always or sometimes spoke the language of the test at home. It suggests therefore that the differences between language groups are not only dependent on language. Indonesia for instance is described as a highly diverse country with more than 600 languages and 200 million people (Baker & Prys-Jones, 1998, p. 375) and yet apparently their pupils do not appear to have been as disadvantaged by writing the test in a second language. A similar pattern was also observed for Morocco and the Philippines in mathematics. In Singapore there does appear to be a difference, but yet those who never speak the language of the test at home, still outperform pupils from 33 other countries.

Looking at the other African countries, the scores of those never speaking the language of the test at home are better in the case of Morocco for mathematics and are comparable for those from Tunisia. This issue needs to be explored further as it appears from the data, that the pupils from other developing countries do not seem to be disadvantaged by writing tests in their second or third language in mathematics or science, however it is not clear why this is. Important lessons for South Africa may lie in the answers.

Results of the English Language Test

In addition to the mathematics and science tests conducted in TIMSS'99, an English language proficiency test was included that aimed to assess pupils' writing related skills and language usage in English. The test comprised 40 items, which were multiple-choice items. Thirty of the forty items had four options, whilst the remaining 10 items had two answer options.

The overall mean score for the language test was 17 out of 40 (42.5%; n= 8349). The minimum score attained was 0 and the maximum score 40. In general, the scores for boys and girls were comparable. As the test was designed for English second language speakers, it is not surprising that native English speakers performed the best of all language groups (25 points out of 40), although one might have expected the scores to have been higher given this fact. The Afrikaans speaking children attained the next highest score with 21 points out of 40. The scores were more or less consistent across the pupils whose main language was an African language.

PLS Exploration of the Contextual Factors on Student, Classroom and School

Three hypothesised models on student, class and school level were analysed using Partial Least Squares (PLS) analysis to explore the direct and indirect effects of

individual variables on all three levels. The results of these analyses were scrutinised and thereafter the class and school level models were combined into one model and reanalysed. The main results are summarised here and the detailed explanation and discussion of these PLS results can be found in Howie (2002).

Student- level factors

Data pertaining to the pupils' home background, their personal characteristics, their aptitude and competencies were explored. A high percentage of variance (50%) in the pupils' mathematics score was explained. Six factors were found to have a direct effect on South African pupils' performance in mathematics, namely the pupils' proficiency in English (*engtest*), their own self concept in terms of mathematics (*selfcnp*), the language pupils spoke at home (*lang*), their socio-economic status at home (*SES*), and whether or not they, their friends and their mothers thought that maths was important (*mathim*) and language of learning in the classroom (*lanlearn*).

School-only level factors

Some important aspects of school quality related to school leadership, parent involvement, school profile, physical resources, human resources, autonomy, learning environment and school administration were explored in the data from the school principal's questionnaire. Two important antecedents related to the type of community and the home language of the pupil, were included in the model. Sixty-two per cent of the variance in the pupils' scores in mathematics could be explained by three factors at the school level, namely, the community where the school was located, the influence that the teachers union have on the curriculum, and an aggregated pupil variable, the extent to which the pupils in the class spoke the language of instruction as their first language.

Classroom-level only factors

From the mathematics teacher questionnaire, a number of classroom level factors were also explored and these resulted in including the following factors in the model: teachers' gender, teaching experience, teachers' level of education, time spent on activities, lesson preparation, teaching load, time on task, teachers' attitudes, success attribution, teachers' beliefs, teaching style, resources, limitations, and class size. In total, this model explained 46 per cent of the variance in the pupils' mathematics scores by seven factors - the teachers attitudes, their beliefs about mathematics, the extent of their teaching and other workload, the size of the class they are teaching, their gender, resources and their dedication towards lesson preparation.

Combined school-class level factors

Because only one class per school was included in the study, class effects could not be studied independently from the school effects. Therefore the school level model and the class-level model were combined and the predictors of mathematics achievement were selected from both models and combined with four aggregated student-level antecedent factors into one model. Therefore factors related to teachers' characteristics, pupils' home background, their aptitude, their attitudes, school quality, teaching requirements, curriculum quality and instructional quality were all explored in one model. Finally, six factors were found that had direct effects on pupils' achievement in mathematics and that explained 27 per cent of the variance in the

mathematics score. These were the location of the school, class size, the attitude of the teacher, teachers' beliefs about mathematics, the teachers' workload (including teaching) and their dedication toward lesson preparation.

Results from the Multilevel Analysis

From the Partial Least Squares analysis, the factors that had a direct effect on math achievement were identified and included into the multilevel analysis. A final variable was included because it was believed to be important from a political perspective, namely the number of pupils enrolled in a school. All variables are listed in Table 1, first column. Ultimately 183 schools and 7,651 South African pupils were included in the multi-level analysis.

In the first step in the analysis the only independent variable included in the analysis was the school of the learner (this is the so-called Null model). The Null-model in Table 1 shows that more than half of the variance in the mathematics achievement scores is situated on the school level (55%) while 45 per cent of the variance is situated at the student level.

In the next step of the multilevel analysis the six student variables were entered successively in the model. The results are summarised in Table 1, column 'Student model – Model 6', in which the numbers represent the regression slopes associated with the various variables (e.g. 5.35 is the regression coefficient associated with 'home language'). Table 1 shows that all regression coefficients are significant, meaning that mathematics achievement tends to be better when the scores are higher on the variables home language, socio-economical status, English tests, self-concept, importance of mathematics and radio language. One explanatory remark should be made: as self-concept has a negative scale, the negative regression coefficient should be interpreted as 'the more difficulties a learner has with mathematics, the lower the achievement score'. Howie (2002) illustrates these data with two examples to illustrate the influence of the language factor. The model predicts a difference of 57 points on the math achievement scale (with national mean score of 275 scale points) between a pupil that has the mean score on the English test (10 out of 40 points) and a pupil 24 points (the average achieved by native English speakers).

As a second example, a pupil who speaks an African language at home and never listens to an English radio station is predicted to score 25.6 points less on the mathematics achievement scale than a student who speaks English at home and always listens to an English radio station. It is therefore very clear that, if one only looks at variables at the student level, there is a strong influence of language on achievement in mathematics.

A final remark on the Student-model pertains to 'difference in deviance' (see Table 1). Given that the deviance is a measure for the appropriateness as to whether a model is a good representation of the reality represented by the data, the fact that the difference in deviance between the Student model and the Null model is highly significant means that the Student-model is a highly significant improvement when compared to the Null-model.

The next phase in the multilevel analysis was to enter the nine school level variables, resulting in the Student-School model, also called model 15 as the model is based on 15 independent variables. The results are summarised in the final column of Table 1.

Table 1: Multi Level Analysis of the South African TIMSS'99 Data with the Math-Test Score as Dependent Variable (Weighted Data)

	Null Model	Student Model	Student-School	Extended Student-school model
Fixed effect	**(Model 0)**	**(Model 6)**	**(Model 15)**	**(Model 16)**
Student level				
Intercept	**288**	**278**	**299.5**	**285.5**
Home language		5.35 **	3.27	2.83
Socio-economic		1.20 **	.88*	0.79
English test		4.07 **	4.00**	3.79**
Self-concept		-6.32 **	-6.29**	-6.20**
Importance of maths		6.39 **	6.35**	6.45**
Radio language		4.75 **	3.95**	3.98**
School level				
Status			-17.27**	-15.24**
Beliefs about maths			-4.46**	-3.68**
Location			8.00**	7.07*
Class language			2.59**	2.55**
Enrolment			.00	0.00
Work time			.52**	0.49**
Class size			-.27	-0.14
Lesson planning			8.02*	7.16*
Teaching time			.10	0.14
Random effects				
School level variance	6 520 (55%)	2 451	1 336	1087
Student level variance	5 342 (45%)	4 570	4 560	4535
	N/A	1 340#	132.80#	1087#
Difference in deviance				
Explained proportion of variance in math achievement				
(when compared with Null Model):				
School level variance		.61	.79	
Student level variance		.44	.50	

Notes: Intercept represents the grand mean of the mathematics scores (differs slightly from the South African mean score of 275 on the international test due to the fact that a number of schools could not be included in the analysis)

N = 7,651 pupils in 183 schools (one class per school);

* t-value > 1.96, this resembles a confidence interval of 95%;

** t-value > 2.58, this resembles a confidence interval of 99%;

The difference between the deviances is significant (p<.001).

In total, 11 of the 15 factors were found to be significant predictors of South African pupils' achievement in mathematics. Enrolment, class size, teaching time and home language appeared not significant in the Student-School model. Once the school-level variables were entered, the effect of home language, which was significant in the

Student model, lost its significance as a consequence of the school level variables such as location of the school. On the other hand, the English test score again appears to be one of the most significant (see Howie, 2002). Another result that can be derived from the full model is that *SES* is no longer significant (see Howie, 2002). The strength and significance of the school-level variables compensated for the pupil variables resulting in home language and *SES* losing their significance in the multilevel analysis.

The quality of the Student-School model as the best one in representing the relationship between the independent variables and achievement can be shown from two other parts of Table 2. At first, the difference in deviance between the Student-School model (model 15) and the Student-model (model 6) is significant indicating that the Student-School model is superior. Furthermore, as can be concluded from the lowest part of Table 1, the strength of the Student-School model is also illustrated by the variance in mathematics achievement explained by this model, being 79 per cent of the variance at school level, and 50 per cent of the variance at student level.

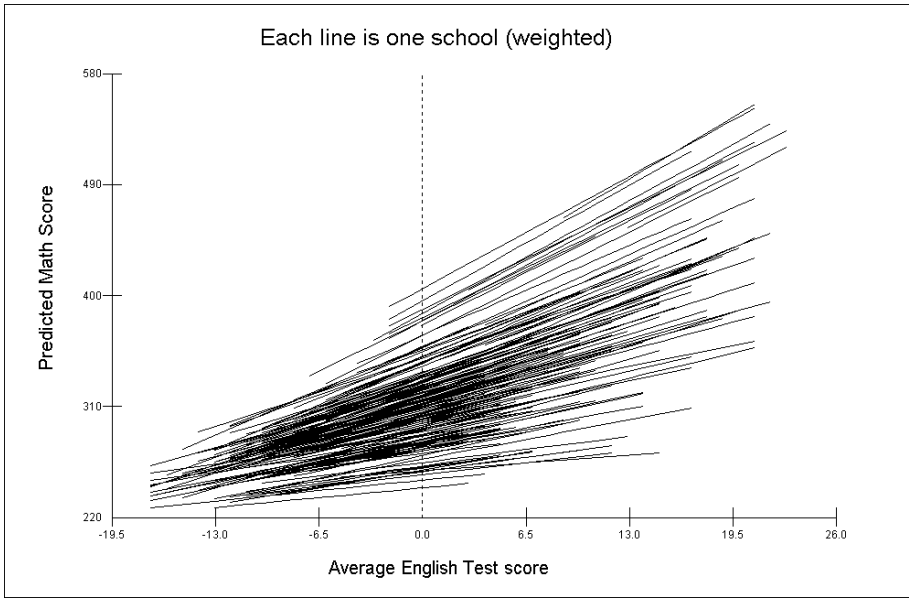
The extension of the Student-School-Model

Due to the amount of the explained proportion variance on school level by the English test score (*engtest)*, the student-school-model was extended with random slopes (i.e. with a random slope for each school) of the average English test score (*engtest*). The results in Table 1 indicate that this model is a significant improvement on the full model (Model 15), as the deviation from the full model is highly significant ($p<.001$; see Table 1). Noteworthy is that the data of Model 16 show that the extension of the full model with random slopes results in it that *SES* is no longer significant. This means that the other school variables explain the variance in *SES* (as was concluded for home language (*hlang)* when discussing the full model). Another observation in the data of Model 16 is that the inclusion of the random slopes (i.e. per school its average value on *engtest* instead of taking the national average) results in lower estimates on all school variables, which shows that language proficiency is related to all school variables in another way.

Figure 2 shows the final model graphically extended with a random slope for each class on the English test (*engtest*) with the other axis representing the mathematics achievement score (resulting in Model 16). If one looks at the pattern of the slopes in Figure 2, it would appear as if the impact of the English test (*engtest*) on mathematics achievement is less in classes with a low average score on the English test. In other words schools where pupils did poorly in the English test, their proficiency hardly made any difference to their mathematics score. Conversely, the better that classes of pupils performed on the English Test, the stronger the relationship of this outcome was with mathematics. In other words, the correlation between the English test and the mathematics score is higher for classes with an on average high score on the English test. There appears to be a curvilinear relationship between English and Mathematics, which means that language proficiency matters more when the English proficiency of classes is higher.

Another observation should be made here. Figure 2 shows that there are schools with a high average score on the English test and yet a low average performance on mathematics, combined with a low correlation between the two variables. This is an indication that there are in addition to English proficiency other variables (either within the model, e.g. location may be a candidate to investigate, or outside the present study) that are related to mathematics achievement. One possibility is that some informal tracking of pupils may be implemented at schools. So, there may be

classes where the majority of pupils speak English at home, but may be grouped into a low ability mathematics class. Possible evidence of this practice was highlighted during the data collection for TIMSS1999 where in at least six school principals prevented field workers from testing the sampled class on the grounds that it was a low achieving group and insisted that the "A" class be tested. Pupils from these schools were subsequently withdrawn from the TIMSS-1999 sample.



Note: the mean of the English test score is 17.01, but set zero in the figure.

Figure 2: Random Slopes Representing the Predicted Score on Mathematics for each School/Class (Based on the English Test Scores)

In conclusion, once all the predictors are added to the model, most of the school-level variance in pupils' achievement scores could be explained in the Student-School model. This is not the case for the student-level variance. A large percentage of the variance on student level (50% of the 45% in the Null model) could not be explained by the predictors (including a number of language related variables) used in this model. This may be due to the fact that other variables that are not included in this study are important as well. For example, cognitive ability was not measured in this study, but was included in the Belgian-Flemish study as a national option. Van den Broek and Van Damme (2001) show that in Belgium this variable explains a great deal of variance at the student level and explains in fact more than any other single variable in their multi-level model. Clearly more research is needed here for South Africa.

However, the predictors did explain a high percentage of the variance between schools for the South African data. This means that a large part of the differences between schools in pupils' mathematics achievement can be attributed to these variables. The Student-School model indicates that significant predictors for how pupils in different schools perform in mathematics are the pupils' performance in the English test, the socio-economic status (to a lesser extent), the pupils' self concept,

the pupils' perception of the importance of mathematics, their exposure to English, how pupils' maths teachers perceive their professional status, pupils' maths teachers beliefs about mathematics, the location of the school, the extent to which English is used in the classroom, the amount of time teachers spend working and the amount of time teachers spend in lesson planning. They are also significant predictors of how well pupils' perform in the same school (within-school variance), but to a lesser extent. Noteworthy is that two of these variables have a negative effect, teachers' perception of their status and their beliefs about mathematics. The stronger the teachers' ideas about mathematics and the perception about the status of the profession are, the poorer their pupils perform in mathematics. This observation should not be looked at in isolation, but in conjunction with the other variables that have a significant effect on mathematics achievement, but further discussion is beyond the scope of this paper.

Conclusions and Reflections

A number of findings were made with regard to the effect of language on South African pupils' achievement in mathematics. This research shows that in South Africa, pupils tended to achieve higher scores in mathematics when their language proficiency in English was higher and were more likely to attain low scores in mathematics when their scores on the English test were low. Those that spoke English or Afrikaans at home tended to achieve higher scores in mathematics. Alternatively, children from homes where African languages were used were more likely to achieve lower scores. Pupils within classes where the pupils and the teachers mostly interacted in the official media of instruction (English or Afrikaans) were more likely to achieve better results in mathematics.

A key finding from this research is that pupils who spoke either English or Afrikaans **at home** achieved higher scores in both the mathematics and the English tests than those who did not. With the increasing heterogeneity of South Africa's schools in culture and language, offering children the opportunity to learn in their own home language is becoming increasingly challenging, as is also illustrated by other (qualitative) studies in South Africa about the relationship between mathematics and language (see e.g. Adler, 1998; Setati & Adler, 2000; Setati, Adler, Reid & Bapoo, 2001).

With regard to language another key finding was that the average English test score was very low and the majority of pupils' English language proficiency was poor. Pupils speaking African languages had very low English language proficiency when compared to other second language learners. Furthermore, there is no room for complacency as native English speakers as a group did not perform substantially better than the Afrikaans-speaking group given that this was an English second language test. The low proficiency may be linked to another finding from the research (see Howie, 2002) that more than two-thirds of pupils had very few books in the home. It is clear that an urgent and intensive intervention is needed in both English as a discipline in the schools and as a medium of instruction. Above all, children need to be encouraged to read in their leisure time, but they also need to be given opportunities to write in English more so than presently so that they become familiar in articulating thoughts and knowledge in English and this should be targeted to all children, not only the second language speakers.

The findings give rise to a **number of reflections** on the relationship between language and achievement in relation to the language policy implemented by the South African government. The difficulty of not being able to communicate fluently

in a common language is leading to increased frustration for the teacher, disorientation on the part of the child, a slow rate of learning, disciplinary problems and teacher centred instruction. Although teachers are aware of the national language policy they may have very different interpretations of it (Setati, 1999). The majority of parents, pupils and teachers perceive English as the gateway to opportunities more globally and therefore many want the pupils to participate in their education through the medium of English. The implementation of the language policy needs reassessing regarding its feasibility and the desires of the community. Unless, some hard decisions are taken, and quickly, by decision-makers, pupils and teachers will continue to struggle and pupils will under-perform in mathematics and other subjects. If, as the majority of parents seem to desire, English is increasingly used then the necessary support mechanisms need to be put into place including intensive language training of second language teachers who will be teaching through the medium of English. On the other hand, if however the decision is made to teach in mother tongue beyond Grade 5, the ramifications are enormous considering that in urban environments, almost all the official languages are found as well as there are increasingly more foreign languages (from elsewhere in Africa and beyond). Segregating children and indeed teachers in terms of language (as clearly offering 11 languages of learning will be impossible within a single school) will be an enormous task and may reverse the cultural integration beginning to take place. Of course, in many rural areas where one language is almost always clearly dominant, it may be more feasible. The danger of considering only one approach in the language policy, albeit an important one, (for example only considering cultural identity or the political perspective); is that the multiple functions of schooling will be ignored. For, the curriculum and pupils' education have to fulfil the desires of society namely, that of educating them towards being a responsive citizen in a democratic society, of attaining certain basic knowledge and skills, being prepared for the workplace and/or further education and acquiring adequate social and interpersonal skills. The government has to take a firm lead in finding the appropriate balance between these perspectives, as if this is left to the schools, the status quo will remain.

In conclusion, the strength of the language component represented in a number of variables that have strong effects on mathematics achievement seems apparent. Moreover, the dedication of the teacher matters with regard to their pupils' achievement and the location of the school is further an important predictor of South African pupils' achievement in mathematics. In the past few years, some progress has been made to address shortcomings in South African schools. Although significant progress has been made with regard to administrative restructuring, policy development and infra-structural improvements nonetheless, the quality of education that the majority of pupils are receiving is far from satisfactory. The study (Howie, 2002) highlighted the most significant predictors of mathematics (of which language is only one) within the scope of the data available and has raised a number of questions and issues that are believed to be pertinent to the future development of South African education, in particular in mathematics. The challenges abound within the education system of this country and in addition to the issues of access and equity the most important challenge awaiting now is that of quality.

This paper has described one case of a national option that was conducted within an IEA study. This research has been able to fill a void in research exploring the relationship between language and mathematics in South Africa. The IEA's TIMSS'99 provided the opportunity for researchers in South Africa to conduct this research, which would not have been possible otherwise given the limited resources within the country. The IEA's technical standards ensured the quality of the sample

and implementation of the TIMSS'99 within this country and others. Furthermore, by allowing a country such as South Africa the possibility to include additional data collection (i.e. the national option) permitted the collection of quality data to pursue national research interests and in this particular project also the development of research expertise within a country with limited expertise. Finally, the authors believe that national and regional options should be greatly encouraged for all IEA studies to increase the relevance of each study for participating countries and for the development of knowledge and skills in the more developing environments in particular.

References

Adler, J. (1998). A language for teaching dilemmas: unlocking the complex multilingual secondary mathematics classroom. *For the Learning of Mathematics*,18, 24-33.

Afrassa, T.M. (1998). *Mathematics achievement in the lower secondary school stages in Australia and Ethiopia: a comparative study of standards of achievement and student-level factors influencing achievement*. Unpublished PhD. Thesis. Adelaide: Flinders University of South Australia, School of Education.

Arnott, A. & Kubeka, Z. (1997). *Mathematics and science teachers: Demand, utilisation, supply, and training in South Africa*, Johannesburg: Edusource.

Baker, C. & Prys-Jones, S. (1998). *Encyclopedia of bilingualism and bilingual education*. Clevedon: Multilingual Matters.

Behutiye, N. & Wagner, J. (1995). *Some aspects of primary school pupils' cognitive development and their relationship to scholastic achievement*. Addis Ababa: Curriculum and Evaluation and Educational Research Division, Institute for Curriculum Development.

Berry, J.W. (1995). Learning mathematics in a secondary language: Some cross-cultural issues. *For the learning of Mathematics*, 5 (2), 18-23.

Cherian, V.I. (1992). *Family influences on scholastic achievement*. Inaugural address delivered on accepting the chair of Psychology of Education at the University of the North on 28 October 1992, Sovenga.

Clarkson, P.C. (1991). *Bilingualism and Mathematics Learning*. Australia, Geelong: Deakin University.

Cohn, & Ressmiller. (1987). Research on effective schools: Implications for less-developed countries. *Comparative Education Review*, 31 (3), 377-399.

Daniel, D. (1995). School attendance and achievement among boys and girls: The case of selected grade six students in Addis Ababa. (pp. 253-264). *Proceedings of the National Workshop on Strengthening Educational Research*. Addis Ababa: Institute of Educational Research, Addis Ababa University.

Deresse, M., Wagner, J. & Alemaychu, M. (1990). *Factors affecting achievement of lower primary school pupils*. Addis Ababa: Institute for Curriculum Development and Educational Research.

Eshetu, A. (1988). *Comparative study of family background of high achievers and low achievers at Entoto technical, vocational and academics schools*. Unpublished BA thesis. Addis Ababa University, Faculty of Education.

Gennet, (1991). *Women in Education: A study of academic performance and participation of female students in high schools of Addis Ababa*. Addis Ababa: Addis Ababa University, Department of Business Education, Faculty of Education.

Georgewill, J.W. (1990). Causes of poor achievement in West Africa School Certificate Mathematics Examinations in Rivers State secondary schools, Nigeria. *International Journal of Mathematics in Science and Technology*, 21 (3), 379-386.

Howie, S.J. (1997). *Mathematics and science performance in the middle school years in South Africa: A summary report on the performance of the South African students in the Third International Mathematics and Science Study*. Pretoria: HSRC.

Howie, S.J. (2001). *Mathematics and Science Performance in Grade 8 in South Africa 1998/1999: TIMSS-R 1999* South Africa. Pretoria: HSRC report.

Howie, S.J. (2002). *English language proficiency and contextual factors influencing mathematics achievement of secondary school pupils in South Africa*. Enschede: Doctoral dissertation, University of Twente.

Howie, S.J. & Hughes, C.A. (1998). *Mathematics and science literacy of final-year school students in South Africa: Third International Mathematics and Science Study*. Pretoria: HSRC.

Howie, S.J. & Pietersen, J.J. (2001). Mathematics Literacy of Final Year Students: South African Realities. *Studies in Educational Evaluation*, 27 (1), 7-26.

Howie, S.J. & Wedepohl, P.T. (1997). *The "crisis" in science and mathematics education in South Africa.* A paper presented at the British Educational Research Association conference, York, England, Sept 1997.

Human Sciences Research Council (HSRC) (1990). *Achievement Test English Second Language Standard 6. Ref. 2564/2.* Pretoria: HSRC, Institute for Psychological and Edumetric Research.

Institute of Education (2000). *MlwiN: a visual interface for multilevel modelling. Version 1.10*. Multilevel Models project, University of London.

Kahn, M. (1993). *Building the base: report on a sector study of science and mathematics education.* Pretoria: Commission of the European Communities.

Maja, B.I., du Plooy, J. & Swanepoel, Z. (n.d.). *Access to Learning: the enabling conditions for successful learning environments*. Unpublished.

Maqsud, M. & Khalique, C.M. (1991). Relationships of some socio-personal factors to mathematics achievement of secondary school and university students in Bophutatswana. *Educational Studies in Mathematics*, 22 (4), 377-390.

Monyana, H.J. (1996). *Factors related to mathematics achievement of secondary school pupils*. M. Ed. Thesis. University of South Africa. Unpublished.

Mpofana, W.S. (1989). *A didactical study of unsatisfactory classroom interaction as a factor related to the high failure rate in senior secondary schools in mathematics in KwaZulu*. M.Ed. University of the Orange Free State. Bloemfontein.

Prinsloo, C., Kanjee, A, Pfeiffer, C. & Howie, S.J. (2001). *Baseline Study Technical Report. Teaching and Learning conditions*. Pretoria: HSRC.

Rakgokong, L. (1994). Language and the construction of meaning associated with division in primary mathematics. In M. Glencross (Ed.), *Proceedings of the Second Annual Meeting of the Southern African Association for Research in Mathematics and Science Education*. Umtata: University of Transkei.

Riddell, A. (1997). Assessing designs for school effectiveness research and school improvement in development countries. *Comparative Education Review*, 41 (2), 178-204.

Sayers, R. (1994). Gender differences in mathematics education in Zambia. *Educational Studies in Mathematics*, 26 (4): 389-403.

Sellin, N. (1989*). PLSPATH, Version 3.01. Application Manual*. Hamburg: University of Hamburg.

Sellin, N. (1990). On aggregation bias. In K.C.Cheung, J.P. Keeves, N. Selling& S.C. Tsoi, The analysis of multivariate data in educational research: studies of problems and their solutions. *International Journal of Educational Research*, 14 (3), 257-268.

Setati, M. & Adler, J. (2000). Between languages and discourses: Language practices in primary multilingual classrooms in South Africa. *Educational Studies in Mathematics*. 43 (3), 243-269.

Setati, M., Adler, J., Reid, Y. & Bapoo, A. (2001). *Incomplete journeys: Code switching and other language practices in mathematics, science and English language classrooms in South Africa*. Language and Education.

Shavelson, R.J., McDonnell, L.M. & Oakes, J. (1987). *Indicators for monitoring mathematics and science education: a sourcebook*. Santa Monica (CA, USA): The RAND Corporation.

Sojourner, J. & Kushner, S.N. (1997) *Variables that impact the education of African-American students: parental involvement, religious socialisation, socio-economic status, self-concept and gender.* Paper presented at the Annual Meeting of the American Education research Association. Chicago, Illinois, March 24-28.

Tartre, L.A. & Fennema, E. (1995). Mathematics achievement and gender: a longitudinal study of selected cognitive and affective variables (grades 6-12). *Educational Studies in Mathematic*s, 28, 199-227.

Taylor, N. & Vinjevold, P. (Eds.) (1999). *Getting Learning Right*. Johannesburg: Joint Education Trust.

Teddlie, C. & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.

Van den Broek, A. & Van Damme, J. (2001). *The effects of School, Class and Student characteristics on mathematics education in Flemish TIMSS'99 data*. Paper presented at the European Conference on Educational Research, Lille, France, 5-8 September.

Vari, P. (1997). *Monitor'95. National Assessment of student achievement*. Budapest: National Institute of Public Education.

Young, D.J. (1997). *A multilevel analysis of science and mathematics achievement*. Paper presented at the Annual Meeting of the American Educational Research Association in Chicago, Illinois, March 24-28, 1997.

11

The Subversive Influence of Formative Assessment

Paul S Black
King's College London

Dedications: *John is certainly a great figure and I have read and admired some of his work. The field of testing and assessment is one in which policy makers and the public take decisions which affect people's lives, too often on the basis of mis-representation of data, inadequate evidence, or no evidence at all. John has been an out-standing figure in working for rigorous evidence, carefully assembled and scrupulously interpreted. The best I can hope for in the work I present here is that I have tried to follow the standards he has set.*

Introduction

Given that I am a keen promoter of formative assessment, my title must read strangely. I choose it to express both a puzzle and a warning. The puzzle has arisen as the group of us at King's College have tried to explore the idea of formative assessment, to develop its practice in schools, and to learn by reflection on the interplay between these two aspects of our work. The practical experience has revealed more significance and depth in what we have been doing than we had anticipated – hence the puzzle. It has also produced a mixture of successes and failures – hence the warning. The purpose of this chapter is to describe and explore these features, as a contribution to the development of a theory of formative assessment: such a theory has yet to be formulated, and it certainly will not be formulated here.

The argument will be developed in five main sections. The first two will describe in turn the experiences that are the basis for the reflections in this chapter, and the

evidence that the work has, by several criteria, been a positive contribution to education. The third section will then explore the dark side, attempting to grapple with and so understand the obstacles that have been and continue to be encountered. This will lead to some more fundamental reflection, in the fourth section, on the interplay between pedagogy, assessment, and learning in the classroom. A closing section will return to the opening theme of puzzles and warnings by looking at what might be needed to take forward further development of formative practice.

The Journey so far

The background

In 1998 we published a review, summarising the results from over 250 articles by researchers from several countries, which established that there was a strong body of evidence to support a claim that formative assessment practices can raise standards (Black & Wiliam, 1998a). At the same time, the published evidence also showed that such practices were only weakly developed in most classrooms: in fact the assessment methods that teachers use are not effective in promoting good learning. In particular, their marking and grading practices tend to emphasise competition rather than personal improvement. However, the published research could not provide recipes for improvement: the reported surveys and experiments lacked the detail that would enable teachers to implement the practices in classrooms.

The project

In response to this situation, the King's team took two steps. The first was to publicise the findings in a 20-page booklet (Black & Wiliam, 1998b). To promote this we organised a press launch and thereby secured media attention. To date over 40000 copies have been sold and the work is widely quoted. Thus it was clear that teachers were interested in what we had to say.

Our second step was to develop the practical implementation of the ideas. Six schools who taught pupils in the age range 11 to 18 years agreed to collaborate with us: each selected two science and two mathematics teachers willing to take on the risks and extra work involved. In second year of the project we added two teachers of English, from each of same schools, so that in all 36 teachers were involved. They were supported by staff from their local (district) education authorities and the project was called the King's Medway Oxford Formative Assessment Project (KMOFAP) to highlight our partnership with those two authorities.

We met the teachers together in a whole day meeting every five weeks, over two years. The findings reported here are based both on records of these meetings, on the observations and records of visits to classrooms by the King's team, on interviews with and writing by the teachers themselves, and on a few discussions with pupil groups (for details see Black et al. 2002, 2003).

Following this project, members of the King's team have responded to numerous invitations to talk to other groups: over three years they have made over 200 such contributions. These have ranged across all subjects, and across both primary and secondary phases. In addition, there has been sustained work with some primary schools. All of this makes us confident that our general findings will be of value to all, although some important details may differ between different age groups and different subjects. Some of our findings have been incorporated into a United Kingdom government initiative in England to improve teaching and learning in the

age range 11 to 14, whilst the project have also been consultants to a national program in Scotland to develop assessment for learning.

The practices developed

These practices will be described here under four headings: oral feedback in **classroom dialogue**, **feedback through marking**, **peer- and self-assessment**, and the **formative use of summative tests**. The account given will be brief – more detailed accounts have been published elsewhere (Black et al. 2003).

For **classroom dialogue** the aim was to improve the interactive feedback which is central to formative assessment. An account of wait time research (Rowe 1974) motivated teachers to allow longer time after asking a question so that pupils would have time to think out responses, and so that all could be expected to become actively involved in question and answer discussions, and to make longer replies. One particular way to increase participation was to ask pupils to brainstorm ideas, perhaps in pairs, for two to three minutes prior to the teacher asking for contributions. Then all answers, right or wrong, had to be taken seriously, the aim being to develop thoughtful improvement rather to evoke the expected answers. A consequence of such changes was that teachers learnt more about the pre-knowledge of their pupils, and about any gaps and mis-conceptions in that knowledge, so that their next moves could address the learners' real needs.

As they tried to develop this approach, teachers realised that more effort had to be spent in framing questions that were worth asking, i.e. questions which explored issues that are critical to the development of pupils' understanding. They also had to focus closely on follow-up activities to formulate meaningful responses and challenges that would help pupils to extend their understanding. Put simply, it became clear that the only point of asking questions is to raise issues about which the teacher needs information or about which the pupils need to think.

To address **feedback through marking**, teachers were first given an account of research studies which have established that, whilst pupils' learning can be advanced by feedback through comments, the giving of marks or grades has a negative effect because pupils ignore comments when marks are also given (Butler, 1988). These results surprised the teachers. They also worried them because of concern about the effect of returning pupils' work with comments but no marks. However, potential conflicts with school policy were resolved as experience showed that the provision of comments gave both pupils and their parents advice on how to improve. It also set up a new focus on the learning issues rather than on trying to interpret a mark or grade. To make the most of the learning opportunity created by feedback on written work, procedures that required pupils to follow up comments had to be planned as part of the overall learning process.

One consequence of this change was that teachers had to think more carefully in framing comments on written work, for it was now evident that these had to identify what had been done well and what still needed improvement, and to give guidance on how to make that improvement. As the skills of formulating and using such feedback were developed, it became more clear that the quality of the tasks set for written homework or class-work was critical : such tasks, alongside oral questioning, had to be designed to encourage pupils to develop and show understanding of the key features of what they had learnt.

For **peer- and self-assessment** the starting point was Sadler's (1989) argument that self-assessment is essential to learning because pupils can only achieve a learning

goal if they understand that goal and can assess what they need to do to reach it. Thus the criteria for evaluating any learning achievements must be made transparent to pupils to enable them to have a clear overview both of the aims of their work and of what it means to complete it successfully. Insofar as they do so they begin to develop an overview of that work so that they can manage and control it: in other words, they develop their capacity for meta-cognitive thinking.

It is not easy to develop self-assessment skills. The first and most difficult task is to get pupils to think of their work in terms of a set of goals. In practice, peer-assessment turned out to be an important stimulus to self-assessment. Peer-assessment is uniquely valuable because pupils may accept, from one another, criticisms of their work which they would not take seriously if made by their teacher. Peer work is also valuable because the interchange will be in language that pupils themselves would naturally use, and because pupils learn by taking the roles of teachers and examiners of others (Sadler, 1998).

However, for such peer-group work to succeed, many pupils needed guidance about how to behave in groups, e.g. in listening to one another, taking turns, and offering affirmation together with constructive criticism about one another's work. A typical exercise would be on the marking of homework. Pupils were asked to label their work with 'traffic lights', i.e. using red or amber if they were totally or partially unsure of their success, and green where they were confident. Then those who had used amber or green would work in mixed groups to appraise and help with one another's work, whilst the teacher would pay special attention to those who had chosen red.

Teachers developed three ways of making **formative use of summative tests**. One way was to ask pupils, in preparation for a test, to 'traffic light' a list of key words or of the topics on which the test would be set, an exercise which would stimulate them to reflect on where they felt their learning was secure and where they needed to concentrate their efforts. One reason for doing this was that teachers had realised that many pupils had no strategy for preparing for a test by formulating a strategic appraisal of their learning.

A second way was to mark one another's test papers in peer groups, in the way outlined above for the marking of homework. This could be particularly challenging when they were expected to invent their own marking rubric, for to do this they had to think about the purpose of a question and about the criteria of quality to apply to responses. After peer marking, teachers could reserve their time for discussion of the questions that give particular difficulty.

A further idea was introduced from research studies (Foos et al., 1994; King, 1992) which have shown that pupils trained to prepare for examinations by generating and then answering their own questions out-performed comparable groups who prepared in conventional ways. Preparation of test questions calls for, and so develops, an overview of the topic.

The teachers' work on summative assessments challenged our expectations that, for the context in which they worked, formative and summative assessments are so different in their purpose that they have to be kept apart; this expectation was strengthened by evidence of the harmful influence that narrow 'high-stakes' summative tests can have on teaching. The finding that emerged was quite different – that summative tests should be, and should be seen to be, a positive part of the learning process. If they could be actively involved in the test process, pupils might

see that they can be beneficiaries rather than victims of testing, because tests can help them improve their learning.

Successful Outcomes

The practical lessons outlined in the previous sections constituted one of the main outcomes of the project, and it is significant that these arose from a close partnership between researchers and teachers in which the teachers contributed at least as much as the researchers in the production of new knowledge about teaching and learning. Two further outcomes are described in this section: the first is about the gains in pupils' test achievements, the second is about ways in which the teachers described the changes to themselves and to their pupils.

Quantitative gains

Although our review of the literature showed that enhanced formative assessment produces gains in pupil achievement, it seemed important to have some indication of the kinds of gains that could be achieved in normal classrooms. By 'normal' we meant that the content of the teachers' lessons would be that of their normal plans, and that the only measures of pupils' performance would be those derived from the normal tests used in their schools. Since each teacher in the project was free to decide the class with which they would work on these ideas, we discussed with each teacher individually how to set up a 'mini-experiment' for each teacher using data available within the school. Each 'experiment' produced an effect size by comparison of learning gains of the experimental class with those of a comparable 'normal' class. Full details are given in Wiliam et al. (2003).

For the 19 teachers on whom we had reliable data, the average effect size was around 0.35. Such improvement, produced across a school, would raise a school in the lower quartile of the United Kingdom national performance tables to well above average. It was clear, therefore, that teachers could improve their pupils' results by working with the ideas that developed in our project.

Qualitative gains

Teachers were asked at the end of the project to write a few pages summarising what they perceived they had learnt from involvement in its work. What they chose to write about was left entirely up to them. The several themes that emerged are illustrated by the quotations below, which are taken from this writing: the names of the teachers and of the schools are pseudonyms, as our policy was to guarantee anonymity.

One concerned the changes in the way they prepared lessons :

> I certainly did not spend sufficient time developing questions prior to commencing my formative training … Not until you analyse your own questioning do you realise how poor it can be. I found myself using questions to fill time and asking questions which required little thought from the students. When talking to students, particularly those who are experiencing difficulties, it is important to ask questions which get them thinking about the topic and will allow them to make the next step in the learning process.
>
> Derek, Century Island

Following from this was their perception of the changes in classroom dialogue. Two teachers commented on different aspects of this change:

> There have been two very positive results from this approach. The most significant one is that because they have to explain their answers each time orally this has carried through to their written work and now they set out their answers fully without being prompted. The second one is with a girl with a statement for being unable to talk or communicate with an adult. Having got used to the extra thinking time she now offers answers orally and will now tentatively explain her answers.
>
> Gwen – Waterford School

> I now use the 'hands down' strategy and this has made a big difference to my classroom discussions. In particular it has broadened the range of participation and removed (at its best) the curious mixture of envy and relief which characterised the mood of a group while the usual people answered questions. In a word discussions are now much more inclusive and much more scary!
>
> Paul – Cornfields School

Another teacher, not directly involved in the project, was encouraged on hearing a talk from an involved colleague, to try promoting oral feedback for his own class; he reported:

> I tried it today with my year 8s, and it works. No hands up, and giving them time to think. I had fantastic responses from kids who have barely spoken in class all year. They all wanted to say something and the quality of answers was brilliant. This is the first time for ages that I've learnt something new that's going to make a real difference to my teaching.
>
> Anon.

The changes clearly had an effect on the pupils' experiences of the classroom, as the following extract indicates:

> They have commented on the fact that they think I am more interested in the general way to get to an answer than a specific solution and when Clare interviewed them they decided this was so that they could apply their understanding in a wider sense.
>
> Belinda – Cornbury Estate School

Linked to these experiences were findings arising from the emphasis on helping pupils to understand the aims of their learning, as two teachers commented :

> It's much more extending this idea of them telling me what they need to learn. I think it helps that you are not telling them what you think they got wrong and they need to go over; they are telling you what things they can't do.
>
> Ceri – Two Bishops School

> I have thought carefully about pupils taking ownership of their own learning. I have now thought more about letting pupils know what the intention of the lesson is and what they need to do to achieve it. This way they have to think what they know and take more responsibility for their own learning.
>
> Angela – Cornbury Estate School

Similar types of change were reported from the practices of peer assessment and the formative uses of tests :

> One technique has been to put the students into small groups and give each student a small part of the unit to explain to their colleagues. They are given a

> few minutes preparation time, a few hints, and use of their exercise books. Then each student explains their chosen subject to the rest of their group. Students are quick to point out such things as, 'I thought that the examples you chose were very good as they were not ones in our books. I don't think I would have thought of those.' Or, 'I expected you to mention particles more when you were explaining the difference between liquids and gases.' These sessions have proven invaluable, not only to me, in being able to discover the level of understanding of some students, but to the students too.
>
> Philip, Century Island School

> They feel that the pressure to succeed in tests is being replaced by the need to understand the work that has been covered and the test is just an assessment along the way of what needs more work and what seems to be fine.
>
> Belinda – Cornbury Estate School

What underlay many of the comments were two themes. One was the vision of empowering students which the teachers had come to value:

> What formative assessment has done for me is made me focus less on myself but more on the children. I have had the confidence to empower the students to take it forward.
>
> Robert, Two Bishops'School

That quotation also signals the other theme, which was a shift in the teachers focus from their teaching to an emphasis on pupils' learning:

> There was a definite transition at some point, from focusing on what I was putting into the process, to what the students were contributing. It became obvious that one way to make a significant sustainable change was to get the students doing more of the thinking. I then began to search for ways to make the learning process more transparent to the students. Indeed, I now spend my time looking for ways to get students to take responsibility for their learning and at the same time making the learning more collaborative.
>
> Tom, Riverside School

What these quotations to convey is the perception of radical change that emerged from the writing of most of the teachers. The overall impression was that the changes amounted to far more than the acquisition of a few 'tricks of the trade'. They were seen by most teachers as having enhanced their professional status and confidence:

The Problems

Doubtful meanings

In a publication for teachers (Black et al. 2002) we included the following extended definition of our idea of formative assessment:

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence.
>
> An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their pupils in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet learning needs.

The detail and precision here have been found essential, because many teachers and researchers seem to have misunderstood the term. One example of misunderstanding is that it is about any assessment conducted by teachers, and in particular that giving a test every week and telling the pupils the marks constitutes formative assessment, which, in the absence of feedback leading to further learning work, it does not. In general, any test or assessment at the end of a piece of learning is too late for formative purposes, precisely because it is at the end, so there is no opportunity to use its results for feedback to improve performance of the pupils involved.

Another example is the belief that it includes portfolio assessment where that is developed with the aim of replacing or supplementing the information which externally imposed tests are designed to provide. Again, there is no formative assessment in such practice except insofar as there is active feedback to change and improve pupils' work as the portfolios are built up.

A more general obstacle has arisen because of the use of the phrase 'Assessment for learning' chosen by some because it seems less technically daunting than 'formative assessment'. The widespread adoption of this phrase to label initiatives by local and central government in England has led to ambiguity in its meaning. An example is a current government emphasis on target setting: the idea is that assessment of each pupil should inform that pupil of the level which his/her work has reached and that this should go with a target for that pupil to work to. The progress up the levels as successive targets are achieved is to be entered in a data-base record which will follow the pupil across and between schools. Whilst this may be useful, to entitle it 'Assessment for learning' in literature and training without making clear that learning is not directly helped by the mere record of such information is mis-leading. In particular, the evidence that formative assessment improves learning cannot be relevant to this activity unless some formative action is associated with it. When, as has happened, one trainer has stated that "all assessment is assessment for learning", the phrase has lost any specific meaning: official adoption has been an accolade of dubious value.

Losing control

Observations made in schools which, since the KMOFAP project, have adopted its lessons, but with relatively light training and support, have been disappointing. Whilst the development of self- and peer-assessment has been quite positive, significant changes in the classroom dialogue between teacher and pupils appear to be rare. It seems that, whilst listening to and accepting the messages about such changes, many teachers are unable to break away from a dialogue dominated by the themselves, with all pupil remarks rarely amounting to as much as a complete sentence, all addressed to the teacher. This is not surprising – during the first stages of the project the teachers said that it seemed risky to increase wait time and give pupils more opportunity to speak, indeed they felt they were losing control. They came to see that such loss was in fact necessary in order to achieve a new relationship with their classes – as one of them expressed it :

> Rather than me just being involved in 'crowd control' and the dispensing of knowledge; there developed a partnership between myself and my students, based on mutual respect and trust; one where all felt comfortable with being challenged and where we could all make mistakes.
>
> Roger Higton p.89 in Black et al. (2003)

Such changes were central rather than marginal. In some recent classroom observations, we have seen teachers encouraging pupils to take time and discuss an issue, but when calling for responses, they react to the first one expressed, by affirming it or by pointing out why it is wrong, and then go ahead with the next item in their planned agenda, so paying no attention to ideas that might have arisen in the discussions with the rest of the class. By contrast, in a different classroom a teacher said very little throughout a lesson, merely setting the agenda, encouraging pupils to argue with one another, refraining from passing judgement, and challenging at critical points to move the argument on. Such a classroom is radically different from the norm where the teacher dominates the discussion. The culture change involved was described by a KMOFAP teacher, reflecting on an entry in a diary which he was encouraged to keep during the project:

> This entry acknowledged that formative assessment requires students to change their approach to their own learning; they had to become more proactive. The ecology of the learning environment becomes based on mutual support and a spirit of co-operation. This gave an environment where students felt safe to give wrong answers and to express, freely, their lack of understanding. Here the classroom ceased to be a habitat where only the brightest survived and flourished, but one where, with careful grouping and good questioning, every student could feel themselves making progress through the lesson.
>
> Roger Higton p.90 in Black et al. (2003)

Pupil resistance

The previous quotation makes clear that pupils as well as teachers had to accept a radical change in the way that they played out their roles in the classroom – there is a challenge to them also. As another diary entry expressed it :

> Students will not invest in its practice if it does not bring them a reward. They are taking risks by allowing themselves to be questioned in a more rigorous manner. . . . Two months later I wrote, "they themselves seem to be becoming more confident learners. Attitudes changing faster in some than others, need to tackle 'quiet-head-down' low achievers. It has become unacceptable in groups to be loud-low-achievers. Loud-high-achievers the norm."
>
> Roger Higton p.90 in Black et al. (2003)

A different teacher experienced pupil resistance in a rather sharp form, when her change of school made clear that it was not only the so-called slower pupils who found it hard to take on a new role:

> The school I now teach in is a grammar [i.e. selective] school and I find that students who have already achieved well in a 'traditional' classroom are not so keen to get involved with their learning on a grand scale! . . .
>
> . . . . The first time I asked a Year 10 top set to work in groups to examine the errors they had made in a test and to help each other to understand fully what was being asked of them, it was unsuccessful. They were not used to the ethos of taking that level of responsibility. That lesson I heard several times 'Why don't you just tell us? . . . I had to work hard to get the year 10 group to see assessment as a way of improving learning, not a series of snapshot judgements about achievement and progress. I needed them to see that their actual test scores were not the issue, what we did about improving their understanding was much more important.
>
> Barbara Eggleton p.86-7 in Black et al. (2003)

The overall conclusion is that a formative environment requires a classroom culture that may well be unfamiliar and disconcerting for both teachers and pupils. The effect of the innovations that are entailed in the adoption of formative practices is to change the rules, usually implicit, that govern the behaviours that are expected and seen as legitimate by teachers and by pupils. As Perrenoud (1991) put it many years before the KMOFAP project:

> Every teacher who wants to practice formative assessment must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils.

The summative barrier

In some respects, the KMOFAP project showed that classroom practice in relation to preparation for summative testing can change. The underlying message was that summative tests can, and should be seen to be, a positive part of the learning process. However, the benefits of the formative approach could only be realised for tests over which the teachers had full control. Matters looked quite different when our project teachers were asked to write about and discuss the interface between their formative work and external test pressures.

Although many of the teachers had become more sensitive and more expert in the framing of productive questions for everyday use, and although they believed that their own tests were more valid than those set externally, they felt nevertheless that these would not have external credibility, whether with colleagues or with the pupils. So there was very little use, in their school tests, of questions set by the teachers themselves. They relied on questions taken from past external tests and from text-books, and often used the same test was year after year. As well as saving time, this reliance left them less open to criticism, for these were indeed the types of tests against which both their pupils and themselves would be judged by parents and the educational bureaucracy.

School requirements for formal test occasions were very variable. In frequency they ranged from none to four or five times year, at fixed times or at variable times. A few believed that they only set tests because their schools required them – as one teacher put it :

> I know a lot more about this class because of the formative assessment . I mean we discuss things, they do presentations, they talk to me, I talk to them, they talk to each other – and I could tell you for every one of this class their strengths and weaknesses.

However, some holding this belief nevertheless realised also that their staff colleagues did not know the pupils in their classes well enough to assess each of them using personal knowledge. In addition, they were concerned that this variability in assessment expertise was exacerbated by the flux of teacher appointments and resignations, and by the frequent use of substitute teachers.

Improved formative practices can generate a lot of data about pupils' progress, and it would seem that a further step might be to record and accumulate this to strengthen teachers' contribution to summative reporting. However, teachers were not able to exploit this potential for synergy and in the practices of recording and reporting and it was hard to find any change from the tradition of recording marks. Several said that it was too hard to review all the data – so they just used the scores on a terminal test to report to the school management, particularly when the need arose at the end of term when they did not have the time to collect and analyse the data.

Another aspect of external pressures on assessment was the need to report to parents. Here, two contrary views were expressed. One teacher explained how he felt he had to use only the test score:

> . . . . everyone is obsessed with numbers rather than what the kids really know. I think parents don't put a lot of weighting onto teacher assessment, they don't even put a lot of weighting onto coursework as a way of measuring their child's ability. Certainly they wouldn't put a lot of store by what I say as opposed to what a test says because at the end of the day society is driven by test results in schools . . at the end of the day if you gave them the choice – they can have what I think or they can have an exam grade – they would take the exam grade every time because that is what they want. They want to be able to compare that with everyone else even though it is not about everyone else.

Other teachers expressed a contrary view. Their experience was that the rich detail provided through good formative assessment was appreciated by parents, in that they and their children could be given specific and helpful advice on how to improve. This different appraisal of the situation may have arisen from a variety of causes, to do with the success of the school as a whole in explaining learning needs to parents.

At the end of one discussion of these issues with a group of the project teachers, one of them summed it up as follows:

> It is a bit depressing that isn't it?

We could only agree with her. It would appear that the development of teachers' assessments can only go so far before it meets a road-block set up by formal testing. It seems hard to see how synergy, between the purposes and methods used for formative and for summative purposes, can be attained until teachers can be given full responsibility for the summative assessments that are required for pupils' certification and for public accountability. To accept and develop such a system would be a major task. But it would not be an impossible one, as the experience of the state of Queensland has shown (Butler, 1995).

Pedagogy, Learning and Assessment

Principles of learning

> Teaching and Learning became inextricably inter-linked in a way that I had not managed before.
>
> Barbara Eggleton p.86 in Black et al. (2003)

One surprising development during the first project year was that the participating teachers asked us to give a seminar on learning theories. In retrospect, we should not have been surprised. We had, after all, stressed that feedback functioned formatively only if the information fed back was used by the learner in improving performance. So what the teachers needed was to be able to formulate feedback that they could judge to be useful. To do that they needed to build up models of how pupils learn.

In the KMOFAP classrooms, as the teachers came to listen more attentively to the pupils' responses, they began to appreciate more fully that learning is not a process of passive reception of knowledge, but one in which the learners must be active in creating their own understandings.

These practices reflect some of the main principles of the constructivist view of learning – to start where the pupils is and to involve the pupils actively in the process. It became clear to the teachers that, no matter what the pressure to achieve good test

and examination scores, learning cannot be done for the pupil; it has to be done by the pupil. Similarly, the focus on peer- and self-assessment, by encouraging pupils to review their work in the light of the goals and criteria, was helping them to develop meta-cognitive approaches to learning. Finally, the involvement of pupils both in whole-class dialogue and in peer-group discussions, all within a change in the classroom culture to which all four activities contributed, was creating a more rich community of learners, one where the social learning of pupils would become more salient and effective.

Thus it could be argued that an underlying reason why development of formative practice is so rewarding is that it helps to ground a teacher's work in some of the basic principles for successful learning.

Motivation and self-esteem

Learning is not just a cognitive exercise. The need to motivate pupils is evident, but it is often assumed that this is best done by offering such extrinsic rewards as merits, grades, gold stars and prizes. There is ample evidence that challenges this assumption.

Pupils will only invest effort in a task if they believe that they can achieve something. If a learning exercise is seen as a competition, then everyone is aware that there will be losers as well as winners: those who have a track record as losers will see little point in trying. Thus, the problem is to motivate everyone, even although some are bound to achieve less than others. In tackling this problem, the way in which feedback given is very important. Many research studies support this assertion.

The work of Ruth Butler (1988), which had been a starting point for encouraging comment-only marking, also showed that those given feedback as marks are likely to see it as a way to compare themselves with others (ego-involvement), those given only comments see it as helping them to improve (task-involvement), and the latter group out-performed the former. Several other studies have made the same point. For example Newman and Schwager (1995) concluded that pupils told that feedback ' …will help you to learn' learn more than those told that 'how you do tells us how smart you are and what grades you'll get'; the difference is greatest for low attainers. More generally, a comprehensive review of research studies of feedback showed that feedback improved performance in 60 per cent of them; in the cases where it was not helpful, the feedback turned out to be merely a judgment or grading with no indication of how to improve (Kluger & DeNisi, 1996).

Feedback given as rewards or grades can focus pupils' attention on their 'ability' rather than on the importance of effort, damaging the self-esteem of low attainers and leading to problems of 'learned helplessness' (Dweck 2000). Feedback that focuses on what needs to be done can encourage all to believe that they can improve. Such feedback can enhance learning, both directly through the effort that can ensue, and indirectly by supporting the motivation to invest such effort.

Whilst the work just quoted refers mainly to day-to-day classroom work, the negative effects of ego-oriented feedback are even more striking when high-stakes tests are involved. Their damaging effects on pupils' motivation and learning work have been strikingly illustrated by a survey of research on testing, motivation and learning published by the United Kingdom Assessment Reform Group (ARG, 2002). This review showed that the use of repeated practice tests emphasises the importance of the tests and encourages pupils to adopt test-taking strategies designed to avoid effort and responsibility and are bound to be detrimental to higher order thinking. The

National Curriculum tests in England have been shown to produce a significant impact on self-esteem. After their introduction, low-achieving students had lower self-esteem than higher achieving students, a relationship which was not apparent before the tests were introduced.

All of this suggests that another underlying reason why careful development of formative practice can be rewarding is that it helps ensure that a teachers' interactions with pupils will enhance rather than damage their motivation and self-esteem. However, this advantage can be annulled by summative test pressures so that changes to reduce such harm are essential.

Other initiatives

The discussion in this section has shown that the introduction of formative assessment practices opened up a far wider agenda than might have been predicted. The fact that links are being claimed to several of the fundamental issues of pedagogy and learning suggests a further question: are these implications of formative assessment practices unique to this particular approach to learning and teaching? It is only possible here to raise the question rather than to attempt to answer it. However, two examples might serve to explain the complex nature of the question.

One is the concept of 'Dialogic teaching' set out by Alexander (2004). His argument for the need to enrich the learning quality of classroom discussion draws on a range of studies of classroom discourse, notably those of Edwards and Mercer (1987) and Bruner (1996). However, he also links his review and recommendations to the work of the KMOFAP project as described above. So this examples is one that might help locate the formative initiatives in a broader context.

A different example is the range of studies variously entitled 'Cognitive acceleration' and 'Thinking skills'. For some of these there is evidence, of the power of their recommended practices to raise attainment, which is as impressive as the evidence for formative assessment (Shayer, 1999). Nothing in the classroom work involved is inconsistent with development of formative work – indeed many of the practices are almost identical. However, what is different is that these interventions are centred in tailor-made classroom practices aimed, not at the usual menu of a curriculum, but as ad hoc exercises designed to develop particular skills in reasoning. Moreover, the work has its own distinctive theoretical basis in the literature on the development of general powers of cognitive reasoning. So here is a different agenda for learning improvement, so different that one can wonder whether a combination of such a program with the more general implementation of formative practices across all classrooms might produce remarkable improvement in pupils attainments.

All of this diversity might be embraced in a comprehensive theory of pedagogy. Given the diversity, it is hardly surprising that no effective theory exists. A theory of formative assessment might be seen as a component of such a theory, but attempts to achieve this limited target have yet to emerge. My colleague Dylan Wiliam and I have tried one approach using activity theory (Black and Wiliam 2004; Engstrom, 1987), for that theory suggests ways of identifying and relating the various elements of change – in teachers' tools, in their roles and outcomes, in the changes in the culture of the classroom as community and in its rules and norms – that are all entailed. However, theories that give more emphasis to the social learning in communities of practice, or, alternatively, which give a linguistic emphasis in the analysis of classroom discourse, may also be contenders, as contributors if not for the leading role.

A Catalyst – Subversive or Benign ?

The main argument of this chapter is incomplete. What it has done is to describe a process of innovation and a set of practices, together with evidence that they improved pupils' attainments and teachers' enthusiasm about their work. Accounting for this success is the puzzle mentioned in my introduction to this chapter. The suggestion is that it is based, albeit as much by the instincts of the innovators as by explicit design, on deep principles of the nature of learning and of pupils' motivation and self-esteem.

However, partly because of this firm under-pinning, an encouragement to develop formative assessment can mislead, for it can entice teachers, by display of the evidence in its favour and by accounts of the feasibility of the practices that they might adopt, into changes in their role and in their relationships with their pupils that are far more radical, and far more demanding than they might anticipate. Insofar as teachers can have both the capacity and the support to weather the challenges that they might unknowingly encounter, the effects might be both subversive, in undermining firmly entrenched traditional practices, yet benign in replacing these with something far better.

However, the dangers are that the many obstacles that lie in the path might lead either to abandonment under stress, or to a superficial adoption which leaves the underlying traditional basics untouched. What I and colleagues now find ourselves emphasising is that the KMOFAP project took two years, that the researchers support and the opportunities for colleagues interactions were far more intensive than in most INSET programs, whether within or outside schools can provide, and that after one year of the project we were still uncertain whether it was going to succeed – so slender was the evidence at that stage that the work in the classrooms was really changing. Only in the second year did new practices begin to bloom. This finding should not be surprising in the light of research findings on the professional development of teachers. What might have to be realised is that formative assessment is rewarding precisely because it catalyses radical change, and for that very reason its adoption is only likely to succeed if the support for implementation is more careful, thorough, sensitive and expensive than is common for reform initiatives.

Acknowledgement

In writing this chapter, I have drawn extensively on the work and writing of my colleagues at King's College, notably Christine Harrison, Bethan Marshall and Dylan Wiliam Jonathan Osborne; many of the ideas and much of the evidence came from them. Information about the King's Assessment Research Group can be found on our web-site: **http://www.kcl.ac.uk/education/research/kallearn.html**

I also acknowledge with gratitude the support for the KMOFAP work provided by the UK Nuffield Foundation and by the U. S. National Science Foundation through their support of our partnership with the Stanford CAPITAL project (NSF Grant REC-9909370).

References

Alexander, R. (2004) *Towards dialogic teaching: rethinking classroom talk*. Cambridge UK : Dialogos.

ARG (2002) *Testing, Motivation and Learning* University of Cambridge School of Education : Assessment Reform Group

Black, P. & Wiliam, D (1998a) Assessment and Classroom Learning. *Assessment in Education*, **5**(1), 7-71.

Black, P. & Wiliam, D. (1998b) *Inside the Black Box: raising standards through classroom assessment* London: nferNelson. See also *Phi Delta Kappan*, **80**(2), 139-148.

Black, P. J.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D (2002) *Working inside the black box: Assessment for learning in the classroom*. London: nferNelson. See also *Phi Delta Kappan* in press 2004.

Black, P.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.

Black, P. & Wiliam, D. (2004) *Theory and practice in the development of formative assessment.* Submitted for publication.

Bruner, J. S. (1996) *The culture of education*. Cambridge MA: Harvard University Press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, **58**, 1-14.

Butler, J. (1995). Teachers Judging Standards in Senior Science Subjects: Fifteen Years of the Queensland Experiment. *Studies in Science Education*, **26**, 135-157.

Dweck, C. S. (2000) *Self Theories: their role in motivation, personality and develop*ment. London: Taylor and Francis.

Edwards, D & Mercer, N. (1987) *Common knowledge: the development of understanding in the classroom*. London: Routledge.

Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research*. Helsinki, Finland: Orienta-Konsultit Oy.

Foos, P. W., Mora, J. J. & Tkacz, S. (1994). Student Study Techniques and the Generation Effect. *Journal of Educational Psychology*, **86** (4), 567-576.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, **27** (1), 111-126.

Kluger, A. N. & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, **119** (2), 254-284.

Newman, R. S. & Schwager, M. T. (1995). Students' Help Seeking During Problem Solving: Effects of Grade, Goal, and Prior Achievement. *American Educational Research Journal*, **32** (2), 352-376.

Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation, in: P. Weston (Ed.) *Assessment of Pupils Achievement: Motivation and School Success*, (pp. 79-101). Amsterdam: Swets and Zeitlinger.

Rowe, M.B (1974) Wait time and rewards as instructional variables, their influence on language, logic and fate control. *Journal of Research in Science Teaching*, **11**, 81-94

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, **18**, 119-144.

Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education* **5**(1), 77-84

Shayer, M. (1999) Cognitive acceleration through science education II: its effects and scope. *International Journal of Science Education*. **21**(8), 883-902.

Wiliam, D.; Lee, C.; Harrison, C. & Black, P. (2003). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education*, **11**(1), 49-65.

12

Diversity of Research on Teaching

Toh Kok Aun
National Institute of Education, Nanyang Technological University, Singapore

Dedications: *John Keeves has been an icon associated with research and has been a pillar of the many IEA mathematics and science studies in the 1970s and 1980s. In his many trips shuttling between Australia and Europe, he has always kept in touch with the many friends he has in Singapore. We will always remember him for his dedication to teaching and teacher education, his leadership role in SISS, and his appointment as external examiner for our post-graduate program in education in the 1990s at the National Institute of Education, Nanyang Technological University, Singapore.*

Abstract: *In this chapter I shall review the rich field of research on teaching accumulated over the past 40 years. It will analyze past, current and emerging research traditions with a brief commentary on their outcomes. While it is not possible to cover all areas of research on teaching in this chapter, it is hoped that the areas touched upon would reflect the efforts of the many researchers, their focus on teacher variables affecting student outcomes, and how this has shifted to a focus on student variables. A new-found concern on teaching for student understanding is rooted in the widespread acceptance of constructivism in many subject areas.*

Introduction

Issues of diversity are part of a common problem that educators in most countries face throughout the world. Many societies have a mix of ethnic groups, languages, religions and cultural traditions. However, despite this diversity, the problem of teachers and teaching are very much the same from one country to another.

Interest in classroom teaching and learning has expanded tremendously in the past 40 years. Much can be learnt from the work of researchers. The field, however, has become much more complicated in terms of the range of hypotheses and factors that it now subsumes. Organizing a historical account of research on teaching is not particularly easy since the period is relatively short and so many developments are overlapping in time. When one particular area of research is at its peak, another is just beginning and yet another is beginning to decline. Additionally, a single study may include interviews with teachers and students, as well as collection of data describing classroom discourse. Thus the different areas of research identified are best seen as heuristics that cut across the field. For the purpose of this article the areas discussed include:

- Teacher personality and teacher-proof curriculum
- Teacher behaviour and school culture
- Time allocation and individual pace
- Teachers' understanding of subject-matter knowledge
- Teachers' understanding of students

Teacher Personality and Teacher-Proof Curriculum

Observation has long been a part of educational research. Some of the first systematic uses of observations in research occurred in the 1960s when the use of intensive observation was deployed for research into teachers' personalities, with the belief that teachers exhibiting certain personality characteristics were more inclined to do well in the classroom (Ryans 1960; Getzels & Jackson 1963). The adage that "teachers are born and not made" has made its rounds and certainly does not help the cause of teacher education. There will certainly be insufficient teachers to man the ever expanding number of classrooms if only "born teachers" are recruited. Not only was this myth debunked in the early 1970s by leading educational researchers, but indeed many teachers are made and the classroom scenario is all the richer for their contributions to the teaching profession.

The mid-1960s and early 1970s were also characterized by the push towards development of teacher-proof curriculum. The interest was generated in response to scientific challenges afforded by the Sputnik launch. Here again the use of classroom observation in research was attempted by the investigators, and what came through was the startling fact that the teachers' use of the curriculum materials was not synchronous with the intentions of the curriculum developers. There was too much of a focus on the curriculum materials in an attempt to make it teacher-proof, and not enough focus on the teachers who provided the necessary interface between the curriculum and the students. It is one thing to provide the curriculum materials; it is another thing to get the teachers to use them in the way intended. The consequent failure to achieve the intended goals of the new curriculum is indeed an eye opener for future attempts at curriculum development. This made educational researchers even more interested in the role of teachers (in this case, teacher behaviour) in the curriculum delivery process.

Teacher Behaviour and School Culture

Researchers would love to demonstrate a link between variation in teacher behaviour and variation in student learning. The area of research into teacher behaviour probably subsumes hundreds of studies with different variables, research methods and findings (Dunkin & Biddle 1974). The work of Rowe (1969) on "wait time" established a link between teachers' willingness to wait for individual students to think and respond, and their subsequent performance. Some 25 years later, work in this area still continues (Duell 1994). Rowe's research has not only identified the importance of teacher behaviour as the fundamental source of influence on student performance, but also established the link between teacher behaviour and the learning process.

The late 1960s also saw the beginnings of interest in school culture (Smith & Geoffrey 1968) as contributing to the learning process. The work has drew continued interest through the 1980s into the 1990s (Philips 1983; Jackson, Boostrom & Hansen 1993). On the sideline, it might be of interest to note that there was a competing group of sociologists challenging the assumption that teachers and/or schools could make a significant difference in the learning process (Coleman et al. 1966; Jencks et al. 1972). Their contentious reports led to a flurry of activity by educators and brought about the process-product paradigm (Eggleston, Galton & Jones 1975) showing that teachers did indeed affect students' learning.

Time Allocation and Individual Pace

Another area of work that captured the interest of researchers is how teachers utilized classroom time. The seminal work of Carroll (1963) was perhaps instrumental in launching the research agenda for academic learning time or time-on-task in the 1970s. Whenever any decline in academic performance is an issue, the question of time allocation springs to the fore. It is a pity though that time allocation studies did not move on to consider variations of time spent on drill versus problem solving at individual student level.

Diversity in the classroom requires educators to accommodate differences in efficiency in internalizing a particular concept. Individual pace, as advocated by Lundgren (1972) and Barr & Dreeben (1983), emerged from amongst the many competing hypotheses concerning how classroom processes might impact on student outcomes. As a result many curriculum materials emerging in the 1970s and 1980s were crafted in an individualized self-paced instruction format.

Teachers' Understanding of Subject-Matter Knowledge

Although there is a dearth of longitudinal studies involving teacher knowledge, there is no lack of studies involving the comparison of novice and expert teachers. Researchers have established distinct differences in the way expert and novice teachers operate (Berliner 1992). The differences in cognitive orientations and successful teaching have set apart the expert from the novice teachers. Teacher educators have known all along that good teaching does not happen overnight, and that initial teacher preparation courses are intended to start them in their career. The transformation from the novice state to an expert is an additive model with dimensions of successful teaching the result of trial and error episodes, or what educators normally refer to as a developmental process.

Within this track Shulman (1986) has argued persuasively on the importance of teachers' conception of subject matter. Oftentimes we have people who are very knowledgeable in their content area, but are somehow unable to present the content in such a way that others can comprehend. This is the important part of teacher knowledge, that of moving "from personal comprehension to preparing for the comprehension of others" (Shulman 1987). The ability of teachers to develop examples and models affects how students understand subject matter. It is what makes a teacher a teacher. To teach is first to understand. Shulman termed this special knowledge which teachers should have as pedagogical content knowledge, or PCK as the acronym. We expect the expert teachers to have developed their PCK to a greater extent than their novice counterparts.

Ho (2004) studied the PCK of novice teachers using a group of pre-service physics teachers and compared them with the PCK of expert teachers. He concluded that the expert teachers, with their better developed PCK, are better able to transform their own knowledge of subject matter in ways which promote their students' understanding, especially for the less able ones. Notable work in this area includes those by Ball (1993), Newmann (1992), Blumenfeld (1992), and Duffy (1993). PCK falls into the additive model framework, although it does not necessarily correlate with years of experience.

Teachers' Understanding of Students

Teachers' understanding of students as students has been given very scant attention. Students bring along with them their emotions, beliefs, attitudes and preferences, and the teacher's knowledge of these attributes is, in many ways, tied to his/her own beliefs, and in turn control how teachers get their subject matter knowledge across. The failure to consider these attributes as contributory to the learning process by researchers has resulted in blame for poor student performance on their lack of commitment, poor attitude, and absence of study skills (Tomlinson 1993), amongst others. Teachers with an inadequate understanding of their students are increasingly recognized as having failed in keeping faith with those who have been put under their charge, as well as caring for them as social beings of the community in which they reside. The neglect of students as social beings is unfortunate from the Vygotskyian perspective as students' affect may provide an important bridge to their thinking. The complete understanding of their needs will enhance the teachers' ability to reach out to them.

Teaching for Understanding

The teachers' ability to scaffold knowledge in ways that help students understand concepts is a combination of two of the areas discussed earlier: (a) teachers' knowledge of subject matter, and (b) teachers' knowledge of students. The combination of these two traditions has spawned interest in the constructivist approach to teaching. It has also benefited from the sum total of accumulated traditions of research, viz.

- that knowledge is constructed
- that knowledge is structured around key concepts
- that prior knowledge influences how new knowledge is integrated
- that knowledge needs to be restructured in the light of new knowledge
- that knowledge is socially constructed

The constructivist tradition is an outcome of the need for teachers to progressively transfer the responsibility for the learning process from themselves to the learners. It stems from the need to have students integrate their existing knowledge with new knowledge. In fact it has become fashionable to talk about constructivism in relation to teaching and learning of almost all subject areas. It is a philosophy of learning founded on the premise that by reflecting on our experiences, we construct our own understanding of the world we live in. The concept of constructivism has its roots in classical antiquity, going back to Socrates' dialog with his followers, in which he asked direct questions that led his followers to realise for themselves the weaknesses in their thinking. Piagetian ideas of assimilation, accommodation and cognitive equilibrium, together with Kelly's (1955) personal construct theory and Bruner's (1960) framework for instruction, have played major roles in the development of constructivist pedagogy. In contrast with the concept of the learner as a passive receiver of knowledge, in a constructivist classroom students are actively involved in the learning process.

If students construct new knowledge out of the experiences they encounter, then it makes sense for the teacher to grasp some parts of their experience and connect them to the knowledge to be taught (Yager 1991; Cheung & Toh 1992). To ensure student understanding requires active student involvement in what is to be learned, giving them as many opportunities as possible to practise what has been learnt. Though the emphasis is on the student, constructivism does not dismiss the active role of the teacher or the value of expert knowledge. It should therefore not be confused with discovery learning. Constructivism modifies that role, so that teachers help students to construct knowledge rather than reproduce a series of facts. The constuctivist teacher provides tools such as problem-solving and inquiry-based activities with which students formulate and test their ideas, draw conclusions and inferences, and pool and convey their knowledge in a collaborative learning environment.

Constructivism transforms the student from a passive recipient of information to an active participant in the learning process. Always guided by the teacher, students construct their knowledge actively rather than just mechanically ingesting knowledge from the teacher or the textbook. When knowledge is actively constructed by the learner, and not passively received from the environment, the student is actively processing the new knowledge. The process of **coming to know** is the result of interaction of the new knowledge with the learner's experience of the world. It is an enormously sensible and rational approach to learning. It has gained fairly wide acceptance in many parts of the world and labelled by Mathews (1993) as the new orthodoxy for teaching.

It is useful to highlight two pitfalls when embracing the constructivist pedagogy. The first pitfall has to do with "radical" constructivism – those who truly embrace constructivism. In the true spirit of constructivism (von Glasersfeld, 1989), and that means "learning is simply a matter of students making sense of the world", any views which students arrive at, for reasons that satisfy them, should be regarded as acceptable. Take the example of a middle school student in the United States who insists she can see an object in an enclosed room when the lights are switched off (Macbeth, 2000). The teacher must recognise that the student is not wrong as far as she is concerned, because after a period of adjustment she is normally able to see objects around her. Most likely she has not encountered complete darkness. She is making sense of the world as she knows it. The teacher cannot be telling her she would not be able to see if there is complete darkness. This brings constructivism into conflict with rationality of science.

The second pitfall has to do with social constructivism. In social constructivism the assumption is that the process of learning is an entirely rational and logical activity. Learning, however, sometimes does not occur in that way. We know, from the work of Posner, Strike, Hewson and Gertzog (1982), that new learning can be brought about when learners are dissatisfied with their pre-existing belief or understanding (as when confronted with a discrepant event), especially when the new knowledge being presented is intelligible (that the learner fully comprehends what the new knowledge means), plausible (it is something that is both believable and consistent pre-existing knowledge of the learner), and fruitful (it is able to provide something of value to the learner, e.g. providing a solution to some unresolved problem).

Teaching therefore involves lowering the status of students' pre-existing views and raising the status of the new knowledge with respect to intelligibility, plausibility and fruitfulness to the learner. Learning involves rejection of the pre-existing view for the new view. However, there are many occasions where the new view is being rejected in spite of all the above efforts being fulfilled. Using the example earlier, the student argues that she knows that given sufficient time for her eyes to adjust, she will be able to see the object even in complete darkness (Macbeth, 2000). After a five-minute wait time in the dark for her eyes to adjust, she still rejects all the evidence that light reflection from an object is necessary for vision. What do we do with a learner who does not wish to 'jump ship' despite all the evidence? This is a case where all the rationality and logical arguments presented are still being rejected. One can have all the respect for constructivism but then personal understanding isn't a science.

Conclusion

We now have the benefit of research work carried out for more than 40 years behind us, but I still have one lament. Despite this rich database, the teaching taking place in classrooms today has not changed that drastically compared to 40 years ago. What happens today is almost exactly what one would see in the 1960s. There is some apparent invulnerability of teachers to change. This is not to say that there were no changes being instituted over this period. The widespread recourse to PowerPoint presentations with multi-media interface or web-based links has replaced the old faithful overhead projector. What an ardent observer would notice, however, is that the instructional style remains unchanged, impervious to the many research findings touting student understanding and constructivism.

There seems to be some form of seemingly stubborn promulgation of the teacher-centred instruction. One would not find much evidence of significant change in instructional style despite the inputs of pre-service training and the many findings from research in educational journals arguing for reformation towards instruction that is more student-centred. The evidence suggests that teaching practices seem remarkably stable at all levels of schooling through the many decades, despite improvements in teacher education and inputs of scholarly knowledge. Here and there one might find an episode or two of constructivism when it suits the teacher's purpose.

I detect a critical tension between the practitioners in the classrooms and the researchers with their vast database armoury prescribing what needs to be done. I detect the tension between knowing the subject matter and knowing how to teach the subject. I detect the tension between fully accepting constructivism and critically questioning its adoption. There is also a larger tension of reform, the tension between hopelessness that things will never change, and an optimism that they will change for the better. I end with some degree of optimism that those at the heart of the

educational process will continue relentlessly to work with the teacher in moulding the future of our citizenry.

References

Ball, D. (1993). With an Eye on the Mathematical Horizon: Dilemmas of Teaching Elementary School Mathematics. *Elementary School Journal*, 93, 373-398.

Barr, R. & Dreeben, R. (1983). *How Schools Work*. Chicago: University of Chicago Press.

Berliner, D. (1992). *Educational Reform in an Era of Disinformation*. Paper presented atthe Annual meeting of the American Association of Colleges for Teacher Education, San Antonio, Texas.

Blumenfeld, P. (1992). The Task and the Teacher: Enhancing Student Thoughtfulness in Science. In J. Brophy (Ed.), *Advances in Research on Teaching,* Volume 3 (pp. 81-114). Greenwich, Connecticut: JAI Press.

Bruner, J. (1960). *The Process of Education*. Cambridge, Massachusets: Harvard University Press.

Carroll, J. (1963). A Model of School Learning. *Teachers College Record*, 64, 723-733

Cheung, K.C. & Toh, K.A. (1992). In the Eyes of the Beholder: Beginning Teachers' Conception of the Nature of Science and Science Teaching. *ERIC Database,* Ascension No. ED 337 448,

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Department of Health, Education and Welfare.

Duell, O. (1994). Extended Wait Time and University Student Achievement. *American Educational Research Journal*, 31, 397-414.

Duffy, G. (1993). Rethinking Strategy Instruction: Four Teachers' Development and Their Low Achievers' Understandings. *Elementary School Journal*, 93, 231-247.

Dunkin, M. & Biddle, B. (1974). *The Study of Teaching*. New York: Holt, Rinehart & Winston.

Eggleston, J., Galton, M. & Jones, M. (1975). *A Science Teaching Observation Schedule*. London: MacMillan.

Getzels, J. & Jackson, P. (1963). The Teacher's Personality and Characteristics. In N. Gage (Ed.), *Handbook of Research on Teaching* (pp. 506-582). Chicago: Rand McNally.

Ho, B.T. (2004). *Development of Pedagogical Content Knowledge (PCK).* Unpublished doctoral dissertation. Singapore: Nanyang Technological University.

Jackson, P., Boostrom, R. & Hansen, D. (1993). *The Moral Life of Schools*. San Francisco: Jossey-Bass.

Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.

Kelly, G. (1955). *The Psychology of Personal Constructs*. New York: Norton.

Lundgren, U. (1972). *Frame Factors in the Teaching Process*. Stockholm: Almquist & Wiksell.

Macbeth, D. (2000). On an actual apparatus for conceptual change. *Science Education*, 84, 228 – 264.

Matthews, M. (1993). Constuctivism and science education: Some epistemological problems. *Journal of Science Education and Technology*, 2, 359 – 370.

Newmann, F. (Ed.) (1992). *Student Engagement and Achievement in American Secondary Schools*. New York: Teachers College Press.

Philips, S. (1983). *The Invisible Culture: Communication in Classroom and Community on the Warm Springs Indian Reservation*. New York: Longman.

Posner, G., Strike, K., Hewson, P. & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211 – 227.

Rowe, M. (1969). Silence, Silence and Sanctions. *Science and Children*, 6, 11-13.

Ryans, D. (1960). *Characteristics of Teachers*. Washington, D.C.: American Council on Education.

Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher,* 15, 4-14.

Shulman, L. (1987). Knowledge and Teaching: Foundations of the New Reform. Harvard *Educational Review,* 57, 1 – 22

Smith, L. & Geoffrey, W. (1968*). The Complexities of an Urban Classroom*. New York: Holt, Rinehart & Winston.

Tomlinson, T. (1993). Educational Reform: The Ups and Downs of Good in Good Intentions. In T. Tomlinson (Ed.), *Motivating Students to Learn: Overcoming Barriers to High Achievement* (pp. 3-20). Berkeley, California: McCutchan.

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese*, 80, 121 – 140.

Yager, R. (1991). The Constructivist Learning Model: Towards Real Reform in Science Education. *The Science Teacher*, 58, 52 – 57.

Part 3

Flinders University Institute of International Education and Beyond

Tony Gibbons
Flinders University

The authors of the papers in this section all acknowledge the impact that Professor Keeves has had on their ideas and their work. They speak from the point of view of both colleagues and students. They speak of his 'visionary intellectual leadership', his 'thirst for the development of knowledge', his patience, his interest and the generosity with his time and ideas. The papers in this section cover a range of topics and different types of inquiry.

Lawson and Askell-Williams are concerned with the quality of cognitive activity and its improvement. Teasdale is concerned with a regional education project aimed at improving the planning and delivery of basic education in 15 small countries in Oceania. Scholten is concerned with student swallowing and brings the power of statistical analysis to the problem of the growth of specific student learning. Thompson uses Rasch scaling to analyse the performance of judges of produce. Alagumalai and Mohyla write about security and privacy concerns for portals and online applications. Gibbons casts a philosophical eye over the concepts of modelling and experiments. Dutney highlights the impact of theological education. Matthews discussed teaching Out of the Unconscious, while Kotte and Lietz discuss factors influencing reading achievement of students in Germany and Finland in the PISA 2000 study.

Thus the inquiries range from policy to philosophy to the behavioural sciences. A broad range in itself but clearly linked by a concern for questions that are important to the provision of education and how to improve it both at the level of the planner and the level of the teacher. However, the range of inquiry evident in these papers must

be taken as an indication of the variety of subjects tackled and the range of intellectual inquiry brought to bear as a result of the breadth and scope of Professor Keeves' interests and mastery.

13

Investigating Good Quality Knowledge about Learning and Teaching

Michael J. Lawson & Helen Askell-Williams
Flinders University

Dedications: *We feel honoured to write a paper for a collection assembled to mark a significant time in the life of John Keeves. One of us, MJL, has worked with John since he came to Flinders University, and HAW has known John as mentor, teacher and supervisor. Our work has benefited immensely from our discussions with John, and we have enjoyed time spent in his company. His contribution to teaching, postgraduate supervision, research and the culture of enquiry and international friendship in the School of Education is enormous. He has made similar contributions across the wider university and to educational research in this State and across the national and international education scene.*

John is a mentor to us and to many other beginning and experienced researchers on a diverse range of research topics. One of the outstanding characteristics of a conversation with John about a complex research design or statistical analysis problem is that he shows immediate interest. We know that he is asked research related questions almost everyday, but, without fail, he engages with our research problems in a supportive way. Even when we don't know something that we should (say, about eigenvalues or regression weights), we aren't made to feel that we are wasting his time. He takes on the problem and works with us to derive a solution. If he isn't sure of the way forward, he always knows someone who is, and advice is sought. John has that quality possessed by truly great teachers: When he is teaching us something, he genuinely makes us feel as though we are teaching him, thus encouraging us to strive even harder. So it is a pleasure to be able to remind others of his contribution to educational

research. In this paper we review some of our research, in much of which John has been involved as co-author or consultant.

Strategies

In the 1960s and 1970s the so-called 'cognitive revolution' in psychology started to make a major impact on instructional psychology. Models of memory such as the modal model of Atkinson and Shiffrin (1968) were examined and refined. One of the key inclusions in such models was the provision for **control** processes. These were seen to be broad strategic processes devised by the individual person in response to task experiences. Both the strategic nature of these processes and their control functions have been of interest in instructional psychology since that time.

Recognition that an individual develops strategies and tactics for completing specific cognitive tasks has been central to development of accounts of learning that adopt a self-regulatory perspective (e.g. Winne & Hadwin, 1998; Zimmerman, 2002). The inclusion of both words in this compound expression is significant in three respects. The specification of a self component recognises the role of the individual in managing learning, not just in an individual study settings, but also in classroom settings where interaction with other students and teachers occurs. In addition, the self is interpreted broadly in this self-regulatory learning (SRL) view, including both cognitive and affective elements. Thus SRL models recognise that students are not only making decisions about how to transform cognitively a particular set of information. Students also make decisions based on sets of beliefs about their own capabilities and of the influences that are likely to be operative in the learning situation. A student's decision making therefore draws broadly upon information about the self, about the situation and about the task: it is emotively hot, cognitively cold and transactively social!

The third point of significance associated with control processes is their controlling, or regulatory, function. In early cognitive models, such as that proposed by Bower (1975), control was manifested in the executive functions of cognitive activity. In the mid-1970s this executive function came into prominence in research on metacognition (e.g. Flavell, 1985) and in studies of mental retardation (Brown, 1978; Lawson, 1980). In the latter research, a point of interest was whether interventions related to the regulation of cognitive activity could add value to more task-strategic interventions, so that the individual would acquire information about both the strategy relevant to the specific task and information about when and how to use that strategy in daily life. Since that time, the proposition that students control their cognitive activity to construct knowledge has become embedded in major curriculum initiatives such as the South Australian Curriculum Standards and Accountability framework (DETE, 2001).

A major interest in our research on the regulation of cognitive activity is a concern to improve the quality of cognitive activity so that the student ends up with more powerful, more generative knowledge (Wittrock, 1990). In the first set of studies reviewed in this paper we focus attention on ways to improve the quality of learners' cognitive knowledge representations. In the final section of the paper we address more directly the question of how to conceptualise and evaluate the quality of learners' knowledge.

Improving Access to Knowledge

When Ken Watson reflected on the short-answer examination performance of his senior geography students at Prince Alfred College, he suspected that many of the students were not effectively accessing knowledge that they had developed during their previous learning activities in class and during homework. Ken worked with his students to review ways to solve examination questions. When he asked his students to write down parts of examination question solutions that they did not know, many of the students stated that they did have the knowledge that was needed to solve the examination questions. This begged the question of why the students did not access their knowledge during the examination. We reasoned that the students were not exerting effective control over their problem solving activity when they worked on the examination problems. This led to our evaluation of a procedure designed to help students make more effective use of their available, but seemingly 'inert' (Whitehead, 1942) knowledge.

The details of our intervention program and its outcomes are reported in Watson and Lawson (1995). Briefly, when students were instructed to, a) be confident as they activated knowledge relevant to the question, b) to inspect their knowledge for possible answers to the set question prior to considering the response alternatives, and c) to reflect critically on the acceptability of their selected response, their performance was substantially improved. Our intervention focussed on strategic behaviours concerned with the students' dispositional state, with procedures for a) problem analysis and representation, b) knowledge activation, and c) the management of cognitive processing activity. Use of the strategies by the students in the experimental group resulted in a statistically significant improvement in their test scores. The experimental effect was obtained within the first week of training, after just four one-hour training sessions, and was shown to be educationally significant. Analysis of students so-called 'thinking aloud' protocols indicated that the observed pattern of strategy change was consistent with the pattern of test results, supporting the assertion that treatment did bring about changes to the strategies used by students in their approach to answering examination questions. However, improved performance required explicit training rather than merely informing the students about the strategy.

The intervention in the Watson study was broadly focussed on dispositional elements as well as on retrieval and management of cognitive processing. Students were encouraged to be confident that they possessed the necessary knowledge and to attribute success or failure to processes that they could control. In addition, self-instructions were used to encourage a strong disposition towards planning and executive control of question answering. In effect, the intervention strategy was directed at getting students to bring the outcome of the problem-representation process under metacognitive, or executive, control, which is especially important for school students who are not yet sufficiently expert to automate parts of the problem solving process..

Improving Encoding of Vocabulary in Second Language Learning

In second language learning, both child and adult learners must acquire a sizeable vocabulary. This is a task at which second-language learners are usually novices, because most of the vocabulary acquisition undertaken in a native language is not explicit learning. We succeed at native language learning without being conscious of

how we do so. The size of the vocabulary acquisition task can be appreciated from an informal estimate gathered by Don Hogben when he and MJL were beginning studies in this area. Don's Italian tutor estimated that for students to read the novel set for the senior undergraduate students of Italian, the students would need to learn the meanings of about 2000 Italian words.

In our observational study of students engaged in learning the meanings of Italian words (Lawson & Hogben, 1996) we found a dominance of strategies that were forms of repetition involving minimal cognitive elaboration. Such repetition is important for learning because it results in strengthening of a memory trace. But it is a rather blunt learning instrument. More effective is the use of elaborative methods such as the keyword mnemonic technique described by Atkinson and Raugh (1975), in which, through the use of imagery, elaboration of the meaning of a new word can provide more direct routes for retrieval of a representation of the word-meaning complex. Hogben and MJL set out to examine whether the beneficial effect of keyword training could be maintained over time in the classroom situation. This issue was of interest because other studies on the effect of keyword training had observed a rapid rate of forgetting in keyword trained students.

The analysis of the data in our keyword study was also of interest, because we wanted to examine the value of the use of the keyword method by experienced students during multiple occasions in a classroom setting. In our study, both immediate and delayed testing of recall for the definitions of concrete and abstract nouns was explored. Students' recall performance was tested on four occasions in order to gain more information about the path of forgetting. The technical difficulty in doing such research was that the students were not seen to be the functional unit of analysis because of their membership of different classes. The solution suggested by John Keeves, and confirmed by his visiting colleague Nic Longford, was to undertake a multilevel analysis that could effectively model the nested structure of the data. This we did, and thus began an interest in our School of Education in the use of the multilevel modelling approach for the analysis of learning growth over time.

In the Lawson and Hogben study, the pattern of recall performance was consistent across all test occasions. The keyword group consistently recalled the definitions of more words than did the untreated control group students. The differences between the groups were substantial on each occasion and the profile of performance of the two groups across time was similar. Both groups showed a decline in recall after the immediate test, this decline being somewhat greater for the keyword students. Thereafter the recall levels of both groups remained stable for the final two test occasions. In the keyword group, performance continued to show a small rise following test 2.

In the multilevel analysis, the coefficients for the estimation of the intercept indicated that the total definition recall was positively influenced by both word knowledge score and by group membership. Students with high word knowledge scores recalled, on average, more word definitions. Of particular importance was the statistically significant effect of group membership, with students in the keyword group recalling, on average, about six and a half more words out of 30 than students in the control group.

The results for slope, related to test occasion, indicated that the total recall of word definition across time was also positively affected by the keyword training. This suggested that the recall level would be predicted to increase slightly across time and that the difference in recall level between the treatment and control groups diverged in a manner that showed greater benefit for the treatment group. The pattern of the

results for slope associated with comparison of immediate and delayed testing showed a statistically significant coefficient for the treatment group, indicating that the major impact of keyword training was at time of the first recall test.

Improving encoding and retrieval through self-explanation

An intervention designed for mathematics learning, which was also directed at control of cognitive activity, was developed by Regina Wong (2002). In this study, our general concern was to explore the detailed effects of a form of study-skill intervention that involved training high-school students to use prompt questions to develop elaborated explanations of the mathematical material they were studying. Our central issues of concern were the effects of self-explanation training, divided into three sets.

1) Whether a positive effect of self-explanation training would emerge in response to a less-directive form of training than had been used in previous laboratory studies.
2) Whether intervention training would affect the accessing and generation of problem-relevant knowledge that was revealed in the students' think aloud protocols during training.
3) The degree of influence of this intervention training relative to other measures of the students' cognitive and affective status, in particular their self-management activity, prior knowledge, and beliefs about mathematics learning. Our analysis of the relative influence on mathematical problem-solving performance among this group of variables was examined using a partial least-squares path analysis (Sellin, 1995).

A similar line of theoretical discussion was developed around the self-explanation procedures investigated by Chi and colleagues (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). In a study of the procedures used by university students during study of a physics text, Chi et al. found that the most effective problem solvers both, generated more explanation statements (termed self-explanations) and, monitored their understanding more frequently, than did the students who were less effective problem solvers.

In our study, when the performance of students receiving the self-explanation training for study of a theorem in geometry was compared with that of students in a no-training control group, the former group showed a substantial significant positive benefit. The improved level of problem solving performance was strongest for problems that required the greatest degree of transfer of knowledge from the problems used in the initial period of study. Analysis of verbal protocol data showed that the self-explanation training was associated with increased frequency of management of cognitive processing, and with increased knowledge access and knowledge generation activity. As was the case in the study by Watson and Lawson (1995), the strategy training in the Wong et al. study was directed at bringing study activity under a greater degree of explicit metacognitive control by the student. When we examined the students' problem solving attempts we found that the requirement to self-explain was associated with more problem-related activity in both searching for, and using, related knowledge and in the students' external constructive activity. A path analysis showed that the beneficial effect of self-explanation training on performance was most strongly associated with increased knowledge-access activity.

The fact that the self-explanation group were more active both in accessing geometry knowledge related to the test problems and in the elaboration of the details given in the problem statements and diagrams was of particular interest in this project because,

in this case, it is not study activity alone that is affected by the self-explanation training. The training also affected use of the knowledge that had been studied on a previous occasion. This pattern of results suggests that the beneficial effect of self-explanation training is not limited to encoding and establishment of a knowledge representation (cf. Chi et al., 1989), but is also operative, at least in the immediate future, when the students are called upon to activate and use this representation.

Although the design of this study did not permit us to make claims about the quality of students' knowledge representations, or of knowledge use, the increased accessing of related knowledge and the greater frequency of elaborative activity could be seen as giving the student a greater chance to identify and exploit knowledge that might be relevant to a particular problem. In each of the three studies described above, our interest was in improving the quality of the students' knowledge representations. That such an effect was obtained is implied by the positive impact on the students' performance of each of the strategy interventions. Yet the information about any change in quality of the knowledge representation is indirect, for we did not have a way of reliably characterising the nature of the quality of students' knowledge representations. Resolving this difficulty has become a major focus of our more recent work.

Characterising the Quality of Knowledge

Throughout the period of John Keeves' work on a broad range of issues of measurement and evaluation in education the problem of evaluating the **quality** of students' knowledge has persistently occupied researchers' thoughts and endeavours. Good quality knowledge is said to promote better recall (Hogan, 1999b; Kirby & Woodhouse, 1994), better learning outcomes (Biggs & Collis, 1982), better problem solving (Carey, 1985; Chi, 1985; Glaser, 1984; Sweller, 1991), critical thinking (Mintzes & Novak, 2000) and better quality actions (Kerr, 1981).

However, identifying the nature of good quality knowledge has not been straightforward. Jacoby and Craik (1979) pointed out that the proposition that more deep and meaningful analyses of perceptual events were accompanied by more durable memory traces was troubled by the fact that "some difficulty has been encountered in specifying exactly what is meant by 'deep' and 'meaningful' " (Jacoby & Craik, 1979, p. 1). In the same year, J. R. Anderson and Reder (1979 p. 385) stated, " there exist no explicit rules … for measuring the 'depth' of a task."

Ten years later, Eysenck (1989) reiterated that "there is no adequate independent measure of the depth of processing" (p. 291). Then, another ten years passed to find Mintzes and Novak (2000) asking, "What does it mean to understand …?" and "How will I know when my students have developed this ability?" (p. 42).

Early depth of processing studies relied upon single indicators of quality, such as encoding information with "meaning." (Craik & Lockhart, 1972). Similarly, Marton and Saljo's (1976) assessment of quality appears to rest upon participants' "conceptions of the intentional content of the [text] passage" (Marton & Saljö, 1976 p. 8).

Moving beyond a uni-dimensional approach, a substantial contribution to identifying the quality of students' knowledge was made by Biggs and Collis' (1982) four-dimensional Structure of Observed Learning Outcomes (SOLO) taxonomy and a multi-dimensional perspective of cognitive (memory) structure proposed by White (1979) and White and Gunstone (1980). White's (1979) initial dimensions of good quality knowledge were a) extent, b) precision, c) internal consistency, d) accord with

reality, e) variety of types of memory element, f) variety of topics, g) shape, h) ratio of internal to external associations, and i) availability.

In related work, Martin et al. (2000) scored the structural complexity of college students' domain-specific concept maps according to six criteria, namely, concepts, relationships, hierarchy, branching, cross links and interconnectedness. By assessing change in the structural complexity of participants' concept maps, Martin et al. concluded that knowledge increases in quality in a stepwise, gradual and cumulative process, with periods of weak and radical change.

Meanwhile, Hogan and colleagues (Hogan, 1999a; Hogan, 1999b; Hogan & Fisherkeller, 2000; Hogan, Nastasi, & Pressley, 2000) conducted a 12-week intervention study called 'Thinking aloud together'. The aims of the study were to develop the metacognitive, regulatory and strategic aspects of knowledge co-construction in a sample of eighth grade students. To assess students' reasoning complexity, Hogan created a rubric containing six criteria: generativity, elaboration, justifications, explanations, synthesis and logical coherence. Hogan found that, although the students could talk about their new-found skills of negotiation and collaboration, when a new problem solving situation was presented, the students did not invoke those skills. Hogan concluded that there was, "a gap between students' metacognitive knowledge about collaborative cognition and their use of collaborative reasoning skills and attainment of well-integrated conceptual understanding" (Hogan, 1999b, p. 1101).

Similarly, McKeown and Beck (1990) investigated the quality of fifth and sixth grade students' knowledge about the historical period leading to the American Revolution, employing measures of correctness of responses, quantity of major ideas, quantity of elaborative ideas, relationships between ideas, and organisation of ideas, to ascertain the quality of students' knowledge. McKeown and Beck found that the students' knowledge was sparse, poorly connected and often confused into cognitive "stews."

The quality of knowledge about learning and teaching

In our research into evaluating the quality of students' knowledge we have focussed upon one particular domain of knowledge, namely knowledge about teaching and learning. Our concern about the quality of learners' knowledge about teaching and learning stems from work by Winne and Marx (1977; 1980; 1982) who drew attention to the mediating effect that learners' perceptions have upon teachers' instructions:

> The emphasis in the cognitive mediational view of teaching is … on students' cognitive interactions with teaching. The ultimate measure of teaching is … whether students think about content in ways that can promote achievement. (Winne, 1985, pp. 673-674)

Winne and Marx (Marx, Howard, & Winne, 1987; 1977; 1980; 1982) sought to demonstrate flaws in the then prevailing process-product paradigm, and to draw attention instead to the key role played by the cognitive activities of learners. The learner came to be viewed as an active agent "continuously involved in cognition about self and environment" (Winne & Butler, 1994, p. 5738). Winne and Marx (1982) demonstrated that when (and if) students attend to teachers' instructions, the students actively interpret those instructions according to their own prior knowledge and motivations. Whereas students' prior knowledge in areas such as maths or science is now recognised as an important starting point in constructivist teaching-learning environments, it is also the case that prior knowledge about teaching and

learning are important variables in the teaching-learning equation, for both teachers and learners.

In their investigations into students' metacognitive instructional knowledge, Elen and Lowyck (1999) observed that students "[did] not seem to have articulate conceptions about the way in which an instructional environment may support their cognitive processing and/or control activities" (p. 157). In order to investigate this premise further, we asked Teacher Education students, "What happens in my university classes that helps me to learn," (Lawson, Askell-Williams, & Murray-Harvey, 2003). This question seemed particularly important to ask of teacher education students for two reasons. First, because the quality of students' knowledge about what helps them to learn would mediate the students' interaction with their own learning processes, which would in turn mediate their interaction with the topic of study. The second reason was that, in the case of Teacher Education students, the quality of their knowledge about their own learning could be expected to mediate their pedagogical interactions with their own (prospective) students.

The teacher education students in our studies listed a range of factors that help them to learn, including, discussion, small group work, hearing different opinions, teachers modelling of good practice, readings, hands-on learning activities, supportive environment and role-plays. In follow up interviews, the teacher education students provided a range of personal theories about their learning, ranging from theories that appear to have potential to facilitate learning, as in Example 1, and theories that appear impoverished, as in Example, 2:

Example 1:

> But somehow to integrate the new knowledge, to sort of append it to the page… or even perhaps to draw all the related pages from my storehouse, my library, and create a new thing, to lay them all out on the table with this new piece and see whether that forms a new pattern that I hadn't had before, which is what I mean by synthesis.

Example 2:

> I don't know what really helps me to learn…I don't know how that helps me to learn, it just does. It's just something I've never questioned, it just helps me…it's just the way I've learned to survive while I'm doing these things. But I don't know how.

In order to investigate further the quality of the participants' knowledge, we created a *generative power* rubric, which was used to classify students' statements about one teaching and learning strategy: class discussions. The rubric contained four levels of generative power, ranging from propositions with no elaboration (level 1), through elaboration (level 2) and implication (level 3), to effects on learning and links with theory (level 4). We employed the rubric to classify statements in the participants' interview transcripts. The results of the classification are displayed in Figure 1, where it can be observed that the generative power of students' knowledge ranged from very low (student F) to high (student H).

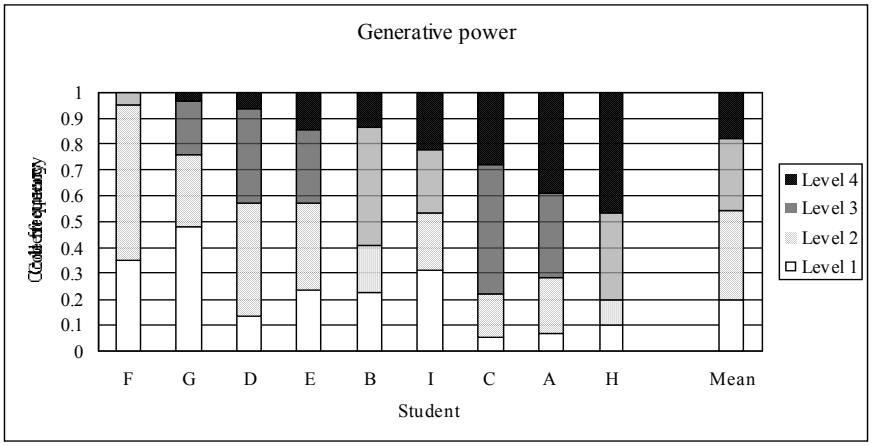


Figure 1: Generative Power Analysis

The results of the generative power analysis suggest variations in the effectiveness of students' engagement with their own learning, and certainly would seem to have implications for teacher education students' ability to generate productive learning experiences for their own students in a key pedagogical technique such as class discussions. Our findings remind us of a point made earlier in this paper in the context of the Watson and Lawson (1995) study: Explicit teaching, rather than simply immersion, would seem to be required in order to inform students about the strengths and weaknesses of different pedagogical techniques.

In a further study, we created a Framework of Quality of Knowledge (the Framework) (Askell-Williams, 2004; Askell-Williams & Lawson, 2003). The Framework contains five main constructs, which were extracted from a review of the works on dimensions of quality by previous researchers, namely *complexity, structure, well-foundedness, contexts and cognitive representations*. We employed the Framework, in conjunction with the coding facility of NUD*IST software (QSR, 1997), to evaluate child-care students' and medical students' accounts of their knowledge about teaching and learning as expressed in in-depth interviews. The data generated by the use of the Framework enabled the use of correspondence analysis, with the aim of finding a low-dimensional representation of the dependence between predetermined categories in a two-way contingency table (Hair, Anderson, Tatham, & Black, 1995). Figure 2 is the graphical display of Dimensions 1 and 2 produced by the correspondence analysis. The display positions the 29 variables and eight participants included in the analysis in relative placement to each other. From Figure 2, it can be seen that Dimension 1 accounts for 49.6 per cent of the variation in participants' responses, and Dimension 2 accounts for 19.3 per cent of the variation.

From Figure 2, it can be observed that Dimension 1 variables range from *prompts, examples, propositions* and *episodic memory* at the left of the chart, to *congruent-program*, *diagrams/flow charts, synthesise, relate components, and cross links* at the right. Taken together, the variables along the Dimension 1 continuum suggest different kinds of *Cognitive Schema* (Anderson, 2000), with *stockpiles of knowledge* at the left-hand pole, to *connected knowledge* at the right.

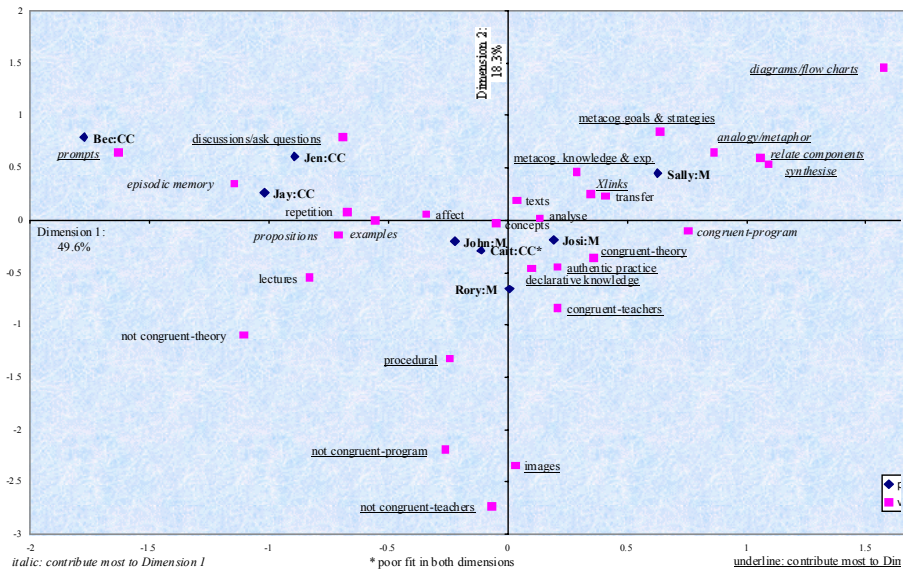


Figure 2: Correspondence Analysis Graphical Display: Dimension 1 and Dimension 2

The positioning of variables in Dimension 2 ranges from prompts, diagrams/flow charts, discussions/ask questions, metacognitive goals and strategies, metacognitive knowledge and experiences, cross links, analogy/metaphor, relate components, and synthesise at the top of the chart, to not congruent-teachers, not congruent-program, images, procedural knowledge, declarative knowledge, authentic practice and congruent with theory and teachers at the lower pole of the chart. These variables suggest a dimension of *Cognitive Representation* (White, 1979; White & Gunstone, 1980), ranging from *Conceptual* or *Abstract representation* at the top of the chart, to *Practical Representation* at the lower pole of the chart.

The correspondence analysis solution also locates participants' scores on the chart. There appear to be substantial individual differences. For example, Bec's (child-care) score is situated to the far left of Dimension 1, and Sally's (medical) score occupies the furthest position to the right. Clusters of participants' scores are also apparent, with the child-care students' scores clustered in the left side, middle to upper quadrant. The medical students' scores are spread out across the remaining three quadrants.

The most interesting part of the correspondence analysis emerges with the *relative* placement of participant and variable scores to the interpreted dimensions. Hence it can be seen from Figure 2 that three of the child-care students' scores are located at the *stockpiles of knowledge* end of Dimension 1, two medical students' scores are located around the centre of *stockpiles* to *connected*, and Sally's score (medical) is toward the *connected knowledge* variables. In Dimension 2, Sally's (medical) and Jay's (child-care) scores are located closer to the *conceptual/abstract* pole, and Rory's (medical) and Josi's (medical) scores are located closer to the *practical* pole.

Correspondence analysis also provides information that permits the creation of participant profiles, which can then be compared to, for example, the profile of a

prototypical Dimension 1 person. The purpose of such a comparison is to go beyond identifying a participant's position on a chart (such as in Figure 2), to consider how the complete set of (in the present case) 29 variables combine to describe a participant's pattern of responses.

For example, Figure 3 is the plot of Sally's profile across all variables, against the profile of Dimension 1.

In Figure 3 it can be seen that Sally's (medical) profile accords quite well with the Dimension 1 profile. Sally scores relatively strongly on *diagrams/flow charts, relate components, analogy/metaphor, synthesise, metacognitive goals and strategies, congruent with program and transfer,* and relatively weak on *images, lectures, procedural knowledge, the three non congruent variables, episodic memory* and *repetition.*

In addition to plotting participants' profiles against dimension profiles, it is possible to create other visual displays from the information provided by the correspondence analysis. For example, a perspective of the relative level, or extent of knowledge across all variables held by each participant, can be gained by plotting Sally's and Jay's (child-care) profiles on the same chart, as displayed in Figure 4.

It is immediately apparent that, with the exception of the variables *propositions*, *lectures, repetition, episodic memory and prompts*, most of Sally's scores are at a higher level than Jay's. Therefore, even when Sally's profile dips as Jay's rises, for example at *metacognitive knowledge and experiences,* and *analyse*, the level of Sally's functionally available knowledge level is still higher than Jay's.

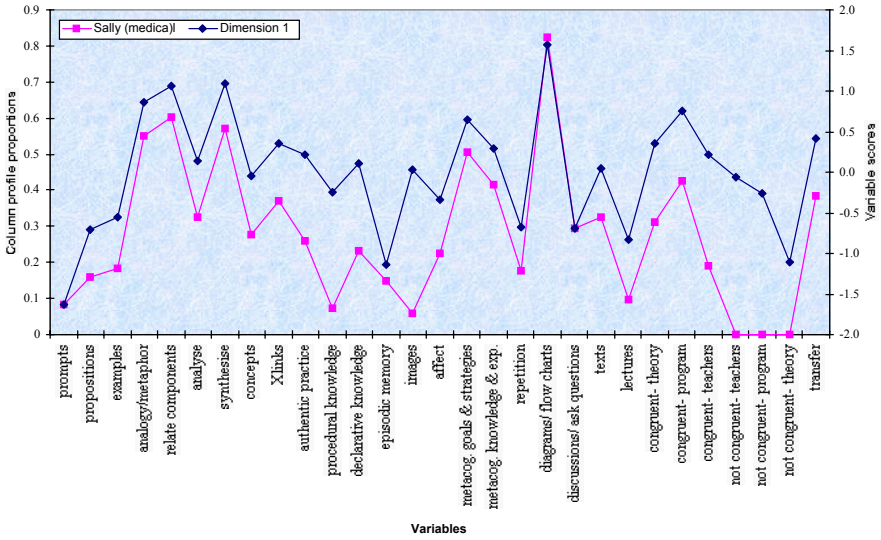


Figure 3: Profile comparison: Dimension 1 and Sally (medical)

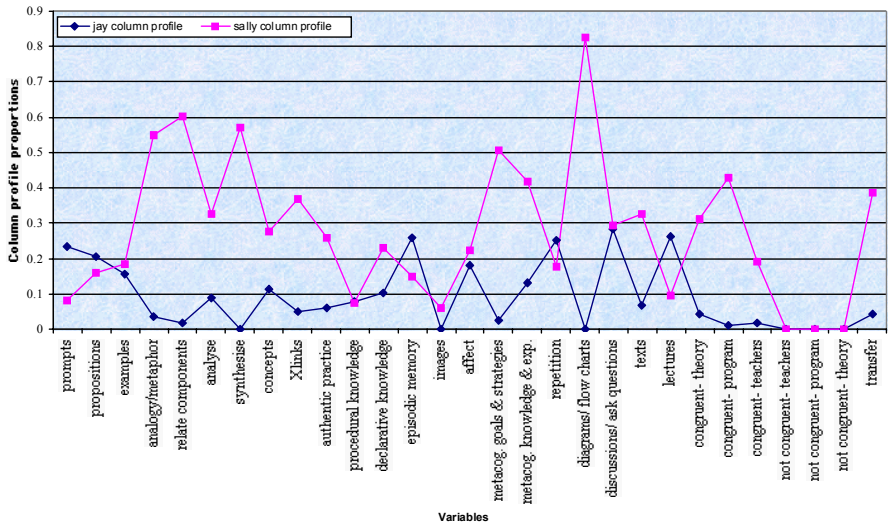


Figure 4: Profile comparison: Sally (medical) and Jay (child-care)

From Figure 4 it is also possible to identify comparative patterns between Sally's and Jay's profiles. For example, whereas Sally's profile rises across the first four variables of the Framework category of *Complexity* (*prompts* to *analogy/metaphor*), Jay's profile falls. As these variables are ordered from less complex to more complex, this indicates that Sally's profile reflects relatively more of the higher-level variables of *Complexity*.

From Figure 2, it seems reasonable to classify Sally as a relatively more connecting person, and Jay as a relatively more stockpiling person. In addition, from the profiles displayed in Figure 3 and Figure 4, it is possible to highlight the variables that contribute to these classifications by observing the peaks and troughs on the chart. Thus, the profiles provide a tool for diagnosis of possible areas of intervention to improve the knowledge about teaching and learning that both Sally and Jay hold. Suggestions might include a program that provides Jay with explicit instruction in strategies for creating connections between pieces of knowledge, such as concept mapping (White & Gunstone, 1992). Jay might also benefit from instruction in metacognitive skills of monitoring and evaluation of progress (Winne & Hadwin, 1998). In Sally's case, the value of imagery as an encoding and retrieval tool (Anderson, 2000) might be a useful addition to her already strong knowledge base about teaching and learning.

Concluding Statement

We noted in the introduction to this paper the benefit to our research resulting from our collaboration with John Keeves. His thirst for the development of knowledge and for the generation of solutions to research problems has inspired us as well as other contributors to this volume. Of further inspiration is the fact that John has generated all of this effect since he is said to have 'retired': He shows no sign of slowing his pace! In this he anticipated the exhortation of our political leaders to the more mature

citizens of our country to stick around in order to be a valuable source of advice and wisdom. We call upon our current leaders to show a similar degree and direction of wisdom to that of John Keeves, in recognising the value of the investment of time, effort and money into research in education:

> Anything so important to our lives as teaching (and learning) is well worth trying to understand. (Kerr, 1981, p. 62, our addition)

References

Anderson, J. R. (2000). *Cognitive psychology and its implications*. (5th ed.). San Francisco: W. H. Freeman.

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385-404). Hillsdale, NJ: Erlbaum.

Askell-Williams, H. (2004). *Teachers' and learners' knowledge about teaching and learning.* Unpublished PhD thesis, Flinders University, Adelaide, Australia.

Askell-Williams, H., & Lawson, M. J. (2003). *Scaling and profiling dimensions of difference in the quality of learners' knowledge about teaching and learning.* Paper presented at the 10th biennial conference of EARLI, Padova, Italy, August.

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2,). New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. London: Academic Press.

Bower, G. H. (1975). Cognitive psychology: An introduction. In W. Estes (Ed.), *Handbook of learning and cognitive processes.* (pp. 25-80). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem in metacognition. In R. Glaser (Ed.), *Advances in instructional psyychology* (Vol. 1,). Hillsdale, NJ: Erlbaum.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (pp. 485-517). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. (1985). Interactive roles of knowledge and strategies in the development of organised sorting and recall. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (pp. 457-483). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11*, 671-684.

DETE (2001). *South Australian Curriculum, Standards and Accountability Framework.* South Australia: DETE publishing. Retrieved on September 15th 2003 from http://www.sacsa.gov.au

Elen, J., & Lowyck, J. (1999). Metacognitive instructional knowledge: Cognitive mediation and instructional design. *Journal of Structural Learning and Intelligent Systems, 13*(34), 145-169.

Eysenck, M. W. (1989). Human learning. In K. J. Gilhooly (Ed.), *Human and machine problem solving* (pp. 289-315). New York: Plenum Press.

Flavell, J. H. (1985). *Cognitive development*. (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39*(2), 93-104.

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

Hogan, K. (1999a). Assessing depth of sociocognitive processing in peer group's science discussions. *Research in Science Education, 29*, 457-477.

Hogan, K. (1999b). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching, 36*, 1085-1109.

Hogan, K., & Fisherkeller, J. (2000). Dialogue as data: Assessing students' scientific reasoning with interactive protocols. In J. J. Mintzes, J. D. Novak, & J. W. Wandersee (Eds.), *Assessing science understanding: A human constructivist view* (pp. 96-124). San Diego, CA: Academic Press.

Hogan, K., Nastasi, B. K., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*, 379-432.

Jacoby, L. L., & Craik, F. I. M. (1979). Effects of elaboration and processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 1-22). Hillsdale, NJ: Erlbaum.

Kerr, D. H. (1981). The structure of quality in teaching. In J. F. Soltis (Ed.), *Philosophy and education* (Vol. 1, pp. 61-93). Chicago: University of Chicago Press.

Kirby, J. R., & Woodhouse, R. A. (1994). Measuring and predicting depth of processing in learning. *The Alberta Journal of Educational Research, 40*(2), 147-161.

Lawson, M. J. (1980). Metamemory: Making decisions about strategies. In J. R. Kirby & J. Biggs (Eds.), *Cognition, Development and Instruction* . New York: Academic Press.

Lawson, M. J., Askell-Williams, H., & Murray-Harvey, R. (2003). *Teacher education students' knowledge about how class discussions help them to learn.* Paper presented at the 10th biennial conference of the European Association for Research on Learning and Instruction.

Lawson, M. J., & Hogben, D. (1996). The vocabulary learning strategies of foreign language students. *Language Learning, 46*, 101 - 135.

Martin, B. L., Mintzes, J. J., & Clavijo, I. E. (2000). Restructuring knowledge in biology: Cognitive processes and metacognitive reflections. *International Journal of Science Education, 22*, 303-323.

Marton, F., & Saljö, R. (1976). On qualitative differences in learning: I-Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Marx, R. W., Howard, D. C., & Winne, P. H. (1987). Students' perception of instruction, cognitive style, and achievement. *Perceptual and Motor Skills, 65*, 123-134.

McKeown, M. G., & Beck, I. L. (1990). The assessment and characterisation of young learners' knowledge of a topic in history. *American Educational Research Journal, 27*, 688-726.

Mintzes, J. J., & Novak, J. D. (2000). Assessing science understanding: The epistemological vee diagram. In J. J. Mintzes, J. D. Novak, & J. W. Wandersee (Eds.), *Assessing science understanding: a human constructivist view* (pp. 41-69). San Diego, CA: Academic Press.

QSR (1997). QSR NUD*IST (Version 4): QSR International.

Sellin, N. (1995). Partial Least Squares Modeling in research on educational achievement. In W. Bos & R. H. Lehmann (Eds.) *Reflections on educational achievement: Papers in Honour of T. Neville Postlethwaite* (pp. 256 -267) Munster: Waxmann.

Sweller, J. (1991). Some modern myths of cognition and instruction. In J. B. Biggs (Ed.), *Teaching for learning: The view from cognitive psychology* (pp. 71-83). Hawthorn, Victoria, Australia: Australian Council for Educational Research.

Watson, K. L., & Lawson, M. J. (1995). Improving access to knowledge: The effect of strategy training for question answering in high school geography. *British Journal of Educational Psychology, 65*, 97 - 111.

White, R. T. (1979). *Describing cognitive structure.* Paper presented at the annual conference of the Australian Association for Research in Education, Melbourne, Australia, November.

White, R. T., & Gunstone, R. F. (1980). *Converting memory protocols to scores on several dimensions*. Paper presented at the annual conference of the Australian Association for Research in Education, Sydney, Australia, .

White, R. T., & Gunstone, R. F. (1992). *Probing understanding*. London: Falmer Press.

Whitehead, A. N. (1942). *The aims of education: And other essays*. London: Williams & Norgate.

Winne, P. H. (1985). Steps toward promoting cognitive achievements. *The Elementary School Journal, 85*, 673-693.

Winne, P. H., & Butler, D. L. (1994). Student cognitive processing and learning. In T. Husen & T. N. Postelthwaite (Eds.), *The international encyclopedia of education* (2nd ed., pp. 5739-5745). Oxford, UK: Pergamon.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.

Winne, P. H., & Marx, R. W. (1977). Reconceptualising research on teaching. *Journal of Educational Psychology, 69*, 668-678.

Winne, P. H., & Marx, R. W. (1980). Matching students' cognitive responses to teaching skills. *Journal of Educational Psychology, 78*, 257-264.

Winne, P. H., & Marx, R. W. (1982). Students' and teachers' views of thinking processes for classroom learning. *The Elementary School Journal, 82*, 493-518.

Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376.

Wong, R., Lawson, M. J., & Keeves, J. P. (2002). The effects of self-explanation training on students' problem solving in high school mathematics. *Learning and Instruction, 12*, 233-262.

Zimmerman, B. J. (2002). Becoming a self-regulated learner. *Theory Into Practice, 41*(2), 65-70.

14

Future Directions for the Reform of Education in Oceania

G R (Bob) Teasdale
University of South Pacific

Dedications: *In the very early days of the Flinders University Institute of International Education (FUIIE), long before it was approved and launched, John Keeves came to me with details of a just-released UNESCO report on the future of education, Learning: the treasure within (Delors, 1996). He suggested that a careful analysis of the report might be a useful intellectual starting point for the new Institute. Little did I realise where this suggestion might lead. I suspect that even John may be a little surprised by the significance of its impact, especially on the work of many of our post-graduate students.*

The above story typifies the role that John has had within FUIIE. Far more than his wise and consistent role as chair of the Institute, and far more than his record of never having missed a meeting of its Board of Management, has been John's visionary intellectual leadership. He has a remarkable capacity to be at the cutting edge of new educational ideas, yet in a way that is deeply grounded in the best of the past, and to share his insights and wisdom with colleagues and students. This, I believe, has been his greatest contribution to the life and work of FUIIE, and certainly the one that I have valued most.

Abstract: *A new regional education project has commenced this year in Oceania, seeking to enhance the planning and delivery of basic education in 15 countries. This paper describes the project and its rationale, noting that is was developed by the Ministers of Education of the region under the aegis of the Pacific Islands Forum, and is accountable to them. It then reflects on the project from a broader international perspective, especially that of Learning: the treasure within (The Delors Report). The priorities of the Ministers of Education are in significant accord*

with current international theorising about education, and analysis of these broader perspectives is adding depth and focus to the conceptualisation of the Project.

Introduction

This paper will review new approaches to the strategic planning and implementation of education in Oceania, and reflect on these from the perspective of *Learning: the treasure within* (Delors, 1996). I use the name Oceania deliberately. We who occupy continents on the rim have tended to view the Pacific Ocean as a vast expanse of water dotted with tiny, isolated islands, their inhabitants disadvantaged by smallness and remoteness. Pacific Islanders are now rejecting this colonial assumption, arguing that they do not occupy "islands in a far sea", but "a sea of islands" (Hau'ofa, 1993:7). Their ancestors clearly did not view the sea as a barrier, but as their livelihood. They were seafarers who were equally at home on sea as on land. They lived and played and worked upon it. They developed great skills for navigating its waters, traversing it in their sailing canoes, and forming "… a large exchange community in which wealth and people with their skills and arts circulated endlessly" (Hau'ofa, 1993:9). In this way the sea bound them together rather than separating them.

The name Oceania captures this holistic sense of people sharing a common environment and living together for their mutual benefit. Many of the inhabitants of Oceania are attempting to reactivate this ethos, seeking ways to help and support each other, rather than constantly turning to the nations on their rim for aid and advice. It is a slow and uneven process, however, much hindered by regional politics, by the insistent pressures of globalisation, and by the continuing impact of colonialism. The latter has successfully divided Oceania linguistically, creating a significant gulf between groups of Anglophone and Francophone islands, and politically, with France and the United States still ruling their colonial empires in Oceania in ways that isolate their people from many regional fora and networks.

As a consequence of the above divisions this paper focuses only on those countries that are politically independent and therefore able to participate in the dominant political and economic policy grouping in Oceania, the Pacific Islands Forum Secretariat (PIFS): Cook Islands; Federated States of Micronesia; Fiji; Kiribati; Nauru; Niue; Palau; Papua New Guinea; Republic of the Marshall Islands; Samoa; Solomon Islands; Tonga; Tuvalu and Vanuatu. To this list should be added Tokelau, which is in the process of achieving self-government in free association with New Zealand, a similar status to that enjoyed by Cook Islands and Niue. Australia and New Zealand also are full members of PIFS, although they tend not to fit quite as comfortably within it, Australia especially so.

Founded in 1971, PIFS brings together heads of governments annually for dialogue and decision-making on regional policy issues, and is administered by a substantial secretariat based in Suva, Fiji. At its meeting in Palau in November 1999 there was considerable debate about human resource needs and the failure of most education systems to satisfy them, thereby perpetuating the region's dependence on highly paid personnel from the rim. Schools and their curricula were criticised for not providing relevant life and work skills, for being too focused on academic success in external examinations, and for not graduating young people who could become productive members of their own villages or urban communities. Accordingly the Forum

directed its secretariat to bring together the Ministers for Education of the region, asking them to deal with its concerns.

The Ministers eventually met 18 months later in Auckland, deliberating on what they referred to as 'basic education', which they defined as all educational provisions for children and youth, both formal and non-formal, except for higher education. The major outcome of the meeting was the development of the *Forum Basic Education Action Plan (FBEAP)* (PIFS, 2001), a short (9pp) but significant document setting out visions, goals and strategies for the future of basic education in Oceania. Its vision is clearly specified:

> Basic education as the fundamental building block for society should engender the broader life skills that lead to social cohesion and provide the foundations for vocational callings, higher education and lifelong learning. These when combined with enhanced employment opportunities create a higher level of personal and societal security and development.
>
> Forum members recognised that development of basic education takes place in the context of commitments to the world community and meeting the new demands of the global economy, which should be balanced with the enhancement of their own distinctive Pacific values, morals, social, political, economic and cultural heritages, and reflect the Pacific's unique geographical context. (PIFS, 2001:1-2)

The Ministers requested the PIFS secretariat to facilitate the implementation of FBEAP, and recommended that they themselves continue meeting on a regular basis to monitor and support this process. Following this first meeting, discussions took place with representatives of the European Union (EU) and a provisional agreement reached that funding for a project to implement FBEAP might be made available under its 9th EDF Pacific Regional Indicative Programme. By the time the Ministers came together for their second meeting in December 2002 these plans were well developed, and a sub-committee of Ministers was formed to finalise a submission. At the same time it was noted that UNESCO had made significant progress in implementing the goals of Education for All (EFA) in Oceania, all Forum countries having completed their EFA Action Plans. The Ministers made clear that they expected FBEAP and EFA activities and programs to be mutually supportive.

The sub-committee of Ministers, under the leadership of the Samoan Minister of Education, the Honourable Afioga Fiame Naomi Mata'afa, developed a proposal, in the form of a Project Financing Agreement, that was accepted by the EU for funding of €8 million over a five year period for a new project to be called 'Pacific Regional Initiatives for the Delivery of Basic Education', abbreviated to: 'The PRIDE Project'. The University of the South Pacific (USP) agreed to manage the Project on behalf of PIFS, and the New Zealand Government, through NZAID, agreed to join as a funding partner with an initial grant of NZ$5 million over three years.

During the latter half of 2003 the USP Institute of Education took responsibility for start-up of the PRIDE Project, including recruitment. I was appointed as Project Director in January 2004, commencing duties just in time to attend the third meeting of Ministers of Education in Apia at the end of that month, and receiving a comprehensive briefing from them of their expectations. Most remaining Project personnel were in place by April 2004, moving into a newly renovated building on the USP Laucala Campus. The Project was officially launched at USP by the Samoan Minister of Education on 14 May 2004 in conjunction with the first meeting of the Project Steering Committee.

The PRIDE Project

Essentially the Project is designed to implement the Pacific vision for education encapsulated in *FBEAP* in the 14 Pacific member states of PIFS, together with Tokelau. Its overall objective is:

> To expand opportunities for children and youth to acquire the values, knowledge and skills that will enable them to actively participate in the social, spiritual, economic and cultural development of their communities and to contribute positively to creating sustainable futures. (www.usp.ac.fj/pride, 2004)

To achieve this objective the Project seeks to strengthen the capacity of each of the 15 countries to deliver quality education through formal and non-formal means. The key outcome will be the development of strategic plans for education in each country. Ideally these plans will be developed following wide consultation with all stakeholders and beneficiaries, including parents, teachers, students, NGOs, private providers, employers and other civil society groups. The Project also will assist countries to implement their plans and to monitor and evaluate the outcomes. Capacity building activities will be provided for educators at national, sub-regional and regional levels. To further support these activities the Project will develop an on-line resource centre to encourage the sharing of best practice and experience amongst countries.

The Project Financing Agreement requires the Project team to work with countries to establish minimum benchmarks, principles and criteria that can be applied to their national education sector strategic plans. These benchmarks will be used constructively and collaboratively to review each country's education plan in order to deliver the best possible support and training, thereby facilitating educational planning and implementation processes. To maximise their utility as a tool for planning, it is important that the benchmarks are characterised by clarity and brevity. Guided by *FBEAP* and the Project Financing Agreement, the Project team has developed the following draft list, wherein 'The Plan' refers to each national education strategic plan.

1) **Living in a globalised world:** The Plan contains strategies for the systematic teaching of literacy, numeracy, ICT and English language, together with life and work preparation skills, thereby equipping all students to take their place in a global world with ease and confidence.

2) **'PRIDE':** The Plan builds on a strong foundation of local cultures and languages, thus enabling students to develop a deep pride in their own values, traditions and wisdoms, and a clear sense of their own local cultural identity, as well as their identity as citizens of the nation.

3) **Vulnerable students:** In order to ensure access and equity, the Plan contains strategies for the teaching of vulnerable students, including those from low socio-economic urban groups, those in remote and isolated areas, female students, those with disabilities, and school drop-outs and push-outs.

4) **Stakeholders:** The Plan shows clear evidence that it was developed using consultative and participatory processes in the broader context of civil society, including parents, students, private providers of education, NGOs, employers and other community and private sector groups.

5) **Articulation:** The Plan seeks to develop a more holistic conceptualisation of education, seeking to achieve more effective articulation: (a) between each level

of education: from pre-school/early childhood through to elementary/primary; from elementary/primary to secondary; and from secondary to TVET; (b) between education and the world of work, not only in the context of paid employment but also of self-sufficiency, self-reliance and self-employment; and (c) between formal and non-formal education.

6) **Financing:** The Plan has been carefully costed and is realistic in terms of current and projected levels of national budgets and donor funding for the education sector.

7) **Research and measurement:** The Plan is based on recent educational data that have been systematically collected, analysed, managed and reported.

8) **Monitoring and evaluation:** The Plan contains a monitoring and evaluation framework that allows outcomes-based judgments to be made about the effectiveness of education provisions at all levels, and in all areas of the curriculum.

9) **Training:** The implications of the Plan for the training of education personnel are addressed and effective training strategies developed, especially for the pre- and in-service education of teachers.

10) **National Development Plan:** The Plan is fully consistent with both the goals and strategies of the most recent National Development Plan.

Once a country has its strategic plan in place it is eligible to apply to the Project for funds to implement key priority areas. These national sub-projects are a key element of the Project, and will account for between 50 and 55 per cent of the total funds available. Funds cannot be provided for infrastructure, major equipment items, or recurrent expenditure, but only for the actual operating costs of action-research projects. In applying for grants, countries will need to demonstrate that sub-projects: (a) are in priority areas identified by *FBEAP*; (b) are able to be completed within the five year time-frame of the Project; (c) are likely to lead to exemplary outcomes that have potential benefits for other countries in the region; (d) have been designed, and will be implemented, in full consultation with all stakeholders; and (e) harmonise with other in-country donor activities and projects.

The task of coordinating in-country activities will be a substantial one. Each country therefore has been asked to nominate a key member of its Ministry of Education staff, preferably someone with responsibility for planning, as national Project focal point. The Project will assist with the provision of office and communications equipment, and where necessary with supplementary staffing support. Specific training programs will be provided for focal points, and regular regional meetings held using audio- and video- teleconferencing facilities.

Capacity building will take a number of forms, including regional, sub-regional and national workshops, study visits and training attachments between countries, and on-the-job training and support. Regional workshops will be limited to two per year, and will focus on key priority areas identified by Ministers of Education. At their request, the first workshop in 2004 is focusing on strategic planning methodologies, and on approaches to data collection, management and reporting in educational planning. The second will consider language policy and education, especially the use of vernacular languages in the early years, and effective approaches to vernacular literacy. Suggestions for subsequent workshops include; (a) the articulation between secondary school and TVET, and between secondary school, TVET and the world of work; and (b) the planning and delivery of effective non-formal education and its

articulation with the formal. Sub-regional and national workshops will be conducted in response to specific requests from national Ministries of Education.

In discussing the PRIDE Project with educators throughout the Pacific and beyond, a frequently asked question is: "How is it different? We have seen many donor-driven education projects and initiatives come and go: why is this one unique?" Their cynicism is justified. The history of educational aid in the Pacific, as elsewhere, is an ambiguous one with at least as many negatives as positives (see, for example, Luteru, 1991; Luteru & Teasdale, 1993; Teasdale & Teasdale, 1999). The present project, however, does have a number of unique features, and there is considerable optimism that it can achieve its goals in ways that others have not. These features include several important aspects:

1) The fact that the Project was designed and approved by the Ministers of Education: the process started with them, not with the donors. It was very clear at their third PIFS-sponsored meeting in Apia in January 2004 that Ministers saw this as their project, and were determined to guide and direct it according to their priorities. Subsequent meetings with individual Ministers have reinforced this view. The donors, in turn, have shown quite remarkable preparedness to allow this to happen.

2) The significance of the acronym: its choice clearly was deliberate, and reflects the wishes of the Ministers. Each country is being encouraged to build its education plans on a stronger foundation of local cultures, languages and epistemologies, thus enabling students to develop deep pride in their own values, traditions and wisdoms, and a clear sense of their own local cultural identity.

3) The strong emphasis on mutual collaboration and support: the aim of the Project is to help countries to help each other. Earlier projects brought consultants from outside the region, and therefore became donor-driven as they responded to donors' priorities and preferences. The PRIDE Project will source most of its consultants from within the region, and already has built up an impressive data-base of qualified people from Oceania. Furthermore, it will fund local educators to go on study and training visits to each other's countries, not to those on the rim and beyond.

4) The encouragement of consultative and participatory approaches to educational planning within each country: there is a clear wish to avoid top-down models of planning and policy-making, and a strong commitment to bottom-up processes involving parents, teachers, students, private providers, NGOs, employers and other civil society groups.

5) And finally, the fact that Ministers want the Project to promote a more holistic and lifelong approach to education, with effective articulation between sectors, and between school, TVET and the world of work.

Conceptualising the Project

The Project team has found it helpful to reflect on its work from a broader international perspective, especially that of *Learning: the treasure within* (Delors, 1996). The priorities of the Ministers of Education in Oceania as expressed in the Project Financing Agreement are in significant accord with current international theorising about education, and analysis of these broader perspectives is adding depth and focus to the conceptualisation of the Project.

From teaching to learning

Ever since the invention of mass compulsory schooling in the early years of the industrial revolution in Europe, the focus has been on the delivery of knowledge to children and youth by adults with the necessary training and/or community recognition. The architecture and routines of the school, and the content and processes of the curriculum, were primarily aimed at preparing the young to be compliant and productive workers. Education was almost entirely teacher-centred: the podium and blackboard at the front of each classroom facilitated control of students and delivery of knowledge. A system of examinations and reporting regulated progression through the school, and provided incentives to acquire knowledge and the formal credentials for having done so. These credentials in turn were linked to subsequent employment. The higher the credentials the more prestigious and well paid the job at the end. It was this system of education that was exported to Oceania during the colonial era, largely by well-intentioned Christian missionaries, and that has proven so resistant to change.

While the above is a very oversimplified account of a much more complex reality, it does highlight the view that the nineteenth and twentieth centuries, educationally speaking, can be characterised as those of the teacher. The teacher was central to educational discourse and process. This has been especially the case in Oceania, and still is in many if not most settings.

The current change in focus to that of the learner, as exemplified in the Delors Report, is highly significant. Even though many might argue that teaching and learning are simply opposite sides of the same coin, and essentially one and the same, the reality is that education is undergoing a profound transformation. The shift in power from teacher to learner is just one element of this. Another significant shift is from the acquisition of knowledge to learning how to learn. And a third is from viewing education as preparation for life and work to a much more holistic process of lifelong learning.

The fact that the Ministers of Education in Oceania have requested the PRIDE Project to encourage a more holistic approach to education, with an emphasis on lifelong learning, is fully in tune with global developments, and has substantial implications. The first challenge is to shift the mindset of teachers, and even more importantly of teacher educators. A second is to transform curriculum content and processes. And a third is to review the deeply entrenched system of external examinations that has maintained the pyramid structure so typical of third world education systems, and that leads to most children being pushed out of an increasingly selective school environment, with implications of failure and rejection.

Tensions and change

Jacques Delors, in his preface to *Learning: the treasure within*, identifies and discusses seven tensions that he believes characterise most learning environments in a rapidly changing world. He revisits these and adds further insights in a later paper (Delors, 2002). Among the tensions he identifies are several that have deep resonance with communities in Oceania, including the tensions between tradition and modernity, cooperation and competition, the spiritual and the temporal, the universal and the individual, and the local and the global.

In neither of the above documents does Delors elaborate on the idea of tension itself. One assumes he is not using the concept of tension in the sense of conflict between opposing factions or ideologies, the kind of tension that can lead to rivalry and war,

but is referring instead to a functional or positive tension. In the context of Oceania this is best explained using the analogy of guitar strings that need to be kept in a constant state of tension if they are to produce pleasing music. One of the tasks of the guitarist is to maintain a functional tension by regularly adjusting and readjusting the strings to ensure harmony. Likewise educators have the constant challenge of helping their learners to achieve a functional balance between the tensions confronting them.

The concepts of tension and balance also are relevant in educational policy and planning, and in curriculum development. Almost every educator I speak with in Oceania believes that the balance is wrong, that the global, the competitive and the temporal have a disproportionate influence in most learning environments. Once again, I find analogy a useful explanatory device. In the realm of visual arts, music, drama and dance in Oceania there are currently some remarkably creative initiatives. Individuals and groups within local communities are creating new forms of expression from the fusion of the traditional and the modern. The Oceania Centre for Arts and Culture at the USP Laucala Campus is playing a significant leadership role here.

By way of example, much contemporary music in Oceania represents a dynamic syncretism of the local and the global. It often has equal resonance with those who celebrate and enjoy the traditional as it has for those who prefer modern western music styles. In the realm of education, whether in policy, planning, curriculum or the classroom, I suggest we can strive for the same dynamic syncretism between the global and the local. Both are important. Young people need to grow up with the skills and confidence to live successfully in a globalising world. Yet it is becoming increasingly recognised in Oceania that they also need to grow up with a clear sense of their own local cultural identity, built on a strong foundation of their own culture and language, and with a deep pride in their own values, traditions and wisdoms.

The PRIDE Project has begun by exploring potential synergies between the global and the local in the strategic planning of education. The first regional workshop involves two participants from each country, all of whom work at a senior level as educational planners. In preparation for the workshop, they have been asked: (a) to identify from their studies and training those western theories and perspectives that they have been able to apply successfully in their own national settings; (b) to reflect on the traditional planning processes that have been used in their own cultures, or are still being used, and to identify the values, beliefs, wisdoms and/or epistemologies that underlie them; and (c) to come to the workshop prepared to share their insights, and to work in small groups to suggest planning processes that syncretise the best of the global and the local. The findings will be fully documented, and participants encouraged to apply any new-found ideas in their work as planners, and to evaluate their utility. In a future workshop we would like to use a similar approach with curriculum writers, especially from the perspective of curriculum process.

The four pillars of learning

One of the most widely recognised and discussed features of the Delors Report has been its notion of four pillars of learning: to know, to do, to be and to live together. While it has been criticised by some in Oceania, Thaman (1998), for example, arguing that it leads to the very conceptual fragmentation that the Report itself so strongly criticises, the idea that all learning is built on these four foundations seems readily accepted by most. The construction of many traditional homes and meeting places in Oceania, for example, is based on four large timber uprights, usually tree- or palm-trunks, one in each corner, these supporting the remaining structure. The idea

that each upright needs to be of similar size or scale in order to ensure structural strength and stability is readily transferred to education, and to the view that all pillars should receive equal emphasis in a child's learning. In reality, however, the representation of each pillar in most education systems in Oceania, as elsewhere, is far from balanced, with 'learning to know' and 'learning to do' occupying disproportionately large parts of the curriculum. As Delors (2002) himself acknowledges, these two pillars have long been self-evident, and are the dominant focus of most education systems.

The 'learning to be' pillar has posed particular challenges for educators. It is the least understood, and the least represented in curricula at all levels. Even though the idea achieved considerable recognition following publication of the 1972 UNESCO report of the same name (*Learning to be*, or the Faure Report), it had not become prominent in education discourse prior to release of the Delors Report. Basically, it has to do with the formation of identity, both individual and collective, and with the achievement of self-knowledge, self-understanding and self-fulfilment (Delors, 2002), and ultimately with the development of wisdom. The full recognition and implementation of 'learning to be' will require "… nothing less than a revolution in education that will be expensive in terms of time" (Delors, 2002:151). Nevertheless, Delors makes clear that we cannot afford to overlook this aspect of learning, for through it people are empowered to learn about themselves, and to become more fully human.

In a deeper way this pillar also has much to do with the nurture and development of spirituality, not just in a religious sense, but also through the broader quest for meaning in life and for explanations of reality, both individual and communal. It is interesting that secular education discourse – that of UNESCO and other international agencies, for example – is starting to emphasise the spiritual, and to advocate a role for education in the spiritual development of children and youth.

From a traditional perspective, this pillar, until the arrival of the Europeans, was a fundamental part of a holistic process of lifelong learning throughout Oceania. It seems that global thinking about education may be coming full circle, returning to the subjective and the spiritual, thereby allowing the peoples of Oceania to reaffirm the legitimacy of their own local ways of thinking, knowing and understanding. It thus reinforces the significance of the PRIDE Project objective, namely to expand opportunities for children and youth to acquire the values, knowledge and skills that will enable them to actively participate in the social, spiritual, economic and cultural development of their communities.

Conclusion

This paper has reviewed new approaches to the strategic planning and implementation of basic education in Oceania, and reflected on them from the perspective of *Learning: the treasure within* (Delors, 1996). Regional priorities are encapsulated in the objective of a new project, Pacific Regional Initiatives for the Delivery of Basic Education. These priorities are in significant accord with current international theorising about education, and analysis of these broader perspectives is adding depth and focus to the conceptualisation of the project.

References

Delors, J. (chair) 1996. *Learning: the treasure within.* Report to UNESCO of the International Commission on Education for the Twenty-first Century. Paris: UNESCO.

Delors, J. Conclusion. 2002 In: *Learniung throughout life: challenges for the twenty-first century.* Paris: UNESCO.

Hau'ofa, E. 1993. Our sea of islands. In E., Naidu, V., & Hau'ofa, E. (eds). *A new Oceania: rediscovering our sea of islands.* Waddell, Suva: University of the South Pacific.

Pacific Islands Forum Secretariat, 2001. *Forum Basic Education Action Plan (FBEAP)*. Suva: PIFS, 2001.

Luteru, P.H. 1991. Development, education and foreign aid in Pacific island countries. Unpublished PhD thesis, Flinders University.

Luteru, P.H. & Teasdale, G.R. 1993. Aid and education in the South Pacific. *Comparative Education*, 29(3), 293-307.

Teasdale, J.I. & Teasdale, G.R. 1999. Alternative cultures of knowledge in higher education in the Australia-Pacific region. In: F.E. Leach & A.W. Little (eds). *Education, cultures and economics: dilemmas for development.* New York, Garland.

Thaman K H. 1998. Learning to be. In: Haw G & Hughes P (eds). *Report of the UNESCO Asia-Pacific Conference on Education for the 21st Century*. Melbourne: Victorian Department of Education.

15

Students' Knowledge of Normal Swallowing: Tracking Growth and Determining Influential Variables

Ingrid Scholten
Flinders University

Abstract: *This study describes the pattern of growth of specific student learning and identifies factors that contribute to these changes. On four occasions 190 speech pathology students from four Australian universities completed a free-response task designed to assess students' learning of core information. Scripts were scored using a valid and robust key. Hierarchical linear modelling was used to measure change in students' understanding over time. Within-student predictor variables included time, whether or not the student had used particular instructional technology and details of clinical practica, and between-student variables included site, grade point average, age, and gender. The most predictive variable proved to be the provision of instruction. Students from one site performed better than did others. Multimedia improved performance, with increased duration of use resulting in higher scores. Male students and those with higher GPAs outperformed their peers at the outset of the study. The discussion of the findings addresses the implications for teaching.*

Purpose

The principal affirmed role of education at all levels, including higher education, is to promote learning, or change in what students know and can do. Such changes are generated and sustained by the activities and resources provided at schools and universities (Willett, 1989). Many teachers and educational researchers desire to improve these educational practices. Determining the "effectiveness of pedagogy and … educational innovation" (ibid, p.346) depends on accurate measurement of individual learning and investigation of the relationship between changes in learning and the educational practices. The correlates of growth are studied "to detect inter-individual differences in the growth parameters that are systematically related to some selected background characteristics of interest" (ibid, p. 413). Nowhere is it more important to understand the myriad of factors that contribute to student learning than in new areas of educational instruction.

Ironically, the largest recognised sub-specialty in the field of speech pathology is also the newest: namely, the management of swallowing difficulties (dysphagia). Practising speech pathologists require a comprehensive theoretical and functional knowledge base to underpin the safe and effective management of people with dysphagia, a potentially life-threatening condition. Because of the rapid expansion of knowledge in the area, this material was previously learned only in the clinical setting, after graduation (Edwards, 1995; Lazarus, 1996). The recent introduction of prescribed coursework within university speech pathology courses has generated only a short history of curriculum planning and development in this topic worldwide. The American Speech-Language and Hearing Association published a guide for curriculum development (ASHA, 1998), but individual syllabi prepared by university teachers have not been exposed to public or professional scrutiny and review.

This paper extends earlier work that determined speech pathology students' knowledge of the normal swallowing process, concepts that are central to developing an adequate knowledge of dysphagia (Scholten, 2000; Scholten, Keeves & Lawson, 2002). The present study investigates the growth in students' knowledge across four occasions that span the 15-week period of formal academic learning and into the following semester. It also assesses the influence on such learning of factors associated with both the student and the instructional environment. The purpose for such an investigation is that the findings might be applied in a targeted way to improve the delivery of teaching in this new and challenging area and, as a consequence, improve student performance. That is, it is not only important to be able to describe the pattern of change but also to be able to provide reasons for any differences, whether in terms of student characteristics or site-based differences. The methods used in this study should be able to be applied more broadly.

Method

One hundred and ninety undergraduate speech pathology students (183 females and 7 males) from four Australian universities participated in the study. Consenting students completed a free-response task designed to assess core topic knowledge. On four occasions, 0) before commencement of lectures in the relevant topic (although after a topic covering basic anatomy and physiology); 1) immediately after the lecture(s) on normal swallowing; 2) on completion of the topic in which the relevant material was covered; and 3) two months later, students were asked to respond to the trigger question: "*Describe the swallowing process in as much detail as possible",* a task that was completed in approximately ten minutes. Demographic information was

also collected to enable multilevel analysis to be undertaken in an effort to learn about the influence of background characteristics. The within-student level predictor variables of interest included testing occasion (TIMEs 0, 1, 2 & 3), use of the CD-ROM, *The Dynamic Swallow* (Scholten & Russell, 2000), and observation of relevant clinical experiences (PRAC). The variables at the between-student level included university attended (SITEs 1, 2, 3 & 4), grade point average (GPA: between 1.00 and 7.00, based on the Flinders University formula), AGE and GENDER. The scripts were collected by the respective teachers and returned to the investigator.

Scoring of Responses

Response scripts were collated and intermixed. No identifying information was visible to the investigator, who scored all responses. Scripts were scored randomly using a scoring key. Inter and intra-rater reliability had been established as above 0.87 (Scholten, Keeves & Lawson, submitted). The quality of students' knowledge, as revealed in their written scripts, was assessed using the SOLO Taxonomy (Biggs & Collis, 1982). The scoring rubric evaluated 12 key features of swallowing, each of which received a score of 0, 1, 2 or 3, depending on the quality of the response (Scholten, Keeves & Lawson, submitted). Rasch analysis was employed to determine a hierarchy of feature difficulty and scale the scores. Scaled scores were normally distributed (mean = - 1.47, standard deviation = 1.31; minimum = - 5.45 and maximum = 1.92; kurtosis = -0.25 and skewness = - 0.35).

Analysis

The questions that arise in this study were answered with hierarchical linear modelling (HLM) (Bryk & Raudenbush, 1992) which provided an integrated approach for investigating the structure and predictors of students' growth in knowledge of the normal swallowing process. HLM is able to describe the pattern of within-student change and the effects on that of between-student characteristics. Additionally, HLM can consider the interactive effects across these levels, overcoming the problems associated with conflation of individual and group effects. In this study changes over time in students' performance (within-student variation) and the influences of student characteristics (between-student variation) on performance formed two levels within the system being investigated. A 3-level model was not pursued because the present study was primarily a study of individual student learning and SITE was of interest only as a general influence on the learning. Students' Rasch scaled score on the test of knowledge of the normal swallowing process was the dependent variable.

Results

The results of the present study must be seen in the context of the initial Rasch analysis which highlighted those components of the swallowing process that students found easier to recall and those that were more difficult. Students typically provided simple answers rather than complete or elaborated responses. Generally students did not develop a deep and integrated understanding of the swallowing process (see Scholten, 2000; Scholten, Keeves & Lawson, submitted).

Descriptive Statistics

Data were available for a total of 190 third year students. There was a considerable withdrawal from the study over time. Prior to the topic commencement 181 students

took part, dropping to 160 participants immediately after instruction, with only 82 students remaining for the third and fourth occasions. Only 41 students completed the test on all four occasions.

In order to perform simple descriptive analysis the continuous predictor variables GPA and AGE were recoded. AGE was coded as either young (24 and under) or mature (25 or over), and students were "banded" into three groups according to their GPA: 1) between 4.00 and 4.99; 2) between 5.00 and 5.99; 3) between 6.00 and 6.99, with two outliers having GPAs below 4.00. Students' scores on each of the 12 key features of swallowing were summed to give a total out of 33 points. Figure 1 presents the overall mean scores for each testing occasion (5.69 +/- 2.29; 10.16 +/- 3.04; 11.51 +/- 3.85; 10.61 +/- 3.29). Detailed descriptive statistics using these raw scores are presented in Table 1.

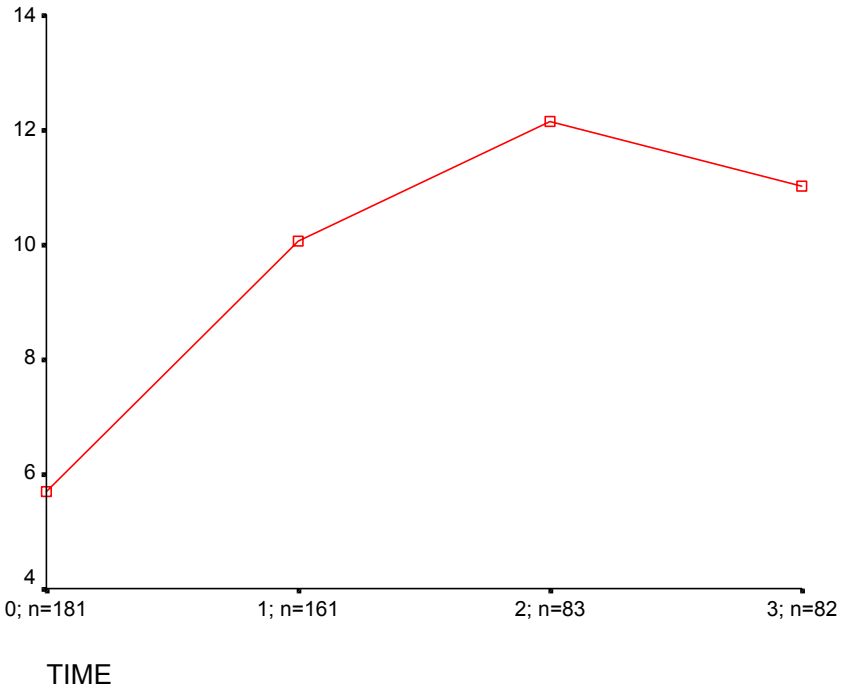


Figure 1. Overall mean scores associated with each data collection point (TIME).

Table 1 Response scores for each of the predictor variables by time

			TIME 0 N = 81	TIME 1 N =160	TIME 2 N = 82	TIME 3 N = 82	TOTAL N N =190
TOTAL		MEAN	5.69	10.16	11.51	10.61	190
		S.D.	2.29	3.04	3.85	3.29	
CD	0	MEAN	5.42	10.32	10.67	8.96	98
		S.D.	2.08	3.02	2.31	3.06	
	1	MEAN	5.55	9.94	10.87	10.70	38
		S.D.	2.05	3.38	4.64	3.01	
	2	MEAN	6.07	11.92	12.92	12.20	15
		S.D.	3.45	4.27	4.19	4.47	
	3	MEAN	6.00	12.22	15.28	13.38	10
		S.D.	3.40	3.11	2.75	3.66	
	Missing	MEAN	6.48	8.25	10.56	10.83	29
		S.D.	2.10	2.38	2.36	1.69	
PRAC	0	MEAN	5.42	10.16	10.66	11.07	110
		S.D.	2.31	3.20	4.06	3.83	
	1	MEAN	5.80	10.88	12.58	10.16	54
		S.D.	2.27	3.46	4.09	3.41	
	Missing	MEAN	6.58	8.32	11.06	10.82	26
		S.D.	2.06	2.47	2.44	1.74	
SITE	1	MEAN	5.25	10.50	9.44	9.19	70
		S.D.	2.16	3.08	2.55	3.26	
	2	MEAN	6.14	8.26	10.77	10.71	35
		S.D.	2.12	2.81	2.63	1.71	
	3	MEAN	5.31	9.46	11.18	9.07	53
		S.D.	2.11	2.61	4.34	1.82	
	4	MEAN	6.53	12.92	13.15	13.72	32
		S.D.	2.75	3.65	3.26	3.77	
GPA	3	MEAN	6.00	7.00	11.51	10.80	2
		S.D.	1.41	2.83	.00	.28	
	4	MEAN	5.22	9.12	10.52	10.06	32
		S.D.	2.24	2.22	3.23	1.94	
	5	MEAN	5.33	10.31	11.47	10.61	102
		S.D.	1.94	3.04	2.17	2.07	
	6	MEAN	7.02	11.07	12.41	11.32	40
		S.D.	2.57	3.49	2.96	2.47	
	Missing	MEAN	5.53	8.45	11.25	8.83	14
		S.D.	2.07	2.21	1.26	2.93	
AGE	Young	MEAN	5.49	10.22	11.56	10.56	159
	<24	S.D.	2.24	3.15	3.66	3.17	
	Mature	MEAN	6.75	9.76	11.00	10.92	30
	>25	S.D.	2.35	4.06	5.18	4.10	
	Missing	MEAN	5.00		13.00		1
		S.D.					
GENDER	Female	MEAN	5.64	10.13	11.41	10.46	183
		S.D.	2.30	3.20	3.83	3.13	
	Male	MEAN	6.71	11.20	15.50	14.67	7
		S.D.	1.80	5.54	2.12	5.51	

HLM Analysis

Difficulties with the administration of the test instrument at one site resulted in some missing demographic data. HLM is unable to estimate missing data at the between-student, or macro-level and researchers either substitute a group mean value or the overall mean value, or exclude cases from the analysis, the method chosen in the present study. HLM also requires that there be at least two data collection points for each participant. Consequently, the data pool was reduced to 151 cases. However, the estimation procedures employed take into consideration the reliability of the data with respect to the number of time points involved. Less weight is given to cases with fewer points and more weight to those cases with more data points, with maximum weight given to cases with all four data points available. Hence, although the data pool is reduced from the total data available, the results obtained can be interpreted with confidence, in so far as data have been included for analysis where the data exist.

Table 2 Descriptive statistics for HLM analyses

Level-1 Descriptive statistics N = 394					
	Code	Mean	SD	Min.	Max.
TIME	0, 1, 2, 3	1.09	1.05	0	3
CD	0, 1, 2, 3	0.44	0.84	0	3
PRAC	0, 1, 2	0.25	0.55	0	1
SCALED SCORE		-1.49	1.37	-5.45	1.92
INTER	0 1 1 1	0.64	0.48	0	1
INTOPIC	1 1 1 0	0.85	0.36	0	1
DELAY	0 0 1 1	0.30	0.46	0	1
Level-2 Descriptive statistics N = 151					
SITE 1	1 0 0 0	0.43	0.50	0	1
SITE 2	0 1 0 0	0.06	0.24	0	1
SITE 3	0 0 1 0	0.32	0.47	0	1
SITE 4	0 0 0 1	0.19	0.40	0	1
GPA	scale 0-7	4.97	0.66	3.00	6.40
AGE	in years	22	5.09	19	48
GENDER	0, 1	0.04	0.20	0	1

Because there are only four testing occasions (TIME 0, 1, 2 and 3), the analysis provides a reliable estimation of only three parameters for the effects of predictors at Level 1, restricting to some degree the examination of the learning process. Consequently, even though information on a number of variables was available, a parsimonious approach was required in considering those variables which might offer the most coherent explanation for change in student performance. Dummy dichotomous variables were constructed for TIME (testing occasions) and SITE to facilitate the exploration of their influence on student learning. Because the effect over time was not linear (with testing occasions spaced differentially at the participants' institutions, depending on when the focused learning took place and whether the content was covered in a dedicated topic or within a broader subject area) several additional dummy variables were created. These proxies for time were: INTER, which separated the first occasion from the others, reflecting the fact that at and after TIME 1 students had been exposed to intervention; DELAY, which separated TIMEs 0 and 1 from TIMEs 2 and 3, to capture the fact that some time had passed since the relevant lectures on normal swallowing; and INTOPIC, which

separated TIMEs 0, 1 and 2 from TIME 3, to reflect the fact that the final testing occurred two months after the completion of the topic.

In the first step of analysis a sufficient statistics matrix was created using software designed specifically for HLM analyses (Bryk, Raudenbush & Congdon, 1996). The descriptive statistics generated in this way are presented in Table 2. The maximum number of iterations for the final analysis was set at 5000, but considerably fewer iterations were required for the successive steps involved in the analysis of the data.

Evolution of Models

The fully unconditional (null) model was first calculated in order to partition variance into within-group and between-group components. This is used to estimate the variance explained as predictors are added to the model to reduce both the unexplained variance and the deviance. No Level-1 or Level-2 predictors are included in the null model and any Level-1 or Level-2 effects are a consequence of only the error terms. The intra-class correlation was low (0.08), indicating that the variance between students was small in comparison with the within-student variance. That is, students were more like other students in their performance at any one time than when comparing their own performance over time. This is consistent with the notion that students' performance changed dramatically as they learned the material in question.

The final model (or models) is arrived at successively by inserting variables of interest into the equations for both Levels 1 and 2 and at times conducting exploratory analyses to learn whether additional variables might enhance the model. The explanatory power of a model can be ascertained by calculating the variance explained and comparing this with that explained in the null model and by examining the change in deviance. The evolution of models in this study is summarised in Table 8, which depicts the variables included at Levels 1 and 2 for each model and provides values for variance estimates and the deviance which both help to evaluate the explanatory power of the model. The most powerful model is that with the lowest deviance. This paper presents three final models because the design limitations of the study would not permit all variables of interest to be modelled simultaneously. The three distinct final HLM models are shown in Table 9 to distil and summarise the main findings of the study. Several key models are described below.

Model 1: The effect of time

After testing the null model, the effect of TIME alone as a Level 1 variable was investigated using the least squares approach to estimate starting values as a reference point for judging the extent to which subsequent analyses refined the model. This simple model assumed a linear relationship between performance and time. Restricted maximum likelihood was the method of estimation. No variables were specified for Level 2.

Level – 1 Model

$$Y = B0 + B1*(TIME) + R$$

Level – 2 Model

$$B0 = G00 + U0$$

$$B1 = G10 + U1$$

Table 3 presents the final estimation of fixed effects with robust standard errors. TIME is significant, indicating that students' scaled scores change significantly across the four testing occasions. The large negative coefficient (- 2.28) for the intercept B0 is due to the very low scores prior to intervention. In Model 1 deviance equals 1214.24 and sigma squared equals 0.99 (see Table 8). Because only about 43 per cent of the variance is explained with this model additional variables were included in the Level 1 model to clarify further the contributions of different components across time. This first model, which assumes students' performance varies linearly with TIME, is limited. Inspection of mean scores on different testing occasions suggests a substantial improvement in performance following instruction, some continued growth during the semester, but a slight decline in performance after the semester (see Figure 1). In order to investigate this more complex relationship, other variables, such as INTER, INTOPIC and DELAY as described earlier, were introduced into the model.

Table 3 Fixed effects for HLM Model 1

Fixed Effect		Coefficient	Standard Error	t-ratio	*p* -value
For INTERCEPT1, B0					
INTERCEPT2,	G00	-2.28	0.08	-29.01	.000
For TIME slope, B1					
INTERCEPT2,	G10	0.72	0.05	15.04	.000

Model 2: Intervention, elaboration and forgetting

The parameters of a more extended model, Model 2, which attempts to depict the observed situation, are specified in Table 4. The effects of instruction and post-topic forgetting were tested using the proxy variables INTER and INTOPIC at Level 1. No Level 2 variables were modelled at this stage.

Level – 1 Model

$$Y = B0 + B1*(INTER) + B2 * (INTOPIC) + R$$

Level – 2 Model

$$B0 = G00 + U0$$

$$B1 = G10 + U1$$

$$B2 = G20 + U2$$

With this improved model the deviance was 1104.77, with sigma squared equal to 0.58 (see Table 8), indicating that there is still much to explain. However, this model offers a credible explanation of the process of learning for individual students, with intervention clearly having a significant positive effect (1.79) on performance. The coefficient for INTOPIC is not significantly different from zero (coefficient = - 0.07, p = 0.53), suggesting that this variable adds little to the model and there is no significant forgetting two months after formal teaching. The model could not be further specified at Level 1 as there were only three degrees of freedom at Level 1 available for the development of models.

Table 4 Fixed effects for HLM Model 2

Fixed Effect		Coefficient	Standard Error	t-ratio	*p* -value
For INTERCEPT1, B0					
INTERCEPT2	G00	-2.60	0.14	-18.18	0.00
For INTER slope, B1					
INTERCEPT2	G10	1.79	0.09	18.88	0.00
For INTOPIC slope, B2					
INTERCEPT2	G20	-0.07	0.11	-0.63	0.53

A final version of Model 2 was reached through an iterative process that substituted DELAY at Level 1 (Model 2A; see Table 8) following the removal of the non-significant INTOPIC (Model 2). This resulted in a model with more explanatory power than others, and it was developed further to include predictor variables at Level 2 for the intercept B0, namely SITE 3, SITE 4, GENDER, and GPA, as suggested in exploratory analyses (Models 2B and 2C; see Table 8). However, the addition of these variables at Level 2, especially SITE 4 for both B0 and B1, picked up variance at that level, leaving less variance to be explained at Level 1, and hence DELAY ceased to be a significant explanatory variable at this level (Model 2D; see Table 8). The accepted model is shown in Table 5.

Table 5 Fixed effects for HLM Model 2C

Fixed Effect		Coefficient	Standard Error	t -ratio	*p* -value
For INTERCEPT1, B0					
INTERCEPT2	G00	-2.76	0.09	-30.39	0.00
SITE 4	G01	0.38	0.19	2.01	0.045
GENDER	G02	0.554	0.15	3.68	0.00
GPA	G03	0.32	0.10	3.2	0.00
For INTER slope, B1					
INTERCEPT2	G10	1.63	0.11	15.13	0.00
SITE 4	G11	0.48	0.20	2.44	0.02
For DELAY slope, B2					
INTERCEPT2	G20	0.17	0.11	1.52	0.13

Model 2C again showed a significant effect attributable to instruction, with a further significant effect due to continued experience within the topic.

Level-1 Model

$$Y = B0 + B1*(INTER) + B2*(DELAY) + R$$

Level-2 Model

$$B0 = G00 + G01*(SITE4) + G02*(GENDER) + G03*(GPA) + U0$$

$$B1 = G10 + G11*(SITE4) + U1$$

$$B2 = G20 + U2$$

This variant of the model is considerably more powerful, with sigma squared equal to 0.45 and a lower deviance of 1058.07 (see Table 8). Seventy four per cent of the variance at Level 1 is explained by this model, but DELAY is no longer significant. When DELAY was no longer specified at Level 1 (Model 2D), the variance explained dropped to only 66 per cent. Deviance had also risen from 1058.07 to 1063.83, suggesting that Model 2C was the more powerful of the refinements even though DELAY was not statistically significant.

Model 3: The influence of multimedia-based supplementary instruction

Because the effect of using the instructional CD-ROM, *The Dynamic Swallow* (Scholten & Russell, 2000a) on students' learning was of particular interest this variable was included in a new model. The variable INTER was retained since it was shown to exert a strong influence on students' performance in previous models. In order to investigate the influence of the program, the variable CD was included at Level 1 in Model 3. Exploratory analysis had suggested that potential Level 2 predictors might include SITE 4, GENDER and GPA, which were included in this and subsequent models.

Level – 1 Model

$$Y = B0 + B1 * (INTER) + B2 * (CD) + R$$

Level – 2 Model

$$B0 = G00 + G01 * (SITE\ 4) + G02 * (GENDER) + G03 * (GPA) + U0$$

$$B1 = G10 + U1$$

$$B2 = G20 + U2$$

Table 6 demonstrates that the inclusion of CD resulted in a somewhat better model than the majority of earlier ones (sigma squared equals 0.54 and deviance equals 1060.00; see Table 8). Variance explained at Level 1 has risen from 43 per cent in Model 1 to 68 per cent in Model 3.

Model 3 indicates that students' starting point, the intercept, is influenced by both their GPA and GENDER, with students having better academic records doing better than their peers and, interestingly, with male students outperforming the females when other factors are controlled. In addition, students from SITE 4 had an initial advantage compared with students from other sites. There is a strong effect (1.67) for intervention (INTER), with students benefiting significantly from instruction about the normal swallowing process. A smaller, but equally significant effect for CD use (0.22) is also observed. Students who use supplementary multimedia improve their scores. The more time that is spent with this form of learning, the better is students' performance, increasing by 0.22 units for each level of CD use (0, 1, 2 and 3). Exploratory analysis suggested that nothing is to be gained by attempting to model interaction effects between other factors and CD use. That is, all students who used the CD improved, regardless of their opportunity for clinical observation, site, GPA, age, or gender.

Table 6 Fixed effects for HLM Model 3

Fixed Effect		Coefficient	Standard Error	t -ratio	*p* -value
For INTERCEPT1, B0					
INTERCEPT2	G00	-4.36	0.52	-8.41	0.00
SITE 4	G01	0.52	0.15	3.50	0.00
GENDER	G02	0.55	0.14	3.90	0.00
GPA	G03	0.31	0.10	3.14	0.00
For INTER slope, B1					
INTERCEPT2	G10	1.67	0.09	18.19	0.00
For CD slope, B2					
INTERCEPT2	G20	0.22	0.06	3.43	0.00

This model could not be further specified and tested because only four time points were included in the data set.

Model 4: The influence of clinical observations

The influence of clinical observations on students' performance was also of interest. Model 4 seeks to test the hypothesis that clinical experience should result in an improvement in students' scores as this experience should enable students to consolidate their dysphagia knowledge. Table 7 shows the effect of introducing the variable of relevant clinical experience (PRAC) at the within-student level. Intervention (INTER) is retained at Level 1 because of its significance in previous models. Exploratory analysis and further unreported models have suggested that SITE 4, GENDER and GPA would be significant Level 2 predictors of the Level 1 intercept and that SITE 4 might interact with intervention. After exploratory analysis, which suggested that no further variables would improve the model, Model 4 is presented.

Level – 1 Model

$$Y = B0 + B1*(INTER) + B2 * (PRAC) + R$$

Level – 2 Model

$$B0 = G00 + G01 * (SITE4) + G02 * (GENDER) + G03 * (GPA) + U0$$

$$B1 = G10 + G11* (SITE4) + U1$$

$$B2 = G20 + U2$$

Table 7 Fixed effects for HLM Model 4

Fixed Effect		Coefficient	Standard Error	t -ratio	*p* -value
For INTERCEPT1, B0					
INTERCEPT2	G00	-2.77	0.09	-30.53	0.00
SITE 4	G01	0.39	0.19	2.09	0.04
GENDER	G02	0.46	0.15	3.02	0.00
GPA	G03	0.32	0.10	3.26	0.00
For INTER slope, B2					
INTERCEPT2	G10	1.65	0.11	15.52	0.00
SITE 4	G20	0.53	0.20	2.66	.01
For PRAC slope, B1					
INTERCEPT2	G20	0.14	0.08	1.77	0.08ns

Although the effect of clinical observation (PRAC) is not statistically significant at 0.05, it might be regarded as marginal since the *p* value of 0.08 is not trivial and provides insight into the process of teaching and learning in addition to the products.

In Model 4 males and higher performing students do better than their peers and students from SITE 4 also have an initial advantage (0.39). The strong effect of classroom instruction is again evident (1.65). SITE 4 also interacts with this intervention, raising questions about the reasons for this improved performance. Variance is traded between Level 2 and Level 1 when variables are included in the Level 2 model and consequently the values for sigma squared and deviance must be observed closely to identify the effect of this trading of variance on the model. Variance explained at Level 1 (66%) is slightly less than the previous model. This model explained more than half of the variance, but was not quite as strong as that arrived at by including CD at the within-student level (Model 3). With a deviance of 1062.14 and similar sigma squared value of 0.58, this model, like Model 3, is considerably stronger than the null model.

Table 8 Overview of models

Model	Level 1 Variables of interest	Level 2 Variables of interest	df	Sigma Squared	Deviance
Null			150	1.72	1364.62
1	TIME		150	0.99	1214.24
2	INTER, *INTOPIC (n.s.)*		150	0.58	1104.77
2A	INTER, DELAY		150	0.44	1094.51
2B	INTER, *DELAY (n.s.)*	B0 = *SITE 3*; SITE 4; GENDER; GPA	146	0.45	1061.27
*2C	INTER, *DELAY (n.s.)*	B0 = SITE 4; GENDER; GPA B1 = SITE 4	147	0.45	1058.07
2D	INTER	B0 = SITE 4; GENDER; GPA B1 = SITE 4	147	0.58	1063.83
*3	INTER, CD	B0 = SITE 4; GENDER; GPA	147	0.54	1060.00
*4	INTER, ***PRAC (n.s.)*	B0 = SITE 4; GENDER; GPA B1 = SITE 4	147	0.58	1062.14

* Models reported in Table 9

** Marginal significance

Italic type indicates non-significant variables of interest

Discussion

A crucial question for educators relates to the course of learning over time. Hierarchical Linear Modeling was used to analyse growth in learning because it is able to describe the pattern of within-student change at the micro level and the effects on such change of between-student characteristics operating at the macro level. Additionally, HLM can consider the interactive effects across these within-student and between-student levels, overcoming the problems associated with melding of individual and group effects. In this study changes over time in students' performance (within-student variation) and the influences of student characteristics (between-student variation) on performance formed two levels within the system being investigated. Variables of interest were modelled in a systematic process that investigated whether or not they contributed to a model that explained more of the residual variance.

Table 9 Summary findings

Final Model	Within-student Predictors	Between-student Significant Predictors
2C	INTER, *DELAY (n.s.)*	B0 = SITE 4; GENDER; GPA B1 = SITE 4
	The strongest in the series of models indicates a persuasive effect of instruction (INTER) for all students. Students' starting point was influenced by both their GPA and GENDER, with students having a better academic history surpassing their peers and, interestingly, with male students outperforming the females when other factors were controlled. Students from SITE 4 also had an initial advantage. Model 2C indicates an interaction between SITE and INTER, with students from SITE 4 responding more to instruction than students from other universities. The effect of DELAY that had been important earlier in the Model 2 series was no longer significant.	
3	INTER, CD	B0 = SITE 4; GPA, GENDER
	There was an overwhelming effect of instruction (INTER) for all students. The effect of using the CD was also positive, suggesting that students who used instructional multimedia in self-directed learning improved their scores, with a stronger effect with increased duration of use. Model 3 again indicated that students' starting point was influenced by both their GPA and GENDER and that students from SITE 4 had an initial advantage.	
4	INTER, *PRAC (marginal sig.)*	B0 = SITE 4, GPA, GENDER B1 = SITE 4
	An alternate model suggests a similar pattern, as higher performing students, males and students from SITE 4 again did better than their peers. There was a strong effect of instruction (INTER) for all students and, like Model 2C, this model indicates an interaction between SITE and INTER, with students from SITE 4 responding more to instruction than students from other universities. Students who were provided with observational experiences concurrent with their classroom learning appeared to surpass their peers, although data from additional students would be required for this factor to be significant. Model 4 was slightly less powerful than Model 3.	

Influence of intervention

The first stage of the current study involved the investigation of changes in students' learning over time. The assessment of students' knowledge revealed a disappointing overall depth of understanding of these important concepts. However, it is encouraging to know that, despite the weak performance, intervention was a highly significant variable, with teaching clearly facilitating student learning without significant short term forgetting.

The second stage of the study helped to elucidate those features that impacted upon the pattern of students' learning over time. Some of these variables operated at the within-student level over time (CD and PRAC) and others focused on differences between students (SITE, GPA, AGE and GENDER).

Influence of CD

Hierarchical linear modelling supported the belief of teachers and experts that use of *The Dynamic Swallow* would exert a positive influence on student learning (Scholten & Russell, 2000b). Because use of the CD could have been a proxy for student motivation, it cannot be explicitly stated that students who use the program for longer periods of time benefit more from this supplementary study than students who spend less time in this self-directed learning, but this association warrants further investigation. It was not expected that there would be enduring effects on learning

with limited exposure of 30 minutes or less, a factor confirmed by the study. Future students should be informed of the potential benefits of spending at least an hour using the program.

The disparate CD-ROM uptake of between 4 and 67 per cent at various sites is of concern considering the positive results for students who did use this form of intervention. Some teachers are better at assimilating instructional technology than others and it is possible that this resource may need to be integrated into courses more effectively in order for its true potential to be realised. In view of the poor uptake at some sites and in recognition of the fact that teachers may not be familiar with the potential of the program, a teachers guide has been added to *The Dynamic Swallow*. The guide was prepared from a constructivist perspective and includes ideas for integrating the program with classroom teaching, including assessment tasks and classroom activities.

Influence of observational experiences

Speech pathology students are required to engage in a series of supervised clinical placements, as it is expected that a clinical practicum in the area will have a significant impact on learning. Although no students in this study were able to participate fully in clinical activities related to dysphagia, a quarter of the students had the opportunity to participate in at least one observational experience. An important finding of this study relates to the trend that even brief relevant observational experiences helped to consolidate learning.

The clinical observations were general in nature and did not require students to activate and apply specific knowledge related to the normal swallowing process to their observations. It is perhaps to be expected that both targeted classroom teaching and learning which took place across the semester and focused self-directed learning using instructional multimedia, would have a greater influence on students' consolidation of knowledge about the normal swallow than one or two observational visits in the field. Although this study did not control for the nature of students' observational experiences some types of clinical practicum are no doubt better at achieving certain aims than are others. It is essential for clinical educators to set clear learning objectives and to plan carefully and select the experiences they intend students to have, taking into account many factors such as educational aims, time available, student readiness and availability of resources. Arranging observational placements requires considerable time and effort on the part of many individuals and it is gratifying to learn that these endeavours appear to be rewarded by improvements in student learning.

Influence of site

Models 3 and 4 demonstrated that students from SITE 4 performed better than their peers from the outset of the study. Reasons for this can only be speculative in view of the data available, but future research aimed at revealing the contributing factors to better performance may provide the key to improving learning of this important content area, with possible implications for learning more broadly.

Students at SITE 4 did not have higher university entrance scores than students from other sites and, although class sizes were small, one other site had a similar class size. It is possible to speculate about the relative effectiveness of teaching and learning at this particular site. Students at SITE 4 may have received previous instruction in the prerequisite area of anatomy and physiology in a manner that resulted in considerably

more retention of content related to swallowing. It is also possible that teaching and learning of normal swallowing at SITE 4 is more effective than that at other sites, supported by the findings of Models 2C and 4 which indicated that students at SITE 4 improved more than their peers following instruction. Interestingly, students at SITE 4 did not demonstrate the decay in learning exhibited by some of their peers from other sites, which strengthens the notion that educational practices at SITE 4 may be more enduring.

Influence of aptitude

The effect of prior aptitude (GPA) is not surprising, with students having higher GPAs outperforming their peers from the outset. However, high ability students' knowledge did not grow at a greater rate than did others, suggesting that with increased resources, including time on task, weaker students could also achieve similar results.

Influence of gender

A surprising finding of the present study is the relative advantage for this material experienced by male students. The seven males who took part in the study appear to have retained more information about swallowing from earlier coursework and related study. There were no outliers to account for this phenomenon and these men were not in the group of highest ranked students. Several reasons might account for this male advantage, including those related to the nature of the content on which students' knowledge was tested. The small male advantage in spatial performance has been well-studied because of the implications for mathematics and science education (Berk, 1991). It could be argued that appreciating the normal swallow demands the ability to develop an internal visual model of the entire process, which is highly integrated and involves continually changing spatial, pressure and flow relationships, features that commonly arise in the traditionally male domain of mechanical engineering. Naturally, the findings of this study should be replicated with a more balanced sample. Nonetheless, the issue of gender relevant pedagogy, taking into account spatial ability, is relevant for all students, and if low spatial ability is a possible impediment to female learning in this domain, there is a case for instructional approaches that compensate for a lower ability in this area.

In summary, intervention and use of the CD-ROM, *The Dynamic Swallow*, are significant variables operating at the within-student level and there is a trend for opportunity to engage in relevant observational experiences to improve students' scores on a knowledge test of the normal swallowing process. Students' aptitude and gender significantly influence students' starting scores at the between-student level, as does being at SITE 4. Finally, there is an interaction between SITE 4 and intervention that is noted in two of the final models, with students at SITE 4 learning more than their peers. There are immediate implications for dysphagia instruction but further research is required in order to determine the reasons for improved performance for students at SITE 4.

References

ASHA (1998). Dysphagia Special Interest Division 13 *Graduate curriculum on swallowing and swallowing disorders*, Rockville, MD: American Speech-Language and Hearing Association.

Berk, L.E. (1991). *Child Development* (2nd ed.). Sydney: Allyn and Bacon.

Biggs, J.B. and Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy.* New York: Academic Press.

Bryk, A.S. and Raudenbush, A.W. (1992). *Hierarchical linear models: applications and data analysis methods*. Newbury Park, CA.: Sage.

Bryk, A.S., Raudenbush, A.W. and Congdon, R. (1996). HLM: Hierarchical Linear and Nonlinear Modeling with HLM/2L and HLM/3L Programs. (Version 4). [Multilevel analysis software]. Chicago: Scientific Software International.

Edwards, L. (1995). Where are dysphagia courses? *ASHA, 37 (*5), 9.

Lazarus, C. How well are we trained to evaluate swallowing? ASHA Division 13 Newsletter, November, 1996, 10-11.

Scholten, I. (2000). What students know and don't know about swallowing. *ACQ: ACQuiring knowledge in speech, language and hearing,* 2(2), 56.

Scholten, I., Keeves, J.P. & Lawson, M.J. (2002). Validation of a free response test of deep learning about the normal swallowing process. *Higher Education, 44,* 233-255.

Scholten, I. and Russell, A. (2000a). *The Dynamic Swallow*. Adelaide: Flinders University.

Scholten, I. and Russell, A. (2000b). Learning about the dynamic swallowing process using interactive multimedia. *Dysphagia, 15,* 10-16.

Willett, J.B. (1989). Questions and answers in the measurement of change. In E.Z. Rothkopf (Ed). *Review of Research in Education, 15.* Washington, D.C: American Educational Research Association.

16

Rasch Scaling and the Judging of Produce

Murray Thompson
Senior Secondary College, Inc

Dedications: *It has been both an honour and a privilege to have worked with, and learned from John Keeves. In many respects, I feel that I have been following in his footsteps. My first memory of him was when, in a somewhat nervous state, I presented my research proposal for my Masters thesis. Such occasions can be daunting and I felt that I had managed to defend my position against some difficult questions, but that my resolve was weakening. John showed genuine interest in my study and, rather than question my plans, was one person who offered some constructive suggestions and who recognised the relevance of my project to the classroom.*

Through this I recognised that John understood the trials of the classroom teacher. It was later that I found out that not only had John taught for many years at Prince Alfred College, where I was teaching, but that he had been at the same school with my father in 1942. Some years later, I asked him if he remembered my father. He thought briefly and then recited my father's full name, a name that very few people know, let alone remember after over fifty years!

Like me, John taught Physics at Prince Alfred College, and as for me, it was Prince Alfred College that whetted his appetite for educational research and led him to ask the questions that he has pursued ever since.

I shall be forever grateful to him for his kindliness, his patience and his genuine interest in my work. His willingness to give so generously of his time and hospitality, and share in his excitement for his work has been inspirational. The following paper is one example of the work through which he has guided me and I am pleased to present these findings.

Abstract: *The training of judges in sport and in industry is a challenging educational problem and recent developments in educational measurement can contribute to the resolution of this problem. The results of the judging of 58 exhibits from one class of a prominent competition were analysed using Rasch scaling. The raw scores were out of 20, graduated in steps of half a point, giving a total of 41 different scores. It was necessary to reduce this scale to an eight point scale, making all scores below 12.5 zero and compressing the remaining scores into seven scores from one to seven with each score combining two of the raw scores. It was found that this compression process does not adversely affect the analysis. This suggests that the traditional judging process tries to apply too fine a scoring system.*

Two programs were used to estimate the harshness of the judges and the level of the exhibits. They were the Quest program and the RUMM program. It was found that there was a good fit to the Rasch model and it was therefore possible to put the quality levels of the exhibits and the harshness of the judges on the one scale and make useful observations about the harshness and consistency of each of the judges which should prove useful feedback for the training and on-going professional development of judges. The RUMM program in particular offers very useful feedback to judges that shows just how they are awarding their grades along the continuum.

Comparison of the raw scores the Quest and the RUMM showed very high correlations, with evidence of a greater spread of the Quest scores than the RUMM scores

Introduction

Imagine a prestigious scholarship or prize to be judged on the quality of a single essay from several hundred candidates by a panel of judges, each assessing the performance of each candidate. What procedures could be put in place to ensure that the most worthy candidate wins the prize? Picture an Olympic Games Diving final, with a diver poised on the 10m tower and a line of judges ready to assess her performance for a possible gold medal. What safeguards can be put in place to ensure that the judging process is fair and not subject to bias of the judges? A row of 200 red wines, lagers, cheeses or olive oils stand awaiting the judgment of a panel of judges. How can the judging panel be certain that their decisions really reflect the quality of the products?

All of these situations have a number of things in common. They involve the complex process of the judgement of a single performance of each of the subjects, assessed by a number of raters. They represent situations of significant advantage to the successful candidate or producer of the product and it behoves the judges and the organizations that they represent to make the judging process as fair, as accountable and as transparent as possible.

As indicated above, this system of judging is not restricted to wine, but includes exhibits and ales, olive oils and dairy products. All of these industries are very important to Australia and all of them have huge export potential, not only for the products themselves, but also for the tourism that they have the potential to generate. An important ingredient in the long-term future success of these industries will be the integrity with which these industries regulate themselves and a vital part of this regulation will be any systems that are put in place to recognise excellence and reward quality products.

Cronbach (1964 pp 506 - 511) outlined some of the problems of making judgements. Referring particularly to supervisors making ratings of their subordinates, he outlined a number of sources of error in the judgement process. The first is 'generosity error' where the rater gives very favourable reports in all but the very worst cases and therefore the reports do not discriminate well. A second problem he cited was 'ambiguity' which is associated with varying interpretations of the criteria. A third source of error that was identified was 'constant error' or 'bias.' For example, one rater may not use the extremes of a scale while another may do so. A further concern is related to 'limited information' which may be available about the individual being judged and the final problem is the 'halo effect' in which an overall opinion might obscure some serious undesirable traits. He suggested that the reliability of ratings can be improved by combining the ratings of several judges. An alternative is to keep records of the ratings of a particular judge to establish and therefore correct any errors. The use of 5- to 7-point scales was suggested as a useful strategy that directs the rater to the kinds of deviation being explored.

Wolf (1997) discussed the use of rating scales, suggesting a number of ways in which their use can be improved. These included the use of multiple raters, and the training of raters. The use of Item Response Theory Methods was also suggested as a way of assigning the values of the scale.

Over the years, there have been developed a number of systems of checks in the judging processes used to rate various products. One particularly common practice is to include several samples more than once to test the consistency of the judgement at various times during the proceedings. However, there is no system in place to allow, either for the huge numbers of tastings required or the other problems of integrity involved. In diving, the highest and lowest scores are removed from the final score to reduce the spread of scores in order to eliminate outlier judgements.

Rasch Scaling

Rasch Scaling has been used in educational measurement to overcome similar difficulties to those described above. Bond and Fox (2001) have outlined the use of Rasch model in a range on measurement problems. Central to the Rasch model is the employment of a unidimensional scale that is used to define both the performance of the students taking a test and the difficulty of the items or questions in the test. This scale is graded in logits.

In Figure 1, the logit scale is shown on the left, ranging from –3.0 to 3.0. This scale is commonly calibrated to make the average difficulty of the items 0. The performance of the students taking the test is represented by an x and the difficulty of each of the numbered questions or items is shown on the right. In the diagram, Item 20 has a difficulty of 1.00. This means that a student whose ability is also rated at 1.00 has an equal chance of getting Item 20 right or wrong.

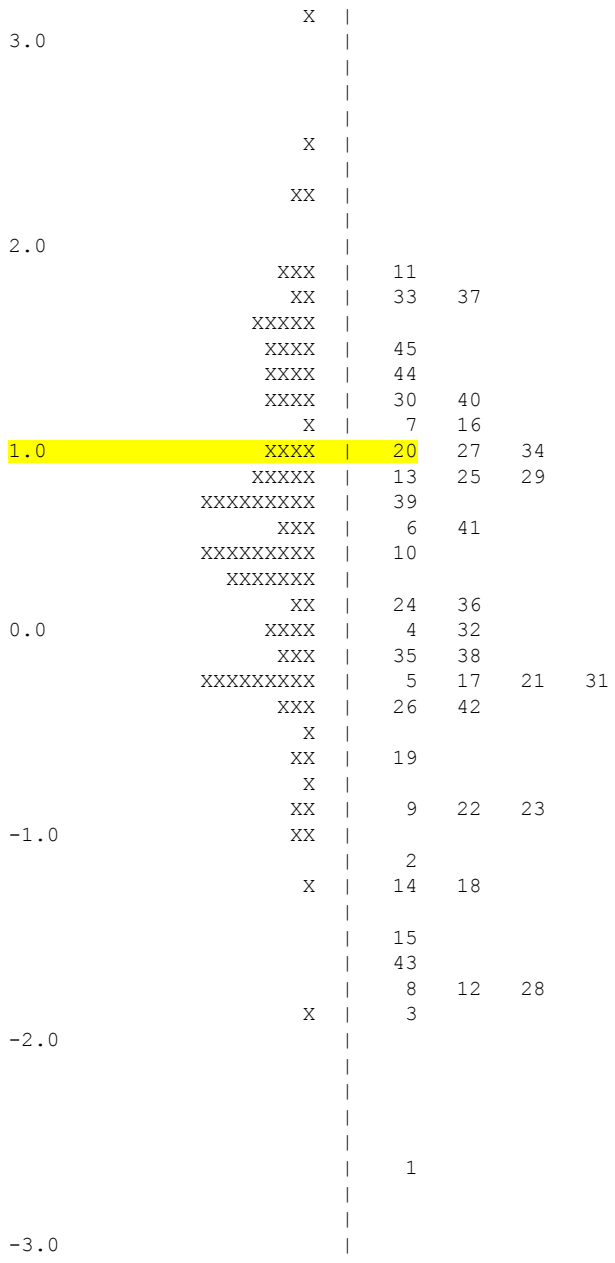


Figure 1: The performance of a group of students plotted on the same scale as the difficulty of the items or questions in the test.

These ideas have been extended to incorporate more complex testing procedures, including allowing for various levels of performance in each question. Thus each item may have several grades attached to it and students may be awarded a score in the range 0-7 for their performance on each question. This is known as the partial credit model and the difficulty of each item is subdivided into several levels. Thus for a particular item it is possible to describe the difficulty in the following manner. If an item has 8 levels, 0-7, then a threshold between two of the levels is set in calibration as that scale distance between two grade scores. Thus, on a 0-7 scale, Threshold 4 would be the level at which a score of either 3 or 4 is equally likely to be awarded. Above this threshold level a 4 is more likely and below this threshold a 3 is more likely.

Once a set of test scores has been Rasch scaled, it becomes possible to use the scales to make important comparisons and connections. For example, it is possible to equate two different tests to one another and therefore make meaningful comparisons between students who take different tests. Similarly, it is possible to make comparisons between markers of essay questions so that essay scripts can be marked by a team of markers, without the need for every member of the team to mark every script. One of the important issues that can be explored using Rasch scaling is the question of judge bias.

Rasch scaling and judging

Recent advances in educational measurement techniques have allowed some very worthwhile analysis of testing. Of particular interest is the application of Rasch Scaling to the area of judgment. This has been used to determine the harshness of essay markers in order to standardise the marking procedures as well as to determine the level of difficulty of questions and the abilities of the students.

Linacre (1999) discussed the use of Rasch models in the measurements of judgements.

> A Rasch model analyses capitalises on the inevitable judge disagreement to construct a sample-free, objective and linear measurement continuum. (p 244)

Liancre (1999) suggested that in a judging system involving the Rasch model, the only requirement is that there must be a connected network of each rater, each candidate and each assessment item.

Wu, Adams and Wilson (1996) discussed the analysis of rater effects in their *ConQuest* software manual. Andrich (1999) has applied such analysis to equate the marks given in essay questions. In this study the effect of question choice on the final mark was explored and a means of equating the marks in several different questions. Thus the analysis was used to compare the marks of students who have selected one question that is seen to be more difficult than another. The point was made that with a minor variation in design, it would be possible to explore the effect of the raters, rather than the difficulty of the question.

Bond and Fox (2001) discussed judged sporting performances, citing particularly the work of Looney (1997) who used Rasch scaling procedures to establish Eastern Block bias in the Free Skate event the 1994 Winter Olympics.

The purpose of this analysis is to investigate the application of educational measurement techniques to some of the problems and challenges associated with the judging of these sorts of products. These problems include:

- variable judge harshness or difficulty,
- judge consistency,
- issues associated with the large numbers of entries,
- training of judges and giving appropriate feedback.

The data to be analysed in this study comes from an international competition for the judging of a particular type of product. The 58 exhibits were judged by each of eight judges. Each judge rated the exhibits out of a total of 20 points with steps of ½ point. Thus there are 41 possible scores that can be awarded by the judges. In practice, the range that is used by the judges is very much less than this. From these exhibits, the lowest score was 7 and the highest was 18.5. It follows that in practice there were only 24 scores that were used by the judges. As well, it was found that scores at the extremes of these margins were rarely used, if at all by some individual judges. In the competition, there is no allowance made for the difference in the harshness of the various judges and that all of the judges were assumed to rate equally.

Problems with the Traditional System

Leniency and harshness

In the traditional system no account is taken of the leniency of the judges. For example, one judge may simply award all exhibits a point higher than the others. Similarly, another judge may be more harsh and so award to all exhibits scores lower than the other judges. This in itself may be seen to even out, but for the sake of consistency between judges, it is desirable for judges to be given clear feedback about how they are awarding points compared with one another on a clearly defined and easy to interpret scale. The final points then can be weighted to take into account the variation in judge harshness.

Consistency

The question of judge consistency is also a difficult one. Allegations have been made which suggest that in some competitions judges associated with particular business interests recognise their own products and award them high points and downgrade the points of direct competitors, (White, 2000). The traditional system has no mechanism for examining the level of consistency of the individual judges. Quite apart from the concern of deliberately rating a particular product up or down, individual judges would be greatly assured by evidence of their consistency. It would be in effect a form of quality control for their work. This problem becomes worse as judges are required to judge more and more exhibits. The confusion and tiredness of the palate must become a concern. With the current analysis of 58 exhibits, the levels of concentration and memory required combined with the confusion of flavours must make this a very demanding task. Even the most experienced judge of a high level of integrity must have concerns about coping with this onerous responsibility.

The Application of Rasch Scaling to Judging

Of interest to judges is the question of how they award their scores for exhibits compared with how their colleagues do. Also of interest is consistency of each judge. Under the proposed model, the judges take a role equivalent to items or questions and the exhibits have a role equivalent to the students. The analysis seeks to determine the level of each of the thresholds for each of the judges and the performances of each of the exhibits, placed on the same scale for easy comparison. The 41 point scale used by judges presents many problems, particularly as most of these scores are not typically used by the judges and, even within the usable range, scores on the extremes are rarely used. Moreover, the computer programs require that a sufficient number of exhibits are associated with each score level for effective calibration and estimation of the threshold levels. Accordingly, the scores were compressed according to the following scale into an eight point scale prior to analysis. This is shown in Table 1.

Table 1 Conversion of the raw scores to an 8 point scale

Score awarded by judge /20	Converted to 8 point scale	Frequency (total 464 = 8 x 58)
≤ 12.5	0	68
13, 13.5	1	43
14,14.5	2	65
15,15.5	3	109
16,16.5	4	95
17,17.5	5	65
18, 18.5	6	19
19, 19.5	7	0

Figure 2 shows the frequency histogram of the scores after they had been transformed to the 0-7 scale. Of particular interest is the generally bell-shaped curve reflecting an essentially normal distribution of the scores.

It is readily seen that this scale compresses the scores at the lower end but retains most of the sensitivity at the upper end where the scores tend to be crowded and where greater discrimination is required. These transformed scores can then be analysed using a variety of Rasch analysis programs. The first program to be used was the QUEST program that estimates the performance of the exhibits and the harshness of the judges using a joint marginal maximum likelihood estimation procedure. This program can be applied to a scale calibrated in a situation such as is presented by the judging of a class of exhibits. This analysis can be carried out in a number of ways so that its application can be explored.

In the situation of calibrating a scale for a group of exhibits, the exhibits are the students and the items become the judges. It thus becomes possible to rate the performance of each exhibit, the harshness of each of the judges by identifying where on the scale their scores lie and the consistency of the judges on scaling on a unidimensional scale.

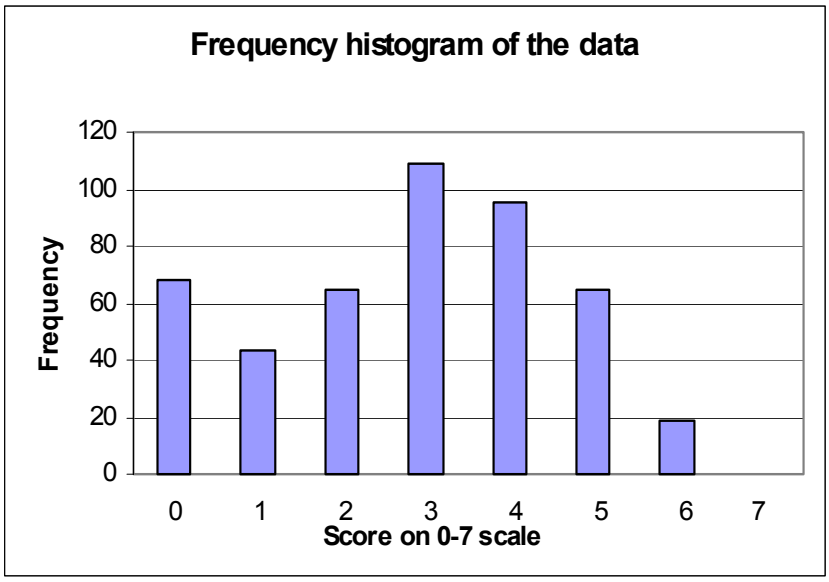


Figure 2 Frequency histogram of the data

Scale calibration

The data were first analysed using the Quest analysis program and a partial credit model. Figure 3 shows the single scale used to rate both the performance of the exhibits themselves and the harshness of the judges. On the left, each of the exhibits is represented by x and it can be seen that the display indicates a well-shaped bell-like curve as would be expected with a normal distribution. On the right are displayed the thresholds in the scale levels for the various judges and between the eight different score levels. These thresholds are the points along the difficulty scale at which the individual judge selects the next highest score. For example, 1.1 is the level above which Judge 1 would award 1 and not 0. Similarly 5.4 is the level at which Judge 5 would is equally likely to award 4 or a 3. Each threshold level is then the change-over scale level for each of the boundaries between the 8 possible scores. Note the differences in scale distance between threshold levels for different judges. Since none of the judges used a score of 7, the threshold value between 6 and 7 has not been estimated for any of the judges. Some judges, for example Judge 2, did not award a score of 6 and so the threshold between 5 and 6 has not been estimated.

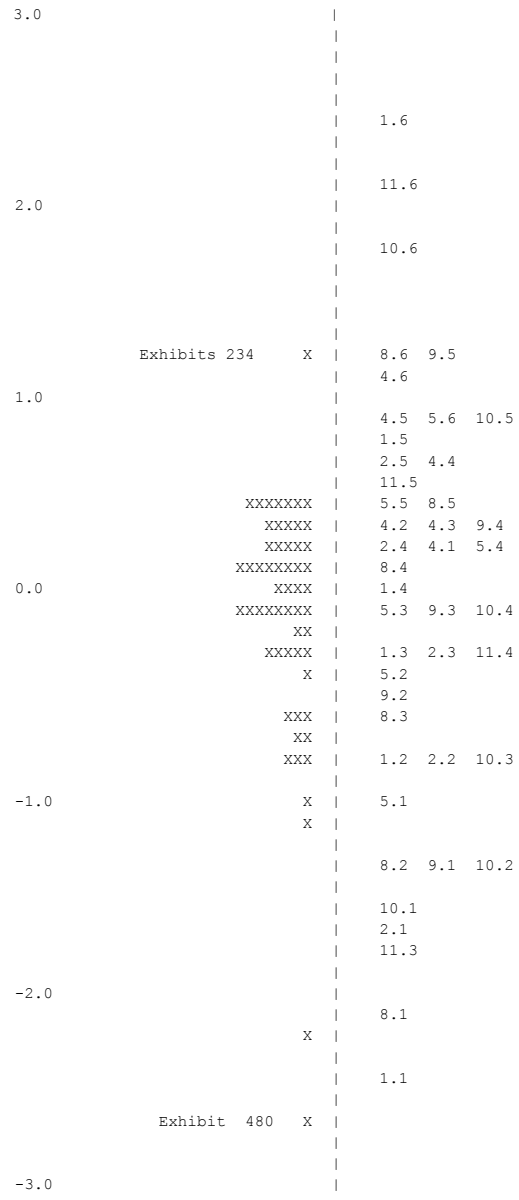


Figure 3: The performance of the exhibits and the thresholds of the judges, plotted on the same scale.

An estimate of the consistency of each of the judges can be obtained using the Infit Mean Square diagram, shown in Figure 4. In this diagram, it is desirable for the values estimating the degree of fit of the judges to the unidimensional scale to be within the dashed lines on either side. The ideal is 1. Deviation to the right indicates that the particular judge has been inconsistent, perhaps by awarding an exhibit with a low score when the other judges have awarded a high score. It can readily be seen that Judge 4 has been the most inconsistent, although this judge is within the acceptable limits indicated by the dotted vertical line. Deviation to the left indicates a higher than might be expected degree of consistency. This is particularly the case for

Judge 2, who is so consistent with the rest of the group nothing much new is being added by his opinion. The actual values of the infit mean squares for each judge are shown, along with the discrimination indices for each judge are shown in Table 2.

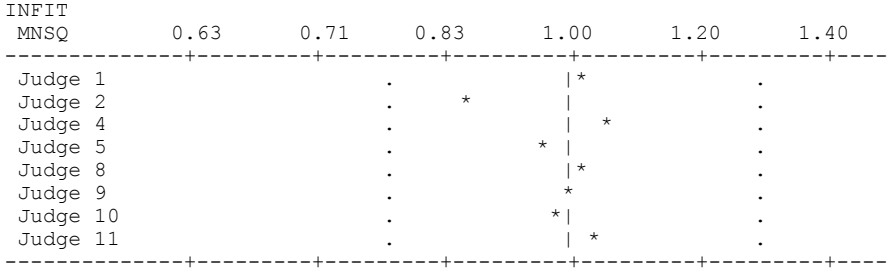


Figure 4: Infit Mean Square data for each judge

Table 2 The Infit Mean Square values and discriminations for each of the judges

Item	Judge	Infit Mean Square	Discrimination
Item 1	Judge 1	1.03	0.58
Item 2	Judge 2	0.84	0.68
Item 3	Judge 4	1.09	0.41
Item 4	Judge 5	0.94	0.64
Item 5	Judge 8	1.02	0.56
Item 6	Judge 9	0.99	0.58
Item 7	Judge 10	0.97	0.55
Item 8	Judge 11	1.06	0.41

In order to summarise these results, it is noted that the judging process does fit the Rasch model well and that the exhibits and the judge harshness can be placed on a single scale which enables estimates to be made of the performance of each of the exhibits and the harshness of each of the judges. Further, it is possible to examine the consistency of each of the judges. At this stage it is possible to comment upon each of the judges by referring to Figures 3 and 4 and to Table 2.

Judge 1 shows a good spread of score across the complete scale, with reasonable spacing of the threshold values. The infit means square value of 1.03 is very nearly ideal, representing a good measure of consistency. Moreover, Judge 1 has a relatively high discrimination of 0.58. Thus, according to this analysis, Judge 1 has assigned the scores with consistency and has used the available range well.

Judge 2 shows a good spread of scores, although since this judge has not used scores of either 6 or 7, the upper thresholds do not appear. Although very consistent, this judge does tend to be harsh, reluctant to award higher scores.

This QUEST analysis shows Judge 4 to be the most inconsistent, but more telling is the extreme bunching of the threshold scores between 0.31 and 1.16. In other words this judge discriminated badly showing very little difference in scores between good and bad exhibits. Not only that, this judge also tended to be quite harsh. In many ways, instead of a steady gradation of scores through the range, there appears to be a sudden jump over a very narrow band.

Judge 5 showed good consistency, with the infit mean square value being very low. The thresholds show quite a good spread, although there is some bunching up of scores in the upper levels. Judge 8 seems to be the model judge, with a good spread of scores through the range, a high level of consistency and good discrimination. Judge 9 is likewise a very good judge with good spread and consistency, although a little harsher than Judge 8, certainly in the upper range where no scores of 6 or 7 were awarded. Judge 10 is rather similar, showing good spread and consistency, but perhaps a little more leniency. Judge 11 did not use scores of zero or one at the lower end, or a score of 7 at the upper end. Nonetheless, Judge 11 showed a reasonable spread of scores within the remaining range, but was not as consistent as most of the other judges and showed more leniency than most of the other judges, with each level tending to be below the levels of the other judges.

The results for the exhibits – The case estimates

The results for each of the judges are brought together to estimate the performance of each of the exhibits. These are the results of the combination of the measures given by each of the judges that take into account their consistent differences in harshness.

Table 3 The Ratings of the top 15 exhibits using the various rating systems

ID	Total Score	ID	Raw Score	Max Score	Estimate or scaled score
234	273	234	37	44	1.23
449	259	466	29	44	0.51
466	257	A43	29	44	0.51
A43	257	220	29	44	0.51
269	256	449	28	44	0.44
38	255	269	28	44	0.44
220	255	A38	28	44	0.44
404	255	404	28	44	0.44
M00	255	M00	27	44	0.38
285	253	A75	27	44	0.38
A75	252	307	27	44	0.38
307	252	256	27	44	0.38
256	251	A70	27	44	0.38
498	250	285	26	44	0.31
442	249	A24	26	44	0.31

Table 3 shows the top 15 exhibits in order of their scores, both using the total score and the calculated score using the scaling system. The total score is the sum of all of the judges' scores out of 20, multiplied by 2 to remove half points. The raw score is the sum of all of the judges' scores after conversion to the 0-7 point scale. The maximum score is theoretically 8 x 7 = 56, but since none of the judges used a score of 7 and scores of 0 or 1 or 6 were not used by a number of the judges the maximum score is stated as 44. The final column is the Rasch scaled score as estimated by the QUEST program. It can readily be seen that the raw score and the Rasch scaled score

result in the same order of the exhibits. The slight variation in the order of the exhibits with respect to their total scores has been a result of the very slight adjustment made in the conversion to the 8 point scale, with a small loss of sensitivity in the scale.

The fit of the exhibits

Just as the infit mean square statistic indicates when an item, in this case the judge, is being inconsistent, it can also happen that an individual exhibit may be judged inconsistently by the group of judges, where for some reason, there may be some disagreement about the individual exhibit. The outfit statistics indicate when there is a significant disagreement between judges. In analysis of the scores of all judges, only five exhibits seemed to have some level of disagreement associated with them. The fit statistics for these five exhibits are shown in Table 4.

Table 4 The exhibits that do not fit into the Rasch model.

ID	Estimate	Error	Infit MNSQ	Outfit MNSQ	Infit t	Outfit t	Judges' scores
311	-0.67	0.31	1.86	6.17	1.4	2.78	1 0 3* 1 4 1 3 2
A60	0.12	0.26	2.51	2.17	2.35	1.71	4 2 5* 1 6 1 3 3
510	-0.31	0.28	1.47	2.37	0.96	1.57	2 1 3* 2 3 0 3 5
291	-0.09	0.27	1.89	1.99	1.57	1.42	5 1 1 1 4 0 6 4
371	0.05	0.26	1.36	1.74	0.81	1.22	2 5 0 3 5 5 2 2
M00	0.38	0.25	2.41	1.62	2.42	1.11	3 4 6* 1 2 5 4 4

These exhibits that do not seem to fit warrant further investigation. For the exhibits 311, A60, 510, and M00, the problem appears to be related to Judge 4. Judge 4 had an average score of 0.91 for this class of exhibits and indeed 72 per cent of the scores awarded by this judge were 0. In each case Judge 4 awarded quite high non-zero scores for these exhibits. The other two exhibits, 291 and 271 showed quite a large variation in scores across the judges. This would indicate some confusion among the judges on these exhibits and these would provide an ideal opportunity for judges to discuss and reflect.

Summary of the Quest analysis

Thus it has been shown that the QUEST program can be used to place the performance estimates of the exhibits and the harshness levels of the judges on a single Rasch Scale using a partial credit model. Moreover, this scale can be used to examine the consistency of each of the judges and provide valuable feed-back for each judge. It can readily be seen that this may prove to be of value for judge training purposes.

To summarise, the scores from each judge out of a possible 20 points have been compressed into an eight point scale and these compressed score have been shown to fit the Rasch model well. This has allowed a ready comparison of the judges, both in terms of their harshness and their consistency. The analysis of these results provides excellent feed-back for the training of the novice judge. Thus the Rasch scaling system provides a ready means for the analysis of such results.

The RUMM Program Analysis

Analysis of the judges

An alternative means of estimating the Rasch scale scores is the RUMM program, which uses a pairwise conditional estimation procedure. The data were analysed using the RUMM program and it was found that there was good fit of the data to the model. This analysis allows for the provision of some valuable feedback of the judgment process. The item characteristic curve shows the eight point score against the location of the exhibits on the logit scale. The graph also shows group mean scores. The exhibits are divided into groups, in this case eight, and the means of the exhibit locations and the expected scores are calculated. These group mean points are represented as the dots on the graph of expected score against exhibit location. The idealised curve according to the Rasch model is plotted as well and a simple representation of how well the judges conform to the model is given by how closely the points lie to the curve. In addition, the residual and Chi square statistics provided give an indication of the consistency of the particular judge. The RUMM analysis also provides category probability curves. In these curves the probability of a particular judge awarding each level from 0 – 7 is plotted against the Rasch scaled logit score. Each level is plotted in a different colour. Clearly, in this graph, the ideal is to have each level being the most likely score awarded in order across the range, with each grade being the most dominant score for a section of the logit range. If the graph of a particular judge does not do this, it may reflect inconsistency or that some scores are under-used. The category probability curves for Judge 1 are shown in Figure 5.

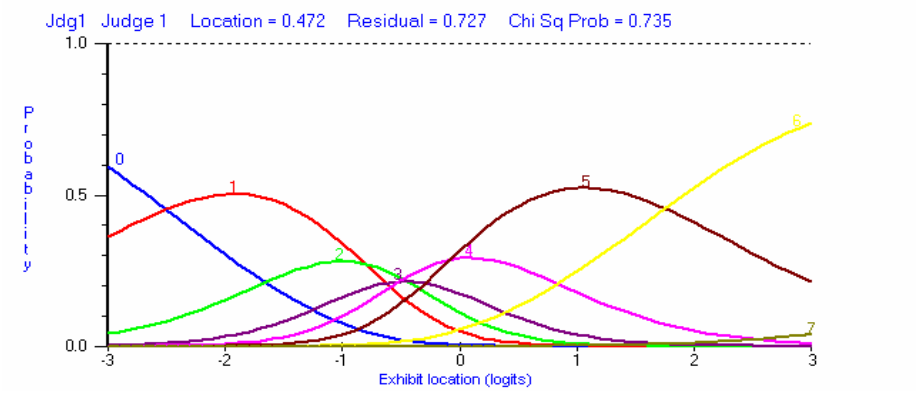


Figure 5: The category probability curves for Judge 1.

These curves indicate an emphasis on the extreme scores used by Judge 1, with scores of 2, 3, and 4 being quite unlikely compared with 0, 1, 5 and 6. The location of this judge is 0.472, which is on the more demanding side and the chi square probability suggests that this judge has been consistent. The item characteristic curve for Judge 1 is shown in Figure 6.

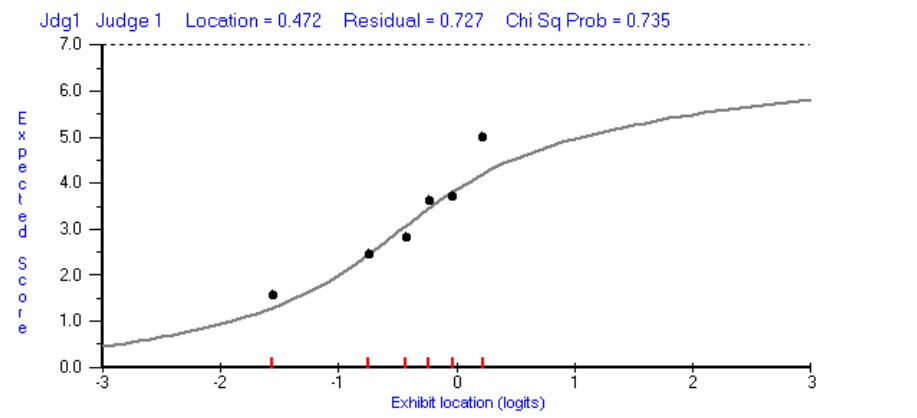


Figure 6: The item characteristic curve for Judge 1.

The reasonably close placing of the group centred mean points, particularly near the centre of the curve shows consistency, but there is some scatter at the extremes of the scale, reflecting the overuse of the lower and higher values. The category probability curves for Judge 2 are shown in Figure 7.

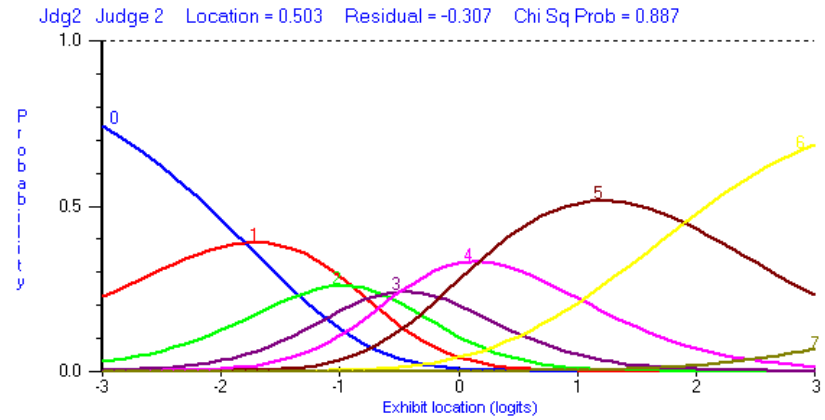


Figure 7: Category probability curves for Judge 2.

These curves show a good spread of scores, with the only criticism being the slightly lower probability of scores of 2 and 3. Otherwise Judge 2 could be described as on the harsh side with a location of 0.503, but nonetheless consistent. The same story is reflected in the item characteristic curve of Judge 2 in Figure 8.

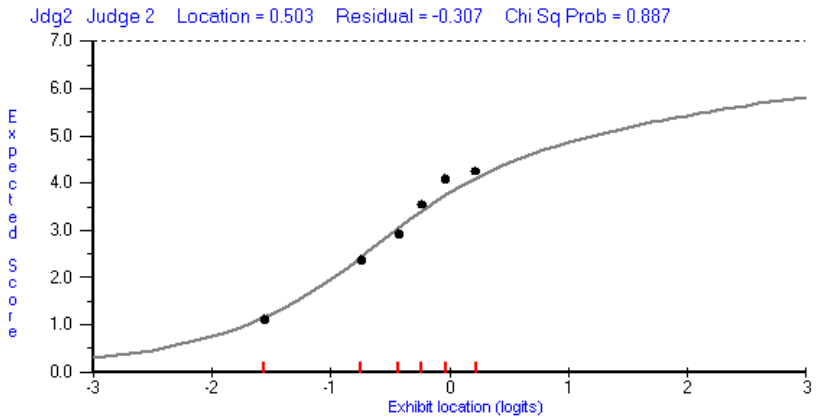


Figure 8: Item characteristic curve for Judge 2.

The Quest program highlighted a concern with Judge 4 and this concern is echoed by the RUMM program. Figure 9 shows the category probability curves for Judge 4.

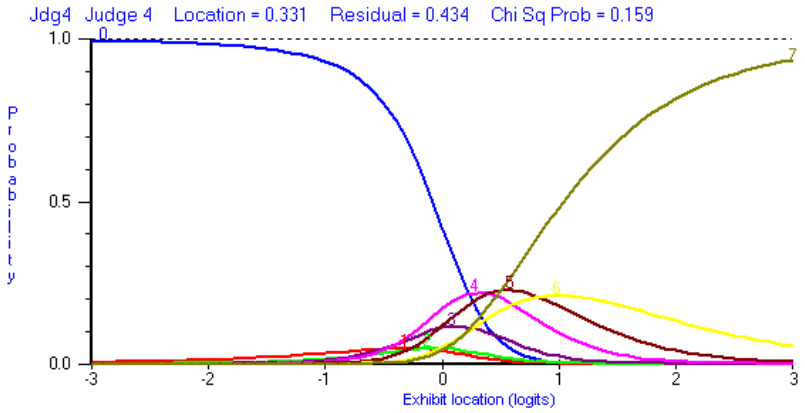


Figure 9: Category probability curves for Judge 4.

The extreme overuse of the score of 0 is most apparent with gross under-use of the middle range. Clearly Judge 4 is not in tune with the desirable characteristics of this class of exhibit. Instead of a smooth progression through the scores there is a very sudden jump. This same conclusion can be reached using the item characteristic curve of Judge 4 in Figure 10.

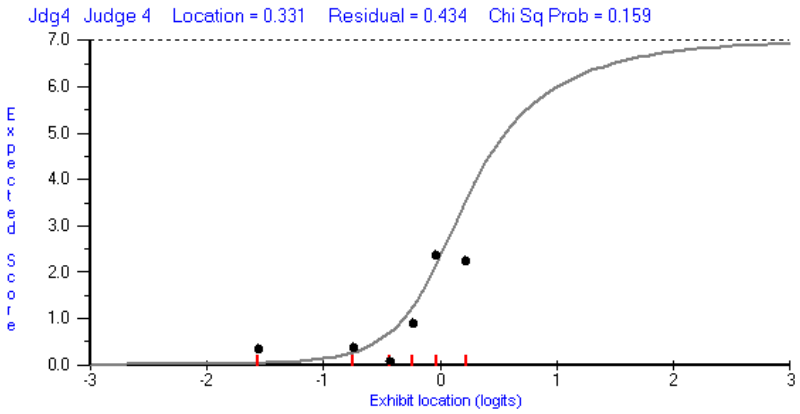


Figure 10: The item characteristic curve for Judge 4.

Particularly noticeable is the very steep nature of the centre of the curve reflecting trigger-like discrimination and the very low scores given for exhibits in the average logit range. If one were to liken the use of each of the scores across the range to the changing of gears on a motorcycle, with each gear having its desirable speed range that allows the motorcycle to accelerate smoothly, then Judge 4 would appear to be revving the vehicle in the lowest gear as high as possible and then changing to the highest gear, forcing the motorcycle to labour. The item characteristic curve does not reflect a smooth transition from one score to the next, but rather a sudden leap. To pursue the motorcycle analogy, Judge 4 would seem not to have a throttle, but rather a trigger!

Figure 11 shows the category probability curves for Judge 5.

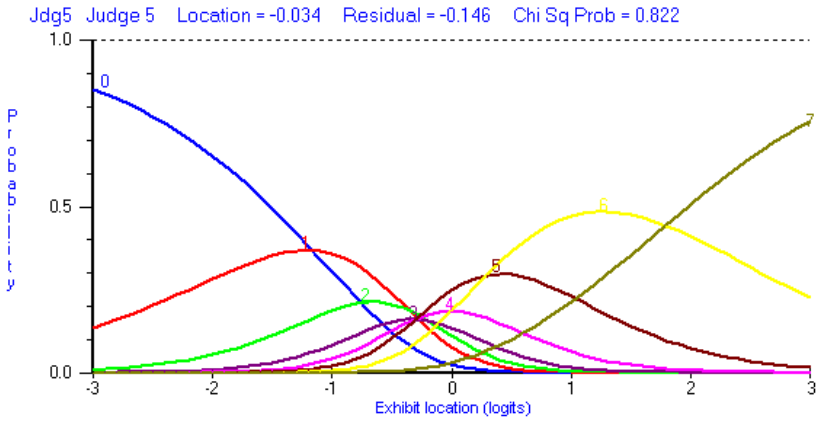


Figure 11: The category probability curves for Judge 5.

Judge 5 is consistent, but has tended to under-use the scores of 2, 3 and 4. The item characteristic curves for Judge 5 are shown in Figure 12.

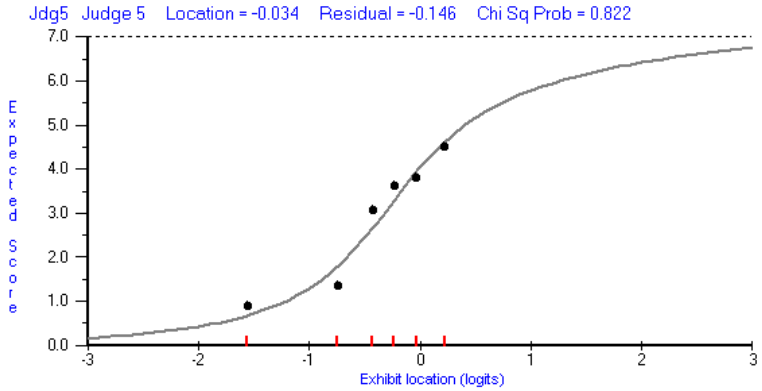


Figure 12: The item characteristic curve for Judge 5.

The item characteristic curve for Judge 5 shows a tendency to underscore at around the area of a logit score of –1 consistent with the under-use of the scores of 2, 3 and 4 referred to above.

The category probability curves for Judge 8 are shown in Figure 13.

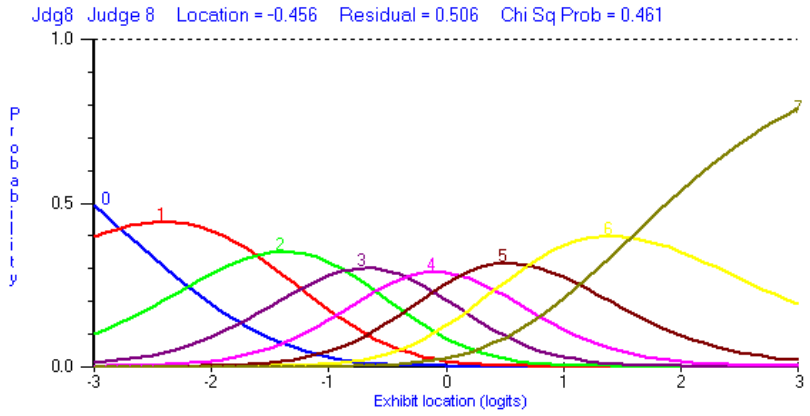


Figure 13: Category probability curves for Judge 8

The Quest analysis indicated that Judge 8 appeared to be a model judge and this is well illustrated by the category probability curves, with each score given prominence through the logit range and a tendency towards leniency. Moreover, Judge 8 has a good level of consistency. The item characteristic curve for Judge 8 is shown in Figure 14.

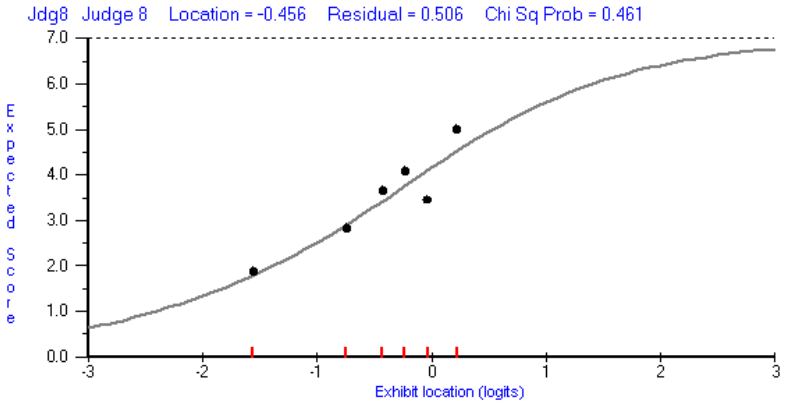


Figure 14: The item characteristic curve for Judge 8.

The closeness of the plotted points reflects a good level of consistency although a low point at a logit value of approximately 0 shows that this judge has underscored slightly in this range. The category probability curves for Judge 9 are shown in Figure 15.

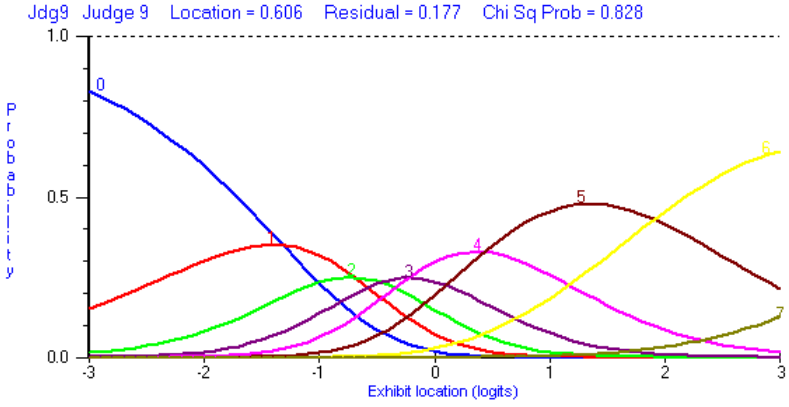


Figure 15: Category probability curves for Judge 9.

Much like Judge 2, Judge 9 has under-represented the scores of 2, 3, and 4 in the middle range, with an over-emphasis of scores of 0 and 5. This judge tends to be on the harsh side but is nonetheless quite consistent. The item characteristic curve is shown in Figure 16.

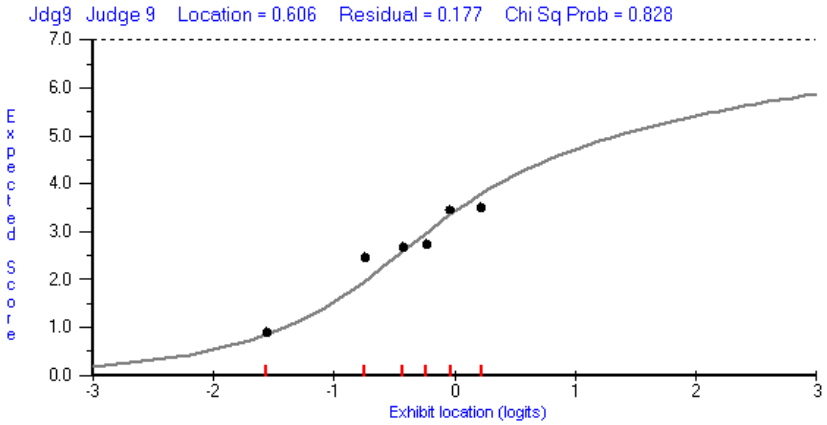


Figure 16: The item characteristic curve for Judge 9.

The category probability curves for Judge 10 are shown in Figure 17.

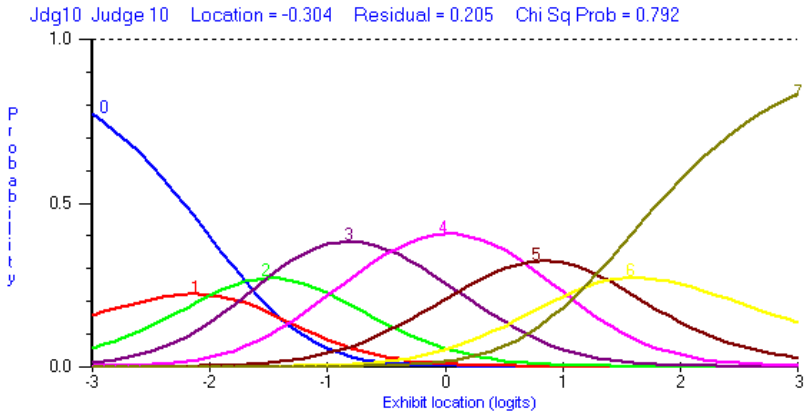


Figure 17: The category probability curves for Judge 10.

Judge 10 tends to overuse the 0 score but generally is a lenient judge showing good spread across the logit range and good inconsistency. The item characteristic curve for Judge 10 is shown in Figure 18.

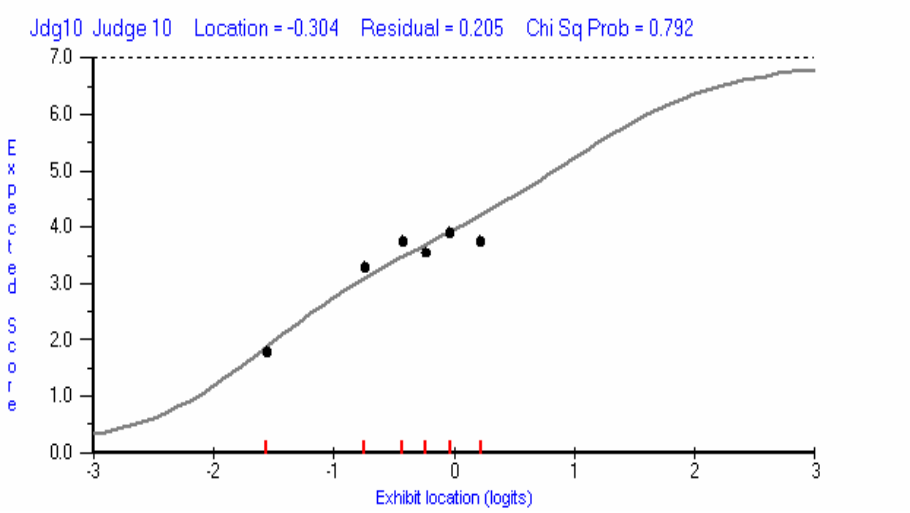


Figure 18: Item characteristic curve for Judge 10.

The item characteristic curve for Judge 10 gives further evidence of consistency with the group means close to the idealised item characteristic curve.

The category probability curves for Judge 11 are shown in Figure 19.

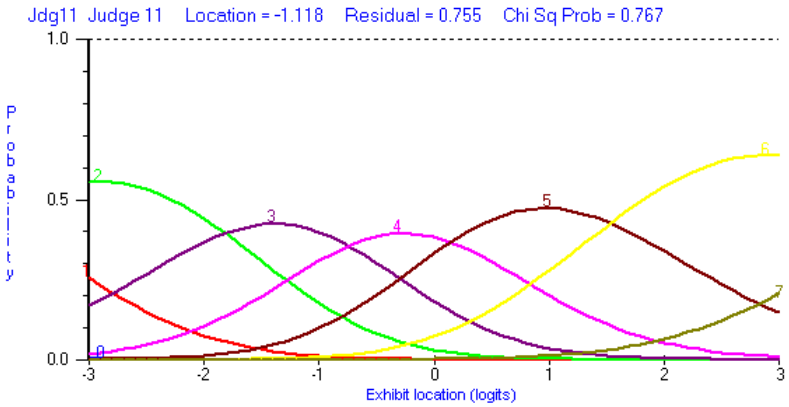


Figure 19: The category probability curves for Judge 11.

It can be seen that Judge 11 has emphasised the middle scores, not using 0 at all and under-using the score of 1. It is not surprising that Judge 11 is the most lenient with a

difficulty rating of –1.118. This has resulted in a broader spread of the scores 2, 3, 4 and 5 across the logit range. The item characteristic curve is shown in Figure 20.

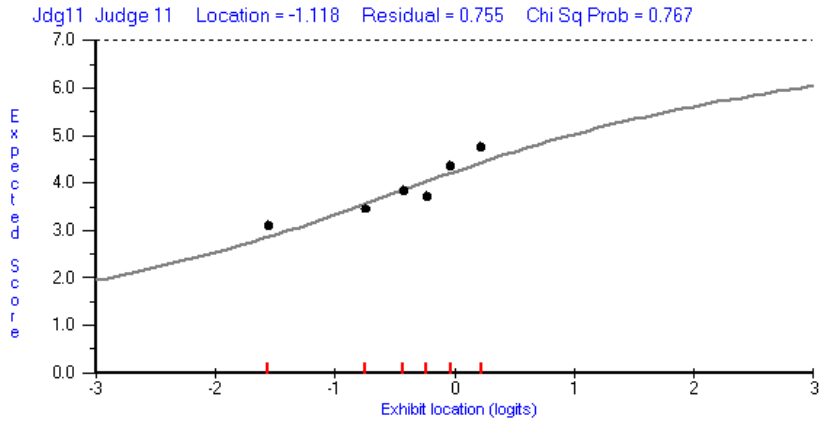


Figure 20: Item characteristic curve for Judge 11.

Analysis of the exhibits

Like the Quest program, it is possible to analyse the fit of the exhibits and see if there are any disagreements between the judges. The program identifies those items for which the residuals are above 2 or below –2. The RUMM program did not identify any exhibits whose residuals were above 2 and therefore did not fit the model; the same exhibits identified in the Quest program had larger residuals than the others. These are displayed in Table 5. It is noted that these are consistent with the QUEST analysis.

Table 5 The exhibits identified by the RUMM program with larger residuals

Exhibit ID	Residual value
311	1.513
A60	1.314
291	1.207
371	1.155
510	0.972
M00	0.828

The Item Map

The RUMM program produces an Item and Person map as well, which in this case is interpreted as a judge and exhibit map. As in the Quest program, this shows the judges and the exhibits plotted on a single logit scale. The RUMM program produces as well an overall item location, in this case judge location and these have been added to the map shown in Figure 21. Figure 22 shows the relative locations of each of the judges.

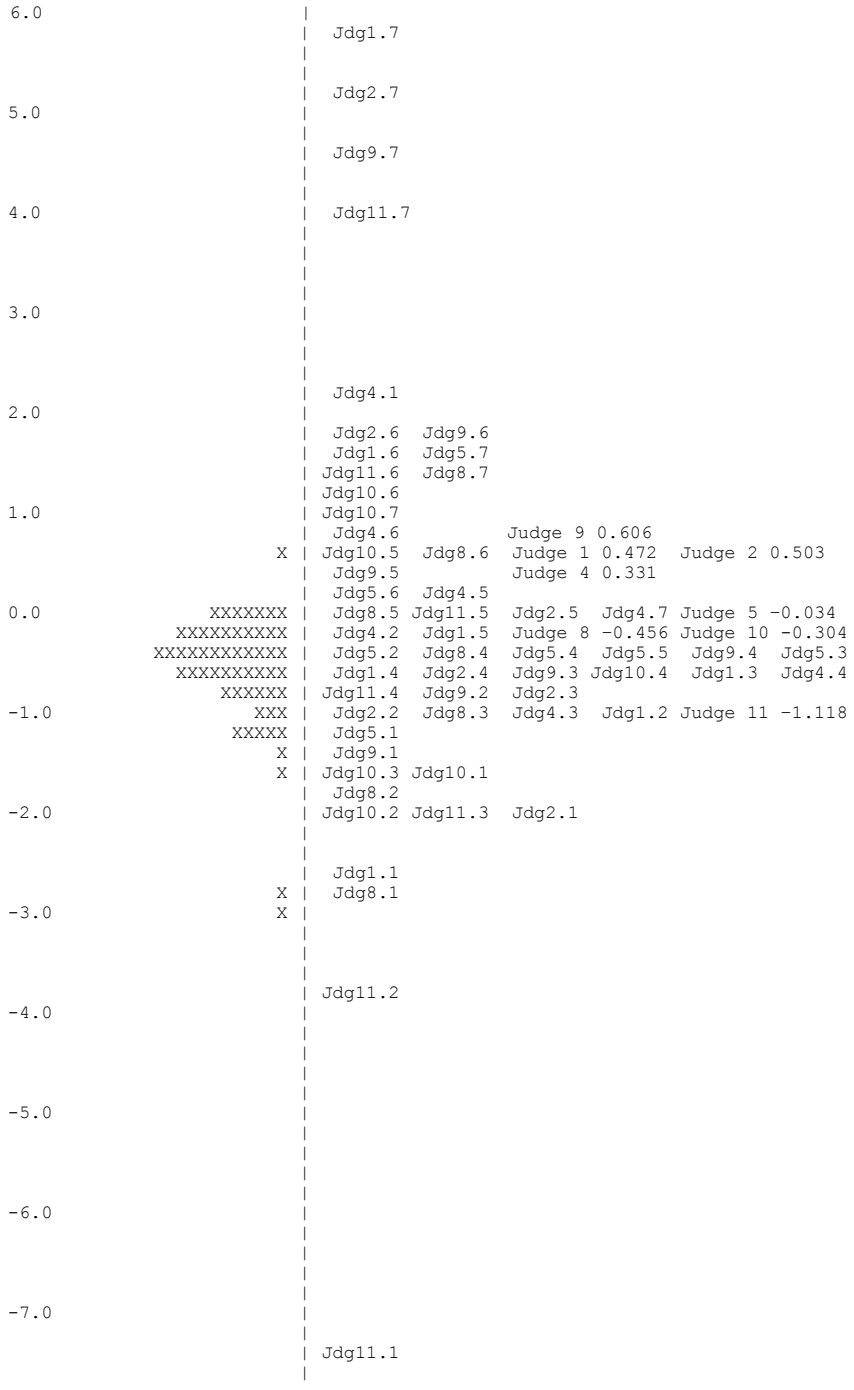


Figure 21: The Judge – Exhibit Map from the RUMM Program with the judges' locations added on the right

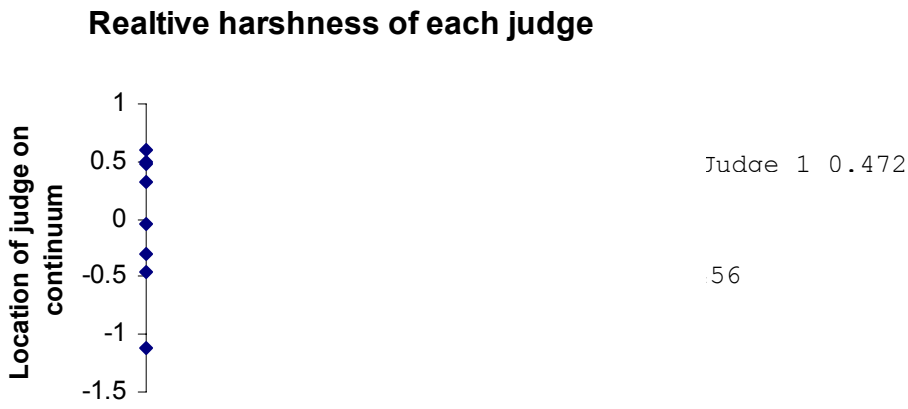


Figure 22: The relative locations of the each of the judges.

A Comparison of the Scoring Systems

The three procedures that have been used to rate the exhibits have been the two Rasch based procedures and the total scores. An important question to be asked is how do the various systems compare with one another? Correlations were calculated between the pairs of scores. Figure 23 shows a graph of the Quest estimate plotted against the raw score total. This shows a high correlation of 0.959.

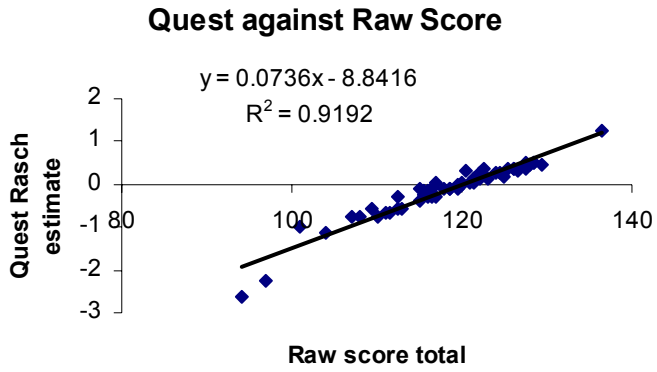


Figure 23: Quest estimate vs raw score total

Similarly, Figure 24 shows the RUMM estimate plotted against the raw score. This yielded a correlation of 0.955.

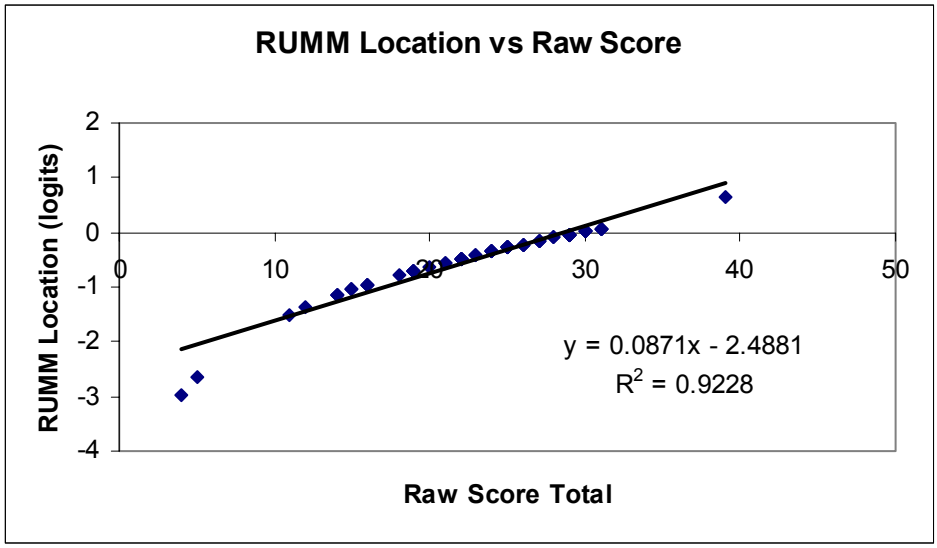


Figure 24: The RUMM estimate plotted against the raw score total.

Of particular interest is the comparison of the RUMM and the Quest programs, with their alternative methods of estimating the locations of the exhibits. Figure 25 shows the RUMM estimates plotted against the Quest estimates.

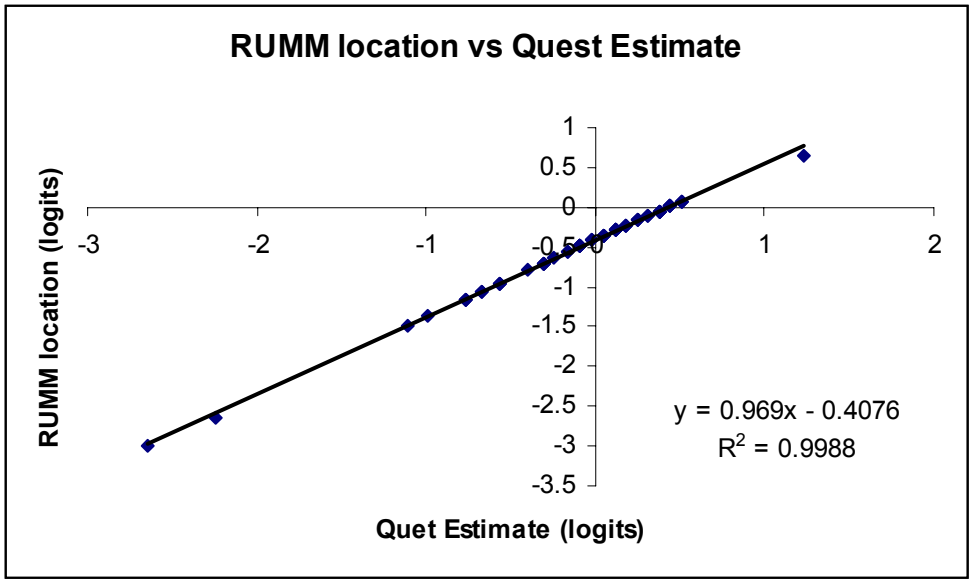


Figure 25: The RUMM estimates plotted against the Quest estimates.

The very high correlation is clear from the graph and is equal to 0.999. Two features are of interest. The slope of the graph indicates that the Quest estimate is more spread than the RUMM estimate. The intercept of the graph suggests that the zero point of the two estimations has been slightly different.

To summarise, the various estimates of the Rasch scaled scores correlate highly with the total scores and with one another. What is important, however, about the Rasch scaled scores is that they provide an interval scaled score. Such interval scores provide powerful means of making connections and comparisons.

Both the Quest and the RUMM programs provided estimates which fitted the Rasch model well and which gave good agreement with one another. For the purposes of feedback to the judging panel, however, the RUMM program, with its graphical displays provides easily interpreted feedback to those not familiar with the complexities of the Rasch scaling process.

In order to reduce the number of possible scores, these data were compressed from the 20 point scale, in increments of ½ point, to an 8 point scale from 0-7, compressing the scores in the lower range, making all scores below 12.5 zero and thereafter, compressing the range, with two possible scores sharing one final score. It has been shown that this compression makes no difference to the analysis. Thus, the Rasch Scaling process could well be used as a means of simplifying the difficult problem that arises when an attempt is made to apply too fine a scoring system, employing a wide range of possible scores, but actually using only a limited subset of them.

Conclusion

The Rasch scaling procedures that have been explored have been shown to yield some useful insights into the judging of exhibits. The application of these educational measurement procedures represents an important use of these powerful measurement tools and indicates that in the wider community and business world, there is much to be gained from the applying the results of educational research findings. Of particular interest has been the ability to put the scores of the judges and the exhibits on a logit scale that has allowed connections to be made between judges. Moreover, this method provides a means of determining both the harshness and the consistency of a particular judge. It has been suggested that this information would be invaluable for the training of judges giving easy to interpret feedback about how a judge is awarding scores.

References

Andrich, D. (1999). Rating Scale Analysis. In G. N. Masters and J. P. Keeves (eds) *Advances in measurement in educational research and assessment,* (pp244-253), Oxford : Pergamon.

Bond, T. G., & Fox, C. M., (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah N.J.: Lawrence Erlbaum.

Cronbach, L.J. (1964). *Essentials of Psychological Testing* London: Harper and Rowe.

Linacre, J. M. (1999). Measurements of judgements in G. N. Masters and J. P. Keeves (eds) *Advances in Measurement in Educational Research and Assessment* (pp244-253), Oxford : Pergamon.

Looney, M. A. (1997). Objective measurement in figure skating performance. *Journal of Outcome Measurement, 1(2),* 143-163.

White, P. (2000). When it comes to awarding medals. *The Advertiser,* October 18, 2000.

Wolf, R. M. (1997). Rating scales In J. P. Keeves (ed) *Educational Research, Methodology, and Measurement: an International Handbook* (2nd ed.), (pp. 958-965), Oxford: Pergamon,

Wu, M. L., Adams, R. J., and Wilson, M. R. (1996). *Conquest generalised item response modelling software* Hawthorn Vic: ACER.

17

Modelling and Experiments

Tony Gibbons
Flinders University

Dedications: *For the last three and one half years I have been fortunate to be supervised by Professor John P. Keeves. My research would not have been possible without his support, encouragement and wise counsel. In attempting to say more I turn to:*

> *As civilized human beings, we are the inheritors, neither of an inquiry about ourselves and the world, nor of an accumulating body of information, but of a conversation, begun in the primeval forests and extended and made more articulate in the course of centuries. It is a conversation which goes on both in public and within each of ourselves. Of course there is argument and inquiry and information, but wherever these are profitable they are to be recognized as passages in this conversation, and perhaps they are not the most captivating of the passages…Education, properly speaking, is an initiation into the skill and partnership of this conversation in which we learn to recognize the voices, to distinguish the proper occasions of utterance, and in which we acquire the intellectual and moral habits appropriate to conversation. And it is this conversation which, in the end, gives place and character to every human activity and utterance. (Oakeshott, 1962, p.199).*

Professor Keeves has mastered the art of conversation.

Models and experiments feature in the work of Professor Keeves. He has pushed me to think more carefully about those two concepts and what follows are some thoughts. They are neither definitive nor final but are offered to stimulate debate.

There is a standard view of the nature of experimentation which is summed up by Kaplan:

> Basically, experimentation is a process of observation, to be carried out in a situation especially brought about for that purpose. The great astronomer Herschel characterized experiment as nothing other than "active observation", which is indeed what it is in astronomy. But no scientific observation, as we have seen, is wholly passive, how much the scientist intervenes before or during the process of observation is a matter of degree. (Kaplan, 1964, p.144)

I am at odds with this view. An experiment is merely a device. It may be a physical device, it may be a thought, but it is a device. It is not observation for observation is what takes place when we observe the device. It may be said that to experiment is to observe but that cannot be. The same device can be observed in different ways, different devices can be observed in the same way. If we lose the distinction between devices and observation then we lose what is useful. Devices are, of course, closely linked to models. We construct a model airplane and place it in a wind tunnel. We vary the speed and direction of the wind and collect the data by some means which has been built into the device. Perhaps we alter the inclination of the wings or the tailplane and collect more data. It might be said that our observation of the device in action leads us to change the inclination of the wings. But this conflates a number of things. The device produces data which are recorded. It is our observation of the record of the data produced by the device that may make us think that it might be interesting to alter the inclination of the wings and see what data are produced. But why should it do that? The observation is of the recorded data and that observation is guided by hypotheses derived from theory. It is the hypothesis or theory that provokes our interest in the recorded data. It is the hypothesis or theory which enables us to observe what has been recorded and reflect upon it. This is what happened when Einstein asked himself what data of the world would he record if he were to travel on a beam of light. This is to say that observing the results of an experiment is a matter of reflection. To observe is to do more than see. Science is an activity in which models play a part, in which experiments play a part, but science is essentially a reflective activity. Reflection in science is an activity governed by the values of intellectual honesty and rigour, respect for evidence and valuing the views of others. These values are the products and the workings of reflection and they are no different from those that should govern the lives of human beings. A practising scientist does not merely observe the device but, to practise science reflects on the device and the data produced.

It is useful to consider the example of Galileo and pendulum motion. For thousands of years before Galileo human beings had observed pendulum motion. Objects had swayed backwards and forwards in a breeze. The rocking of a boat could set objects suspended in that boat into motion. Chandeliers had swayed above the heads of intelligent and perceptive human beings. As Matthews remarks (1994, p.111) people were searching for a reliable way to measure time because it so obviously held the answer to many of the problems that beset everyday life. Yet no one, apparently, saw isochronic motion in the pendulum before Galileo. The crucial change that Galileo made was to introduce idealised conditions for the swinging pendulum. If the swinging bob can be considered a point mass which is subject to no air resistance, no

friction, no elasticity and so on, then, it is possible to introduce mathematics. Mathematics can work in an idealised system and the results can be referred back to reality (Gibbons, 1979). This brings out the notion that the experiment is not merely a device but a device conceived in a special way by the experimenter. All devices used by the experimenter are constrained. That is, the experimenter sees the device and uses it not as a realistic object, whether tangible or not, but as an idealised object. This affects the data that are produced and therefore the observations that can be made.

The idealised system is essentially a model. As Matthews remarks:

> There is, however, a problem of idealization hiding fundamental mechanisms in the world. Keeping one's eye on the essential property is scientifically commendable, provided that it is the essential property, and that concentration on it does not blind one to other significant influences or properties. (Matthews, 2000, p.247)

What is the nature of the problem to which Matthews points? In Euclidean geometry, 'any two points lie on one and only one straight line'. This together with other axioms constitutes the formal system of Euclidian geometry. The structure of the geometry lies in the relationships between the axioms and their deductive consequences. A model in this context is any entity or set of entities which constitute an interpretation of the geometry, its theorems and axioms, and in which the theorems and axioms are true. Thus, geometrical diagrams of the various Euclidean theorems are entities which, viewed ideally, reproduce the relationships that hold between the axioms. The term 'model' in this context is one that occurs in the interpretation of a deductive system. It has been called 'formal' (Kaplan, 1964) and 'logical' (Hesse, 1967). This sense of 'model' occurs in science where mathematical theories are applied in scientific practice. The commonest use is probably in the application of probability to the behavioural sciences. In this case a mathematical theory containing probability axioms is used to generate a 'formal' or 'logical' model.

This use of the term 'model' is in contrast to its use in common parlance. A model here represents a likeness, a replica. Quite often it is also thought of as a miniature version of the thing modelled. Models are smaller and therefore it is possible to see far more, usually all, of the aspects of the model at once. Common parlance here supposes that the model is a physical system. So a model of a building project can be viewed with what might be termed 'a God's eye view', something which is normally impossible. However, to scale down is necessarily to lose something. An engineer viewing a model of a bridge accepts that when the bridge comes to be built it cannot be built with the materials and structural strength of the model, it would fall if it were. A model of the solar system is not a miniaturized copy but a model and therefore it lacks something of the real thing, the gravitational effect, the spatial void and so on. This is the case in planetariums. In the same way when we talk of persons modelling their behaviour on another we are not talking of copying that behaviour but adapting the perception of another's behaviour. A model is not a copy and it is a mistake to treat it so. It would be to treat the model as identical in all respects to the object modelled except in the matter of size. A model may however be a replica. In which case a model may be said to condense while resembling significantly that of which it is a model.

The terms ‘resembling’, ‘like’, ‘analogous to’, ‘similar’ and others all occur in talking about models. (Hesse, 1967, p.355) makes a distinction between formal and material analogy.

- Formal analogy is that of ‘structure or isomorphism between model and system’. There may also be a formal analogy between two materially dissimilar models of the same formal system.
- Material analogy is that between the object modelled and its replica. The replication may consist in some material identity of substance or shape but not of size.
- Models may be analogous both formally and materially.
- All analogies imply differences, negative analogy, and similarities, positive analogy.

The ‘formal’ model mentioned above is one that has formal analogy. The model bridge of the engineer has material analogy.

Consider the use of a model in science. If we imagine a system under investigation then it is normal scientific practice to seek another well-established system which is sufficiently analogous and use that system to explain the system under investigation. Refracted light is produced by shining white light through a piece of glass. If we imagine light as a stream of particles or corpuscles which behave according to Newton’s laws of motion then it is possible to explain the refraction. The rectilinear propagation of light becomes deducible from Newton’s laws of motion provided we imagine light to be analogous to corpuscles. The phenomenon is explained by reference to a familiar theory. However, all analogies are both negative and positive. Hesse (1967, p.356) argues when the negative analogy of a model is removed it may then properly be called a theoretical model in science. She uses the example of the billiard ball model of gases in which the size and colour of the balls is not intended to be part of the analogy. Size and colour are part of the negative analogy. A theoretical entity is left when the negative analogy is removed and this serves as the model.

There are dangers in the discard of the negative analogy. Kaplan (1999, p.88) notes two dangers of using the term model, one the opposite of the other. Kaplan does not refer to Hesse and does not mention the notions of positive and negative analogy. However the dangers to which he points can be viewed as ways in which negative analogy can disrupt the successful use of a theoretical model:

- attaching significance to features of the model which are only part of the structure of the model and not of that which it is modelling; and
- omitting something from the construction of the model which is significant in that which is being modelled.

To which might be added:

- inserting as part of the model something which does not exist in the object being modelled.

This is the problem to which Matthews refers.

Removing the negative analogy produces a model in which, in addition to the positive analogy, there will be analogies that are not at present evident and part of the heuristic of the theoretical model lies in the investigation of this excess of analogy. Hesse (1967) terms this excess, ‘surplus meaning’.

Kaplan approaches the concept of 'model' in science differently:

> Structure is the essential feature of a model. In an interpretation of the model, a correspondence is specified between the elements of the model and those of some situation, and between certain relations holding within each set of elements, so that when two elements of the model are in a certain relation the corresponding elements stand in the corresponding relations, and vice versa. A set of elements related in certain ways is a system; a structure is what is shared by corresponding systems (or it may be identified with the set of all possible systems corresponding to a given one and thus to each other). (Kaplan, 1999, p.88)

Kaplan then goes on to describe a physical system which may be a model boat or airplane or shopping mall. Where the model is used in a way which preserves its structure then, he argues, the results show what would happen to other systems having the same structure. This is to make the point that models are about correspondence of structure and not about content, a point which he had made earlier (Kaplan, 1964). He continues:

> When the correspondences are only suggested rather than being explicitly defined, the symbolic system is an extended metaphor, intermediate between a metaphor and a model is an analogy, in which correspondences are explicit but inexact. All three have roles in the actual conduct of inquiry; the view that only models have a place in science makes both terms honorific.
>
> In another honorific usage "model" is a synonym for "theory" or even "hypothesis". The term is useful only when the symbolic system it refers to is significant as a structure - a system which allows for exact deductions and explicit correspondences. (Kaplan, 1999, p.88)

While it is possible to see the distinctions that Kaplan is making they appear to be distinctions that can rarely serve a useful purpose in the practice of science. Hesse uses the term 'analogy' and this covers any system which bears a resemblance, is like or is a replica of another system. There does not appear to be a need to complicate matters further. Parsimony operates.

There is no doubt that theoretical models have a significant part to play in the practice of science. However, there has always been debate about the relationship between theory and theoretical models. Are models merely a device to aid in the understanding and production of theory or are they a necessary part of the theory? This is another rerun of the debate between positivist and realist theories of science. On the positivist side there is an insistence that all non-logical terms in a theory should have an explicitly empirical definition. This makes models dispensable when the theory is fully developed. On the realist side there is a view, which I accept, that theoretical models are an essential part of the theory (Campbell, 1957). The corpuscular model that enables an explanation of refracted light has its origin and interpretation in Newtonian mechanics. There the corpuscles do not have an explicitly empirical definition. Thus the positive analogy of the model becomes part of the explanation and part of a theory of reflection and refraction. This will be true wherever concepts in the positive analogy of the model do not have an explicitly empirical definition. For instance, any explanation that relies on particle mechanics, as the Gas Laws do, relies on a theoretical model which contains definitions that are not explicitly empirical. The point here is that the experiment is something which the scientist constructs or imagines and, using what has been constructed, he or she attempts to

investigate. The investigation will, hopefully, yield some worthwhile results. Experiments and the observations that are made of them are processes employed in the investigation of the relationship between hypotheses and theories and the real world.

What counts as worthwhile? The point of an experiment is to investigate an idealised or constrained system in the hope that it will throw some light on the nature of the real world. The biologist who injects material into a living cell is operating in a constrained and isolated world. The physicist who experiments with a system, whether tangible or not, is operating in a world that is imagined. Whatever results are achieved in experiments have to be checked against reality for the point of the experiment in the first place was to throw some light on that reality. This creates a problem. The data are the result of working with a system that is constrained and therefore the real results from an unreal system. The problem arises in referring the data back to the real world. Consider the swinging bob again. A typical school experiment is to set up a swinging bob and measure the period of the swing (T) for various lengths (l) of the bob. A graph may then be plotted of T against l. If the experimenter is working from the formula $T = 2\pi\sqrt{l/g}$ where g is the gravitational constant, then, theoretically, the graph should be a straight line. It never is. With a nod to people such as Gauss the best straight line is drawn. The drawing of the straight line is an affirmation that, within the constraints of the experiment, there is correspondence between the formula and the world. That affirmation rests upon the accuracy of the experimental data. If it were impossible to obtain data that gave a spread of readings that justified the drawing of the best straight line then doubt is cast on the formula and the theory on which it rests. If more and better experimental technique leads to a greater approximation to a straight line then there is a justifiable assumption that there is correspondence with the real world. In practice this led to the formula for the compound pendulum. In practice, technology, when using scientific work, incorporates an estimate of the percentage error into its work and designs and builds accordingly.

There is one type of experiment that has played an important role in modern science and an examination of which brings out a great deal that is important in the role of experiments in science, the thought experiment. I would suggest that any paradigm contains within it a fatal flaw the effect of which will lead to modification or even discard of the paradigm. There was just such a flaw in the Newton's work on gravitation and the cosmos. Newton made two crucial assumptions, that both space and time were absolute. By the first he meant that space is to be thought of as an infinite linear plane. This is not our experience unless we confine our experience to our immediate surroundings. By the second he meant that time was the same from any vantage point. This was the belief that underpinned the creation of Greenwich Mean Time. Liebniz, Newton's contemporary, criticised both assumptions. He took both space and time to be relative. The criticism had little effect and the Newtonian assumptions were undisturbed for more than two hundred years. Michelson's work in the nineteenth century produced anomalies for the Newtonian paradigm but such was the weight of that paradigm that it was not doubted. Time was an absolute. However, there were also problems with the Newtonian interpretation of Maxwell's work on electromagnetism. Enter Einstein and perhaps the most famous thought experiment of the twentieth century. What would happen if I rode on a beam of light looking back at a clock? Time would stand still for me. At that point I am in a time and space capsule. In that capsule I can discover the same laws as a person standing beside the clock but

the values that I will get are different. The value that remains constant is the speed of light. Time is relative. From this comes other thought experiments. Imagine the earth is a flat disc and imagine identical clocks set to the same time placed at the centre of the disc and on the turning rim. The clock at the centre will show more time passed than that on the rim.

Matthews offers a succinct account of the use of thought experiments in schools:

> Galileo offers a model of thought experiment for the classroom. He begins with familiar circumstances, he conceptualizes these in the old theory, in thought he extends the familiar circumstances, and then sees whether the old conceptualizations are adequate to the new situation: where they are not, he proposes new conceptualizations and new theories. (Matthews, 1994, p.103)

This account is an account of rationality at work. The thought experiment is not confined to science but appears wherever rationality is at work. Searle's, 'The Chinese Room', is one such example (Searle, 1999) as indeed are the replies of his many critics.

The experimental device is that conjured up by the imagination. It might be Einstein's ride on a beam of light, Newton's rotating bucket of water, Galileo's dialogue on falling bodies, Liebniz's imaginary experiment which attacked the Cartesian idea of momentum or Born's train of electrons riding on a crankshaft. In each case the experimenter has asked a question of an imaginary situation. These questions are often – if X is true then what would happen if…? or, what would the world be like if…? The first question is usually aimed at exploring the current theory and, perhaps, revealing its flaws. The second question is usually aimed at attempting to formulate a new theory. These are classic instances of human reflection at work. Note that the questions are not peculiar to science.

It is obvious that the human imagination can conjure up an infinite variety of scenarios and, thus, there must be some guidance placed on the production of thought experiments and their evaluation. This is derived from the theory that is being considered in the first place. However, this cannot rule out inconsequential thought experiments and their results. The results of the thought experiment must eventually be checked against the real world. In so far as this is possible this is a matter of checking the predictions from the experiment against actuality. So, Einstein's prediction from the General Theory of Relativity that the gravitational field of the sun would cause light to bend was tested by Eddington during the solar eclipse of 1919. The prediction was confirmed.

Matthews (1994, pp.99-105) gives a useful account of thought experiments and their importance not only for science but for science teaching. He argues that there has been a neglect of thought experiments by teachers of science and that this is surprising given their importance in the history of science. He comments that Mach is especially notable for his advocacy of the importance of thought experiments in science teaching and that Mach stressed that the use of thought experiments with students enabled the teacher to assess the grasp that students had of the concepts. This simple but important point appears to have been neglected not only by Mach's contemporaries but, according to Matthews, by present day science textbooks and teachers of science. Clearly a thought experiment is an experiment with concepts. To ask students to imagine looking back at a clock as they rode on a beam of light and describe the clock after a period of time is to ask students to play with the concepts and that play will be revealing of the students' grasp of the concepts.

In science teaching the use of computer systems to replicate historical thought experiments is now possible. There is now the possibility of developing computer systems that enable the teacher and his or her students to develop and work with thought experiments. This is an important development.

Thought experiments are the reflective scientist at work. Thought experiments in the teaching of science are an aid to the understanding of concepts and to the development of reflective thinking. But then, to teach science is to attempt to develop reflective thinking. And it must not be thought that this applies only to science.

References

Campbell, N.R. (1957) *Foundations of Science; the philosophy of theory and experiment. (Formerly titled: Physics, the elements)*. New York: Dover Publications.

Gibbons, J.A. (1979) Curriculum Integration. *Curriculum Inquiry*, 9:4, 322-332.

Hesse, M. (1967) Models, and Analogy in Science. In Edwards, P. (Ed.) *Encyclopedia of Philosophy,* New York: MacMillan.

Kaplan, A. (1964) *The Conduct of Inquiry: methodology for behavioral sciencce*. San Francisco: Chandler for Science Research Associates.

Kaplan, A. (1999) Scientific Methods in Educational Research. In Keeves, J.P. (ed.) *Educational Research, Methodology and Measurement: An International Handbook*. (pp. 112-118) Oxford: Pergamon.

Matthews, M.R. (1994) *Science Teaching*. London: Routledge.

Matthews, M.R. (2000) *Time for Science Education*. New York: Kluwer Academic.

Oakeshott, M.J. (1962) *Rationalism in Politics, and other essays*. London: Methuen.

Searle, J. (1999) The Chinese Room. In Wilson, R.A. & Keil, F. (Eds.) *The MIT Encyclopedia of the Cognitive Science,* Cambridge, MA: MIT Press.

18

Theological Education and the Identity of the Uniting Church in Australia

Andrew Dutney
Flinders University

Introduction

The Uniting Church in Australia was formed in 1977 by the union of the Australian Congregational, Methodist and Presbyterian churches. Although the immediate round of negotiations that culminated in this union began in 1957, with the formation of the Joint Commission on Church Union by the national governing bodies of the three denominations, these churches had been involved in earlier series of unsuccessful negations almost without interruption from 1901. The impediments to an earlier union had been numerous of course, but one of the questions that inhibited the process was the question of principle, Will this united church be more and not less truly a part of the one church of God? All Christian denominations wrestle with this question. It is the question of authentic ecclesial identity; the alternatives being a slide into sectarian, schismatic, or even heretical identity. Obviously this would be an undesirable outcome for a denomination of the Christian church. To wind up the Congregational, Methodist and Presbyterian ways of being church in favour of embarking on the formation of a new, united way of being church involved an acute presentation of this risk. The Joint Commission on Church Union gave it close attention in their two substantive reports on *The Faith of the Church*, in 1959, and

The Church: Its Nature Function and Ordering, in 1963. In this short paper I will examine the way in which theological education and theological scholarship were understood to play a crucial role in securing an authentic Christian identity for the Uniting Church. In particular, I will show how that role was written into the *Basis of Union*; the foundational statement of Uniting Church theology and polity and, indeed, the document upon which the three denominations voted, deciding that amid all the uncertainties and risks of union the *Basis* represented an acceptable way of securing an authentically Christian identity for the new denomination.

"A genuinely educated ministry"

In an article on his "hopes and fears" for the Uniting Church, published just before union, Davis McCaughey wrote:

> At the centre of the Church's life [ought to be] a concern for the integrity of the faith … One condition of survival for this tradition is the presence among ministers and people of men and women with a passionate interest in the substance of the faith. Perhaps we should call it, a life-long love affair with the integrity and substance of the faith. It is not given to everyone to have this, but it should be given to some; and it should be **a mark of the church** that this gift is cherished. A sign of its presence among us will be **a genuinely educated ministry**: men and women who lovingly and reverently care for what the Christian faith has meant at various periods in its history, and for what it might yet mean in the intellectual, imaginative and cultural context of their own day. (McCaughey, 1977, p.19, my emphasis)

McCaughey's statement was an endorsement of a characteristic feature of the Reformed tradition, in which the academic nature of the training of ministers is a matter of religion. In that tradition the need to maintain a high standard of theological education is integral to the doctrine of ministry.

Even so, McCaughey's statement reflects a remarkably elevated view of the importance of theological scholarship for the Christian identity of the Uniting Church. He refers to three groups of people: "ministers", "people", and "men and women with a passionate interest in the substance of the faith…a life-long love affair with the integrity and substance of the faith". It is this last group, theological scholars, that are the real focus of what he has to say. They correspond generally to the those referred to as "teachers", "doctors" or "lecturers" in John Calvin's sixteenth century polity. It involved a fourfold structure of ministerial leadership: pastors, teachers, elders and deacons. The teachers' ministry was to teach "the faithful in true doctrine, in order that the purity of the Gospel be not corrupted either by ignorance or by evil opinions" (Calvin, 1964, pp.62-63). Calvin intended that the teachers would share with the pastors, elders and deacons in the government and leadership of the church, and work most closely with the pastors in leading the congregation into a true understanding of Biblical faith by exercising their teaching ministry among the people of the church. However in the context of late twentieth century Australia the teachers would most commonly have been members of the faculties of theological colleges, teaching theological students, doing research and only exercising a ministry among "the faithful" as time, inclination and opportunity allowed. McCaughey acknowledged that not every one can be a scholar, nor need they be. But he insisted that some members of the church "should be" scholars and that cherishing their presence "should be a mark of the church".

For McCaughey to suggest that an appreciation of theological scholarship "should be a mark of the Church" was to say that it stands alongside those other activities known technically in theology as 'the marks of the church' (the preaching of the Word of God, the right administration of the sacraments and the maintenance of discipline) as a sign that the Uniting Church is a valid embodiment of the one church of God. It is highly unlikely that he was seriously proposing a fourth mark of the church but, rather, using the technical term in a novel way to add emphasis. Yet in making this rhetorical association McCaughey was not actually exaggerating the Reformed commitment to scholarship. It is a matter of Reformed religion that its congregations are served, as McCaughey said, by "a genuinely educated ministry: men and women who lovingly and reverently care for what the Christian faith has meant … and for what it might yet mean". And this "genuinely educated ministry" has been made available to the church by the faithful work of the scholars who taught them in the church's theological colleges and continue to resource them in their ministries. It has been said that if Reformed churches "did not lay stress on an Apostolic Succession [as episcopal churches do] they did on an Academic Succession" (Ainsley, 1940, p.149). McCaughey's comments, just before the inauguration of the Uniting Church, bear this out.

Theological education in the *Basis of Union*

The *Basis of Union* did not refer to theological education as such. However, it did describe the nature of the church and of ministry in such a way that it would be impossible for the Uniting Church to organize its life as the kind of community that the *Basis* envisaged without making careful, effective provision for theological education.

In **paragraph 5** the *Basis of Union* addressed the authority of the Bible, requiring that "When the Church preaches Jesus Christ, her message is controlled by the Biblical witnesses." That use of the plural reflected a recognition that the Biblical testimony to God's self-revelation is provided by numerous human voices. The implication of this was that the different voices needed to be distinguished from each other, evaluated in the light of major Biblical themes and values, and interpreted for the contemporary context. The church's proclamation of Jesus Christ can only be "controlled by the Biblical witnesses", that is, when the hermeneutical sciences are applied to the task. So, when paragraph 5 went on to commit Uniting Church ministers "to preach from these [Biblical witnesses] and to administer the sacraments of Baptism and the Lord's Supper as effective signs of the Gospel set forth in the Scriptures" there can be no doubt that it was being intended that these ministers would need to be educated. They could not preach or administer the sacraments in the way envisaged by the *Basis* without theological education. And, it should be noted, this is precisely the connection with the education of the ministry that was understood when the earliest Reformed churches identified "the Word of God purely preached and heard, and the sacraments administered according to Christ's institution" as the "marks" of the true church (Calvin, 1960, 4.1.9, p.1023; Ainsley, 1940, pp.43-44). In 16th century Switzerland it was humanistic education and in late 20th century Australia it was, quite specifically, an education combining study in the university and the theological college that would enable a man or woman to "preach Jesus Christ" in a message "controlled by the Biblical witnesses". That was understood: the necessity of a genuinely educated ministry.

And the need for theological education was made even clearer in the paragraph dealing with the ancient creeds of the church. In **paragraph 9** of the *Basis of Union* the Uniting Church received the Apostles' and Nicene Creeds "as authoritative statements of the Catholic Faith, framed in the language of their day and used by Christians in many days". That is, even as the authority of the creeds was being affirmed, it was with the caveat that the creeds do not speak our language. It is not merely that they are written in Latin or Greek, but that they use concepts and make assumptions that are culturally foreign to us. So paragraph 9 went to say that the Uniting Church "commits her ministers and instructors to careful study of these creeds and to the discipline of interpreting their teaching in a later age", commending the creeds to ministers for use in "instruction in the faith … and in worship". The *Basis of Union* thus envisaged the study and interpretation of the creeds as an on-going task of the minister; the scholarly teaching elder, engaged in the lifelong ministry of finding informed, creative and effective ways to give the people of his or her congregation access to the ancient Christian faith which was their heritage. It was what McCaughey later called "a genuinely educated ministry". The church could not ask this of ministers without also making adequate provision for their theological education, both prior to admitting them to the ministry and then, in a supportive way, throughout their ministry.

Paragraph 10 of the *Basis of Union* dealt with the authority of Protestant confessions of faith in the same way. Key doctrinal texts of the three uniting denominations were named: the Scots Confession of Faith (1560), the Heidelberg Catechism (1563), the Westminster Confession of Faith (1647), the Savoy Declaration (1658) and John Wesley's Forty-Four Sermons (1793). It was understood from the outset that these texts spoke with different, sometimes conflicting voices. The divisions between the denominations had not been over nothing. It was also understood, therefore, that appeals to the authority of these confessions could not consist in any kind of thoughtless proof-texting. Their authority had to be of a different kind. So paragraph 10 committed the Uniting Church to use these documents to "learn … from the witness of reformation fathers" and to "listen to the preaching of John Wesley". "**Listening**" and "**learning**" expressed a relationship to inherited doctrinal traditions that was quite different to the more familiar relationship of simply **conforming** or **agreeing** to the requirements of past generations; a more adult, sophisticated and (most importantly) ecumenically achievable relationship. And again, paragraph 10 required the Uniting Church to "commit her ministers and instructors to study these statements" in order that they could continually offer "the congregation of Christ's people" access to their Protestant heritage in informed, creative and effective ways. There could be no eluding the *Basis of Union's* requirement that Uniting Church Congregations be served by a genuinely educated ministry.

The Teacher, Doctor or Lecturer in the *Basis of Union* (almost)

The Joint Commission on Church Union's own awareness of the importance of informed, scholarly leadership in the church was evident in each of their reports (JCCU, 1959; JCCU, 1963). But it was only in 1968 that the Commission recognised the need to make this explicit in the *Basis of Union*. Davis McCaughey's redraft of the *Basis of Union* was circulated among the commissioners in March 1968. The "dynamic, listening, ready-to-obey" tone of the redraft (Gillman, 1968) won the

approval of the commissioners and also excited the suggestion of "a considerable number of points … which he would take into account when re-writing this re-draft over the Easter period in time for circulation to members for the May meeting" (JCCU Executive, 1968). Two of the comments on paragraph 10 are of relevance here. The first was from the Rev Henry Wells, the secretary of the Congregational Union of Australia. He wrote to McCaughey:

> This is, I think, a helpful para. Re Reformation Confessions and Wesley's sermons. However (before plunging into para.11), should not something more be said? Should there not be some recognition of the unfolding knowledge of the meaning of scripture and the recognition of new insights? At the very least, I suggest there would need to be amplification of the last paragraph to include such – or, better still, some new short paragraph or sub-paragraph. (Wells, 1968)

The second response was from Maynard Davies, a leading Congregational layman, who wrote:

> Clause 10 concludes with references to the need for a constant appeal to Holy Scriptures and the responsibility for exposition of the faith. The fancy and sub-Christian sects invariably appeal to the Bible for their justification. We shall have to define our attitude more clearly. Could we not say that we give thanks to God the Holy Spirit who has raised up many devoted scholars of all communions in the past hundred years, through whose work we have acquired insights which shall guide our teachers in faithful interpretation of Holy Scripture. We shall not misuse the Bible to write our own ticket to life here or hereafter. We shall be very wary of isolated proof-texts considered apart from the whole Gospel. Nor shall we dismiss modern scientific thought because it has things to say about the nature of man and of matter which can neither be proved nor disproved by reference to the Scriptures written two thousand years ago. Could we not admit gratefully that some modern biological authorities are widening the horizon of thoughtful Christians beyond that dreamed of by our founding fathers, and that the faith of the Church can no more stand aloof than it can stand aloof from the startling changes in the economic and social structures of humanity. How can we express our conviction that we are involved through Christ in the modern world, yet would not disregard our heritage?
>
> I wish I had the gifts necessary to say all this more adequately. To me, Teilhard de Chardin's conception of evolution as essentially sacramental is the most significant theological revelation of this century. Yet to read clause 10 as it stands the impression conveyed to our constituency is that the Uniting Church is unwilling to launch its ship into the deep waters on which modern man is currently battling for life. It is a ship on dry dock we have, with barnacles on her bottom. (Davies, 1968)

In an almost joyous response, Davis McCaughey offered a new paragraph for the JCCU's consideration that immediately commended itself to the commissioners who met after Easter. (It is most important to recognise that this paragraph was included in the *Basis of Union* quite specifically for the purpose of moderating any inclination in the Uniting Church to return to confessionalism, still less to leave the door open to any kind of fundamentalism. There is no other reason for it being in the *Basis of Union*.) McCaughey's response to Henry Wells and Maynard Davies became paragraph 11 in the *Basis of Union*:

> The Uniting Church acknowledges that God has never left his Church without faithful and scholarly interpreters of Scripture, or without those who have reflected deeply upon, and acted trustingly in obedience to, his living Word. In

> particular she enters into the inheritance of literary, historical and scientific enquiry which has characterised recent centuries, and thanks God for the knowledge of his ways with men which are open to an informed faith. She lives within a world-wide fellowship of Churches in which she will learn to sharpen her understanding of the will and purpose of God by contact with contemporary thought. Within that fellowship she also stands in relation to contemporary societies in ways which will help her to understand her own nature and mission. She thanks God for the continuing witness and service of evangelist, of scholar, of prophet and of martyr. She prays that she may be ready when occasion demands to confess her Lord in fresh words and deeds.

In McCaughey's original draft, however, the last two sentences did not appear. Instead, the suggested paragraph concluded with the words:

> … and she will seek to maintain centres of learning and research in the prayer that nothing may be lost which God would teach his Church. (McCaughey, 1968).

McCaughey's intention in May 1968 was plain. To be an authentic part of the Christian church the Uniting Church required the presence and work of what he would later call "a genuinely educated ministry". The education of that ministry required in turn the presence and work of scholars "with a passionate interest in the substance of the faith" to teach and continually resource the ministry. Indeed he would later say that "it should be a mark of the church that this gift is cherished" in the Uniting Church. But already in 1968, in his draft of a new paragraph for the *Basis of Union*, he was suggesting that the identity of the Uniting Church as part of the universal Christian church would require the maintenance of theological colleges, "centres of learning and research", and the foundational theological statement of the Uniting Church ought to contain an explicit undertaking to do just that. While the concluding statement that the JCCU finally adopted has much to commend it, a certain clarity on this aspect of the Reformed tradition was lost when McCaughey's plainly stated, pragmatic undertaking was passed over.

However, McCaughey's desire for the church to appreciate the importance of theological scholarship and the maintenance of "a genuinely educated ministry" did not abate after the minor amendment to his 1968 draft of paragraph 11. The *Basis of Union* of 1971 had several appendices dealing with specific matters. They did not form part of the document on which the churches voted, but were intended to give some indication of where the church might head. These appendices were received by the first Assembly, in 1977, and referred to the relevant Commissions of the Assembly. In a section "Concerning the Ordering of the Ministry" McCaughey's proposal for the conclusion of paragraph 11 appeared again:

> The Uniting Church undertakes to support and develop centres of theological study and teaching in as many of the University centres as is necessary, where the University itself does not provide adequately for such study; and she further undertakes to secure and support scholars in the main fields of theological study, in order that they may give their full time to this work. (UCA, 1977, p.30)

The statement was never adopted by the Assembly in such a way as to prevent the church from ever withdrawing its support for theological education and theological scholarship. As it happens it does describe very well the actual practice of the UCA in the first 25 years of its life. And that practice is one that is legitimated in the doctrine of ministry that is expressed in the *Basis of Union*.

The question that the Uniting Church is now trying to deal with in a third major review of theological education in less than a decade is whether the present system of theological education is the only means by which it can fulfil its commitment to "a genuinely educated ministry". Changes in the church's circumstances since the time of union have required a review and core activities and their relative priority, including theological education (Hughes 1997; Kaldor, Dixon et al. 1999a, pp.13-16; Dutney 2002, pp.5-13; Bellamy and Castle 2004). As I have shown in this paper, there is a great deal at stake; not least the claim of the Uniting Church to be an authentic participant in the one church of God. The Uniting Church cannot be the church envisaged in its *Basis of Union* unless it continues to find ways to ensure that it served by "a genuinely educated ministry".

References

Ainsley, J.L. (1940). *The doctrines of ministerial order in the Reformed churches of the sixteenth and seventeenth centuries*. T&T Clark, Edinburgh.

Bellamy, J. and K. Castle (2004). "2001 church attendance estimates." *National Church Life Survey Occasional Paper No.3*.

Calvin, J. (1960). *Institutes of the Christian religion*. J.T.McNeill (ed.). The Westminster Press, Philadelphia.

Calvin, J. (1964). Draft ecclesiastical ordinances, September and October 1541. In J.K.S.Reid (ed.) *The Library of Christian Classics Vol. XXII. Calvin: Theological Treatises* (pp.58-72). SCM Press Ltd, London.

Davies, M. (1968). *Comments on Revised Basis of Union dated March 1968*. Held in the Mitchell Library, Sydney. ML MSS 2733 ADD-ON 1034 Box MLK 01081.

Dutney, A. (2002). *'A Genuinely Educated Ministry': An Historical Study of the Place of Theological Education in the Inherited Traditions of the Uniting Church in Australia*. School of Education. Adelaide, Flinders University.

Gillman, I. (1968). *Letter responding to a draft of the Basis of Union from Rev Prof Ian Gillman, Emanuel College, Brisbane, to Rev Prof J Davis McCaughey, Ormond College, Melbourne, dated 6.v.68*. Held in the Mitchell Library, Sydney. ML MSS 2733 ADD-ON 1034 Box MLK 01081.

Hughes, P. (1997). *Religion in Australia: facts and figures*. Melbourne, Christian Research Association.

JCCU. (1959). *The faith of the church.* The first report of the Joint Commission on Church Union of the Congregational Union of Australia and New Zealand, the Methodist Church of Australasia, and the Presbyterian Church of Australia. The Joint Board of Christian Education, Melbourne, 1978.

JCCU. (1963). *The Church: Its Nature, Function and Ordering*. Part one being the second report of the Joint Commission on Church Union set up by the Congregational Union of Australia, the Methodist Church of Australasia, and the Presbyterian Church of Australia. Part two being the Proposed Basis of Union for these churches. Uniting Church Press, Melbourne, 1984.

JCCU Executive. (1968). *Minutes of the Joint Commission on Church Union Executive*, March 29, 1968, 4. Held in the Mitchell Library, Sydney. ML MSS 2733 ADD-ON 1034 Box MLK 01081.

Kaldor, P., R. Dixon, et al. (1999a). *Taking stock: A profile of Australian church attenders*. Adelaide, Openbook Publishers.

McCaughey, J.D. (1968). *Typescript draft of Basis of Union with handwritten insertions*. Held in the Mitchell Library, Sydney. ML MSS 2733 ADD-ON 1034 Box MLK 01081.

McCaughey, J.D. (1977). The Uniting Church in Australia: Hopes and fears. *St Mark's Review*, 89, 18-22.

UCA. (1977). *Minutes of Proceedings of the First (Inaugural) Assembly of the Uniting Church in Australia, 22nd June, 1977*. The National Office, Sydney.

Wells, H. (1968). *Letter responding to draft of the Basis of Union from Rev Henry Wells, Eastwood, NSW, to Rev Prof J. Davis McCaughey, Ormond College, Melbourne, dated 13th May 1968*. Held in the Mitchell Library, Sydney. ML MSS 2733 ADD-ON 1034 Box MLK 01081.

19

Teaching Out of the Unconscious: The Role of Shadow and Archetype

Robert Matthews
University of Adelaide

> *An understanding heart is everything in a teacher, and cannot be esteemed highly enough. One looks back with appreciation to the brilliant teachers, but with gratitude to those who touched our human feeling. The curriculum is so much necessary raw material, but warmth is the vital element for the growing plant and for the soul of the child.*
>
> *Carl Jung (1875 - 1961)*

Introduction

This work is an extension of that done by Guggenbuhl-Craig (1971), who developed the notion of a split archetype to understand the interplay of ego and shadow in the helping professions. His focus was in the profession of psychotherapy. Here I extend into the field of education, into the profession of teaching. The split archetype is an expression belonging to analytical psychology that arises from the work of C. G. Jung. I will outline the meaning of this expression below, as it is central to the understanding of unconscious aspects involved in teaching as argued herein. Suffice to say that, unlike sociological inquiries into power and teaching which are limited to conscious interplay only, an analytical psychological approach views the role of archetypal and shadow elements arising from the **unconscious** as an essential participant in the interaction between teacher and student. It is hoped that such a

'deeper' articulation of classroom interaction may form a basis for developing both a program for inclusion in teacher training and an intercession program for practising teachers.

Many new teachers enter the profession with an 'understanding heart'. They are committed young people with the wellbeing of their students upper most in their mind. During their teacher training program they often come into contact with teachers who seem to carry an opposite psychological energy, those, for whom it may be said, whose hearts have hardened. Such teachers appear to be engaged in a battle with their students, as if it is 'us against them'. Their classrooms are governed by force of will. Trainee teachers are often repelled by these encounters. They turn away thinking, "I will never teach like this." But perhaps it would be wiser to think, "there but for the grace of God go I." For perhaps these teachers who have hardened did so over time and were indeed once very much like these aspiring teachers. Some, who turn away today, may well grow hardened tomorrow.

The remainder of this chapter is an examination of the hardening that many teachers appear to develop through their professional lives and what possible actions can be done to redeem the situation. The examination begins with an archetypal understanding of the 'healthy' teacher. This understanding is this expanded to illuminate how a 'healthy' teacher may become closed in attitude and behaviour with their students. Several avenues to resolve the situation are then offered.

Teaching as a Helping Profession

Why do people become teachers? There are many answers to this question but the most keenly felt are those concerned with the helping of students. Such teachers carry a heart felt desire to contribute to their student's positive growth. They are passionately committed to repeatedly assist students through the transition from unknowing to knowing; the never-ending cycle from ignorance to mastery, each 'round' seeing the student to have accumulated an increment more learning. This is the daily grind of a classroom. There are many trainee teachers who possess a heart-felt commitment to give their energies over to the learning process of their students.

Teaching is a specialised profession, directed to maximise a student's illumination. Whether the curriculum was designed for purely vocational reasons or to encourage the individual's growth in their potential is irrelevant to defining teaching as a helping profession. Teachers are professionals trained with specific skills to perform specific tasks. They are entrusted to this work by the parents of the students and by the State. There are certain legal requirements that teachers must abide by, some in addition to those of the general public. A teacher's character is required to be in good standing; police checks are carried out to attest a new teacher's fitness. The State also requires that teachers be trained in first aid and in raising inquiry when it is suspected students are significantly 'at risk' in either their physical or mental person. Teachers play a significant role in assisting the development of children in our society. This is evident by the special requirements society places on them. Teaching is clearly one of the helping professions.

Of course not all teachers possess a deep conviction to assist in the development of their students. Perhaps they see teaching purely as a means of earning a living. They may function quite well in the technical aspects of teaching but show little thought of the pedagogic relationship so crucial to the minds of 'committed' teachers. Curiously enough it is not the cynical or indifferent teacher who is most effected by energies arising from their unconscious, rather it is the 'committed' teacher who is most at

risk. The former moves through their daily routine, not being too deeply touched by events. Whilst the latter does their all for the potential of their students and is deeply touched. The irony however is that by the very commitment to teaching as a helping profession this teacher will face the greater destructive problems from their unconscious. It is a psychological law of Jung's that whatever energy is brought to focus in the ego there will at the same time be constellated the opposite energy in the unconscious. The committed teacher who consciously wishes to see his students reach their full learning potential, constellates in his unconscious an energy wishing to depotentiate and humiliate his students.

This chapter is for those teachers who are most committed to teaching as a helping profession, for they will activate the archetype of the teacher most strongly. They will realise its positive effects in consciousness. It will orientate them and reward them with the joy at seeing their students succeed. But it will also burden them with a strong shadow. Jung regarded our psyche as a totality. It is a difficult view to accept, but we are never without our shadow. It cannot be avoided by concentrating on the good. Indeed this only intensifies it. The one-sided view of the puritan fosters destructive behaviour to arise from the unconscious (examples would be the religiously zealous witch hunts of the last millennium or the dark excesses of Nazi Germany). The tragedy in teaching is not the indifferent teacher; rather it is the committed teacher who has hardened over time. This teacher who was once so open with his students, now feels alienated to them. Where once he took satisfaction in their achievements now there is only indifference. This is the teacher who confides in the staff room "how bad students are these days, it wasn't like this when I begun teaching." But perhaps it is not the students that have changed rather it is the teacher.

The Shadow of Teaching

> In reality, the acceptance of the shadow side of human nature verges on the impossible. Consider for a moment what it means to grant the right of existence to what is unreasonable, senseless and evil! Yet it is just this that the modern man insists upon. He wants to live with every side of himself – to know what he is. That is why he casts history aside. (Jung, 1932, para. 528)

What is the shadow of the psyche?

When contents of the psyche are pushed out of the 'light' of consciousness, to be held in the 'darkness' of the unconscious, it may be said that these contents have fallen into shadow. There are certain ways of being that are approved of (and their opposite disapproved of) by society, by the collective one lives in. One learns to keep contents that meet with repeated disapproval beneath the level of conscious awareness, they remain in the collective shadow. A similar development occurs for each individual. The opposite of one's personal values, contents about oneself that are disliked (inferiorities for example), likewise are concealed from consciousness. These form what is known as the personal shadow. Because the shadow is taking up values typically in opposition to the ego, when the shadow does break into consciousness it is often experienced unpleasantly as something other than oneself. The reflex response by the ego is to repress shadowy contents; this process itself is usually unconscious. The natural flow of energy of these contents has been blocked in the repression.

But there is an avenue of expression open to shadow contents and that is via projection. The darkness we resist to see within is often recognised in others. Many of

the prejudices we struggle with in our classrooms are forms of projection. A teacher who pairs up two students they know will not get along together may be acting out of an unconscious vengeance because yesterday one of the students was disobedient to the teacher (personal shadow) or a teacher ridicules a trainees attempt at including curriculum requirements into a lesson plan because the teachers at this school all dislike the State curriculum (collective shadow). A student who teases another for doing badly in a test may well be projecting his or her own sense of academic inferiority (personal shadow) or the taunting seemingly motivated by a student's skin colour (collective shadow). Often in group dynamics such as a classroom, there will be a student who carries the role of a scapegoat. There is an unspoken agreement that this student will carry certain negative projections of the group.

The shadow projections are by their nature difficult to see. They emerge from our unconsciousness and so require a special effort of awareness to make them fully conscious. However if one can be persuaded that it is important to look for these shadow projections then it is precisely our ego's resistance to them, experienced as negative affect, that will give them up. Bernie (1989:140) suggests that "when we see something in a student or colleague which displeases us we can generally decide" what action to take and that is an end to it. "However, when we find we cannot leave it alone, when it gets under our skin and stays there, we can conclude that we are projecting." With this realisation we now have "an invitation to search for an aspect of ourself which is not getting recognition in our self-image or adequate expression in our behaviour."

As we tend to deal with the shadow side by repression and/or projection, it is a real strength of character that a teacher not succumb to their shadow's energy and leave their projections onto their students unchecked and unquestioned. And an even greater step is to try and hold the projection itself. This profound moral effort should be acknowledged and applauded. It requires a realisation that whatever is wrong in the world is also within. It also requires a commitment to become more conscious (less repressed) and to "project nothing, or as little as possible, onto others. This, however, deprives us of the ability to say 'they' are wrong and must be fought against, which deprives us of one of our main ways of dealing with the shadow." (Earl, 2001, p. 281) It is understandable then that "most of us do not want to either become conscious or withdraw our projections. Unsurprisingly, chaos, coercion and conflict flourish unrestrained." (Jung, 1938, para. 140)

When shadow elements 'break into' our behaviour it is often in association with the power drive. It was recognised early on in psychoanalytic research that the human psyche contained what was called a drive or instinct to power. The power drive (like all psychological drives) has both a positive and negative aspect. The desire to succeed in life is of course a healthy part of one's psyche, but the collective places limits on ambition and when one goes too far you can be judged a tyrant or power hungry. Accordingly we learn to check this drive and much of it becomes held in the unconscious. When shadow elements do appear in our lives they often include a release of this concealed drive to power. These elements may be said to belong to the power shadow.

The power shadow is of particular concern in teaching due to the leading role a teacher is required to take in the classroom. In our everyday lives our sense of 'acceptable' behaviour and feelings of guilt are in most instances adequate to suppress power driven shadow contents from being expressed. However there are two types of situations for teachers where they are likely to appear. One is during the experience of fear and the other the experience of righteousness. During these

moments it is easier to banish from consciousness any dissenting voice to the power drive.

Many trainee teachers feel very anxious as they approach their first time in front of a class. They have a fear of not being able to keep their students on task and quiet, of losing control of their students. The realisation that a class is getting dangerously 'out of their control' is usually accompanied by a rising feeling of fear in the teacher (and a sense of pleasure in many students). In moments of fear the intensity of shadowy contents increases and often overwhelms our conscious defences. Suddenly we find ourselves yelling at our students or perhaps issuing severe detentions in haphazard fashion. We experience ourselves as being out of our own control. And when the lesson is over our first thought is, 'what do I do if this happens again?' Some trainee teachers experience this as a loss of self, as a loss of conscious control and have a sense their power shadow has emerged. The conscious feeling of weakness in the face of resistance from their students has constellated the opposite feeling of strength in the teacher's unconscious and powerful feelings of aggression surface. Other teachers do not seem to experience much awareness of this inner process. For whatever reason they are less able to find awareness of these psychic processes, perhaps they are already too caught up in their power shadow. This attitude is indicated by comments such as, 'I can't wait to get back into the classroom and show those … they deserve all they get and next time I'll be ready for them.'

The other time where the power shadow is likely to appear is in moments of righteousness. There are times when teachers think they are very much in the right when the shadow can again be unknowingly expressed. Only this time it is even harder to sense what has happened. As Guggenbuhl-Craig (1971, p. 10) says of those committed to helping others, "the power drive is given freest rein when it can appear under the cloak of objective and moral rectitude. People are the most cruel when they can enforce cruelty to enforce the 'good'."

The most common illustration of the power drive emerging in moments of righteousness would be in the desire for approval. Many teachers give a lot of themselves to their students - their time, care and consideration. The shadow side of this pedagogic love is an unconscious demand to be loved or approved of in return. Often a teacher may be in the situation where they have gone 'the extra mile' for a student, only to see that student take the teachers effort for granted. The shadow reflex by the teacher is often a demand for something back. The conscious thought may be something like 'well I gave to you, it is right that you give back to me.' Their unconscious under this cloak of 'it is my right' is now free to get up to all sorts of vengeful or spiteful antics with the student.

Note I am not suggesting that in general classroom relationships are dominated by shadow. I think there are moments when this is true, difficult moments where shadow projections dominant classroom dynamics. For the most part interactions in the classroom are genuine encounters between people.

The Archetype of Teaching

An archetype is Jung's concept of our psyche's 'objective' inner nature. This is a hypostasised concept legitimised by the careful analytic work of Jung and his followers. It finds support in the symbolic collective works of Religion, myth and folk tales. The idea that our psyche has a nature other than merely the bio-chemical is baffling from a materialist viewpoint. However as DNA encodes the basic physical structure of our body (including the brain), according to Jung archetypes encode the

potential of psychic behaviour – of our thoughts, feelings and imaginations[1]. For Jung the energic reality of the mind (traditionally called spirit) is of equal validity to the reality of matter. Jung described the concept of archetype as follows:

> The archetypes…are pre-existent to consciousness and condition it, appear[ing] in the part they actually play in reality: as a priori structural forms of the stuff of consciousness…As an attribute of instinct they partake of its dynamic nature, and consequently possess as specific energy which causes or compels definite modes of behaviour or impulses. (Jung, 1995, p. 380)

Archetypes are activated when one encounters a typically human situation (Stevens, 1994, p. 48). In the situation of the family for example, one sees "[a] mother or father react[ing] archetypally to a son or daughter, a man react[ing] archetypally to a woman, etc.." Is it really surprising that the human psyche has developed a means to stimulate thoughts and images consistent with the requirements of these recurring situations?

> In the examples given above the archetypal situation is an interaction between two people. It appears that archetypes which govern relationships between people "have two poles, so to speak. The basic situation of the archetype [is] a polarity…Each of us is born with both poles of the archetype within us. If one pole of an archetype is constellated in the outside world, the inner and opposite pole is constellated as well. A child awakens maternal behaviour in its mother. In the psyche of every woman there is the inborn potentiality of motherly behaviour within the mother-child situation, which in some mysterious way must mean that the child is already contained within the mother. (Guggenbuhl-Craig, 1971, p. 89)

Does the act of teaching have such a typical pattern of behaviour? Learning has been an essential activity carried out by humans for millennia. The phenomenon of mass education is only very recent, but the need for one human to learn from another has always been with us. Whether the subject matter of instruction is how to light a fire, perform a high religious ceremony, solve a problem in calculus or write in the latest computer language, we have always needed to transmit skills and knowledge from person to person. Since teaching is a recurring situation that humans encounter, it follows that an archetypal basis would have developed to support its activity. It is reasonable to believe that there is an archetype which governs teaching, it is the archetype of the 'knowing adult – unknowing child'.

That a teacher is to act as a knowledgeable adult is self-evident. A teacher's content knowledge gives him a confidence from which to teach and is a security for his students. The surety that a 'teacher knows what he is on about' is very important for students. It gives them what one might call a pedagogic trust (in terms of content at least). In archetypal terms, it also will constellate the 'unknowing child' in his students. They sense in their teacher's knowingness a future possibility of what they might know. The point made by analytical psychology is that running parallel to this external duality of teacher (knowing adult) and student (unknowing child) there is a related internal duality occurring (being sourced by the archetype). There is both a knowing adult and an unknowing child active in the teacher's psyche and likewise in the student. These are the two poles of the archetype that exist in tension to each

[1] The notion of archetype is not to be confused with the exact thought or image, but rather it is the potential for such psychic elements and not the elements themselves. In other words, the thoughts, feelings and images will vary from person to person, but that there is a common intention in behaviour running through this variation is the Unifying aspect that speaks of what Jung termed the Archetype.

other. It is only through this tension that the interest to learn will come into being for the student and the desire to teach be activated in the teacher.

Why is tension so essential? Energy will only arise (interest if you like) in the process of learning provided an inner tension exists in the psyche of teacher or student. In other words, tension and flow of energy in the psyche are interdependent. A teacher or student who has lost their respective pole, will also suffer a loss of energy; the teacher who becomes only a knowing adult will lose interest in teaching, the student who becomes only an unknowing child will lose interest in learning. If the teacher and/or the student is no longer in touch with this knowing-unknowing archetype the learning process suffers and the student-teacher interaction suffers. The teacher no longer has any sense of his own unknowingness he cannot relate to the student. Indeed he will lose interest in taking the student through the learning process. Instead he becomes frustrated that the student is not more adult like. He may demand to the student, "it's so simple, I can't make it any clearer for you." He has lost the ability to identify with the state of unknowingness that is the target of his teaching. The student likewise needs to have an inner tension with adult knowingness. They need to sense that they can become knowing in what is taught. This will only happen if their knowing-adult is constellated in the learning process. A teacher who has lost connection with his students will not constellate the knowing-adult in his students. It is not only the conscious connection that is broken in 'bad' teaching. There is also this unconscious connection that can only be constellated when the archetype is functioning in teacher and student.

A teacher may keep in healthy contact with the teacher archetype by keeping an eye on a sense of their own 'childishness'. The child quality of playfulness is essential to a good teacher. The act of problem solving is assisted by a playful attitude. The teacher must not loose the spontaneity in their teaching "and must let himself be guided somewhat by his own interests. He must not only transmit knowledge but also awaken a thirst for knowledge in the children, but this he can only do so if the knowledge-hungry, spontaneous child is still alive within him (Guggenbuhl-Craig, 1971, p. 105)."

When our ego is acting in connection to an archetypal basis, there is a surety and pleasure which arises. The archetype is an inner object. Relation to it orientates and energises. At moments of feeling truly alive, there is most likely some archetype active in our psyche. When teachers talk about what guides them in their work, their love for their work etc. they are describing their ego's connection to the archetype of teaching (see Edinger, 1992, p. 3). When the archetype of teaching is active in teachers they are exerting a power not driven by some will to dominate their students but a will to drive the learning process itself. The teacher is 'correctly' animated and students are constellated into the learning process. Their psyches become caught up in the dynamic. This is possible because the archetype is a primal factor. Everyone shares this factor. Archetypes speak to everyone in fundamentally the same way. It is right to see them as a safe point from which to exert power. Our subjective selves cannot endure being a point of power without destructive unconscious energies arising. But an archetype being of our inner objective self can act as a stable psychological centre of power. You see the ego/subject is too easily exhausted and changeable to hold as a focus of power. But an object, be it inner or outer, can never really be exhaustible. Archetypes in this sense are our eternal touchstones and energisers.

The Split Archetype of the Teacher

The environment of the classroom is not the deciding factor in whether or not the archetype is split. Ultimately it is the teacher's psychological development and character that decides.

It is not always easy for a teacher to hold the tension between the two poles of the teacher archetype. Our ego does not want to be in confusion between the two states, it prefers to have a singular view of adult correctness. But when this is done, a splitting of the archetype occurs. The pole of the unknowing child is repressed in the teacher's psyche; it continues to operate in the unconscious, inevitably leading to a disturbance. With the many stress points a teacher faces in the average classroom, there are numerous opportunities for repressed contents to be projected onto students and elsewhere (other teachers, the administration[2]). The projection "of one pole of the archetype onto the outer world may bring momentary satisfaction. But in the long run it means that the psychic process is blocked (Guggenbuhl-Craig, 1971, p. 91)." As the split in the archetype deepens, the teacher's behaviour will harden.

> A hardening of the teacher is evidenced by a lack of childlike qualities such as playfulness in their demeanour. The teacher cannot sense his own childishness. It is not difficult to find classrooms that are governed by teachers who "seem to have lost every trace of childishness, who have even fewer childish traits than the average healthy adult." The teacher comes to see all the awkwardness of being a child in the student and none in themselves – they are in command, they are always right, they cannot be challenged, questioned or doubted. They have become 'only-teacher'. The unknowing children appear almost as an enemy. These teachers are often heard to complain that children know nothing and do not wish to learn. "Their nerves are easily torn by their students' childishness and lack of self-control. For this kind of teacher children are the Other, that which he himself wishes never to be. Such teachers derive a certain pleasure from demonstrating their power over children, tormenting them and keeping them in line with carefully calculated mathematical 'averages' (Guggenbuhl-Craig, 1971, p. 105)."

The teacher, whose unknowing child has been split off, believes that there is nothing for him to learn in the classroom. They are only-knowing, there is no desire or reason to learn anything new. There is no interest to learn new things about their students. They perceive their students as negative objects, as 'stupid' or 'lazy', a base expression of power. They no longer need to grow as better teachers, learning improved methods of delivery and examining new ways of classroom management. All problems in the classroom are due to the negative child qualities of the students – it is their inability to learn. The only-knowing teacher no longer needs to learn anything himself, but can 'happily' project the shadow of unknowingness onto his students.

In such a situation students will struggle with their own learning. They become caught up in the teacher's only-knowing position. All responsibility to learn is given over to the teacher. They are the authority in what is done next and how to do it, there is no inner adult constellated in the student to take responsibility. It is often said that authoritarian teachers manifest students who cannot direct their own learning. This may be due to the split archetype in the teacher creating students who believe that only the teacher need be active in the process of learning.

---

[2] For an insightful account of how the school institution can create conflicts that attract shadow projections from both teachers and students, see Hawley (2001).

Closing the Split Through Power

The expression of the power shadow by teachers who have split off their unknowing-child pole into their unconscious is not to be seen as wholly negative. It may indeed result in frustrated outbursts, even tirades, by the teacher **against** his students or some other expression of the power shadow, but it also expresses the possibility of reunification. Indeed the outbursts of power can be interpreted psychologically as unconscious attempts to bring the two poles of the archetype back into contact. Even if the contact is one of antagonism between the two poles, it is still some kind of contact and opens a possibility of a healthier relation within the teacher between knowing and unknowing.

The 'tyrant' teacher is at least still actively engaged with his students. The 'removed' teacher, the one who appears to cruise through lessons on minimal effort and concern has "either repressed one pole of the archetype so severely that it can no longer be projected, or else he has never really been concerned with the basic problem of [teaching] and his choice of profession was superficial (Guggenbuhl-Craig, 1971, p. 96)." For this type of teacher the archetype of the teacher is more psychological remote.

Closing the Split Through Reflection

Teacher training programs are increasingly valuing the role of 'teacher reflection'. These days trainees are routinely educated in the practices of both biographical and critical styles of reflection (Valli 1993). The biographical writing of journals for example gives opportunity to reflect on experiences during teaching practice. It is here that trainees will come to experience the dynamic processes of the classroom and if their choice of profession was not merely instrumental they will experience the archetype of teaching within themselves. Listening to trainee teachers debriefing after a stint of teaching practice, I am often fascinated by their stories. Many will say what an intense experience it has been. It is obvious they have been through something that touched them. Many are full of questions and paradox, they are trying to hold the poles of the archetype and find some resolution. They want to be related with their students, they want so much to help them learn, but they also see the need for authority, to be a knowing other. They are in conflict and I think this is how it should be. I am more concerned by the trainees that return from teaching practice and believe they know it all. Here I sense a splitting in the archetype occurring. They are already on their way to being the only-knowing one. They often show little interest in their training course afterwards. They tell stories of how much better they are at teaching than their mentor teachers were and boast of how much content they covered with their students – as if quantity were the only measure of learning.

Writing and discussing biographical reflections for trainees and for practising teachers is a very useful means to calm and examine the conflicts that naturally arise for many teachers in their practice. Ego can easily get tangled by intense experiences especially those where shadow elements are brought forth. The on-going complexities of classroom dynamics offer teachers, even those highly experienced, such occurrences on a daily basis. The calm and clarity restored to the teacher though reflection will not give them the 'great solution' that solves all future problems in the classroom. What it does give the teacher though, is the psychic balance to stay in conscious contact with the archetype, to hold the conflict between the poles experienced by consciousness. The alternative is to have the archetype torn apart and

the teacher become the inflexible adult in the classroom who has lost the unknowing-child to their unconscious.

This kind of reflection will only go so far as we can bear to go ourselves. It is natural to protect oneself from embarrassing disclosures to others. In group reflective work we resist sharing experiences that show our humiliation or feelings of inferiority to others. Colleagues are often cautious not to loose face with one another, nor to risk giving a rival sensitive information that may be 'used against one' at a later time. There are limits for all of us in what we will feel comfortable disclosing to others. Another difficulty lies in the fact that we are dealing with our unconscious. It is not easy for reflection to touch deeply enough to bring our unconscious shadow elements into awareness. A last difficulty concerns the dynamic properties of shadow material itself. Whenever discussions are raised concerning issues that touch the shadow, these contents become active and participants may well unconsciously resist and frustrate the dialogue unconsciously.

There is one more avenue that can reunite the split. It is perhaps the most powerful of all. Through relationship our shadow can be transformed.

Closing the Split Through Eros

It is a teacher's relationship with the students that is hardened when the archetype is split. It is also relationship that can heal what has been split apart.

There is something so powerful about simply sitting with students and talking with them. When you just simply talk to your students it challenges any negative objectified views of them that your shadow may have been feeding you. Through relatedness a more balanced perception arises. Often this interaction is enough to deal with the sorts of shadow problems many teachers have, but many times it is not. When the archetype becomes split in a teacher, relationship with the students also suffers. If a teacher has 'shut-down' with his students, it will require relationships outside of the classroom to touch their negative shadow contents.

Relatedness is such a fundamental part of our lives. It challenges us in ways we cannot easily turn from without a regrettable hardening occurring. When those we love challenge us, our shadow is most vulnerable. For example a teacher's wife complains to him about the lack of time he spends playing with their children and how so much of what he says to them are organisational commands. But his wife is not a student.

> Finally all that is left him is to confront his own dominating behaviour and to feel regret at the pain he is causing. Such a confrontation can become very complicated and difficult, but as he gradually begins to think more about himself, a very small part of his shadow has been brought closer to home. (Guggenbuhl-Craig, 1971, p. 150)

Indeed it is through relationship that most of a person's psychological development occurs throughout life. It is so important for a teacher to have relationships that are of 'his equal', i.e. not just relationship with his students, but with colleagues and friends whose opinions cannot be dismissed out of hand.

> The point is that he must actively, painfully and joyfully come into direct contact in his dealings with his fellow men. He must somehow find a way to once more expose himself to the most difficult challenges. He must be shaken. The senile "I know, I know" must become the Socratic "I don't know." (Guggenbuhl-Craig, 1971, p. 152)

Our chief concern is overcoming the split in which he lives. This can only be done if some of the shadow contents arise into the conscious field. With each such act of awareness the grip of the unconscious shadow holding the unknowing-child pole is loosened and the split teacher archetype begins to close.

The ability of relationship to touch the unconscious shadow contents of a student is of equal importance. Trainee teachers will often be thrown by encounters with 'difficult' students. Repeated strong resistance by a student can be a daunting situation for the best of us. In the face of continual aggressive and defiant behaviour there may seem little that can be done and we steel ourselves to endure or else we fall into shadow ourselves. Often the student is repeatedly ejected and a pattern is set. Relationship with this student has shut down. The student's shadow has reduced the teacher to a negative object and is to be defied. Intractable students often display almost a complete submersion into the unconscious of their knowing-adult pole, becoming all-child. There is a barrier of rage that resists the teacher's instruction touching the student's interest. The student's desire to know has been 'switched off'. Possible reasons for their resistance are many but often stem from life experiences outside of the school and thus may have little to do with the teacher (being figures of authority, teachers make excellent hooks for student's projected contents).

What can be done to reunite the split in the archetype of the student? The one-to-one approach to deal with intractable students is the most common technique suggested by teachers and in classroom management textbooks. The archetype of teaching provides an explanation as to why building a one-to-one relationship with the student is so effective. If the teacher can reach the student through genuine human interaction, then the shadowy resistance may be lifted. Genuine interaction creates moments of connection that are incongruent with the shadow projections – "perhaps this teacher is not such a bad bloke after all." If the shadow projections can be arrested in this way then the knowing-adult pole may once again be activated and the student will again be open to the teacher's knowing words. Back in the classroom the teacher's instruction will not be so resisted and will constellate the desire to become a knowing-adult in the student. The classroom now has a positive meaning for this student and his shadow material will not be such a problem.

A different shadow problem occurs in students when their failure with learning becomes projected onto the subject content itself. Feelings of failure tend to be suppressed into shadow to protect the ego from feelings of inferiority. Students that feel a repeated sense of failure in a particular subject will build inferiority contents in their shadow as a consequence. As this continues, the student increasingly resists the teacher's attempts to engage the student in the learning process. The knowing-adult is difficult to constellate in the student. As the teaching archetype becomes increasingly split in the student, they become more prone to negative projections onto the teacher and increasingly difficult to relate with. One way to address this situation is to give work to the students that they can succeed in. The feeling of success will touch their unconscious sense of inferiority and in time quell their projections. Once again allowing the knowing-adult to be constellated through the teacher's action.

Implications for Teacher Training and Development

The profession of teaching being one of helping others means that the act of teaching faces a significant threat from the shadow. The conscious desire to help our students, constellates an opposite unconscious desire to thwart them. This opposition places dangerous psychological pressure on our maintaining a healthy connection between the two poles of the teacher archetype within our psyche.

There should no longer be well-educated teachers who believe in all seriousness that they can practise their profession like engineers, purely technically and objectively. There should be no teachers who think that only their pupils are childish, while they themselves have put all that behind them. Such fundamental changes in attitude pose tremendous challenges to the educators who train teachers.

In the preparatory training for teachers and in the staff room at schools there is much talk of the difficulties created by students, but hardly any mention of one's own dark side. Part of the training of teachers should emphasize that the problems in the breakdown of learning in the classroom are one's own as well. Trainee teachers should be encouraged to develop awareness in the ways their shadows (and their students) can be expressed in the classroom with devastating effect. If the archetype of the teacher becomes split, the psychological wellspring of the committed educator is dried up.

Only a raising of shadow material to consciousness can reunite the split. In this context, much can be done to encourage trainee and experienced teachers in certain reflective practices that target contact with their shadow contents. The other possibility is the fruitful effect of interaction between one's friends and family, between those whose opinions we cannot easily ignore. It is not easy to hear criticism from others, but we are more likely to listen if it comes through the voice of friendship. Relationships bring us to a more honest view of ourselves, including our shadow aspects.

This is the hope, that the teacher whose unknowing child has fallen into darkness, leaving them a hardened only-knowing figure to dominate a classroom through power, may find their wellspring of teaching again. That they may close the split and find that part of themselves that was lost. Then they will remember the reason they chose to teach – to help students in this fundamental human process called learning.

References

Earl, M. (2001) Shadow and Spirituality, *International Journal of Children's Spirituality*, 6 (3), 2001.

Edinger, E. (1992) *Ego and archetype,* Boston & London: Shambhala.

Guggenbuhl-Craig, A (1971) *Power in the Helping Professions*, Spring, Texas.

Hansen, D. T. (2001) The Moral Dimension of Teaching, in V. Richardson (Ed.) *The Handbook of Research on Teaching,(*4th Edition), American Educational Research Association.

Hawley, R. (2002) Souls in boxes: the problem of organisation, In J. Livernois and N. Cater (Eds.) *Spring: A Journal of Archetype and Culture, 69.*

Jung, C. G. (1932) Psychotherapist or the clergy? *Collected Works Volume 11: Psychology and religion: West and East*, London: Routledge and Kegan Paul.

Jung, C. G. (1938) Psychology and religion. *Collected Works Volume 11: Psychology and religion: West and East,* London: Routledge and Kegan Paul.

Jung, C. G. (1995) *Memories, dreams, reflections*, (5th Edition), Fontana Press.

Lampert, M. (1990) When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching, *American Educational Research Journal*, 27(1), 29-63.

Neville, B. (1989) *Educating psyche: emotion, imagination and the unconscious in learning,* Collins Dove.

Stevens, A. (1994) *Jung: a very short introduction*, Oxford: Oxford University Press.

Valli, L. (1993) Reflective teacher education programs: an analysis of case studies. In J. Calderhead, J. and Gates, P. (Eds.) *Conceptualizing Reflection in Teacher Development,* (pp. 11-21), London: Falmer.

20

Collaboration over the Net: HTML & Java, the Necessary Tools

Sivakumar Alagumalai and Jury Mohyla
Flinders University

Dedications: *John has guided us, through his passion for learning, into a number of innovative research studies. This article highlights our journey into the future, and in the area of privacy and security in the use of information and communications technology.*

Abstract: *Educational use of internet has been increasing at an exponential rate. Teacher's infobases, net-tests and research net-surveys are in practical use in institutions across the globe. However, against the effective use of internet is also the risk of hackers who disrupt education's trust into the technological arena. This paper attempts to illustrate the state-of-art use of the internet in education, especially in enhancing sound pedagogical principles, cyber methodologies and collaboration. A teacher's infobase used to gather and disseminate information on lesson plans and target sections of multimedia titles (softlets), a net-test administered for a multimedia and net-surveys used to gather information for research are some current uses of internet that are discussed. It also reviews current hacking procedures, especially through IP and domain-name specific hacking and 'point-of-presence' hacking techniques, and highlights possible risk-areas in both application software and server software. Use of newer versions of hypertext-markup-language, compiled modules and java applets are highlighted. Firewall security protocols and implications for education are discussed.*

Introduction

Information technology (IT) "will profoundly alter work, education and social interaction over the next 20 years – far more than it has in the past two decades" (Heales, 1997, 48). He further indicates that IT has been vital to the 'expansion' of education and the Internet has enabled education to go beyond classrooms and borders. Although researchers may argue that the explosion in the use of the Internet has been triggered in the 'economics arena', education is equally fast in extending its tentacles into adopting pedagogical sound technology under its wings.

Articles in the daily newspaper (The Australian, 8 Dec 1997, pp 38-48) like "Educator plugs push technology", "Education is the key to IT future" and "Intranet boom lift Net software growth over C/S," highlight some of the impacts of the Internet on education and vice-versa. Henderson (1997) points out that the ease of information gathering over the Internet will enrich and encourage learning experience. This becomes more pertinent with education heads towards the direction of flexible delivery (Anderson and Alagumalai, 1996; Alagumalai and Anderson, 1997). This is further coupled with faster CPUs, enhanced network protocols, multi-threading multi-tasking server software and high-tensile fibre optic transmission cables (such as Digital Subsriber Lines) and connectors (Tebbutt, D. (1998) *InfoGeek*, Technology for Life in the *Weekend Australian*, April 4-5, p.13).

However, the use of Internet is not without its security risks. Headlines like "ISP hack designed to destroy businesses" and "Net hackers 'nasty' and getting worse" (The Australian, 8 Dec 1997, pp 33 and 42) paint a grim reality awaiting unsuspecting web administrators. A growing number of these 'web administrators' are educators by training who have taken an interest about the Internet and its associated hardware/software and may not have insights into detailed and 'hidden' technical problems.

This article reviews common educational use of the internet and evolving trends in the use of educational databases, with special reference to the use of Teacher's Infobase (Alagumalai and Anderson, 1997 and 1998) and Net-tests and Survey research (Alagumalai and Anderson, 1997 and 1998). It also examines the security issues in both hardware and software associated with educational databases and databases in general. Various security measures are reviewed and plausible solutions are discussed in light of enhancing better confidentiality and authenticity to educational use of the Internet.

Education and the Internet

Freeman (1997) purports that the Internet and WWW applications are the latest of technological innovations which offer great potentials to teaching, especially for re-engineering higher education. He further cites examples which indicated that more than 96 per cent of students indicate positive impacts of learning through the Internet, further supported by a positive improvement in grades.

The Internet has opened up discussions in almost all fields of study and allows scientist to access data, participate in collaboration and compare findings. Finegold and Pundak (1997) purport that the Internet enables students, teachers and experts to work together at a distance, each in accordance with his or her individual constraints. They further indicate that 'real-time' comparisons of weather influences in different regions, levels of humidity, growth and withering of plants and much more are now available to students everywhere. Thus the Internet has enabled a whole new era of

learning - where it is no longer imperative that students study only in fixed age groups at regulated times.

Numerous studies (Miller and Olson, 1994; Freeman, 1997; Anderson and Alagumalai, 1997) highlight the positive interactive features of the Internet, and the availability of information at the fingertips. As argued by Hahn and Stout (1994, p.2), "the Internet is, by far, the greatest and most significant achievement in the history of mankind." Anderson and Alagumalai (1996), Anderson and Alagumalai (1997), and Alagumalai and Anderson (1997) highlight the pedagogical implications of the use of Internet in education, especially when it is coupled with a cyberspace curriculum and associated parallel testing mechanics. Following these arguments, inherent differences between classrooms that advocate traditional methods and those that use the Internet are evident. Table 1 summarises these differences.

Table 1. Differences between traditional classes and classes that use the Internet

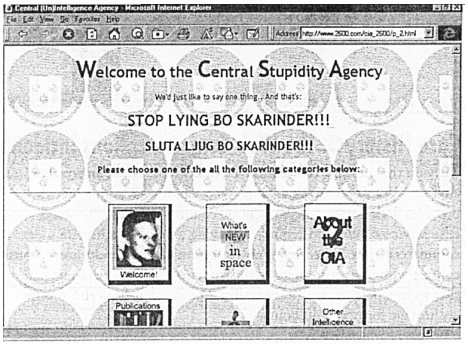
Factors/Styles	Traditional Classes	Internet-based Classes
Learning environment	Confined by space (class, school, laboratories)	Borderless
Time constraints	Fixed curricula time	No time constraints
Class-size	Restricted by school/district	Limitless
Primary knowledge source	Teacher	Anyone and everyone
Age range	Fixed by administrators	No age limit
Learning initiator and facilitator	Teacher	Student and/or Cyber-expert
Materials used for learning	Paper, Pencil, Board etc..	Computer technology
Information storage source	Library	Remote databases
'Freshness' of information	Depends on last edition	Dynamic and latest
State of information	Static	Dynamic and transient
Learning sequence	Linear	Parallel
Evaluation method	Paper and pencil	Net-tests and projects
Lesson plan and pedagogy	Linear and structured	Androgogic and dynamic

These advantages of using the Internet are reflected in changes in educational policies and government initiatives. For example, in the *Goals 200: Advance America Act* specifically indicates that the “growth of the Internet has given us the chance … to place information at anyone’s electronic fingertips.” Similar aims and advantages are highlighted in the *National Education Goals: Building a Nation of Learners*. In South Australia, the Department for Education and Children Services (DECS) has initiated the DECSTech2001 project. It seeks to equip schools, teachers, students and the broader community the tools and skills necessary for the next century. Details of this project are available at the *DECS Homepage*.

Thus, with the potentially rich information source on the Internet and its dynamism with regards to its evolving framework, security, integrity and authenticity are intricately tied to its presence. The next section examines some of the security problems and its plausible causes.

Network and General Security

As the use of the Internet grows, malicious destructive side of the Internet industry is emerging (Riley, 1997, p.33). With numerous commercial and government sites coming under hacker's control, educational databases and its related pedagogies are not too far off. Kottler and Fincher (1997) illustrate how hackers can attack even the most 'secure' and 'public-sensitive' sites. Figures 1 and 2 (Kottler & Fincher, 1997, p. 535) help highlight the risks involved.



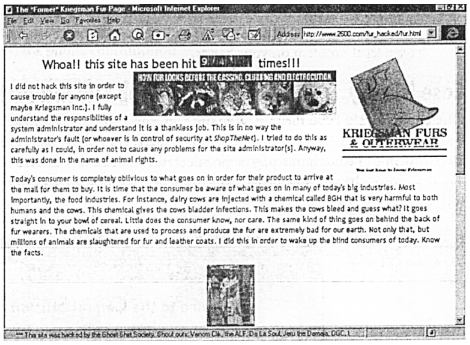


Figure 1 The hacked CIA homepage **Figure 2** The hacked Kriegsman Furs site

"For many organisations, security might best be defined as the safeguarding of corporate resources. These resources can include such things as corporate secrets, computer data, or employer information" (Swank, Kittel & Spenik, 1997). Thus, when an institution or organisation provides access to its information through its network, it also couples the potential for 'misguided intentions'. These illustrations highlight why such an emphasis is placed on web security.

Kottler and Fincher (1997) argue that website security is a combination of host and network security. "The host security model depends on the host operating system to provide security, and the network security model works across the operating system" (Kottler & Fincher, 1997, p. 536).

Morrison and Fletcher (1996) indicate that one of the basic premises of computer security is to contain information so that one can identify where it came from, who modified it and where the information can go. They refer to this concept as an 'information bucket'. Namespace, method interfaces, inter-process communication and memory as examples of information buckets. With reference to these information buckets, they argue that "the problem comes when the bucket boundaries overlap" (Morrison and Fletcher, 1996, p.741). According to them, information has 'leaked' when two buckets can read and write a common file. Thus, leaking information buckets are an indication of potential security problems.

This problem of 'leaks' is compounded when are various levels of access. Swank, Kittel and Spenik (1997, p. 58) note that "there are many levels of security access controls in the operation of a WWW database site. Each level has a specific relationship to the client/server applications that depend on its existence."

In effect, the security of any system can be gauged by:

1. how many information buckets there are;
2. how many overlap there is between buckets; and
3. how trusted the programs are protecting those data channels (especially when information flows between buckets) (Morrison and Fletcher, 1996, p.742).

The basis of most of these communication over the Internet if through the HyperText Transport Protocol (HTTP). Associated with this protocol are passwords and related authentication procedures. However, as argued by Swank, Kittel and Spenik (1996, p.58), "in basic HTTP authentication, passwords are passed over the networks as unencoded test. Although this is not human-readable text, anyone with a knowledge of network packets can easily trap and decipher a password." (*see* http://www-security-faq-html for details) However, Franks et al., (1997) indicates that for a most purposes, such authentication serves its basic functions and its strength may vary depending on the implementation.

Intricately tied to this authentication are the allowed access-levels and need scrutiny is security is to be ascertained. Global access, Per-directory access and Host-filtering are three transactions occurring over the network and their details can be summarised as follows:

- Global access (GA) — Httpd GA security maintained based on directives in global configuration file. Directives determine host- and user-level access to directories on WWW. GA configuration file is read-only when web-server is initiated. Thus require restarting of server after every configuration file modification.
- Per-directory access (PDA) — PDA configuration enables users with write access permission to configure access to documents. Such configuration does not require 'root access' privileges on the system. Unlike GA security, PDA configuration files are read and parsed for each access.
- Host-filtering (HF) — HF enables specified hosts to access a specified document tree. Hosts with access are identified in the configuration file by a domain name or IP address. Web administrators can use HF so that only corporate hosts can access new development.

These access-levels highlight the hardware-software duality of security issues. Although intricately tied to each other, with major modifications and adaptations occurring in the software domain, it can be argued that in advancing information over the network, its would be wise to have an understanding of both interfaces. Table 2 summarises some hardware and software security features and their drawbacks (Graham, 1995; Kottler and Fincher, 1997; Kittel and Spenik, 1997).

Table 2 Security features involving hardware and software

Hardware security	Software security
Firewalls – combination of computers, routers and software. Can be configured so that only traffic that meets specified rules can pass through it. Firewall keeps TCP/IP packets from entering the local network from the outside and thereby protects the LAN. Firewalls are either host-based or router-based. *Drawbacks* Cannot protect from inside attacks Slows access because of data filtering Expensive *Packet-filtering routers* – hardware device that connects different networks by routing packets between them. Routers read a packet's information and can therefore filter packets. Filtering of packets based on its address of origin and protocol used. *Proxies* – dual-homed machines. Proxies that are set up correctly can provide a good level of security for most sites.	*Passwords* – although 90% of websites have included password access to their information, it is recommended that one should review password policy periodically and monitor compliance. *Encryption* – enables confidentiality, integrity and irrefutability for the information (Trcek and Blazic, 1997). Key-based encryption are most widely used and include *symmetric systems* such as Data Encryption Standard and *Public key systems*. The later employ two keys and the Pretty Good Privacy is an example implementation of a public key system that is available on the Net. *S-HTTP* – or Secure-HTTP is an extension of the HTTP protocol that provides secure transfers using encryption. Kottler and Fincher (1997, p.544) indicate that "it will be used a lot in the future as commerce on the Net becomes more common." *iHTML* – a secure language used with ODBC compliant structures (Website 2.0) *DHTML* – dynamic-HTML for a more interactive and secure publishing and database integrity *SSL* – Secure Sockets Layer, introduced by Netscape can secure other TCP/IP protocols.

Apart from these available security features, Microsoft has introduced its *Private Communications Technology* (PCT) which is similar to Netscape's SSL except that it has a relatively better client authentication. Furthermore, in developing operating systems, Microsoft has incorporated its *NT Security* into its latest NT server software. Kottler and Fincher (1997, p.544) purport that one can build any network system with the NT system and guarantee full security. They further indicate that NT conforms to the US government's C2-level security guidelines. More of this would be discussed in the later paper.

It is imperative that plugging all possible security holes is of the highest priority and ensure a secure server. Generally most servers allow search of directories and users can access and view all documents in a web document domain. Thus temporary files left in the 'Recycle Bin' can be a potential risk and could contain pertinent information about the domain and server. Swank, Kittel and Spenik (1996, p.61) argue that web administrators should closely monitor and manage both domain access and document placement. They suggest that one method of controlling the document domain is to enforce strict document security so that only the administrators can store and modify documents in the web domain. This highlights that physical security of both the server and its related document would be the highest priority prior to considering the security risks mentioned above.

Educational Databases and Security

A preview of educational sites and the associated statistics of access rates can be a mind-boggling experience. Rarely are information made available without a database engine running in the background. Knowledge-base sites rely primarily on ODBC compliant structures and most educational sites use Common Gateway Interfaces (CGIs) to allow access to the information in these databases. Although it is acknowledged that the security concerns highlighted in the previous section are important, practitioners in education grapple with even more basic security practicalities.

Work done by Alagumalai and Anderson (1998a) illustrate the importance of an educational infobase for teachers and educational practitioners. They also argue that changes in teaching and learning through the use of the Internet require parallel assessment procedures and illustrate that net-tests and net-surveys are the way to go in the future (Alagumalai and Anderson, 1998b). Although conceptually novel and practically of immense value, the CGIs employed and database used are open to security risks, quite basic to the ones mentioned earlier.

Primary to these CGIs are Server Side Includes (SSI) and need to addressed first. SSIs are specially formatted HTML commands that enable web authors to accomplish tasks such as incorporating standard document texts in their web documents, executing CGI scripts, executing operating system commands to name a few. A standard *exec* SSI entry can be represented thus:

Visitor Odometer: *<!-- #exec_cgi="/cgi-win/counter" -->

Here the CGI script (/cgi-win/counter) is executed by the server, and the resulting text/graphics is displayed in place within the document. Figure 3 illustrates a commonly used counter.

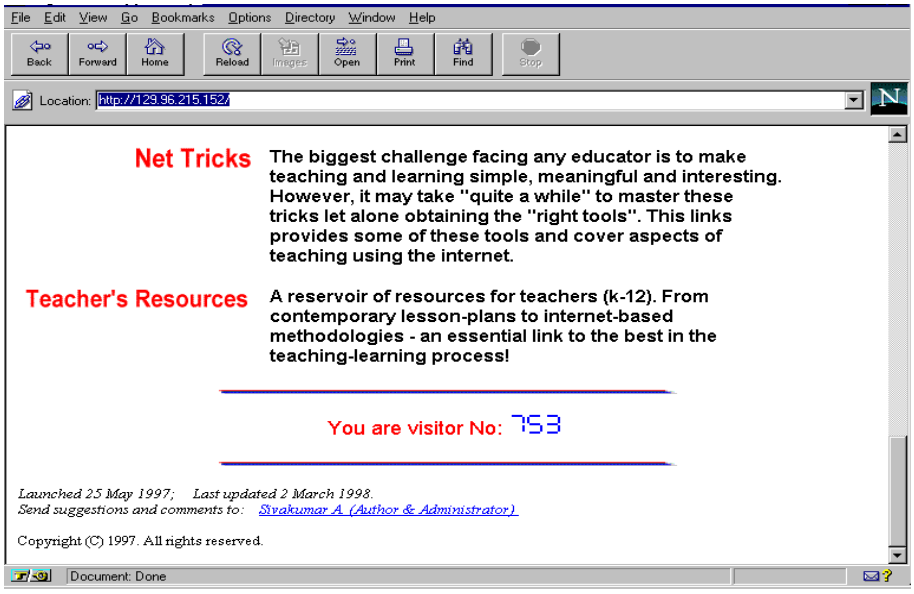


Figure 3 Web counter used at the PROF1903 Flexible Delivery Site

However, if */cgi-win/counter* is configured accidentally with 'write' permissions, so that anyone can edit and save the documents, the *counter.exe* file can be changed to execute a delete command that may delete the contents of the page being browsed or in the extreme case all files on the system. Swank, Kittel and Spenik (1996, p.62) advise that "turning off the *exec* form of the SSIs makes good sense."

The CGI scripts in the *counter* example are executed by the server process supporting the client request for execution – and are vulnerable and must be monitored closely. Careful planning and strategically placing files in secure directories can circumvent potential problem. Graham (1995, p.202) argues that "truly secure communication requires negotiated encryption of the data being sent from client to server and from server to client." Negotiated encryption here means that the server and client must negotiate the scheme to be used and must also exchange information allowing each of them to send encrypted information that can be decoded by the receiver. Graham (1995) argues that this requires both client and server contain encryption/decryption software and they both understand the same encryption schemes. Although a sound piece of advice, in reality it may be beyond the realms of a novice web administrator whose primary interest is setting up an information database. Furthermore, most educational database are set up for disseminating information and restrictions on access violates the true sense of education and learning.

Eddy and Haasch (1996, p.128) however indicate that web developers using CGIs should "limit the application's use of system resources, such as printers, network resources and directories." Traditionally, most SSIs using CGIs have the directories the *exec* files displayed in the HTML document. The development of 'virtual serves' by server software developers (Webstar for Mac; Wesbite Pro 2.2) allow for directories to be 'masked' thus preventing malicious access and manipulation.

Swank, Kittel and Spenik (1996) indicate that a good practice for web administrators is to ensure that all CGIs are approved for implementation before generation. Code checks and script security has also to be ensured. Furthermore, only the web administrator should have write permission to any CGI-executable directory.

This cautious statement brings into perspective the random 'downloading' of compiled CGIs and being used with documents. Counter programs, Guestbooks, LED scroll panels, 'Pseudo-SQL' database programs and Password Authentication programs need to be used sparingly. These *exec* programs exist side by side with sites that require payment for similar services. It would thus be advisable, as argued by Swank, Kittel and Spenik (1996), to be able to script one's own CGIs, test it within the domain, before being placed for public consumption (and abuse!). One would never be certain of what is being given away by having commercially available 'shareware' as SSIs.

Database access control is dependent on the database management system (DBMS) software in use. Standard database security is inherent to most database platforms. Thus to have viable control over database access, self-scripted gateway software should be used. Self-scripted CGIs are far more reliable and safe as compared to their commercially available counterparts. Scripting languages such as Pearl, J++, Active X and Java contain support for accessing a database. The introduction of iHTML by O'Reilly (1997) brings an even more secure interface between HTML and the database in use. The output of these CGI commands in then can be manipulated to present output in the format desired by the web administrator.

Swank, Kittel and Spenik (1996), and O'Reilly (1997) argue that CGIs enable the web administrators to have greater control over access to the data repository by

specifying database accounts to use when accessing the database. Kottler and Fincher (1997) demonstrate this restriction of access capacity through Microsoft's Visual InterDev software. Coupled with the NTFS file system of the NT Windows operating system, consistent security is ensured by eliminating unwanted hackers.

Recent research on web security have indicated that a read-only interface to a database query, for example, can use a database account with read-only privileges. This removes the possibility of accidentally deleting data by disallowing write access. It is however encouraging to note that developers of server software and those associated with database and queries are beginning to take a keener interest in this problem. Association through parallel authentication procedures between Microsoft products like Windows NT 4.0, SQL Server and Microsoft IIS 3.0 are paving the way to a more secure web publishing and data acquisition. The next section highlights some of these recent developments that are revolutionising web access and associated securities.

Java, iHTML and the Future of Educational Sites

One positive way of ensuring a secure system is starting from the rudiments, i.e. having the appropriate basic disk partitioning and underlying file formats. Kottler and Fincher (1997, p. 544) "strongly recommend NT File System (NTFS) over the File-Allocation Table (FAT) file system for the servers. By using NTFS on the server, one can restrict access to folders and files based on the rules set out in the security plan. With FAT you cannot!" They further advocate that the NTFS file permissions and user group profiles form the basis for security on NT servers. These are also the foundation for any Web site's security. Thus prior to installing any web server software, having an NTFS would lay the basis for primary security. With the demands placed on Microsoft by the various NT consumers, Microsoft has committed itself in providing even better products and thus better security features.

Databases too have evolved from the dBase III days, and multiple access privileges and facilities are available in currently available database engines and software. Access 97 with its JET engine has relational-database features at the 'read only', 'write', 'read and write' levels (Rahmel & Rahmel, 1996). One way to handle access privileges is to control application access at the document level using specially created database user accounts. Thus for database access documents that require only query capabilities, one can set up an account with read-only privileges. Another account can also be easily created to handle transactions.

An alternative way to handle access privileges is to use a temporary table to hold database transactions (inserts, updates, appends, and deletes) for batch processing. Swank, Kittel and Spenik (1996) indicate that using such a queuing system gives web and database administrators the capability to review transactions before processing them.

The emergence of scripting languages like Perl, C++, VisualBasic, J++ and Java enable codes to be compiled. Compiled codes not only hide the details of the whole document, but also 'masks' evidence of any SSIs. To do this one needs to add a few lines that send out HTML tags and lets the server know it is being sent as an HTML document. The example below helps illustrate this feature:

```
#!/bin/sh
echo “content-type: text/html”;
echo;
echo “<HTML>”;
echo “<HEAD><TITLE>My Test Page</TITLE></HEAD>”;
echo “<BODY>”;
echo “<H1>This is my test page. </H1>”;
echo “</BODY></HTML>”;
```

Similar ‘encoding’ can be done with complied C++, VisualBasic and Perl scripts. Forms and in fact any HTML document can be made secure through this encoding and compiled form (Amundsen & Smith, 1996; Eddy & Haasch, 1996).

Following closely to HTML being encoded by compiled scripts, iHTML, made available through O’Reilly (1997), also has advanced features. They include secure server-side bindings which allow databases to be accessed by built-in codes by the web administrator. The administrator could able and disable database access by clients through these built-in functions. According to O’Reilly (1997), iHTML is set to revolutionise e-commerce as it offers high-level encryption and security. Database integrity is claimed to be preserved even with multi-thread access and relational databases optimally perform with numerous access hits.

The newest language on the block, JAVA, carries with it all the ‘bests’ in security. Morrison and Fletcher (1996, p.736) indicate that “there are no known security vulnerabilities in Java.” However, they are quick to add that the form of attack that is going to be used the most against a system using Java is the ‘Trojan horse’ attack. They claim that associated malignant applets can cause the following:

- disclosure of information;
- compromising the information integrity; and
- denial of service.

A detailed look at these malicious processes along with the built in features of Java may shed light on its prowess. With regards to 'denial of services', some resources of the system will no longer be made available by the applet(s). In extreme cases, large amount of CPU time and memory can be gobbled up and may end with a system crash. Morrison and Fletcher (1996, p.739) indicate that "a great deal of care has gone into creating and testing the Java exception handlers that identifies and halts faults in Java applets thus averting a crash." They further claim that steps have been taken to control that a Java applet cannot modify files on the client system.

Information integrity could be compromised when malicious codes could modify financial statements and which may cause legal suits and auditors notices to follow. Information integrity becomes extremely crucial especially when dealing with medical patient's database records. Imagine the case of a malicious applet altering the medication and diagnostic records of medical records.

Along similar lines, disclosure of information is a possible treat to database security. Malicious program can send important information to the hackers. The program could use email or communicate with an Internet server. The consolation offered by Java developers is that "precaution has been taken to ensure that an applet cannot read other files on the system - as it is difficult for a malicious program to know what is sensitive and what isn't" (Morrison and Fletcher, 1996, p.736). In effect, Java seems to offer educators the relatively best security to launch their websites.

Conclusion

As government projects and initiatives move into making information technology part of the teaching and learning environment, educators need to look beyond just being providers of information the traditional print-media way. Internet and educational databases seem to be the basis for future pedagogies and methodologies (Anderson & Alagumalai, 1996).

This article highlights the security risks involved in using networked information transfers, especially the specific consideration to be given to hardware and software securities. Furthermore, evolving development trends in software do give hope to better security for information on websites.

A good secure system would have included the four-tier approach in security measures. First by having a secure basic operating system with NTFS format and accompanied by a reliable web server like O'Reilly's Website Pro, which has inbuilt restricted access like SSL and iHTML that only allows access to specific information. Next databases holding information should have good 'handshake' mechanics with both operating systems and server software to only allow restricted 'read' or 'write' access. Finally, only documents and applets that have the security properties indicated above should be used.

However, it should be noted that these four-tier security measures are time-based and hardware specific and may be obsolete in the future. In summary, educators and web administrators of educational sites must and need to keep up with and 'be above' the changing trends in the information technological world - so as to make proactive improvements and refinement, rather than to moan and then react when security is infringed and damages had been inflicted to their trade 'tools'.

References

Alagumalai, S. (1996) SOLO, *Rasch, Quest and curriculum evaluation*. Paper presented at the 10th Joint Conference of the Australian Association for Research in Education and the Singapore Educational Research Association, Singapore, November 25-29, 1996.

Anderson, J. & Alagumalai, S. (1996). *From text-based pedagogy to cyber-based methodology*. Paper presented at the 10th Joint Conference of the Australian Association for Research in Education and the Singapore Educational Research Association, Singapore, November 25-29, 1996.

Amundsen, M. & Smith, C. (1996) *Database Programming with VisualBasic 4*. Indianapolis, Indiana: SAMS Publishing.

Anderson, J. & Alagumalai, S. (1997). HTML: The next language of communication for information technology in open learning. *UNICORN: Journal of the Australian College of Education, 23*(3), 11-20.

Alagumalai, S. & Anderson, J. (1998) *INFOBASE: A Practical Teacher's Database*. Paper to be presented at the 15th Australian Computers in Education National Conference, Adelaide, July 5-8, 1998.

Alagumalai, S. & Anderson, J. (1998*) Educational Revolution: Net-tests and net-surveys*. Paper to be presented at the 15th Australian Computers in Education National Conference, Adelaide, July 5-8, 1998.

DECS Homepage (1998) *DECSTech2001 Plan*
http://www.nexus.edu.au/strategic/decstech/index.htm

Eddy, C., & Haasch, B. (1996). *Web Programming with Visual Basic*. Indianapolis, IN: Sams Publishing.

Finegold, M., & Pundak, D. (1997) Stars on the Net: Conducting student projects in astronomy by Internet. In Fisher, D. and Rickards, T (Eds.) (1997) *Science, Mathematics & Technology Education and National Development*. Proceedings of the 1997 International Conference on Science, Mathematics and Technology Education, Hanoi Vietnam, 6-9 January, 1997.

Franks, J. et al., (1997). *An extension of HTTP: Digest Access Authentication* (Jan, 1997)
http://ds.internic.net/rfc/rfc2069.txt

Freeman, M (1997) Flexibility in access, interaction and assessment: The case for web-bases teaching programs. *Australian Journal of Educational Technology, 13*(1), 23-39.

Goals 2000: *Advance America Act (1997).* http://putwest.boces.org/standards.html

Graham, I.S. (1995). *HTML Sourcebook.* New York: John Wiley & Sonc Inc.

Gurewich, N. & Gurewich, O. (1995) *Master Visual Basic 4*. Indianapolis, IN: Sams Publishing.

Heales, J. (1997) Education is the key to IT future. *The Australian* 9 Dec 1997, p.48.

Henderson, L. (1997) Educator plugs push technology. *The Australian* 9 Dec 1997, p.43.

Hilvert, J. (1997) Net hackers 'nasty' and getting worse. *The Australian* 9 Dec 1997, p.42.

Kottler, J., Fincher, G., and et al., (1997) *Microsoft Visual InterDev: The Comprehensive Solutions Package UNLEASHED*. Indianapolis, IN: Sams Publishing.

Long, G. (1997) Multimedia multiplying. *The Australian* 9 Dec 1997, p.43.

Miller, L. & Olson, J. (1994) Putting the computer in its place: A study of teaching with technology. *Journal of Curriculum Studies, 26*(2), 121-141.

Morrison, M, and Fletcher, M. (1996) *JAVA UNLEASHED*. Indianapolis, IN: Sams Net Publishing.

National Education Goals: Building a Nation of Learners. (1998).
http://www.negp.gov/WEPBG110.htm

O'Reilly (1997) *Website Pro 2.0* (Computer Software). CA: O'Reilly Publishers.

Rahmel, D. & Rahmel, R. (1996) *Developing CLIENT/SERVER Application*. Indianapolis, IN: Sams Publishing.

Riley, J. (1997) ISP hack designed to destroy businesses. *The Australian* 9 Dec 1997, p.33.

Scott, C., Shannon, B., Font, F., & Hatfield, B. (1995) *VisualBasic 4 UNLEASHED*. Indianapolis, IN: Sams Publishing.

Swank, M., Kittel, D., & Spenik, K. (1996) *WWW Database Developer's Guide with Visual Basic*. Indianapolis, IN: Sams Net Publishing.

Tolhurst, W.A., Pike, M.A., Blanton, K.A., & Harris, J.R. (1994) *Using the Internet*. Indianapolis, IN: QUE Corporation.

Trcek, D. & Blazic, B.J. (1997). Certification authorities in a global network – Procedures and guidelines for a public key infrastructure. *Australian Computer Journal, 29*(2), pp. 41-47.

Walters, C. (1997) Intranet boom to lift Net software growth over C/S. *The Australian* 9 Dec 1997, p.38.

21

Factors Influencing Reading Achievement in Germany and Finland: Evidence from PISA 2000

Dieter Kotte and Petra Lietz
Causal Impact, Germany & International University Bremen, Germany

Introduction

The Organization for Economic Co-operation and Development (OECD) embarked on its Programme for International Student Assessment (PISA) in 2000. In a three-year cycle, it was planned to compare the performance of 15-year-old students across member countries. The first cycle in 2000 focused on Reading, the second cycle in 2003 on Mathematics and the third cycle in 2006 will concentrate on Science with the other two subject areas featuring as minor components.

The information gained from PISA is the basis on which the OECD develops some of the educational indicators used in its annual publication 'Education at a Glance' (e.g. OECD 2002, 2003a). In particular, indicators are derived from the data obtained in the PISA surveys that are considered to reflect the quality of educational output for each country. These output indicators, in turn, are linked to indicators of educational input such as educational spending in each country.

Such bivariate correlational analyses between input and output indicators may go some way to highlight differences across countries. However, more complex analyses are required to address which factors operate similarly or differently across countries to influence achievement at the student level. The complexity of such analyses stems from the fact that contextual information, for example about students' home environment, their attitudes and expectations or about teacher- and school-related

factors, needs to be taken into account at the appropriate levels in order to understand what leads to differences in performance within and across countries.

Without analyses of this kind, measurement programs such as PISA have the tendency to become an international horse race with a focus on relative performance levels without making use of the full potential of the contextual data that are collected and that can be used to explain differences in student achievement.

Finland was the highest scoring country in the PISA-2000 data collection (mean score: 546; standard deviation: 89). In contrast, Germany (mean score: 484; standard deviation: 111) performed well below the OECD average of 500 (Adams & Wu 2002, OECD 2001, 2003b). Much has been speculated about possible reasons for this difference in performance and many suggestions as to measures to be taken in Germany as a consequence (Hofmann & Lock 2002, Kultusministerkonferenz 2002)

However, this article is aimed at examining the way in which student- and school-level factors operate to influence student achievement in the two countries in order to base the current discussion about possibilities of improving the performance of German students on available evidence rather than on data-free claims and speculation.

The data

The data used to examine ways in which background factors operate to influence student achievement in Finland and Germany have been taken from the publicly available PISA website (www.pisa.oecd.org). As the study did not obtain data from teachers, it was possible to investigate variables only at two levels, namely the student and school level. Table 1 lists the number of cases in the international PISA data sets for Finland and Germany.

Table 1: Number of cases for Finland and Germany

	Germany		**Finland**	
	student level	school level	student level	school level
number of cases	5073	219	4864	155

The analyses

First, descriptive statistics were calculated for each variable measured in the student and school questionnaire to check for missing data and to examine the frequency distributions. Then, correlations were calculated between all student and school background variables and reading achievement. As the main aim of the analyses was to compare patterns of effects of factors influencing reading achievement in Germany and Finland rather than to optimize explained variance within each country, variables considered for inclusion in the analysis had to be available from the data sets of both countries.

Based on results from previous studies of student and school factors influencing student achievement in reading (Elley 1994, Lietz 1996, Lundberg & Linnakylä 1993, Purves 1973), single-level path models - separately for Germany and Finland and separately for the student and school level - were developed based on prior research

on school achievement (Keeves 1991, Kotte 1992, Lietz 1996). The models were then analyzed using partial least squares (PLS) analyses using PLSPATH (Sellin 1990).

These exploratory PLS analyses served two purposes. First, they examined which of the hypothesized relationships between the predicting constructs as well as between the predictors and reading achievement in each country emerged in the two data sets. Secondly, PLSPATH allowed identifying to what degree a construct operated directly and indirectly to affect reading achievement in each country. Only those relationships between constructs with a direct effect $\geq|0.10|$ on any other construct in the model were considered sufficiently substantial to be retained in the subsequent hierarchical analyses. For reasons of space limitation, complete results of these preliminary path analyses cannot be reported here. However, Table 2 provides details of those student- and school-level constructs which emerged as having a substantive effect on reading achievement either directly or indirectly.

Table 2: Overview of student- and school-level constructs extracted for further analyses using PLSPATH

Construct	**Variable(s) used to form construct (and PISA variable name)**	**Coding/comment**
Student-level constructs		
COMPUSE	computer usage to learn school material (IT05Q03), for programming (IT05Q04), spreadsheets (IT06Q03), educational software (IT06Q05)	scale/factor score based on the four variables; high value denotes more intense usage of computers by student
EFFORT	index comprising questions related to students' perception of how their teacher wants them to deliver and perform (ST26Q02, ST26Q03, ST26Q04, ST26Q15 and ST32Q03)	scale/factor score based on the five variables; high value denotes higher standards of school work
FEMALE	student's gender (ST03Q01)	1=not female, 2=female
GRADE	Grade student is enrolled at (ST02Q01)	8=Grade 8, 9=Grade 9, 10=Grade 10
READACH	reading achievement (PV1READ)	Rasch scaled reading score: range: 0-1000, midpoint 500
READINT	PISA index denoting engagement in reading based on responses to nine questions ranging from 'I read only if I have to' to 'Reading is one of my favourite hobbies' (JOYREAD) PISA index denoting interest in reading based on responses to 3 questions ranging from 'I read in my spare time ' to 'When I read I sometimes get totally absorbed' (INTREA) borrow books from library (ST38Q01) DIVREAD	scale/factor score based on the four indexes; high value denotes high interest in reading

(continued)

Table 2: Overview of student- and school-level constructs extracted for further analyses using PLSPATH (continued)

Construct	Variable(s) used to form construct (and PISA variable name)	Coding/comment
Student-level constructs		
SELF	range of PISA indexes and variables comprising: verbal (SCVERB) and mathematical (MATCON) self-concept, general academic ability (SCACAD), control strategies (CSTRAT), elaboration activities (ELAB), perceived self-efficacy (SELFEF), effort and perseverance (EFFPER), ability to memorize (MEMOR), instrumental motivation (INSMOT), control expectation (CEXP), competitive learning (COMLRN), study time management (ST32Q01)	scale/factor score based on the range of indexes; high value means high degree of self-concept/learning strategies
SES	mother's main job (BMMJ) father's main job (BFMJ) educational level of mother (MISCED) and father (FISCED) number of books at home (ST37Q01)	scale/factor score based on five variables; high value denotes high socio-economic status
School-level constructs		
ADMIT	admission of students based on recommendation by feeder schools (SC07Q03) and on student assessment (SC07Q01, SC07Q02)	1=never, 2=sometimes, 3=always
IGNORE	learning hindered as teachers are ignoring students' needs (SC19Q07)	high value denotes strong hindrance of students to study properly
IMPACH	importance attached to achievement as reflected in principal's perception of frequency of student assessment through assignments/projects/ homework, freq. of performance reports to parents, teacher valuing academic achievement (SC09Q01, SC09Q05)	scale/factor score; high value denotes high importance attached to achievement by the school
INSTGRP	group students for institutional purposes (SC18Q03)	1=yes, 2=no
LOWACH	policies against low achievers enforced (SC10Q01, SC12Q02, SC18Q03), e.g. extra lessons or relegation	low value denotes highly active policies against low achievers
P_GIRLS	percent of girls enrolled at school	
PRIMARY	school contains primary section (Grades 1-4) (SC05Q01)	1=yes, 2=no
PRINCIP	frequency of principal reporting on students' performance (SC17Q02)	high value denotes high frequency of reporting
SCHSIZE	size of school	number of students enrolled in school
STUDBEHA	student behaviour; hindrance of students to study properly, eg. due to disruptions or lack of discipline (STUDBEHA)	high value denotes strong hindrance of students to study properly
UPSEC	school contains upper secondary section (Grades 11-13) (SC05Q11)	1=yes, 2=no

Note:
All variables were standardized with a mean of 0 and a standard deviation of 1

Subsequently, the student- and school-level constructs were entered into a two-level hierarchical linear model which, in turn, was analyzed separately for Finland and for Germany using Hierarchical Linear Modeling (HLM) software (HLM-5; Raudenbush, Bryk, Cheong & Congdon 2000). For details regarding the statistical algorithms of HLM-5, see Raudenbush & Bryk (2002).

The model was refined based on interim results 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, Lietz & Kotte 2000) was considered to be not sufficiently substantive and, hence, removed.

Results

In this section, results of the two-level HLM analyses of the two data sets are reported, first for Finland then for Germany. The country-specific discussions are followed by a comparison of results for the two countries.

Two-level HLM model of reading achievement for Finland

In Finland, a hierarchical model consisting of six student-level factors and two school-level factors emerges from the HLM-analyses. Table 3 presents the final estimation of fixed effects of each factor on reading achievement in terms of its γ-coefficient (ie., standardized regression weight) together with the corresponding standard error, the t-ratio as well as the significance level (ie., p-value). It should be stressed that no interaction effects whereby school-level factors might impact on the relationships between student-level variables and reading achievement were found.

Table 3: Final estimation of fixed effects; two-level HLM model for Finland for reading achievement

Fixed Effects on READACH	γ-coefficient	standard error	t-ratio	p-value
Level 1/student-level effects				
READINT	0.28	0.02	15.694	0.000
GRADE	0.21	0.01	17.380	0.000
SES	0.20	0.02	11.744	0.000
FEMALE	0.20	0.01	14.631	0.000
SELF	0.19	0.02	11.672	0.000
COMPUSE	0.12	0.02	7.359	0.000
Level 2/school-level effects				
P_GIRLS	0.13	0.03	4.009	0.000
IGNORE	0.09	0.03	2.852	0.005

In Finland, the strongest factor influencing positively reading achievement is reading interest (γ=0.28). This finding confirms the importance of this relationship reported previously (e.g. Elley 1994, Lietz 1996) whereby students who read more frequently and not only non-fiction books but also magazines, comic books, fiction, email, web pages and newspapers perform at a higher level in reading. In addition, socio-economic status (γ=0.20) has a considerable influence on reading.

Furthermore, a gender effects operates at the student as well as the school level in Finland. Thus, not only do girls achieve at a higher level than boys as indicated by the effect of FEMALE on reading achievement (γ=0.20) but also attendance at a school

with a higher proportion of girls (P_GIRLS: γ=0.13) is linked to higher performance in reading. It should be noted that – like all other effects in one HLM analysis – these relationships emerge while the impact of other factors, such as, for example, reading interest, are being taken into account.

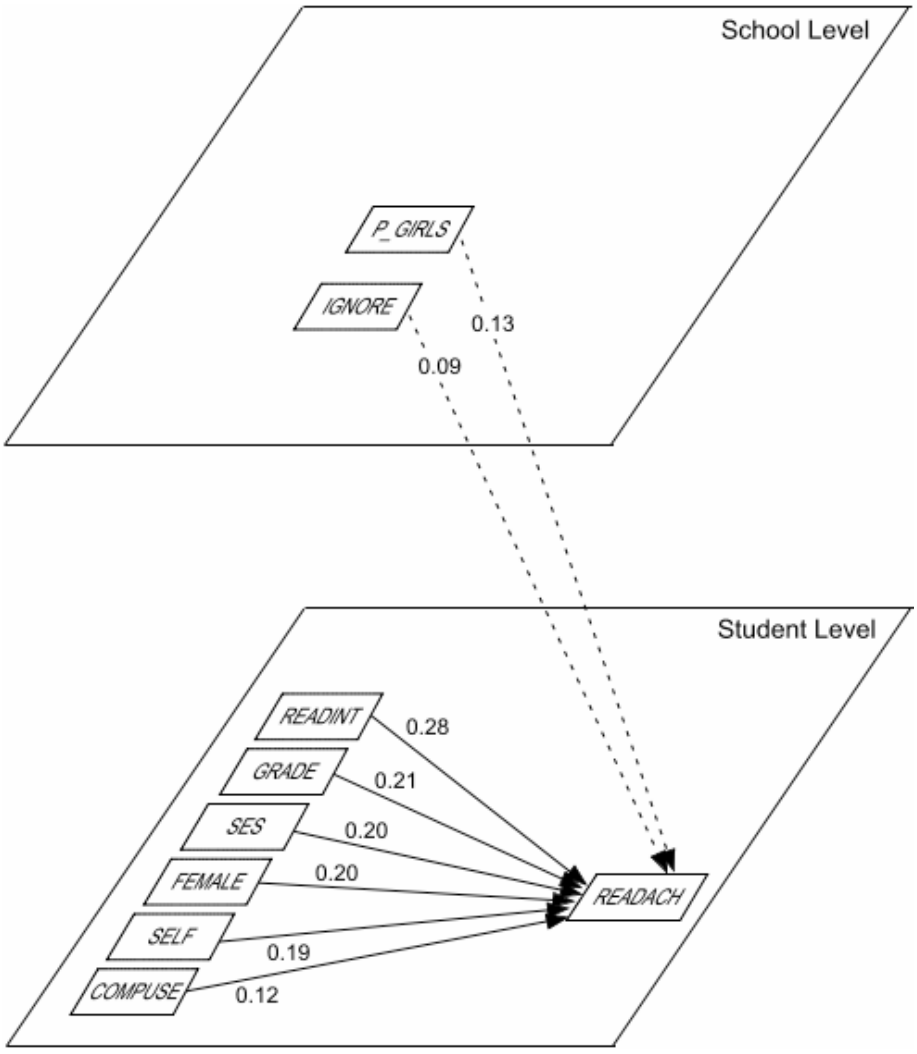


Figure 1: Final two-level HLM model for Finland for reading achievement

The three other significant predictors of reading achievement at the student level are GRADE (γ=0.21), SELF (γ=0.19) and COMPUSE (γ=0.12). Thus, 15-year-old students who are enrolled at Grade 9 rather than Grade 8 are better readers. This is not surprising as the students in the higher grade will have had more exposure to reading (at school) which, in turn, is bound to increase reading achievement (Lietz 1996). Likewise, reading achievement is influenced positively by students with

greater self-confidence in their verbal and mathematical ability who are convinced that they can perform well if they put their mind to it. Finally, using a computer regularly contributes to a higher performance in reading.

The only other school-level factor with an effect on reading achievement in addition to P_GIRLS (γ=0.13) is IGNORE (γ=0.09). This result indicates that students read marginally better in schools in which the principal reports that learning may be somewhat hindered as a result of teachers not addressing students' needs adequately - a result that is somewhat puzzling. An examination of the wording of the question and response options (SC19Q07) reveals that 91 per cent of all principals report some degree of hindrance as reflected by the frequencies by which the response options were chosen, namely 'not at all' (8%), 'very little' (57%), 'to some extent' (34%), 'a lot' (1%). Thus, this effect might reflect the tendency of Finnish principals to be more critical of their teachers than appears warranted given the performance of Finnish students in reading. Alternatively, the finding could support the hypothesis that schools run by principals with higher expectations regarding their teachers provide a context which fosters student achievement.

The fact that factors operate to influence reading achievement mainly at the student level in the Finnish model might be explained by the relative homogeneity of the Finnish school system in relation to other countries (OECD 2001; see also Table 5 below). It should also be kept in mind that the model did not include any teacher-level variables as teacher background information was not obtained in the PISA 2000 survey.

In summary, students in Finland can expect a higher score in reading achievement when they…

- spend more time on reading and read more diverse materials;
- are enrolled in the higher grade;
- come from a home with a higher socio-economic status;
- are female;
- have greater self-confidence;
- use computers more often for educational and programming purposes;

and when their school…

- has a higher proportion of female students; and
- is run by principals with higher expectations regarding the way in which teachers in their schools meet students' needs in their instruction.

Two-level HLM model of reading achievement for Germany

In Germany, results of the HLM analyses reveal that four factors directly have an impact on reading achievement at the student level while five factors operate at the school level to influence performance in reading. The final estimates of fixed effects together with their standard error and corresponding significance statistics are given in Table 4. In addition, Figure 2 illustrates the direct effects of student- and school-level constructs on reading achievement of 15-year-old students in Germany.

Table 4: Final estimation of fixed effects; two-level HLM model for Germany for reading achievement

Fixed Effects on READACH	γ-coefficient	standard error	t-ratio	p-value
Level 1/student-level effects				
GRADE	0.20	0.02	12.328	0.000
SELF	0.06	0.02	3.741	0.000
EFFORT	0.05	0.02	3.036	0.003
FEMALE	0.05	0.01	4.234	0.000
Level 2/school-level effects				
LOWACH	0.27	0.04	6.584	0.000
UPSEC	-0.25	0.04	-5.648	0.000
SCHSIZE	0.22	0.05	4.434	0.000
STUDBEHA	-0.13	0.04	-3.467	0.001
PRIMARY	0.12	0.04	3.103	0.002

Results show that reading achievement (READACH) is influenced directly at the student level by GRADE (γ=0.20), SELF (γ=0.06), EFFORT (γ=0.05) and FEMALE (γ=0.05). All these effects - with the exception of GRADE - are relatively small which reflects the lower importance of student-level factors compared with school-level factors when explaining differences in student performance in Germany because, in Germany, relatively less variance is associated with the student level when compared with other OECD countries (OECD 1998; see also Table 5 below).

The effect of GRADE on reading performance (γ=0.20) reflects that 15-year-old students who are enrolled in higher grades (ie., Grade 9 or 10) are better readers. Reasons for this might include the fact that these students have attended school one year longer than their counterparts in Grade 8 and the fact that 15-year-olds in Grade 8 are likely to have repeated a grade during their schooling as a result of their overall lower performance.

Furthermore, students' self-concepts and attitudes both impact positively on reading achievement (γ=0.06). Thus, students with greater confidence in their abilities with a greater inclination to persevere and better study time management perform at a higher level. Likewise, female students in Germany are better readers than their fellow male students (γ=0.05). German students who experience teachers with high standards regarding school- and homework tend to perform slightly better in reading (γ=0.05).

The effects of school-level variables on READACH are, in general, stronger than the effects of student-level variables in Germany. Thus, LOWACH (γ=0.27), UPSEC (γ=̃0.25), SCHSIZE (γ=0.22), STUDBEHA (γ=̃0.13) and PRIMARY (γ=0.12) contribute significantly to explaining differences in achievement between students. The effects of the two constructs UPSEC and PRIMARY may be perceived as reflecting the different school types in Germany: The negative effect of UPSEC indicates that if the school in which a student is enrolled, in addition to the lower secondary section - the level of schooling attended by German 15-year-olds -also contains an upper secondary section, such a school is, most likely, an academically-oriented 'Gymnasium' with students achieving at a higher level. Conversely, if there is no upper secondary level found in a particular school this affects reading performance of 15-year-old students negatively: they are, most probably, students attending a 'Realschule' or 'Hauptschule' - school types with an overall lower performance (Baumert et al. 2003a).

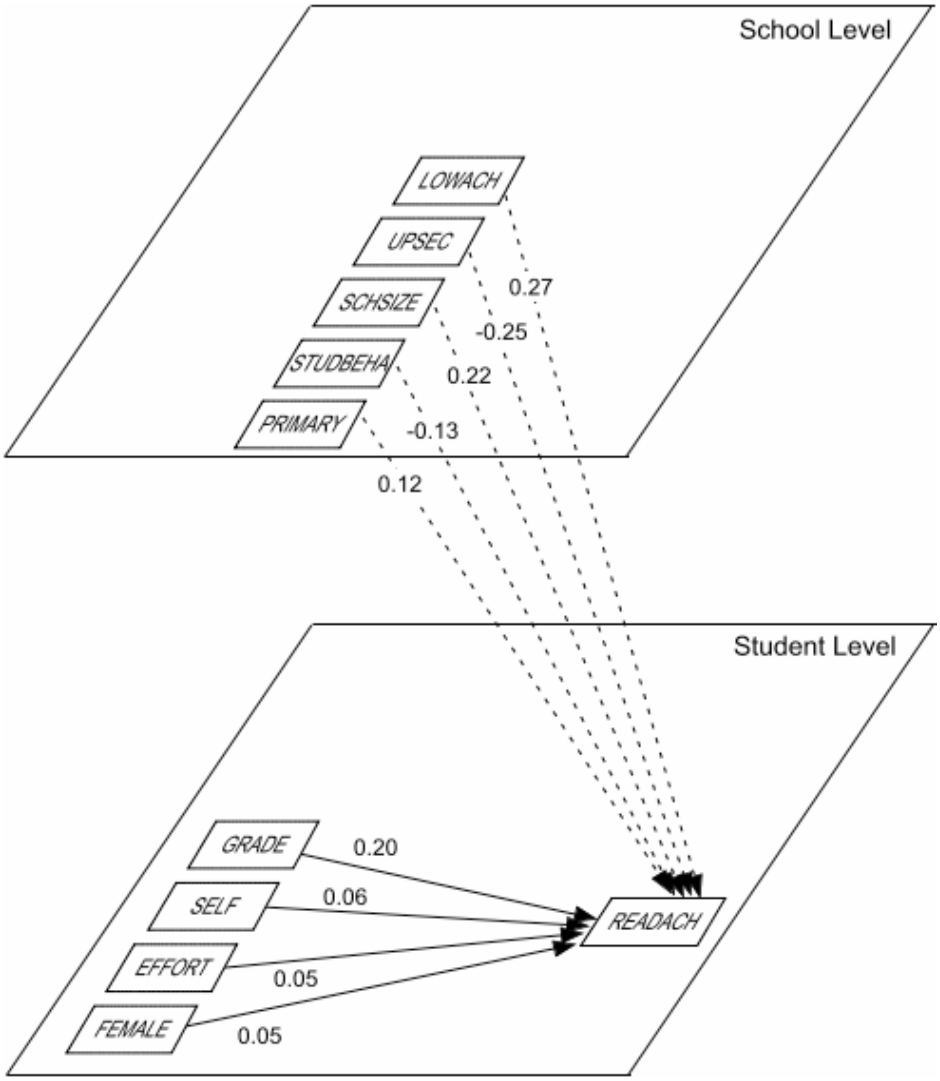


Figure 2: Final two-level HLM model for Germany for reading achievement

A similar line of argument applies to the effect of PRIMARY on reading performance (γ=0.12). The positive effect means that students who attend schools without a primary school section perform at a higher level than students enrolled in schools with such a section. Again, it can be argued that this reflects the effect of the segregated school system in Germany. Schools which cater for the more vocationally-oriented students (i.e. 'Hauptschule' or 'Realschule') frequently contain a primary school component whereas the academically-oriented high schools (i.e. 'Gymnasium') do not include the German primary Grades 1-4.

The strong effect of LOWACH on reading achievement (γ=0.27) may be understood in light of the above discussion. Schools that have to deal more often with low achievers, either by transferring them to another school, through ability grouping or

by providing special training in the language of instruction – reflecting a migrant background – are, more likely, to be a 'Hauptschule' or a 'Realschule' (Baumert et al. 2003a).

The negative school-level effect of STUDBEHA ($\tilde{\gamma}$=0.13) indicates that students in schools where the principal reports a stronger hindrance of learning due to students inadequate behaviour reading performance is negatively affected. It should be noted that this effect emerges, when the factors relating to the differences between academically- and vocationally-oriented schools in Germany are taken into account.

It can be concluded that school-level factors play a more pronounced role than student-level factors with regards to achievement in reading for students at the age of 15 years. At that age, in Germany, filter and segregation processes have taken place to allocate students to a particular school type.

At first glance, it looks as if the type of school fosters or inhibits reading performance in Germany per se. However, the two factors reflecting school type (ie., UPSEC, PRIMARY) should be regarded as 'placeholders' for socio-economic status of a student's home. For example, Baumert, Watermann and Schümer (2003b) report that, in Germany, the 15-year-old child of a parent of the highest occupational level is seven times as likely to attend an academically-oriented 'Gymnasium' than his or her fellow student from a working class background.

Thus, the strong effects of the two school-level factors on reading achievement suppress the effect of student-level SES. Such an effect has also been pointed out by Baumert et al. (2003b) who analysed several models using the extended German PISA national data set. As soon as school-type was included in the analyses, the effects on performance of socio-cultural structural characteristics in general and socio-economic index in particular decreased considerably.

In summary, students in Germany can expect a higher score in reading achievement when they

- are enrolled in the age-adequate grade and are, thus, non-repeaters;
- have sufficient self-confidence in their abilities and positive attitudes towards studying;
- perceive their teachers to have high standards with respect to students' work;
- are female;

and when their school…

- is academically oriented;
- has to take fewer measures to support lower achievers; and
- reports little or no disruption of school-life due to inappropriate student behaviour.

Table 5: Estimation of variance components and explained variance – two-level HLM models for Germany and Finland

FINLAND	Estimation of variance components between	
	students ($\hat{\sigma}^2$)	schools ($\hat{\tau}_\pi$)
number of cases	4864	155
fully unconditional HLM model	0.88	0.17
final two-level HLM model	0.68	0.14
Variance at each level:		
Between students	$\frac{0.88}{0.17+0.88} = \frac{0.88}{1.05} = 0.84$	
Between school	$\frac{0.17}{0.88+0.17} = \frac{0.17}{1.05} = 0.16$	
Proportion of variance explained by final two-level model:		
Between students	$\frac{0.88-0.68}{0.88} = \frac{0.20}{0.88} = 0.23$	
Between school	$\frac{0.17-0.14}{0.17} = \frac{0.03}{0.17} = 0.18$	

GERMANY	Estimation of variance components between	
	students ($\hat{\sigma}^2$)	schools ($\hat{\tau}_\pi$)
number of cases	5073	219
fully unconditional HLM model	0.55	0.64
final two-level HLM model	0.52	0.23
Variance at each level:		
Between students	$\frac{0.55}{0.64+0.55} = \frac{0.55}{1.19} = 0.46$	
Between school	$\frac{0.64}{0.55+0.64} = \frac{0.64}{1.19} = 0.54$	
Proportion of variance explained by final two-level model:		
Between students	$\frac{0.55-0.52}{0.55} = \frac{0.03}{0.55} = 0.05$	
Between school	$\frac{0.64-0.23}{0.64} = \frac{0.41}{0.64} = 0.64$	

Comparison of results for Finland and Germany

A comparison of the results of the two-level HLM analysis of factors influencing reading achievement in Finland and Germany reveals both, similarities and differences. At the student level, GRADE, SELF and FEMALE influence reading achievement positively in both countries. In other words, students who are non-repeaters, have greater confidence in their academic abilities and their abilities to control their learning and who are female, perform at higher levels in reading.

While the socio-economic status of the student's home emerged as a direct effect in Finland, any effect of student-level SES on achievement in Germany was suppressed by type of school factors, namely UPSEC and PRIMARY.

While interest in reading - overall higher for Finnish students - was found to be the strongest student-level factor influencing reading achievement in the Scandinavian

country, this factor failed to emerge as a significant factor in explaining differences between students in Germany.

It became evident that school-level factors were stronger and more numerous in number in Germany than they were in Finland. In contrast, student-level factors dominated the two-level HLM model in Finland whereas the direct effects of such factors in Germany, with the exception of GRADE, were rather small and scarce. It is, therefore, important to take a closer look at the variance proportions of each of the HLM models.

As indicators of the appropriateness of the HLM models for the two countries, the respective variance proportions for each level have been calculated. In Table 5 the variance estimates for the unconditional models and the final models are presented, together with the proportions of the variance available at each level and the variance explained at each level.

An examination of the variance proportions for the unconditional model reveals that in Finland more of the variance is associated with the student level (84%) whereas in Germany most of the variance is associated with the school level (54%) which is in line with previous results of German and Finnish school achievement data (OECD 1998).

The explained variance reflects the available variance at each level. Thus, in Germany, the model explains a greater proportion of the (larger) available variance at the school level (Germany: 64%; Finland: 18%). The reverse applies in Finland where the model explains more of the (larger) available variance at the student level (Germany: 5%; Finland: 23%). However, this result whereby only five per cent of variance can be explained at the student-level, yet a considerable 46 per cent of the variance in Germany is associated with the student-level, requires further attention. It might be interesting, for example, to investigate the extent to which - maybe - teacher- and teaching-related factors affect reading performance of 15-year-olds. However, the current international data sets do not allow such analyses as no data were obtained from teachers in the PISA-2000 study.

Overall, the two-level HLM model for Finland is able to explain only 22 per cent of the total variance - the inadequacy here clearly being lack of identifying appropriate student- and teacher-/teaching-related variables. The German model accounts for 37 per cent of the total variance and is, therefore, a slightly better representation of the many relationships and underlying effects which seem to operate in influencing reading achievement for 15-year-olds.

Conclusion

As could be shown using hierarchical linear modelling, a range of factors contributed to the explanation of differences in student reading achievement as assessed by the OECD PISA study. While the two-level HLM models identified some communalities of the way in which reading achievement is influenced in Germany and in Finland, substantial differences in explaining reading achievement in the two countries remain.

GRADE, which may be perceived here as a substitute for scholastic aptitude in that it reflects whether or not a student had repeated a grade, SELF and FEMALE were identified as factors common to both countries at the student level. While reading performance was influenced considerably by interest in reading, no significant effect emerged from reading interest on reading achievement in Germany.

Much of the differences between the two education systems apparently stemmed from the fact that in Germany the larger proportion of the variance was associated with the school level whereas most of the differences in performance between students in Finland were associated with the student level. Thus, in Finland remedial action will have to revolve around providing increased assistance at the individual student level.

In contrast, the findings for Germany underpin the strong segregating effect of the placement of students into the different school types in that country. If this observation serves as an artefact for the differences in socio-economic status remedial efforts may be difficult to introduce. However, since about 46 per cent of the total variance in Germany is found at the student level – which itself is not directly related to SES – options exist to improve reading performance. While a lot of the current political debate in the German media focuses on the extension of school hours ('Ganztagsschule'), results of the current analysis do not provide evidence in support of such a policy change. Instead, the evidence suggests that student performance in reading benefits from a strong emphasis by schools on achievement and on high expectations and standards with respect to student work.

In order to provide more detailed recommendations, further investigation could focus on introducing teacher and instructional variables into the model to examine the way in which such factors contribute to explaining differences in student performance. Whereas the PISA-2000 data collection did not include such variables, recent efforts to develop this assessment program further appear to address these issues.

References

Adams R., & Wu M. (eds) (2002). *Programme for international student assessment (PISA): PISA 2000 Technical Report.* Paris: OECD.

Baumert, J., Artelt, C., Klieme,E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., & Weiß, M. (Hrsg.) (2003). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* . Opladen: Leske + Budrich.

Baumert, J., Watermann, R., & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. *Zeitschrift für Erziehungswissenschaft, 6(*1), 46-71.

Elley, W. B. (1994). *The IEA Study of reading literacy. Achievement and instruction in thirty-two school systems*. Oxford: Pergamon Press.

Hofmann, M., & Lock, S. (2002). Von Dakar bis PISA: Zukunftsinvestition Bildung. *Entwicklung und Zusammenarbeit*, Nr. 8/9, August/September, 2002, 244-245. Retrieved April 14, 2003, from http://www.dse/zeitschr/ez8902-7.htm.

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.

Kultusministerkonferez vom 17./18. 10. *2002 PISA 2000 - Zentrale Handlungsfelder. Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern (Stand: 07.10.02)..*

Lietz P. (1996). *Reading comprehension across cultures and over time.* Münster/New York: Waxmann.

Lietz P., & Kotte, D. (2000). *The importance of economic literacy*. New York: Peter Lang.

Lundberg I., & Linnakylä P. (1993). *Teaching reading around the world.* Hamburg:

International Association for the Evaluation of Educational Achievement.
Organisation for Economic Co-operation and Development (1998). *Education at a glance*. Paris: OECD.
Organisation for Economic Co-operation and Development (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.
Organisation for Economic Co-operation and Development (2002). *Education at a glance*. Paris: OECD.
Organisation for Economic Co-operation and Development (2003a). *Education at a glance.* Paris: OECD.
Organisation for Economic Co-operation and Development (2003b). *Reading for change: Performance and engagement across countries: Results from PISA 2000*. Paris: OECD.
Purves, A. C. (1973). *Literature education in ten countries: International studies in education II.* Stockholm: Almqvist & Wiksell.
Raudenbush S. W., & Bryk A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (Vol. 2).* Thousand Oaks, California: Sage
Raudenbush S. W., Bryk A. S., Cheong, Y. E., & Congdon, R.T. (2000). *HLM5: Hierarchical linear and nonlinear modeling*. Lincolnwood, Illinois: Scientific Software International:
Sellin N. (1990). *PLSPATH version 3.01. Program manual*. Hamburg: University of Hamburg.

EPILOGUE

LIFELONG LEARNING AND THE PLACE FOR ICT: Learning and Research for the Twenty-first Century

John P. Keeves
Flinders University

It is an honour and a privilege for me to write about this subject for this volume that marks the occasion of my eightieth birthday. Before looking forward to consider the topic of 'Learning and Research for the Twenty-first Century', I find it of considerable interest and importance to look back over the past 40 years. In 1964, I was working on four projects at the ACER:

(a) the development of a new primary school mathematics course and the preparation of a guide book for primary school teachers;
(b) the planning and conduct in Australia of the First IEA Mathematics Study;
(c) supporting the planning, development, and evaluation of new science courses at all levels of schooling;
(d) the evaluation of new educational materials that catered for individual differences in the classroom.

The electronic computer suddenly entered my life and work. Twenty years earlier at the University of Adelaide I had been introduced to new fields of learning in new courses in matrix algebra, statistics, nuclear energy, and semi conductors. These four fields came together in the electronic computer, with huge demands for electrical energy, and the facility to calculate statistical parameters through the use of matrix

algebra in computer-based estimation. A few thoughts come quickly to mind in relation to these four projects.

1. *Primary school mathematics*

   The introduction of decimal currency and subsequently decimal weights and measures was essential for computer-based calculation and freed the primary school arithmetic curriculum of about one and half years of curricular time. It was clear that this time should be replaced with an introduction to mathematical ideas, and the subject of mathematics was introduced in primary schools.

2. *First IEA Mathematics Study*

   All data collected in this study were obtained on mark sensed answer sheets or had to be recorded on punch cards and sent in four separate sub-sample packages for processing with an IBM 360 computer at the University of Chicago. The computer output with correlation and regression estimates transformed the ways in which I thought about educational research.

3. *New Science Curricula*

   Following the great advances in science in the first 60 years of the twentieth century, science curricula needed very substantial redevelopment starting at the primary school level. Moreover, in order to maintain the necessary flexibility to incorporate future developments and to cater for students who learnt at different rates and with different teaching experiences it was essential to move from the use of hard cover text books to carded material or small booklets that could in the longer term be computer based.

4. *Evaluation of New Educational Materials*

   The new educational materials that were being developed in the United States and United Kingdom not only involved new curricula, but also sought to cater for individual difference. W. C. Radford the Director of the ACER had argued at an ANZAAS congress in 1959:

   > The proper handling of the individual differences between children still remains the most challenging problem in education … . But whatever our views on the matter of the contributions of ability and development, of nature and nurture, we know that in any normal teaching situation in a classroom, a teacher is faced with a class in which children have different preparations for the work to be done and the experiences to be offered, have different capacities to read, to figure, to manipulate materials whether concrete or abstract, to memorize and to express themselves, and different emotional drives to succeed, to participate, and to contribute. (Radford, 1961, p.3)

This work led me to evaluate carded reading materials produced by Science Research Associates (SRA) and Encyclopedia Britannica in Chicago and to review developments in the field of programmed instruction with linear or branching programs. The types of materials ranged from reading laboratories, language laboratories, programmed text books, mechanical and electronic teaching machines with sequentially planned programs, to cybernetic computer based machines, where the students steered their own way through the learning experiences provided.

In 1964, I wrote in a review:

> The possibility that a more sophisticated electronic device may ultimately be extremely useful has not been ignored by research workers in the computer field or by the electronic industry. (Keeves, 1965, p. 6-7)

And I concluded with the following statements:

> More recent developments in branching programs, the rapidly extending field of computer-based experiments in individual instruction, and a number of the fundamental issues regarding learning raised by programmed instruction, have therefore now begun to attract much of the interest that was first shown in experiments in the use of actual programs. (Keeves, 1965, p. 6-7)

In early 1965, I had the opportunity to spend five weeks travelling around North America and England to investigate developments in all four fields in which I was working. Meeting IEA colleagues in Chicago completely changed my ideas about educational research, and attendance at an AERA Meeting in Chicago at which Ben Bloom was President showed me the size and quality of the research enterprise in North America. I learnt that IBM had bought out SRA and engaged Ralph Tyler as their educational consultant. However, computer-based developments were slower than expected and IBM subsequently abandoned this work.

Nevertheless, Georg Rasch's ideas about measurement in education were exciting some of the graduate scholars with whom I worked in Chicago and were starting to give rise to new ideas about evaluation and assessment for the measurement of achievement and attitudes. I was fortunate to be able to return to Chicago and London for brief periods in 1966, 1967 and 1968 to keep abreast of these developments and to read a draft of Bruce Choppin's communication to *Nature* on item banking that would enable more flexible approaches to the assessment of student learning.

New Approaches to Assessment

Seven years later in 1971, just before I went overseas to work in Stockholm for a year, the Headmaster of a large Melbourne school, B. W. Hone, who had led the movements towards the rebuilding of science laboratories, the development of new science curricula and subsequently the provision for every secondary school in Australia of a new library took me to lunch and discussed with me his vision of flexible programs of assessment to measure student learning and to guide individualized teaching. In Stockholm in 1972, we argued over cups of coffee and lunches about criterion referenced assessment, Rasch scaling and large scale monitoring of educational achievement programs that were becoming possible with the increasing power of computers. Torsten Husén and Neville Postlethwaite who were guiding the IEA enterprise, Gilbert Peaker, who had undertaken a review of Rasch's work for the British Ministry of Education and David Walker who had worked in Edinburgh with D. N, Lawley, who had developed in the early 1940s the basic ideas of item response theory and maximum likelihood estimation, led these discussions. In 1973, I went back to North America to review developments in the field of item banking in Chicago and Los Angeles, but was disappointed to learn that the field was still restricted by the lack of power of the computers available.

At ACER we continued to develop item pools with a strong diagnostic component at all levels of schooling that helped to improve greatly the assessment materials employed at the school level. Alan Larkin and John Connell, my colleagues at Flinders University, were among the workers on this project. However, neither Rasch

scaling, nor computer storage and delivery were sufficiently far advanced to assist with the implementation of the ideas for assessment that we had in mind.

In 1974, the ACER was commissioned by the Chairman of a Commonwealth Government Committee of Inquiry to conduct a testing program, that would be Australia-wide at the 10- and 14-year old levels, that would be based on the ideas of criterion-referenced assessment and Bayesian statistics, and that would assist in raising performance in literacy and numeracy in Australian schools. This work gave rise to two major developments that are continuing to the present time:

(a) the Basic Skills Testing Programs that are operating in all states of Australia and are Rasch scaled, now that appropriate computer programs have been developed;
(b) the Longitudinal Studies of Australian Youth, that were able to build on measures of student achievement at the 10- and 14-year old age levels.

These studies have opened up the field of monitoring educational outcomes in Australia.

Meanwhile, the design of computers was advancing, until by the early 1980s desk top computers were becoming available and main frame computers had increasing power, but were still unable to support the widespread use of item banking and more flexible approaches to learning and student assessment that were needed in schools.

The Ideas of Lifelong Learning

Ideas that are at the present time transforming education world-wide are those of lifelong learning and recurrent education. Torsten Husén in Stockholm has had a guiding hand in this work through educational reform in Sweden, through UNESCO, and its institutes both in Paris and Hamburg, and more recently through OECD. The ideas of lifelong learning and the monitoring of educational change were frequently argued over lunch in Stockholm and it is not surprising that the OECD monitoring programs associated with *Education at a Glance* and PISA were introduced and are being run by former students of Torsten Husén and Neville Postlethwaite from Stockholm and Hamburg. It is to Sweden that we turn today to see lifelong learning in operation in its most highly developed forms, and to OECD to lead the development of monitoring educational outcomes.

Education in South Australia

A little over 20 years ago, I had the opportunity to consider, among many other issues, the possibilities associated with the advancement of lifelong learning and the use of computers in education. The two ideas came together in the following words.

> The possibilities of fostering non-traditional modes of education and the development of non-formal approaches to education … are likely to be greatly enhanced by new systems of communication. …Such possibilities will probably be developed by educators working outside the school system. … Nevertheless, it may be necessary … to recognize and undertake new roles as technology is developed for education … to make non-formal methods of education a reality through the use of telecommunication networks and computer based systems of learning and instruction. It is now possible that the new frontiers of education will be in the field of non-formal education and outside the educational institutions of schools, colleges and universities as we know them. (Keeves, 1982, p. 36)

So little thought was being given in Australia in the early 1980s to the use of computers in education that the ACER asked Jonathan Anderson to undertake a review of possibilities in this field. This led to the publication of a monograph *Computing in Schools: An Australian Perspective* (Anderson, 1984). Jonathan continues to work on these issues for UNESCO (Anderson and van Weert, 2003).

During the past 20 years the fields of lifelong learning and information and communications technology have emerged as separate domains, growing and changing and being guided largely by the worlds of commerce and industry, rather than being directed and developed together in the fields of formal education.

The Advancement of Neuro-Science

In 1964, I decided to buy, although I had access to the best educational library in Australia at the ACER, two books, one by McVicar Hunt titled '*Intelligence and Experience'* and the other Flavell's work '*The Developmental Psychology of Jean Piaget'* and shortly afterwards I was given a copy by the author of '*Stability and Change in Human Characteristics'.* These books changed my ways of thinking about the brain and human development, leading me to accept the proposition that the effects of the educational environment in developing the operations of the brain continued throughout life. This was a bold proposition that is only now being supported by the emerging field of neuro science. I await with great interest the next OECD publication, that Denis Ralph has contributed to, due in a few months time and continue to search for an informed work that summarises what is known in the field and its implications for education, because I believe that the field of neuro science will have substantial implications for education at all levels.

Looking towards the Future

The introduction of information and communications technology into schools during the past two decades has been largely divorced from the teaching of the school curriculum, even in the subjects of science and mathematics, where teachers might be expected to be more familiar with ICT. Moreover, science and mathematics teachers have been reluctant to be concerned with issues that involve the application of their disciplinary knowledge to technological problems, or to consider the ways in which ICT might enhance the presentation and learning of disciplinary knowledge. As a consequence, the focus of technology teaching has been largely on ICT and often merely on the development of word processing skills, and the use of spread-sheets and the internet, without recognition of the many ways in which electronic computers are transforming the conduct of mathematical and scientific inquiry, or the ways in which computer based technology is being applied in industry and commerce. The time would appear to have come for mathematics, science and technology teachers to reshape their curricula in order to integrate the use of technology, not only into the processes of teaching and learning, but also into the processes of doing mathematics and science, with recognition that the dynamic power of computerized technology is capable of changing modes of thinking and learning by students in schools, in working life, and in lifelong learning and personal development through searching for relevant information and using that information in new ways. The establishment of the Australian Science and Mathematics School (ASMS) at Flinders University is an important step in this direction in South Australia.

A case must be argued most strongly for greater awareness in educational circles of the growing potential for the educational use of computers and ICT not only in the classroom, but in the fields of mathematics and science investigation and inquiry, as

well as its uses in the home, non-formal and informal learning environments, and in the workplace in many different forms. Sivakumar Alagumalai is starting to argue this case rigorously and well in Australia, and in particular, in South Australia.

This leads to the consideration in our teaching in university classes of such uses as:

(a) individualizing instruction,
(b) diagnostic assessment with immediate feedback and informed correctives,
(c) reflective thinking in learning with different modes of immediate verification,
(d) experimental design that involves the testing of models by simulation,
(e) inductive thinking by the control of experiences provided or through simulation,
(f) enrichment experiences through access to video presentations,
(g) collection, storage and presentation of data,
(h) computing where complex calculation is required,
(i) searching the internet for information and selecting information for appropriate use, and
(j) robotic performance of complex tasks in gathering data.

The Investigator Learning Laboratory

The challenge during recent months has been to try to extend the different uses of computers and ICT in our classes in the Investigator Learning Laboratory. Some of our research efforts have been largely informal. However, several small pilot studies are in progress.

1 A review of research (by Ting Seng Eng) into the gains in achievement arising from teaching through computer assisted learning in the classroom has reported modest gains, but with a considerable reduction in time involved in learning. Little research of this type has been reported in Australia.
2 The systematic observation and analysis (by Julian Forbes and I Gusti Ngurah Darmawan) of students learning at the computer when solving computer presented problems and the analysis of video and audio records to examine hypothesized relationships.
3 Planning an evaluation of the learning experiences provided in mathematics at the Noel Baker Mathematics Centre by Anthony Harradine.
4 Developing an Item Bank in the fields of literacy and numeracy at the Grade 3 and Grade 5 levels for computer adaptive diagnostic and remedial instruction. (by Jonathan Carpenter, Hungi Njora and Ting Seng Eng).
5 The development and use of simulation and games exercises (by I Gusti Ngurah Darmawan) to assist with the learning of statistical ideas and principles.

We argue that the major problems associated with computer assisted learning do not seem to have been adequately investigated, namely: Do students change in the ways in which they think and work as a consequence of learning in a computer assisted environment using ICT that provides instruction through (a) visual presentation, (b) audio presentation, (c) textural presentation, and (d) dynamic control of the content and pace of learning? I believe that information that would assist investigation might be made available from video recording as well as audio recording, and computer based records of computer use. However, the field of neuro science may well be capable of providing information on the workings of the brain in such learning

situations. There is need for research into these problems with students of all ages and all levels of intellectual capacity.

With rapid changes in the fields of lifelong learning and the use of computers and ICT in education, there is an urgent need to monitor in a systematic way not only the inputs to education, but also the outcomes of education at all levels of learning in schools, universities, technical and further education centres, non formal education in industry and commerce, and informally through museums and the mass media. Only with sound knowledge of the changes that are taking place, through innovation and reform can appropriate plans be laid for the future. This is the important principle of design-based research.

In conclusion, I would like to express my deep appreciation to the University, to the School of Education, and to the Flinders University Institute of International Education for providing opportunities and facilities for undertaking teaching and research, the two basic pillars of university life, over the past 14 years.

References

Anderson, J. (1984*) Computing in Schools: An Australian Perspective*, Australian Education Review No. 21, Melbourne:ACER.

Anderson, J. and van Weert, T. (2003) *Information and Communication Technology in Education: A curriculum for schools and programme of teacher development*, Paris:UNESCO.

Keeves, J. P. (1965) *Programmed Instruction in Australia, 1965*, Information Bulletin No. 2, 1965, Series No: 44, Melbourne:ACER, p.4.

Ibid p.6-7

Keeves, J.P. (Chair) (1982) *Education and Change in South Australia. Final Report. Committee of Enquiry into Education in South Australia*. Adelaide:South Australia, Government Printer, p. 36.

Radford, W. C. (1961) Individual differences and classroom organization, *Australian Journal of Education,* 5, 3.

Shannon Research Press
Adelaide, South Australia
ISBN: 1-920736-11-5