

*Research Conferences*

*2004 - Supporting Student Wellbeing: what  
does the research tell us about social and  
emotional development of young people?*

---

Australian Council for Educational Research

Year 2004

---

Measuring the Social Outcomes of  
Schooling: What Does ACER Research  
Tell Us?

Margaret Forster  
ACER

# Measuring the social outcomes of schooling: What does ACER's research tell us?



**Margaret Forster**

*Australian Council for Educational Research*

Margaret Forster is Research Director of the Assessment and Reporting Research Division at ACER. Ms Forster has extensive experience in the area of assessment and reporting and works as a consultant nationally and internationally. She has direct experience in the development of assessment materials for a number of assessment programs including the New South Wales Basic Skills Test, the Queensland Core Skills Testing Program and the Western Australian Monitoring Standards in Education Program. She conceptualised and co-authored the first Developmental Assessment Resource for Teachers (DART English Upper Primary) and is co-author of the ACER Assessment and Values Questionnaire, and the widely acclaimed Assessment Resource Kit (ARK) materials.

Recent national consultancies on the revision and implementation of assessment and reporting frameworks include work with the Western Australian Curriculum Council and Education Queensland. Recent international consultancies include work for the World Bank Institute, the Peruvian Ministry of Education, AusAID in Papua New Guinea, UNICEF, the Chilean Ministry of Education, the Hong Kong Curriculum Development Institute and the Hong Kong Institute for Education.

For the past eight years the Australian Council for Educational Research (ACER) has been at the forefront of work in Australia to refine conceptualisations of the social outcomes of schooling, and to explore ways to measure the social and emotional development of young people.

Contexts for ACER's work in this area include: consultancies for individual schools; the development of questionnaire instruments as part of fee for services provisions for schools; assistance for state ministries of education; and development work for a variety of tertiary assessments. For example, in 1998, ACER worked with one school to formally assess and monitor the moral, ethical, social and emotional development of its students through secondary school (Forster & Masters, 2002). In 2001 ACER began work with the Education Department of Western Australia (EDWA) to develop instruments to address the social outcomes of schooling within EDWA's system-wide monitoring program that collects evidence of student achievement at Years 3, 7 and 10. These instruments measure interpersonal skills (collaboration, conflict resolution, and communication skills); intrapersonal skills (feelings in relation to self and self-management skills); and social, moral and ethical development, with the intention of reporting the achievements of students from Year 3 to Year 10 on a single scale. And ACER is currently working with the South Australian Education Department to define 'student wellbeing' across the compulsory years of schooling.

This paper reflects on ACER's research into the conceptual and practical

challenges of refining definitions of social and emotional growth in a way that allows the definitions to be operationalised as valid, reliable and useful measurement instruments.<sup>1</sup>

## **The development process**

In exploring ways to measure growth in the social outcomes of schooling in general, and social and emotional development in particular, ACER's approach is consistent with the approach taken in literacy, numeracy and subject areas of the curriculum. We begin by mapping the intended variable or underlying dimension we wish to measure. That is, we begin by articulating what it means to get better, improve or grow in that area of learning. What would a low level of achievement along the variable look like? How would we recognise high-level achievement?

Once we are clear about the intended variable, we select the most appropriate assessment method for collecting evidence of student achievement along that variable. Should we collect evidence through a written test (e.g., multiple-choice or extended response)? Would a performance assessment be more appropriate (e.g., a spoken response or a dance sequence)? That is, we decide which assessment method will provide us with the best evidence of the learning outcomes of interest.

Next we develop assessment tasks that address the variable. In this way, we give our initial conceptualisation explicit meaning (or definition). Each task we develop is easier or harder or of the same level of difficulty as the other

<sup>1</sup>This paper draws on the work of a number of ACER staff including Geoff Masters, Prue Anderson, Julian Fraillon, Doug McCurry, Jennifer Bryce, and Neville Chiavaroli.

---

tasks we are developing along the single underlying variable. Thus the tasks we develop represent our expected ordered definition of the intended variable. They are an operationalisation of the intended variable.

In some areas of learning, the conceptualisation and operationalisation of the variable of interest are relatively straightforward. In the case of the two variables 'social' and 'emotional' growth the task is extremely challenging.

### **Challenge 1: conceptualisation – understandings or behaviour?**

In conceptualising both social and emotional growth ACER's approach has been first to draw on the relevant literature (e.g., Helmke, 1994; Harter, 1990; Marsh, 1989; Goleman, 1995; Kuhl & Kraska, 1994; Mayer & Salovey, 1993; Cacioppo & Gardner, 1999; Eisenberg, 2000; Zahn-Waxler et al., 1992). Next, depending on the context for the work, researchers examine the evidence about what schools can reasonably have influence over (e.g., Lee & Robbins, 1995; Rutter et al., 1980; Mooij, 1999; Battistich et al., 1997; Helmke, 1994; Ainley et al., 1998) and the social outcomes embedded in state and territory key learning area documents (Curriculum Council, 1998; Curriculum Corporation, 2000). Our research suggests that while there is reasonable consensus in the literature and in state and territory documents about what the 'social outcomes' of schooling are not (e.g., not diagnostic assessments of psychiatric illness), there is no clear single definition of what they are, except as broad clusters: emotional (intrapersonal), social (interpersonal),

and physical/behavioural.

A fundamental challenge for test developers is the relationship between understanding and behaviour. Do we conceptualise social and emotional growth as increasing sophistication of social and emotional understandings or insight (that may or may not be acted out in increasingly appropriate social behaviour and emotional response)? Or do we conceptualise social and emotional growth as increasingly appropriate, acceptable or valued social behaviour and emotional response (regardless of whether the behaviour or response is driven by mimicry, genuine insight, particular values, or personality)? Is it possible to behave in an appropriate way and not have sophisticated understandings or insight? Is it possible to have sophisticated insight and behave or respond inappropriately? In conceptualising *growth*, do we focus on behaviour or the understanding? Which instruments will provide us with a more valid measure of social and emotional growth: those that address behaviour or those that address understanding?

In undertaking research into the development of valid and reliable measurement instruments that can be used over time to monitor growth, ACER has addressed both conceptualisations. Both approaches have their strengths and weaknesses.

### **Challenge 2: operationalising 'understanding'**

A conceptualisation based on understanding describes variables of increasing sophistication of reasoning about and insight into social or emotional interactions. The lower end

of these variables is illustrated by students' understanding of the surface motivations driving and impacting on people's behaviour; the higher, by their understanding of the varied and sometimes contradictory emotions and motivations that drive and affect people in complex social and emotional situations.

Two kinds of paper and pen assessments have been developed to gather evidence of achievement along these social and emotional variables: scenarios with multiple-choice questions, and scenarios with short open-ended questions. Scenarios include video extracts, written dialogues, and extracts from novels. The scenarios present small vignettes that highlight issues, problems, or social and moral dilemmas that require differing levels of complexity of social reasoning to interpret and possibly resolve.

The continuing refinement of this approach addresses a single underlying issue: What is it that is being valued in students' responses? We know that behaviour is more or less appropriate given a cultural context. We know also that a range of behaviours (acquiescent through to challenging) can be appropriate, depending on circumstance. We also know that in the context of a paper and pen test, it can be difficult to provide sufficient context for a single defensible response, unless formulaic scenarios are presented that require formulaic, knowledge-based responses.

To address this issue, wherever possible, longer extracts (a paragraph) are used rather than brief ones (one or two sentences). Longer extracts provide sufficient context to reduce the number of plausible interpretations of a situation. Depending on the focus of

---

the question, we also sometimes use open-ended responses rather than multiple-choice, and scoring guides allow for alternative 'right' answers based on plausible reasoning, as well as for partial credit scoring of the quality of answers. For example, partial credit scoring sometimes gives additional credit to responses that appreciate the values that underpin people's varying perspectives on the situation, rather than responses that appreciate a single perspective only.

### **Challenge 3: operationalising 'behaviour'**

A conceptualisation based on *behaviour* describes variables of increasingly appropriate social or emotional responses and interactions. The lower end of these variables is illustrated by students' limited repertoire of behaviours, or inappropriate behaviours and reactions; the higher by a sophisticated repertoire of behaviours and reactions that can be observed in complex social and emotional situations.

Two kinds of paper and pen assessment and one performance assessment have been developed to gather evidence of achievement along these social and emotional variables: Student and teacher questionnaires (e.g., Masters & Forster, 2000) provide evidence of students' attitudes and values as generalised expressions of their likely behaviours; small group discussions provide direct evidence of how students behave in constrained contexts; and constrained teacher judgments provide evidence of students' behaviour.

A common challenge to the questionnaire assessment of values is

that respondents may provide what they know is the 'best' answer regardless of what they really think. In ACER studies, where data are reported at the group or subgroup rather than individual level, questionnaires for students are completed anonymously and in some studies every 10th or 11th questionnaire statement is worded negatively to identify students who are simply 'agreeing' or 'disagreeing' with every statement, regardless of its content. A computer check highlights these students and their data are discarded. (Responses to the negatively worded items are not included in the analyses.)

A particular challenge to questionnaire assessment in some ACER work has been the intention to report student achievement from Year 3 to Year 10 on each variable on a single scale. Our data suggest that Year 3 students, on average, have higher levels of optimism, self-confidence and self-efficacy than students in Year 7, and students in Year 7, on average, have higher levels than students in Year 10. This finding does not mean that the measures are invalid. It may be true that, given their relatively limited real life experience, the relatively sheltered contexts in which they relate to others, and their relatively limited cognitive development, Year 3 students are more optimistic and self-confident and have stronger self-efficacy than Year 7 and 10 students. It is also possible that this steady decline between Year 3 and 10 in optimism, self-confidence and self-efficacy is how students grow in the long run towards a stronger view of self as they experience and reflect on a greater range of social and emotional interactions.

An advantage of teacher reflections on student behaviours is that teachers have less reason to provide the 'best answer'.

On the other hand, they may lack the opportunity to observe student behaviour in a broad range of contexts. ACER data using teacher reflections suggest that the pattern observed for students' self-evaluations of optimism, self-confidence and self-efficacy is mirrored by teachers' evaluations of students' ability to empathise with others and to behave in a cooperative way.

Observing students in the context of small group discussion is an attractive approach because it provides direct evidence of their interactive skills and understandings; yet it brings its own challenges. In most test development contexts, researchers are careful to select material that is not confronting or emotionally challenging for students, so that students will be able to demonstrate the best of what they know and understand without distraction or distress. For example, stimulus passages to which students might have a strong emotional reaction are not included in reading tests.

ACER research suggests that measures of social and emotional interaction in the context of small group discussion distinguish between the behaviours of Year 3 and Year 7 students but are less successful at distinguishing between the behaviours of Year 7 and Year 10 students. A possible explanation for this finding is that students with more sophisticated negotiation and mediation skills are unable to demonstrate them given that sufficiently challenging and complex stimulus are excluded from a study for fear that they might be too emotionally charged and disruptive.

ACER is now trial testing a new instrument, based on teacher judgments, for collecting evidence of student's feelings about self, and of self-

---

management behaviours. The initial conceptualisation of the variable and the development of the instrument are built on feedback from teachers who were asked to describe observable behaviours that they judged to represent high, medium and low levels of positive feelings about self and self-management skills. The descriptions that teachers provided have been used as the basis for a set of items that operationalise the intended variable.

The items are grouped under the aspects of 'autonomy', 'enterprise' and 'self-awareness' and have been written with either a predominantly primary school context or a predominately high school context. Some are unique to a single year level and other items are being administered to two or more year levels. Items consist of a context and a set of item specific, described response categories. For example, three response categories for the context 'accesses own materials' read: 1. 'The student consistently locates personal and school related items efficiently as they are needed (such as jumpers and jackets, lunch boxes, stationery and books)'. 2. 'The student locates personal and school related items such as jumpers and jackets, lunch boxes, stationery and books with some effort or inefficiency'. 3. 'The student frequently is unable to locate their personal and school related items and usually spends time and effort searching for them.'

Researchers intend that this instrument will provide teachers with sufficient context for them to make reliable judgements of students' behaviours, and that the mix of year-level specific items and items that are administered across year levels will provide a solid foundation for a single reporting scale across the years of schooling.

## Conclusion

What does ACER's research into the measurement of the social outcomes of schooling in general, and the measurement of social and emotional growth in particular, tell us? It tells us that, in principle, we can approach the measurement of these dimensions in the same way as we approach the measurement of any other underlying variable, although the work in this area is particularly challenging and we need to remain vigilant in reflecting on what we are valuing in students' responses.

ACER research also tells us that we need to be cautious about the inferences we draw from these measures. In particular, we need to be cautious about the kinds of understandings we infer that students have, and the relationship between those understandings and behaviour. Understanding social interactions, and the capacity to articulate these interactions, is not necessarily an indication of likely behaviours. Nor is it necessarily an indication of self-insight.

Finally, we need to be cautious about generalising from the sub-domains of social and emotional growth that have been addressed to a larger domain of student wellbeing which may be defined in different ways.

The presentation accompanying this paper is structured around examples of tasks designed to collect evidence of students' social and emotional growth. Discussion will focus on the way in which the tasks assist us to define the underlying variables (that is their conceptual strengths and weaknesses), and their usefulness for providing schools and systems with information that will assist in the support of students' development.

## References

- Ainley, J., Batten, M., Collins, C. & Withers, G. (1998). *Schools and the Social Development of Young Australians*. Camberwell, Vic: Australian Council for Educational Research.
- Battistich, V., Solomon, D., Watson, M. & Schapps, E. (1997). Caring school communities. *Educational Psychologist*, 32 (3), 137–52.
- Cacioppo, J. & Gardner, W. (1999). Emotion. In J. Spence, J. Darley & D. Foss, *Annual Review of Psychology*, 191–214. Palo Alto, CA: Annual Reviews.
- Curriculum Corporation. (2000). Curriculum Corporation seventh national conference: conference papers' pages. Carlton South, Vic: Curriculum Corporation (Australia).
- Curriculum Council. (1998). *Curriculum Framework for Kindergarten to Year 12 Education in Western Australia*. Perth: Curriculum Council.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. In S. Fiske, (ed.), *Annual Review of Psychology*, 665–9, Palo Alto, CA: Annual Reviews.
- Forster, M. & Masters, G. (2002). Accepting the challenge: assessing the moral and ethical outcomes of schooling in S. Pascoe, (ed.), *Values in Education College Year Book 2002*. Deakin West, ACT: The Australian College of Educators.
- Goleman, D. (1995). *Emotional Intelligence*. New York: Bantam.
- Harter, S. (1990). Adolescent self and identity development, in S.S. Feldman & G.R. Elliott, (eds), *At the Threshold: The developing adolescent*. Cambridge, MA: Harvard University Press.

- 
- Helmke, A. (1994). Self-concept, development of. In T. Husen & T. Poselthwaite, (eds), *International Encyclopedia of Education*, 9, 5390–4, London: Pergamon.
- Kuhl, J. & Kraska, K. (1994). Self-regulation in learning. In T. Husen & T. Poselthwaite, (eds), *International Encyclopedia of Education*, 9, 5400–7, London: Pergamon.
- Lee, R. & Robbins, S. (1995) The relationship between social connectedness and anxiety, self-esteem and social identity. *Journal of Counselling Psychology*, 45 (3), 338–45.
- Marsh, H.W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, 81, 417–30.
- Masters, G. & Forster, M. (2000). *Attitudes and Values Questionnaire*. Camberwell Vic: Australian Council for Educational Research.
- Mayer, J.D. & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, 17, 433–42.
- Mooij, T. (1999). Promoting prosocial pupil behaviour: a multilevel theoretical model. *British Journal of Educational Psychology*, 69 (4), 469–78.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1980). *Fifteen Thousand Hours*. London: Open Books.
- Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E. & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, 28, 1038–47.s