

Developing Formal Mathematical Assessment for 4- to 8-Year-Olds

Brian Doig
Deakin University

The assessment of children in their years before school and their first years of school has been, traditionally, informal. Further, assessment of children's mathematical skills at this level has been infrequent compared to social, emotional and physical assessments. However, there are contexts where reliable, valid, standardised data from assessment in mathematics are required. This paper outlines the development of two assessment tools for mathematics that were originally developed for such contexts. Item Response Theory (IRT) analyses enabled the construction of assessment forms that address the range of abilities of 4- to 8-year-old children, and provided the scales used for constructing formative and summative reports of achievement. A description of the development of the assessment tools and the IRT analysis that provides the reporting formats are presented together with some research uses of the tools.

This article describes the development of two mathematics assessment tools suitable for use at the pre-school level, where formal assessment is rare. The article also describes how two issues in classroom assessment that challenge the development of assessment tools at this level were overcome. These issues are: the wide range of mathematical understandings of children of this early age; and the need to provide reporting to teachers that will assist in planning for appropriate mathematical learning experiences for the children assessed.

The purposes of this article are: to demonstrate the possibilities for standardised assessment in the early years; to show how Item Response Theory (IRT) analyses can provide reporting formats for assisting early years professionals; and to describe some examples of two assessment tools in research contexts. While there are some necessary differences in the detail of these assessment tools, the development of one parallels the other. In some of the following sections both tools are described separately, and in other sections the tools are discussed together. When examples are used, the source assessment tool is indicated.

Background

The dominance of constructivist approaches to mathematics learning in the early years of schooling has begun to change perspectives on effective practice for young children. The earlier Piagetian notions of stages of development are giving way to the realisation that effective learning takes place when a rich, supportive environment offers challenge and relevance (e.g., Cook, 1996; Doig, McCrae, & Rowe, 2003). Further, there is a growing awareness among early years professionals of the wide range of mathematical

capabilities of children entering pre-school and school (Aubrey, 1997; Bottle, 1998; Doig et al., 2003; Groves & Cheeseman, 1993; Munn, 1994; Nixon & Aldwinckle, 1997).

The Australian Council for Educational Research (ACER) conducted a study examining the relationship between age of entry to school, school structure, curriculum, teacher expectations, and student outcomes in language and mathematics (Curriculum and Organisation in the Early Years of School, 1997-1999). For the purposes of this study, it was necessary to have measures of developmental progress that were applicable to children at the pre-school level and in the early years of schooling, were easy to administer and score, and provided readily interpretable results. The budget for the study prohibited the use of individually administered instruments, and it was considered unlikely that teacher observations over an extended period of time would provide reliable data. As no suitable instruments meeting these criteria could be found, it was necessary to develop measures specifically for use in the study. The two different tools that were created were: the Who Am I? developmental assessment material (de Lemos & Doig, 1999a, 1999b); and I Can Do Maths (Doig & de Lemos, 2000).

Who Am I? was developed from previous research on the use of copying tasks for the assessment of developmental level and school readiness (de Lemos, 1973, 1980; de Lemos & Larsen, 1979; de Lemos & Mellor, 1991). This work was subsequently used as a basis for developing a measure of school readiness based on copying tasks (Larsen, 1987). While other less familiar or regular figures could have been included in Who Am I?, Piaget and Inhelder's (1956) research linked the stages that they observed in children's ability to copy regular geometrical forms to cognitive development. These earlier studies indicated that copying tasks were strongly associated with subsequent school achievement, and provided a reliable measure of development.

Measures of spontaneous writing were included in Who Am I? as indicators of developmental levels, because there is a link between children's early attempts at writing and their growing understanding of the way in which spoken sounds are represented by print (Ferreiro & Teberosky, 1982). The links between this form of writing and emergent literacy is supported by the work of Clay (1993) and Hannavy (1993), while the work of Snow, Burns, and Griffin (1998) has shown that letter recognition is strongly related to later achievement in reading.

The final task, in Who Am I?, asks children to draw a picture of themselves. This well-known developmental task has been used also as a measure of developmental level by Brenner (1964), de Lemos (1973), and Harris (1963).

In a similar manner to Who Am I?, the development of I Can Do Maths was influenced by the understanding that children come into pre-school and school with a wide range of experiences and understandings fostered by parents. For example, in an English study of 3- and 4-year-olds'

mathematical knowledge prior to pre-school or school, it was found that the “children showed considerable knowledge and some consistent patterns of responding ... [and] the findings are unlikely to result from children noticing the numerals unaided and inventing their own ideas about what they mean” (Ewers-Rogers & Cowan, 1996, p. 23). Other examples include those from the work of Gelman and Gallistel (1978), who reported that “children as young as two years can accurately judge numerosity provided that the numerosity is not larger than two or three” (p. 55), and Zill, Collins, West, and Hausken (1995) who found that children of ages 3 to 5 had a wide range of mathematical skills and urged pre-school teachers to maintain children’s engagement to further develop these skills and understandings.

Research shows that children make great progress in terms of curriculum content during their first year at school. Suggate, Aubrey, and Pettitt (1997) tested children on rote counting, counting objects, and reading, writing and ordering numbers. Tymms, Merrell, and Henderson’s (1997) study of children’s development during the first year of school also showed a “massive difference to the attainment of pupils in Reading and Maths” (p. 117), after allowing for pupil background factors. Stewart, Wright, and Gould’s (1998) study showed that “progress [in mathematics] was made by the majority of students and syllabus expectations were not only reached but exceeded by many of these students” (p. 562).

Although some earlier experimentation had shown that young children can cope with written response formats (Doig, 1995), some of the children in the Curriculum and Organisation in the Early Years of School, 1997–1999 project were very young (3 years of age), and it was decided that questions be presented orally to reduce the reading and writing loads on the children. Item content was based on the content of the national profiles in mathematics (Australian Education Council, 1994) in which the early levels focus on concepts and skills in Number, Measurement, Chance and Data, and Space.

Group administration of the assessment items was used to reduce the time required for administration, although this meant that children would need to record their own responses in some way. Further, two different assessment forms were used at different year levels to shorten the time required of the children, and to provide the most appropriate set of questions.

In all, a set of 150 questions was constructed from which a final set of 47 items was selected for the published version of *I Can Do Maths*. This set was broken into two sub-sets, with the second set containing some harder items that were only administered to children in their second and third year of school. The identification of these harder items was determined in discussion with early years practitioners.

As with *Who am I?*, the *I Can Do Maths* items are administered orally in a lock-step fashion; that is, all children worked on the same question at the same time, and advanced through the questions at the same pace.

These questions were in two formats: either they had a disguised multiple-choice response format, or they asked for a simple, written, numerical response. Figures 1 and 2 show the two different question formats.

Put a ✓ on the tallest tree.

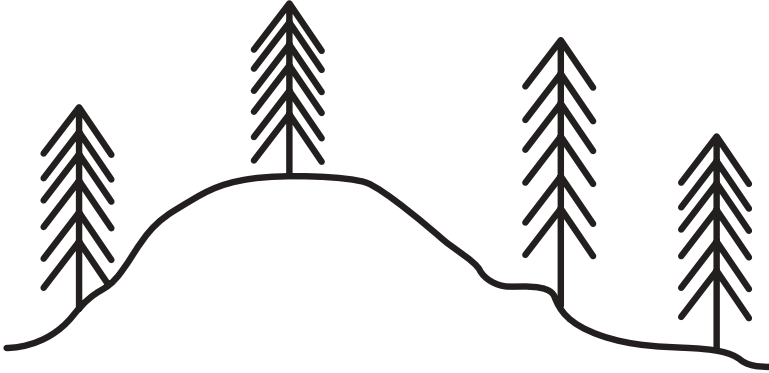
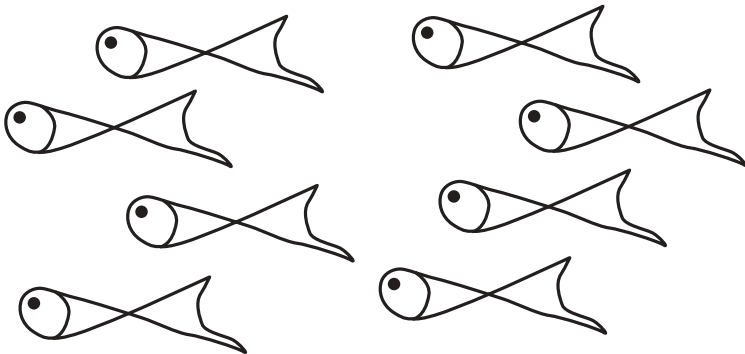


Figure 1. A question with a disguised multiple-choice format.

Count how many fish there are.



Write the number of fish there are. _____

Figure 2. A question requiring a written, numerical response.

Reporting Requirements

As the achievements of children in the early years would be useful to early years professionals, reporting the results of assessment in a clear and comprehensible manner was of paramount importance. It was decided that three different reports would be provided: a normative report, showing how children assessed were placed with respect to other children of that age, or in

that year of schooling; a report that presented diagnostic information for professional use; and, finally, a descriptive report for parents.

The range of reports envisaged for both assessment tools suggested that an IRT analysis would be more fruitful than traditional approaches in that it would enable, by using a Rasch analysis (Rasch, 1960): the use of ramped questions and developmental scoring for the youngest children; the use of equated forms in the data collection; the establishment of developmental scales that would trace children's progress across the age group in the project sample; and the provision of formative (diagnostic) reports to teachers (Doig, 1992), and descriptive reports to parents.

Data Collection

The data for the development of *Who Am I?* and *I Can Do Maths* were collected from a sample of pre-schools, schools, and children from across Australia. The children attended a total of 84 schools and 47 pre-schools, including some attached to primary schools. These sites were selected at random from all states and territories, with the exception of Tasmania. While not proportionally representative in terms of state, the sample covered a wide range of sites throughout Australia. From each of the participating pre-schools and schools, one class at each of the relevant year levels (pre-school to Year 2) was selected. This provided a total sample of over 4000 children, with about 900 children at each of the pre-school and pre-Year 1 levels, and about 1200 children at each of the Year 1 and Year 2 levels.

Who Am I?

Data analysis

Children's responses to the *Who Am I?* tasks were sorted into a series of categories, established on the basis of actual responses, that is, like responses were put together. These categories were ordered by reference to expected developmental progression as suggested by the research literature. This same literature was also used to develop the scoring criteria. The process was repeated for each *Who Am I?* task. See Adams, Doig, and Rosier (1991) for another example of these processes being used for categorising free-response data.

Responses, once categorised, were analysed using Masters' (1982) Partial Credit Model that provides estimates of the ability needed to achieve that category of response. That is, it is not assumed that all questions are of equal difficulty, nor that the achievement categories form a set of "steps" that require the same amount of development to achieve them. For another example of scoring and analysis of responses that views response categories as partly correct, see also *Tapping Students' Science Beliefs* (Adams et al., 1991; Doig & Adams, 1993).

The Partial Credit Model form of analysis provides a probabilistic relationship that places children's ability and the category difficulty on the same scale (see Bond & Fox, 2001, for an explanation of Rasch scales). In

addition, questions were grouped into four sub-scales: Copying (circle, cross, square, triangle, and diamond); Symbols (name, numbers, letters, words, sentence); Drawing; and Total (the total of the other three scales).

Because of the possibility of bias in interpreting responses to Who Am I? tasks, a sense of the inter-rater reliability was required to indicate the consistency with which the same result would be obtained if the child’s response was scored by different people. To obtain this, the same set of 30 booklets was marked by 21 different raters, all of whom were experienced teachers. The results of this exercise indicated a satisfactory level of agreement between the different raters with no more than one score category difference on any task across the group of raters.

The partial credit analysis provided an estimate of reliability for Who Am I? of 0.91, indicating a high level of internal consistency for the tasks. Due to the somewhat novel nature of the Who Am I? assessment tool, some care was taken to ensure that Who Am I? was valid with respect to its content and construct. In Who Am I?, tasks focus on aspects of children’s development that are directly related to the objectives of the early years of school curriculum. Data from the sample provided correlations of about 0.6 between scores on Who Am I? and scores on the Literacy Baseline (Vincent, Crumpler, & East London Assessment Group, 1996), which are similar to other reported correlations between developmental measures (Tymms, 1999).

Table 1

Percentage of Children Achieving Highest Level on Who Am I? Tasks, by School Level (de Lemos & Doig, 1999a, p. 22). (Reproduced by permission of the Australian Council for Educational Research Ltd.)

Task	Pre-school and Pre- primary Mean age = 4:11	Pre-Year 1 Mean age = 5:9	Year 1 (QLD & WA) Mean age = 6:1	Year 1 (Other states) Mean age = 6:9	Year 2 (QLD & WA) Mean age = 7:1	Year 2 (Other states) Mean age = 7:8
Name	21	63	81	92	97	99
Diamond	5	17	36	41	74	65
Numbers	3	30	60	85	98	97
Letters	12	66	67	94	95	98
Words	<1	20	34	71	86	90
Sentence	<1	11	26	58	83	84
Drawing	2	3	4	12	33	35
Total N	866	915	411	924	311	888

Construct validity is based on an accumulation of evidence relating to the assessment and what it measures. Such evidence would include correlations with other relevant measures, studies of changes in performance over time and, particularly in the case of a developmental measure, with increasing age. In this case, evidence of developmental progression is indicated by the increase in mean score according to both age and school level, as indicated in Table 1, that shows the proportion of children by school level who achieve the highest level on some of the key Who Am I? tasks. For example, the Name task shows a substantial increase in higher performance once children enter school (21% in the highest score category at pre-school level and 63% at pre-Year 1 level). Again, the percentage of children in the highest score category on the Diamond task increases on entry to school, and continues to increase with more schooling. The change between Year 1 and Year 2, for children not in Western Australia or Queensland, shows a smaller increase (41% to 65%) than the increase for Western Australia or Queensland (36% to 74%) over the same two-year period. This difference may be accounted for by curriculum differences between the states. Similarly, the large difference on the Words task between Queensland and Western Australian Year 1 students (34%) and other states (71%) is likely to be due to the extra year of schooling for the latter group.

I Can Do Maths

Data collection

The final set of 47 questions for I Can Do Maths was selected on the basis of providing the widest coverage of the curriculum content, and question difficulty. Different assessment forms were used at different year levels to shorten the time required of the children, and to provide the most appropriate set of questions. The second (harder) set was administered only to children in their second and third year of school.

Data analysis

Responses to the I Can Do Maths assessment questions were scored as correct or incorrect and were analysed using an IRT Rasch analysis (Rasch, 1960; Wright & Stone, 1979) that gives estimates of the ability required of the child to obtain a particular Total Score. This analysis provides a probabilistic relationship that places children's ability and the question difficulty on the same interval scale, thus allowing direct comparisons between different raw scores in terms of the ability. As part of the analysis, questions were grouped into sub-scales (Number, Measurement, and Space). These scales are the basis of the reports for I Can Do Maths.

For the development of the published version of I Can Do Maths (Doig & de Lemos, 2000), the 47 project questions were made into two sets—Level A (30 questions) and Level B (33 questions)—and printed in separate booklets. The two booklets have some questions in common, and these common or link questions allow children's ability estimates to be placed on

a single scale, and thus allow their mathematical development to be monitored over the age range 3 to 9 years (see Kolen, 1999, for an explanation of test equating).

Reporting Results of Assessment

While question development is an important aspect of assessment construction, the framework for reporting children's performances is equally critical. In the case of assessment tools such as *Who Am I?* and *I Can Do Maths*, the use of IRT analyses of the data provided opportunities for reporting both summative and formative (diagnostic) results. For *Who Am I?*, two forms of report are provided: an Individual Profile that contains both summative and normative information, and a diagnostic map, or DIAMAP (Doig, 1992).

For *I Can Do Maths* there are three forms of report provided: diagnostic, descriptive, and normative. These reports are based on the underlying scales constructed by the Quest Rasch analysis (Adams & Khoo, 1993, 1996) of the original project data. The project data for questions included in Level A provided the scales for Level A (for children up to the beginning of their second year at school) and for Level B (for children in their second year of schooling).

Diagnostic Reports

A particular feature of the scales constructed using Rasch analysis is that the likelihood of a particular response to a question can be calculated for a child with a specific Total Score. This enables a DIAMAP to be constructed (Doig, 1992).

In a DIAMAP, questions that lie below children's Total Score lines are expected to be easy for them, while those above this line are expected to be too difficult. In other words, the further a question is below the line, the more likely it is that it will be answered correctly, and the further above the line, the less likely it is that it will be answered correctly.

To use a DIAMAP, a line is drawn across it at the child's Total Score level. A circle is then drawn around each assessment question that the child answered correctly. Some of these marked tasks may be above the child's Total Score line and some below it. Once the Total Score line has been drawn and each correctly answered question marked, there are four conditions of diagnostic interest: tasks expected to be correct that are correct, tasks expected to be correct that are not correct, tasks not expected to be correct that are correct, and tasks not expected to be correct that are not correct. There are two conditions above the child's score line and two below it, and in each section of the DIAMAP there is an expected and an unexpected condition.

A child's specific strengths are shown by the correctly answered questions above the child's Total Score line. When these questions are within a particular curriculum or topic area, such as number or addition, then this may indicate strength in that area. Individual questions, on the other hand,

may reveal particular strengths within an area: for example, a particular strength in simple addition within Number, but not strength in other aspects of Number.

Specific weaknesses are shown by the questions answered incorrectly lying below a child's Total Score line. Again, when these questions are within a particular curriculum or topic area, this may indicate a general weakness. Individual questions, on the other hand, may reveal a particular weakness within an area: for example, a particular weakness with shape recognition within Space, but not in Space generally.

The DIAMAP shown in Figure 3 is for 6-year-old Daryl, who was assessed using the Level A assessment form mid-year in his first year of school in the Northern Territory. Daryl's Total Score was 18. The circled question numbers on the DIAMAP show which questions he answered correctly. The DIAMAP score line shows a reasonably clear division between questions that he could and could not answer successfully, which is expected in a DIAMAP.

Of the Number questions below the score line (expected to be easy for

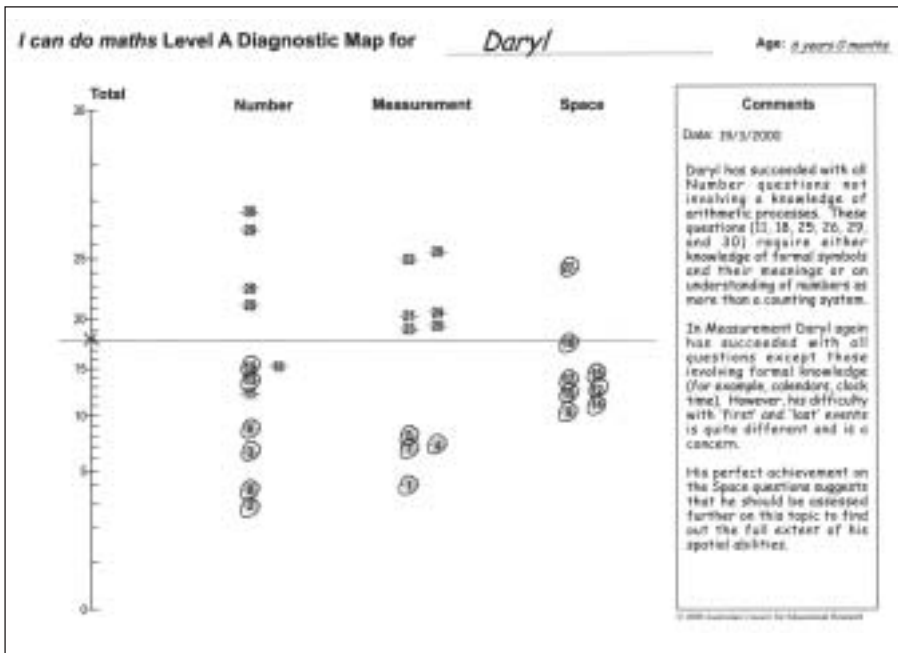


Figure 3. An example of an I Can Do Maths DIAMAP (Doig & de Lemos, 2000, p. 17).

(Reproduced by permission of the Australian Council for Educational Research Ltd.)

Daryl), all are correct except question 11 (add 5 and 4) and question 18 (identify the number 65). Daryl’s Measurement successes lie below his score line as expected. Daryl’s achievement on spatial questions is very good, and well beyond what is expected of a child with his overall score. The overall picture is of a competent child in counting and measurement, but not yet familiar with the conventions of mathematics that enable success with formal work. For a child at mid-year in the first year of school, it is likely that he has not as yet been exposed to the formal aspects of the curriculum.

Daryl’s DIAMAP alerts us to a possible problem with first and last (questions 20 and 21), but more importantly shows that a reliance on a score alone would not provide a complete picture of Daryl’s abilities, nor provide suggestions for future learning experiences.

The Diagnostic Map for Who Am I? is interpreted as that for I Can Do Maths, except that the questions are scored at several levels rather than simply correct or incorrect. In Figure 4, each highest level of response has been circled and a line drawn across the Diagnostic Map at Ronnie’s Total Score (22). This line divides the assessment levels (scores) on the Who Am I? tasks into those he is expected to achieve (below his Total Score line) and those he is not expected to achieve (above his Total Score line).

As can be seen in Figure 4, most of Ronnie’s highest assessment levels

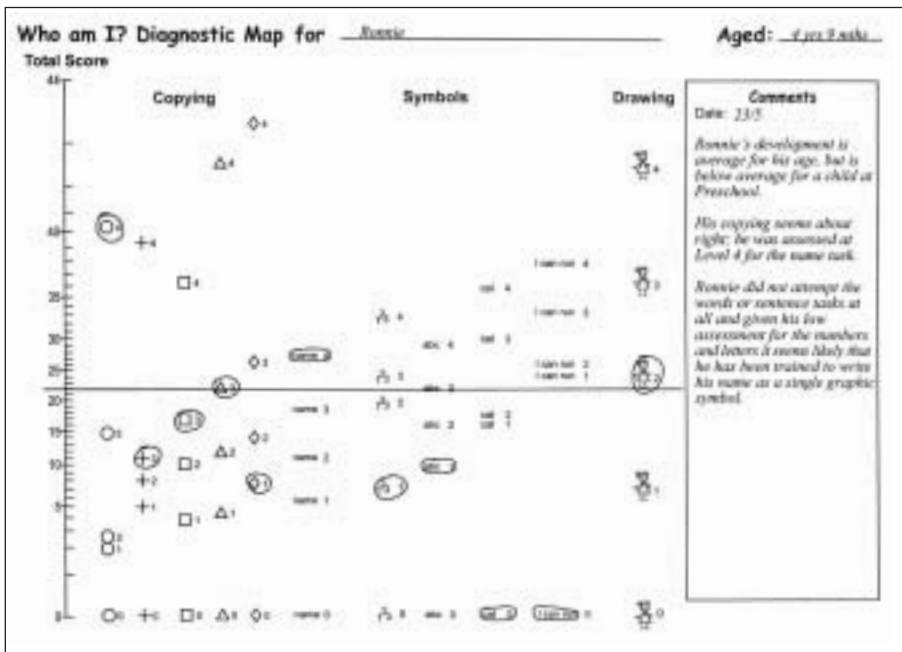


Figure 4. An example of a Who Am I? DIAMAP (de Lemos & Doig, 1999a, p. 16).

(Reproduced by permission of the Australian Council for Educational Research Ltd.)

are for the Copying tasks. The high assessment at Level 4 for the Circle and Name tasks is in contrast to Ronnie's results for the Symbols tasks (Numbers – Level 1, Letters – Level 1, Words and Sentence – Level 0).

Individual Reports

Figure 5 shows an Individual Profile for reporting the results of Who Am I? The interpretation of this report is the same as for any normative report. These normative comparisons provide a guide to the performance expected of children and allow for variation in performance between children stemming from individual differences. These norms also allow teachers to determine where each child is in relation to other children, as a basis for grouping children for different types of activities. The shaded bands show the expected range of scores for the middle 80% of children in each of the distinct state school structures across Australia.

Maryanne's score on the Symbols scale (9) shows that her development

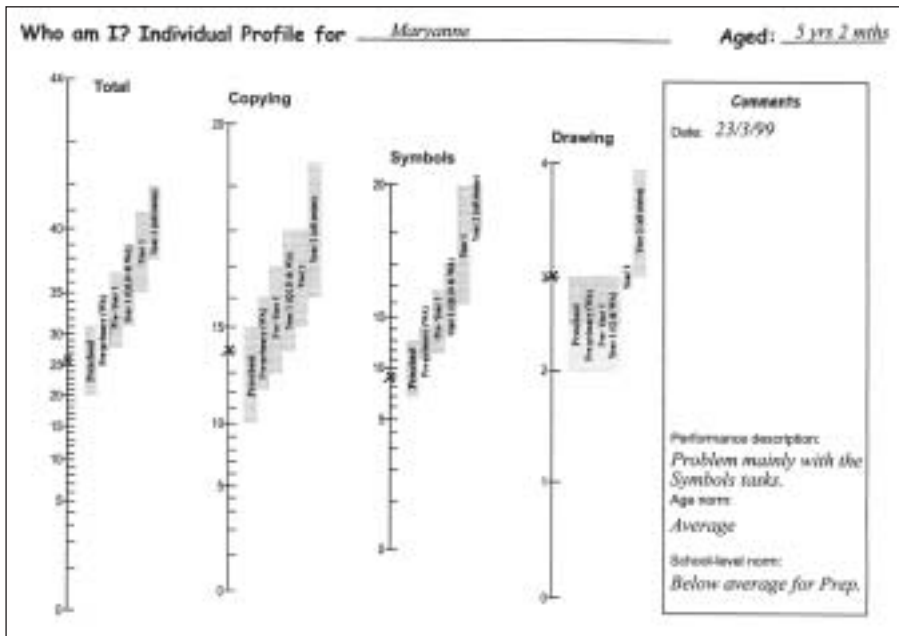


Figure 5. An example of a Who Am I? Individual Profile (de Lemos & Doig, 1999a, p. 13).

(Reproduced by permission of the Australian Council for Educational Research Ltd.)

in this area is below that expected of children at her stage of schooling. Her Copying and Drawing scores (14 and 3 respectively) place her just within the expected range for a Preparatory year child in Victoria and her Total Score (26) places her below the expected level for children on entry to school.

An I Can Do Maths individual report is interpreted in a similar manner

(see Doig & de Lemos, 2000, for examples of I Can Do Maths individual reports).

Descriptive Reports

The interpretation of an I Can Do Maths Descriptive Report, as shown in Figure 6, parallels that of a DIAMAP insofar as descriptions of performance below the Total Score line are likely to have been achieved, while those descriptions of performance above the line are yet to be achieved. Note that where there is more than one description of ability given, the higher description on the scale subsumes lower descriptions. The example in Figure 6 is for Daryl, in his first year of school.

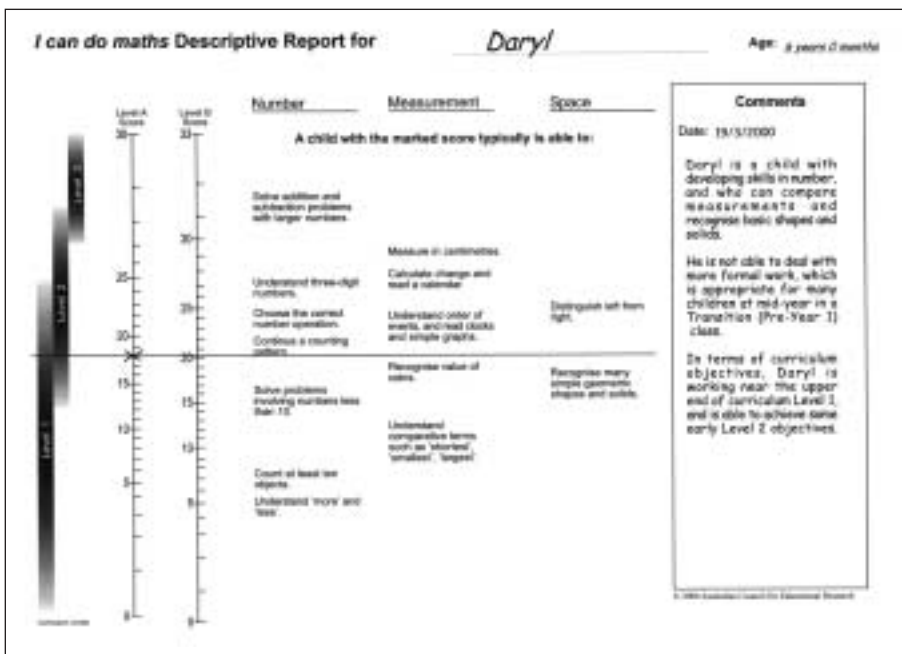


Figure 6. An example of an I Can Do Maths Descriptive Report. (Doig & de Lemos, 2000, p. 19).

(Reproduced by permission of the Australian Council for Educational Research Ltd.)

Other Uses of These Assessment Tools

Both Who Am I? and I Can Do Maths have, in a sense, come back to their origins. These assessment tools started as instruments for use in a research project, were refined and developed into classroom assessment tools, and have now been employed for data collection in both Australian and overseas research studies. Some of these studies are described below.

Comparative Research

Who Am I? has been used to compare pre-school children's development in three different cultural groups: Chinese children in Hong Kong, Anglo-Indian children in India, and Australian Indigenous children. The results of this series of studies are given after the description of each sample group (de Lemos, 2002; de Lemos & Doig, 2000).

In the study of pre-school development in Hong Kong, the children were of Chinese origin and had Cantonese as their mother tongue, although with varying levels of English proficiency. The sample of 60 children was between 3 and 6 years of age, and at different levels of pre-school.

The administration of Who Am I? was undertaken by the children's pre-school teacher. Children's responses were in a mixture of Cantonese and English and scripts were scored by a native speaker of Chinese, with adaptations for Chinese script. That is, the original scoring criteria were maintained by, for example, equating English letters with simple Chinese characters, and English words with complex Chinese characters.

The Indian sample consisted of 249 children, from Himachal Pradesh (near the border of Tibet), and were from mainly rural communities with Hindi as their mother tongue. The children's ages were from 4 years to over 10 years, with most aged between 5 and 6 years. Who Am I? was administered in Hindi, under the guidance of the National Institute of Educational Planning, New Delhi. The children responded in Hindi, and as with the Hong Kong group, adaptations were made for language variations.

Australian Indigenous children were administered Who Am I? at two sites: a pre-school and a primary school. Children were assessed at the end of their pre-school year and at the end of their first year at school. As well, Who Am I? was administered to a sample of 523 urban Australian pre-school and first year of school children. All of these Australian children were administered Who Am I? in English.

A summary of the results of these administrations for children at a similar age and level of schooling is given in Table 2.

Table 2

Comparison between International and National Groups On Two Who Am I? Scales (adapted from de Lemos & Doig, 2000)

Group	N	Mean age	Copying mean	Symbols mean	Total mean
Hong Kong	19	6.0	17.7	18.5	39.2
India	249	5.9	13.3	10.7	25.9
Indigenous Australian	60	5.5	14.0	8.8	25.9
Urban Australian	523	5.9	15.8	15.3	33.9

A comparison of the Total Score means displays the differences between groups from rural (India and Indigenous Australian) and urban (Australia and Hong Kong) areas. However, it is clear that there appears to be a substantive link between performance on 'copying' tasks and 'symbols' tasks for Hong Kong and Australian children, but not so clear for the rural groups (India and Indigenous Australian), although questions about familiarity and, or, exposure to print for these two groups may arise. More complete details of these studies can be found in de Lemos and Doig (2000), and de Lemos (2002).

The National Longitudinal Survey of Children and Youth

Canadian data on Who Am I? have been collected from over seven hundred 5- and 6-year-olds at pre-school level as part of the North York Community Project in Ontario. The North York study is, in turn, part of the broader Canadian National Longitudinal Survey of Children and Youth that is being conducted by Statistics Canada and the federal government department of Human Resources Development Canada. This national study is following the development of children from birth to early adulthood, and includes both health and educational aspects. The study is designed to monitor the impact of factors that influence children's social, emotional, and behavioural development (Statistics Canada, 2004).

Initial data from the North York Community Project indicated that Canadian children's performance was similar to that of Australian children at a comparable level of schooling (pre-Year 1). Further data were collected from a sample of 12,000 4- and 5-year-olds and these suggested that Who Am I? scores are less sensitive to differences in a child's home language than the Peabody Picture Vocabulary Test that was administered at the same time as Who Am I? (North York Early Years Action Group, 2000). A more extensive report on the continuing Canadian use of Who Am I? is found in de Lemos (2002).

Project Good Start

Project Good Start (Doig & Rowe, 2002) had as one of its aims to raise the awareness of educators and the general community to the considerable achievements of young children (Doig et al., 2003). This on-going study is examining the development of the mathematical skills and understandings of some 3000 children in their year-before-school and during their early years of school.

Thomson (2004) reported that most children in the project were attaining Level 3 or Level 4 on the majority of the copying tasks in Who Am I?, although there were significant gender differences. For example, the item asking children to draw a circle, 73% of boys and 86% of girls were assessed

at Level 3 or 4 and for a similar item asking children to draw a triangle, 48% of boys and 60% of girls reached level 3 or 4. Children's performances on I Can Do Maths surprised pre-school staff and parents alike. Results on I Can Do Maths showed that 40% of the children in the year prior to entering school were able to identify a cylinder, and that 30% could solve "4 more than 5" and "2 less than 6" (Peck, 2003). Further findings from this project are reported in Thomson (2004).

Re-analysis of I Can Do Maths data

In a further use of I Can Do Maths, the data used in its development were re-analysed to present a picture of children's development in the areas of Number, Measurement, and Space. Doig and de Lemos (2000, 2003) argued that the hops, steps, and jumps in the development of both individuals and groups as they grow as mathematicians can be either hurdles or gateways to development. The re-analysis demonstrated that good assessment, combined with good analysis, might reveal more than simply children's achievement of curriculum content. In the re-analysis, the I Can Do Maths questions were put in order of their Rasch difficulty estimates (see Doig & de Lemos, 2000, 2003, for details of the original Rasch analysis). The percentage of children responding correctly to each item was then plotted against the corresponding item. These percentages revealed that, as one would expect, the proportion of children responding correctly increased as they proceed through school. Figure 7 gives the details of the Space re-analysis by Doig and de Lemos (2003).

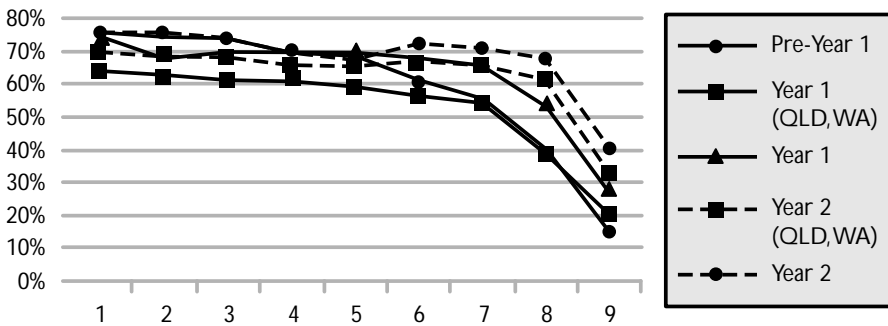


Figure 7. The Space questions from I Can Do Maths, in order of difficulty (Doig & de Lemos, 2003, p. 276).

There appeared to be little difference in children's learning trajectories, that is, the harder questions were difficult for all, not just children in the lower year levels. The most difficult question involved distinguishing left from right (Question 9).

This similarity across year levels raises the issue of appropriateness of curriculum content. For all but two questions, children's performances between pre-Year 1 and Year 2 differed little, suggesting an under-estimation of younger children's spatial abilities. Again this appeared to be the case for Measurement. The re-analysis implies that items dealing with comparisons of attributes such as length or area show little difference in correct response levels across the early years of school, suggesting that children are not being challenged by the curriculum.

In Number, too, there appeared to be little difference between the year levels for a number of curriculum aspects. Differences did exist, however, between children in the later years of schooling, once more formal arithmetic questions appeared. For example, the question "Tom had 5 gum-nuts and found 4 more. How many does he have now?" produced about a 10% difference between the earlier and later year levels. The Doig and de Lemos (2003) re-analysis identified clearly that there exist some apparent problems with curriculum expectations vis-à-vis children's abilities.

Discussion

The purposes of this article were: to demonstrate the possibilities for standardised assessment in the early years; to show how IRT analyses can provide reporting formats for assisting early years professionals; and to describe some examples of these assessment tools in research contexts.

In this article, descriptions were provided of the development of two mathematics assessment tools designed for use at the pre-school and early years of school levels, where formal assessment is rare. In Australia, at the present time, there is widespread emphasis on interviewing children as they enter school, with a view to providing appropriate learning experiences for individual children. The seminal work of Wright and his colleagues (e.g., Wright, 1991, 1994, 1999; Wright, Martland, & Stafford, 2000) and the subsequent development of the Count Me In suite of early and later years programmes, is an outstanding example of the clinical interview genre in assessment. Clinical interview techniques have been used and promoted for many years as the best way of assessing children's mathematical abilities, a view with which the author agrees (Hunting & Doig, 1997). As both Count Me In and the Early Numeracy Research Project (Clarke, Sullivan, Cheeseman, & Clarke, 2000) have shown, this approach can have profound effects on both teaching and learning in the early years.

However, the time needed for interviewing means that programmes tailored to children's needs are not commenced for some time after the interview has taken place, which can make the interview data out-of-date due to the rapid development of children in their first year of school. Thus, the necessity for a group assessment is obvious if one considers that the planning and implementation of appropriate learning experiences should occur close to the time of assessment. The two examples discussed in this paper show that for very young children it is possible, through oral administration, to use

more formal, standardised forms of assessment. This raises the possibility of large-scale studies of children's mathematical development, not possible with interview or one-on-one methodologies. The Canadian longitudinal study, for example, is a case in point. Cross-national studies provide valuable insights into our own, and other's, educational practices, and the use of standardised tools are a necessity in this form of research.

The cross-cultural validity of Who Am I? has been, in part, confirmed by the Canadian longitudinal survey through comparisons with the Peabody Picture Vocabulary Test. Who Am I? shows little bias due to language factors. Similarly, the Hong Kong and Indian research demonstrate that Who Am I? can be adapted to non-Roman alphabets and ideographic writing. This appears to be unique in mathematics assessment at any level.

The Who Am I? and I Can Do Maths forms of assessment combined with Rasch scaling have three advantages over other approaches to assessment. First, they allow monitoring across years of development, the reason for the use of I Can Do Maths in Project Good Start, where this feature allows researchers to track the development of children as they pass from pre-school to the early years of school. Second, they enable the construction of reports that provide both summative and formative information. The examples of reports in this paper clearly show how appropriate analysis can be effective in informing educators about the strengths and weaknesses of the students across a range of years. Finally, the re-formatting of I Can Do Maths data, as illustrated by Doig and de Lemos (2003), demonstrates the power of the Rasch scale that defines a difficulty for each question for the entire sample age range.

Conclusion

The use of these assessment tools by researchers internationally suggests that there exists a need for standardised, early years mathematics assessments. While not all early years professionals need or want such tools for their particular contexts, there are others whose interests lie in mapping children's mathematical abilities. Quality assessment tools provide a means of achieving a mapping over time, place, or culture. Further, tools such as these provide a language for discussion about contemporary issues, such as the pre-school to school transition, which benefits practitioners and researchers and which, in turn, should benefit the children we serve.

References

- Adams, R. J., Doig, B. A., & Rosier, M. (1991). *Science learning in Victorian schools: 1990, ACER Monograph No. 41*. Melbourne: Australian Council for Educational Research.
- Adams, R. J., & Khoo, S.-T. (1993). *Quest: The interactive test analysis system*. Hawthorn: Australian Council for Educational Research.
- Adams, R. J., & Khoo, S.-T. (1996). *Quest: The interactive test analysis system, Version 2*. Camberwell: ACER.

- Aubrey, C. (1997). Children's early learning of number in school and out. In I. Thompson (Ed.), *Teaching and learning early number* (pp. 20–29). Buckingham: Open University Press.
- Australian Education Council. (1994). *Mathematics – A curriculum profile for Australian schools*. Carlton: Curriculum Corporation.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bottle, G. (1998). *A study of children's mathematical experiences in the home*. Paper presented at the 8th Annual EECERA Conference, Santiago de Compostela, Spain.
- Brenner, A. (1964). *Anton Brenner developmental Gestalt test of school readiness*. Los Angeles: Western Psychological Services.
- Clarke, D., Sullivan, P., Cheeseman, J., & Clarke, B. (2000). The Early Numeracy Research Project: Developing a framework for describing early numeracy learning. In J. Bana & A. Chapman (Eds.), *Mathematics education beyond 2000* (Proceedings of the 23rd Annual Conference of the Mathematics Education Research Group of Australasia, pp. 180–187). Sydney: MERGA.
- Clay, M. (1993). *An observation survey of early literacy achievement*. Auckland: Heinemann Educational.
- Cook, D. (1996). Mathematical sense making and role play in the nursery. *Early Child Development and Care*, 121, 55–66.
- de Lemos, M. (1973). The development of spatial concepts in Zulu children. In J. W. Berry & P. R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural Psychology* (pp. 367–380). London: Methuen & Co.
- de Lemos, M. (1980). School readiness and school achievement: Research findings on some of the factors related to school readiness and later school achievement. In The Australian Pre-School Association (Ed.), *Children, the community's concern* (Proceedings of the 15th National Conference of the Australian Pre-school Association, Sydney, May, 1979, pp. 113–124). Canberra: The Australian Pre-School Association Inc.
- de Lemos, M. (2002). *Patterns of young children's development: An international comparison of development as assessed by Who Am I?* Retrieved 13 December, 2004, from <http://www11.sdc.gc.ca/en/cs/sp/arb/publications/research/2002-002358/SP-520-09-02.pdf>
- de Lemos, M., & Doig, B. (1999a). *Who Am I?: Developmental assessment*. Melbourne: Australian Council for Educational Research.
- de Lemos, M., & Doig, B. (1999b). *From research to practice: Who am I? – A tool for school entry assessment*. Paper presented at the Third Warwick International Early Years Conference, Coventry, UK.
- de Lemos, M., & Doig, B. (2000). *Assessment in the early years: An Australian perspective*. Paper presented at the Symposium on Assessment in the Early Years at the 6th European Conference on Educational Research, Edinburgh.
- de Lemos, M., & Larsen, P. (1979). *School readiness and school achievement: A longitudinal study. Final report*. Hawthorn, Vic: ACER.
- de Lemos, M., & Mellor, E. (1991). *Determining readiness for school: A joint parent and teacher approach*. Paper presented at the National Conference of the Australian Early Childhood Association, Adelaide.
- Doig, B. A. (1992). *DIAMAPS-Self-scoring kidmaps*. (ERIC Document Reproduction Service No. ED346146)

- Doig, B. (1995, April). *A format for all seasons: Developing prompted-response items for younger children, non-readers, limited writers, and others*. Paper presented at the International Objective Measurement Workshops, University of Chicago.
- Doig, B., & Adams, R. J. (1993). *Tapping students' science beliefs*. Hawthorn: Australian Council for Educational Research.
- Doig, B., & de Lemos, M. (2000). *I can do maths*. Melbourne: ACER Press.
- Doig, B., & de Lemos, M. (2003). Hops, steps and jumps: Mathematical progress in the early years. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity* (Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia, pp. 270–277). Geelong: MERGA.
- Doig, B., McCrae, B., & Rowe, K. (2003). *A good start to numeracy*. Commonwealth of Australia. Retrieved 26 October, 2003, from <http://www.dest.gov.au/schools/publications/2003/index.htm>
- Doig, B., & Rowe, K. (2002). *Project Good Start: Effective numeracy practices in the year before and the first year of school*. Paper presented at the National Numeracy Information Forum, Canberra.
- Ewers-Rogers, J., & Cowan, R. (1996). Children as apprentices to number. *Early Childhood Development and Care*, 125, 15–25.
- Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Exeter, NH: Heinemann.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Groves, S., & Cheeseman, J. (1993). Young children's number concepts – The effect of calculator use on teacher expectations. In B. Atweh, C. Kanes, M. Carss, & G. Booker (Eds.), *Contexts in mathematics education* (Proceedings of the 16th annual conference of the Mathematics Education Research Group of Australasia, pp. 327–333). Brisbane: MERGA.
- Hannavy, S. (1993). *Middle infant screening test*. Windsor, UK: NFER-Nelson.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough draw-a-man test*. New York: Harcourt, Brace & World Inc.
- Hunting, R., & Doig, B. (1997). Clinical assessment in mathematics: Learning the craft. *Focus on Learning problems in Mathematics*, 19(3), 29–48.
- Kolen, M. J. (1999). Equating of tests. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 164–185). Amsterdam: Pergamon.
- Larsen, P. (1987). *Copying skills* (Research Edition). Melbourne, Victoria: ACER.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Munn, P. (1994). The early development of literacy and numeracy skills. *European Early Childhood Education Research Journal*, 2(1), 5–18.
- Nixon, D., & Aldwinckle, M. (1997). *Exploring: Child development from three to six years*. Katoomba, New South Wales: Social Science Press.
- North York Early Years Action Group. (2000). *Initial report on the North York community project (unpublished)*. Ottawa: Human Resources Canada and Statistics Canada.
- Peck, R. (2003). The numeracy skills of pre-schoolers. *The Common Denominator*, 197, 2–3.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.
- Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Statistics Canada. (2004). *National longitudinal survey of children and youth*. Author. Retrieved August 24, 2004, from www.statcan.ca/english/sdds/4450.htm
- Stewart, R., Wright, B., & Gould, P. (1998). Kindergarten students progress in the Count Me In Too Project. In C. Kanas, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times* (Proceedings of the 21st annual conference of the Mathematics Education Research Group of Australasia, pp. 556–563). Brisbane: MERGA.
- Suggate, J., Aubrey, C., & Pettitt, D. (1997). The number knowledge of four to five year olds at school entry and at the end of their first year. *European Early Childhood Education Research Journal*, 5(2), 85–100.
- Thomson, S. (2004). Numeracy in the early years: Project Good Start. *Australian Primary Mathematics Classroom*, 9(4), 14–17.
- Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Reading Research*, 22(1), 27–36.
- Tymms, P., Merrell, C., & Henderson, B. (1997). The first year at school: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3(2), 101–118.
- Vincent, D., Crumpler, M., & East London Assessment Group. (1996). *Reading progress tests, stage 1*. London: Hodder & Stoughton.
- Wright, R. J. (1991). An application of the epistemology of radical constructivism to the study of learning. *Australian Educational Researcher*, 18(1), 75–95.
- Wright, R. J. (1994). A study of the numerical development of 5-year-olds and 6-year-olds. *Educational Studies in Mathematics*, 26, 24–44.
- Wright, R. J. (1999). Professional development in recovery education. In L. P. Steffe & P. W. Thompson (Eds.), *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld* (pp. 135–151). London: Falmer.
- Wright, R. J., Martland, J., & Stafford, A. (2000). *Early numeracy: Assessment for teaching and intervention*. London: Paul Chapman.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Zill, N., Collins, M., West, J., & Hausken, E. G. (1995). Approaching Kindergarten: A look at pre-schoolers in the United States. *Young Children*, 51(1), 35–38.

Author

Brian Doig, School of Scientific and Developmental Studies, Faculty of Education, Deakin University, 221 Burwood Highway, Burwood, VIC 3125. Email: badoig@deakin.edu.au