Improving Learning

2 June, 2011

Ms Kerryn Riseley, Executive Officer
Education and Training Committee
Parliament House
Spring Street
EAST MELBOURNE 3002

VIA EMAIL: etc@parliament.vic.gov.au

Dear Ms Riseley,

Re: Parliamentary Inquiry into the Education of Gifted and Talented Students

Please find attached a submission from the Australian Council for Educational Research to the Victorian Parliament's Education and Training Committee's Inquiry into the Education of Gifted and Talented Students.

While the terms of reference of the Inquiry are broad, ACER has provided a submission which focuses mostly on one aspect of the terms of reference: the identification of gifted and talented students.

It is ACER's submission that the identification of students with high academic potential is key to the proper targeting of programs and pathways aimed at providing appropriate teaching and learning opportunities for academically able students. In order to identify these students, teachers, schools and systems need to use effectively targeted, research-based assessment instruments as part of the formal process of identification.

Should you require any further information or explanation of our submission, please contact Ralph Saubern, Director Assessment Services on 03 9277 5681 or Saubern@acer.edu.au.

Yours sincerely,

Professor Geoff N Masters
CHIEF EXECUTIVE OFFICER

Australian Council *for* Educational Research Limited
19 Prospect Hill Road (Private Bag 55) Camberwell VIC 3124 Australia
t +61 3 9277 5555 f +61 3 9277 5500 w www.acer.edu.au
ACN 004 398 145 ABN 19 004 398 145

Australian Council *for* Educational Research Ltd

Issues in assessment for the identification of academically able students

Submission to the Parliamentary Inquiry into the Education of Gifted and Talented Students

Australian Council for Educational Research

May, 2011

Contact:
Mr Ralph Saubern
Director, Assessment Services
Australian Council for Educational Research
Private Bag 55
Camberwell VIC 3124

Contents

Executive Summary

- The proportion of Australian students performing at the top levels in international comparison studies such as PISA and TIMSS has declined and is lower than in other comparable jurisdictions. This has lead to an overall significant decline in the literacy achievement and mathematical literacy achievement among Australian students.

- Strategies for addressing the needs of academically able students can be properly targeted by using effective tools to identify students with high academic potential.

- There is considerable debate on both the definition of a "gifted student" and the most appropriate methods for identifying such students. Traditionally, observational techniques and IQ tests have been widely used. Both approaches have limitations. Formal methods that address a broad range of gifted abilities also have limitations.

- It is essential that assessments of academic and non-academic ability have appropriate research-proven validity, reliability, sensitivity and reference ranges, and are applied appropriately.

- ACER's work focuses on the identification of academically able students through the development and use of assessments which measure higher order reasoning ability at the appropriate target level. ACER test instruments are underpinned by a Rasch measurement approach allowing the construction of a measurement scale.

- In addition to the technical requirements of appropriate test construction, there are a range of practical requirements for the effective and equitable administration of high stakes selection tests.

- It is ACER's submission that effectively targeted, research-based assessment instruments should be part of formal processes that schools and systems use to identify students for selection and placement into programs and streams for academically able students and that such instruments should be administered using the highest quality administrative procedures.

About ACER

The Australian Council for Educational Research (ACER) is an independent, not for profit research organisation established in 1930 with over 340 staff in offices around Australia and internationally. ACER has a long history of developing and administering research-based assessments to meet complex client needs in the collection of student data. It provides assessment services for a range of school, university, government and non-government clients.

Significant current programs include school monitoring and achievement tests, selective entrance school tests, scholarship selection tests, international surveys and comparison assessments, senior secondary scaling and moderation tests, university entrance tests and medical admissions tests.

Appendix 1 lists some ACER programs.

ACER is pleased to make a submission to the Parliamentary Inquiry into the Education of Gifted and Talented Students drawing on its long experience of the assessment of student ability and achievement.

Introduction

The Australian report on OECD Programme for International Student Assessment (PISA) 2006 (Thomson & DeBortoli, 2008) described a significant decline in average reading literacy scores of Australia's 15 year old students between PISA 2000 and PISA 2006 and a significant decline in the average mathematical literacy scores of female students. The decline in reading literacy scores was confirmed by PISA 2009 (Thomson, De Bortoli, Nicholas, Hillman & Buckley, 2011) and in addition a significant decline in the overall mathematical literacy score for all students was reported.

Australia was the only high achieving country in PISA to show a significant decline in reading scores and one of ten to show a significant decline in mathematics scores.

Further examination of the PISA 2009 data found that there was no significant difference in the proportion of Australian students achieving at the lowest proficiency levels in either reading or mathematics. Rather the decline has come from the top of the achievement distribution, with the proportion of students achieving at the highest two proficiency levels declining significantly (Thomson et al., 2011).

Figures 2.1 and 2.2 in Appendix 2 show the proportion of Australian students performing at the highest levels of PISA 2009 in mathematical literacy and reading literacy. Sixteen per cent of Australian students, compared to 50 per cent of students in Shanghai and 36 per cent in Singapore, achieved these standards in mathematical literacy; 13 per cent of Australian students, compared to 19 per cent of students in Shanghai and 16 per cent in New Zealand and Singapore, achieved these standards in reading literacy.

These findings about trends in mathematics achievement have been confirmed by data from the IEA Third International Mathematics and Science Study (TIMSS) (Thomson, Wernert, Underwood & Nicholas, 2008). At Year 8 level TIMSS identified a large decline in the proportion of students achieving at the high and advanced benchmarks, from 33 per cent of students in 1994 to 24 per cent of students in 2007 (although only a small decline in the proportion of students at the very highest level).

A number of initial theories have been advanced to explain this decline at the top levels of performance. While the explanation for the decline and appropriate strategies to arrest and reverse the trend have not yet been determined, there are existing in many educational systems and individual schools specialised programs and pathways aimed at addressing the teaching and learning needs of academically able students. These programs can be properly targeted using effective assessment tools to identify students with high academic potential.

Assessment of giftedness

There is no clear and common definition of a gifted student (Robinson, Zigler and Gallagher, 2000). However, students who show extraordinary academic or creative ability (e.g. in the arts or music) may be classified as gifted (U.S. Dept. Ed., 1993).

Typically, gifted students learn rapidly in certain domains, quickly grasping and applying concepts. In these domains, they think critically and creatively, and use advanced techniques to problem-solve. They may have a good memory and vocabulary, an independent approach to tasks, and sophisticated interests in their preferred areas. They may have advanced leadership and interpersonal skills. They may display perfectionism, hypersensitivity and over-excitability.

Ideally, 'Appropriate identification of talent provides gifted students with access to appropriately challenging academic programs and services' (Jarosewich, Pfeiffer and Morris, 2002). Such identification is not always made, leading to inappropriate educational programs for gifted students. Further, behaviour problems and certain learning disabilities may confound identification (Pfeiffer, 2001). In certain circumstances, gifted students may underachieve due to psycho-social factors.

Traditionally, gifted students have been identified by:

1 *Observation of students and their work* – These traditional approaches including review of classroom work, observation of behaviour, interviews, examination of portfolios and auditions can be affected by inconsistent and subjective judgments.

2 *IQ tests* – Typically, these assess *g* (general intelligence), fluid intelligence, Gf (general ability to think logically and solve novel problems), and crystallized intelligence, Gc (ability to apply skills, knowledge and experience, which is affected by learning). Models of intelligence typically present Gf and Gc as interacting subsets of *g*. IQ tests may provide assessments of other factors. Generally they correlate with educational and job success. Common assessments of IQ include the Wechsler series, the Stanford-Binet series and the Raven's Progressive Matrices series (for fluid intelligence). The tests vary in approach, sensitivity, reference ranges and ceiling scores, and have other limitations.

3 *Reasoning tests focussed on particular broad abilities* – These commonly assess Quantitative Reasoning and Verbal Reasoning. In addition, they often assess Abstract Reasoning, in a manner similar to IQ tests (Next in Sequence, Complete the Pattern, Middle of Sequence). Interpersonal Reasoning or Writing Ability may also be assessed. These tests tend to rank students and norms can be developed related to Rasch scales. Such tests have significant correlations with IQ tests but tend to use more authentic contexts. Examples include the Selective High Schools Entrance Test and the Co-operative Scholarship Testing Program, which are discussed elsewhere.

4 *Achievement Tests* – These may be direct assessments of curriculum-based skills such as Mathematics or English tests, or they may be cross-curricular and generic, such as the GAT, which, though not aimed at identifying giftedness, does assess and rank skill levels related to and underlying success in Mathematics/Science and Humanities/Social Science. Again, these tests tend to rank students and norms based on Rasch scales can be developed.

5 *Multifactorial teacher-admistered assessments* – Although these assessments may be preferred because they focus on a wide range of aspects of giftedness, including non-academic elements such as creativity and leadership, older tests of this kind have technical inconsistencies and other problematic aspects (Jarosewich et al, 2002). A newer test, the GRS (Pfeiffer-Jarosewich, 2003), is still being evaluated.

Superior academic ability has been identified by the five traditional assessment types. Ability in arts, music and in the interpersonal domain has most effectively been identified by the first and fifth of these. Oddly, IQ and academic achievement tests have at times been used in selection of students in relation to non-academic areas (Alvino, McDonnel and Richert, 1981).

In addition to the models of giftedness related to the assessment approaches described above (e.g. the concepts of IQ, broad-ability reasoning skill and multifactorial abilities), various models for understanding intelligence and other aspects of giftedness have been developed and applied in educational contexts. Some of these have been used as the basis of the development of gifted education programs and others have become widely used in general classroom teaching and learning. However the lack of rigorous, research-based assessment methods specifically associated with some theories restricts their use in the formal identification of giftedness though they can be seen as explanatory. These include:

- *Developmental Approach* – Developed by Francoys Gagné (e.g. Gagné, 1993, 1995), this theory makes a distinction between “giftedness” and “talent” and considers the factors or “catalysts” that may facilitate the development of giftedness into talent. Gagne’s model includes five domains of aptitude: intellectual, creative, socio-affective, sensorimotor and others.
- *Multiple Intelligences* – Developed by Howard Gardner (e.g. Gardner, 1999), this theory typically postulates several intelligence modules, for example: linguistic, logico-mathematical, spatial-visual, interpersonal, intrapersonal, bodily-kinaesthetic, musical and naturalistic.
- *Triarchic Theory* – Developed by Sternberg (e.g. Sternberg, 1997), this theory postulates three types of mind components: *metacomponents* (executive processes guiding problem solving and decision making), *performance components*

(processes that carry out the actions dictated by the metacomponents) and *knowledge-acquisition components* (used to obtain and selectively combine relevant information).

Some issues in giftedness assessment

While all of these approaches have been used in the identification of academically able students there are significant concerns with three of the approaches that preclude their adoption as the basis of the development of effective assessment tools.

Observation of students and their work

Whereas observation of student work, behaviour, interview and examination of student output can lead to the identification of giftedness, this approach can be seen as subjective and not amenable to wide-ranging comparisons.

IQ testing

Early IQ approaches provided a single score for *g*. IQ score was related to the number of standard deviations the score was from the mean score of a population of the same age. Typically, 1 SD corresponded to 15 IQ points and giftedness was variously defined - e.g. as scores of at least 2, 3 or 4 SD above the mean of 100. Important early work in this area was done by Binet and Spearman. More recently multifactorial approaches to IQ have predominated. As well as *g*, a set of scores may be provided – e.g. Gf, Gc, Gq (quantitative reasoning), Grw (reading and writing ability), Gv (visual ability), Gsm (short term memory). Gq and Grw can be seen as related to school learning. Important researchers in relation to this approach include Wechsler, Guilford, Thurstone, Cattel-Horn-Carroll, and Renzulli. The multifactorial approach is exemplified by Carrol, 1993.

Identification of academically-gifted students using IQ tests can be problematic. For example, different IQ tests have different reference ranges, sensitivities and ceilings, as do different reasoning and achievement tests. A new version of the same test can have a different reference range, which may be in part a response to the Flynn effect (which relates to the general increase in population IQ scores over time, Flynn 1987, 2009). Some IQ tests seem not to be sensitive in the gifted range and have low ceiling levels.

IQ tests have been accused of bias against minorities. Results may be affected by nutrition, health, psychological state, educational and cultural background. In addition, it has been postulated that lack of interest in IQ-style items by some students may lead to low scores that do not reflect ability.

Multifactorial teacher-administered assessments

Jarosewich et al, 2002 state that ‘The evaluation should consist of a comprehensive, multifactorial assessment, using psychometrically sound measures and incorporating information from a variety of sources.’ They support the use of well-designed, teacher-administered rating scales assessing a range of aptitudes. However, they indicate that highly validated instruments for non-academic aspects of giftedness have not been readily available. They do suggest that tests such as GATES (Gilliam, Carpenter and

Christiansen, 1996), GES-2 (McCarney and Anderson, 1989) and SRBCSS (Renzulli, et al, 2002) have useful approaches but have technical and other flaws.

The new Gifted Rating Scales (GRS, Pfeiffer-Jarosewich, 2003) teacher-administered instrument assesses Intellectual ability, Academic ability, Creativity, Artistic talent, Leadership ability and Motivation. It is meant to be used in conjunction with IQ tests or other tests of intellect or academic ability. This test appears to have overcome some problems with older tests but there have been criticisms such as the following:

- Gifted students do not always work well with teachers and there may be a negative halo effect and other biases.
- Too much emphasis may be put on a commercial published test applied by lay people.
- Evaluation studies so far have mostly been done by people associated with the test (e.g. Pfeiffer, Petscher and Kumtepe, 2008).
- Criteria such as Leadership and Motivation are often counter-correlated with IQ, so academically able students may be screened out.

Broad ability approach to the identification of academically able students

ACER has long history of developing and administering reliable, research-based tests that identify academically able students using broad ability approaches. See Appendix 1 for a list of relevant programs.

Test constructs

ACER uses various test constructs, each focussed on the particular test purpose and target population. For the identification of students of high academic abilty, the ACER focus is on assessing higher order, cross-curricular reasoning skills that are characteristic of such students. The approach is consistent with the modern IQ approach of assessing *g*, Gf, Gc, Gq, Grw etc, and the multiple intelligences approach of Gardner and Gagne's model.

There is evidence that IQ scores correlate significantly with higher-order, cross-curricular reasoning tests (Frey and Dettermann, 2004, Deary, Strand and Smith, 2007) but compared with IQ tests these can been seen as more authentic in style and demand, often utilising real-world contexts and tasks that are more acceptable to students.

Typical components in such tests are:

- *Quantitative Reasoning:* Consistent with assessment of Gardner's logico-mathematical intelligence and Gq, this component assesses ability to deal with non-specialist quantitative problems and is a traditional approach. Items can assess ability to comprehend and apply quantitative information using deductive and inductive reasoning; infer quantitative relationships by recognising patterns and generating hypotheses; select, transform and synthesise information to enhance problem solving; critically evaluate quantitative models and arguments; and generate, apply and evaluate solutions to problems.

- *Abstract Reasoning:* Consistent with assessment of Gf, this component assesses abstract reasoning skill using items in Next in Sequence and Complete the Pattern form that require students to hypothesise and identify rules and relationships. Other item types can also be used, including Middle of Sequence in which the sequence is jumbled and needs to be reordered before solution. Not being language dependent, such items are particularly useful for NESB students.

- *Verbal Reasoning:* Consistent with assessment of Grw and Gardner's linguistic intelligence, this component is concerned with the interpretation and understanding of ideas and language. It can address the recognition of explicit and implicit meaning, including comprehension of relationships between concepts presented in a verbal context, plausible reasoning, the drawing of conclusions from arguments presented, and the ability to discriminate and make judgments in a verbal context.

- *Writing:* This component assesses the generative and productive thinking of a candidate. Writing tests contrast with the receptive and reactive nature of IQ tests and contrast with standardised, objectively scored tests of Verbal, Quantitative and Abstract Reasoning. In a writing assessment candidates have to produce ideas rather than react to and analyse ideas put before them. They are best thought of as performance tests of thinking rather than literacy tests. Writing based assessments correlate well with performance in school subjects.

- *Interpersonal Reasoning:* This construct relates to aspects of Gardner's Personal Intelligence. It is a new construct formulated by ACER, based on the work of Gardner and Mayer, Salovey and Caruso (1999, 2002). It tests ability to reason about people, in particular the ability to identify emotions and feelings, to explain, understand and show insight into actions relating to these feelings and to predict feelings, responses or behaviour in given situations.

Validity of such tests is indicated by research on ACER's Co-operative Scholarship Testing Program (CSTP) in Victoria that found that Level 1 scholarship winners (based on CSTP performance) had a median ENTER score in the top 1.5% of all Year 12 students (ACER internal research).

Measurement model

High standards for test reliability and validity are achieved by trial testing items to ensure they match the target student ability range and have appropriate statistical properties. If items are not trial tested, they may not target the population correctly and fail to discriminate reliably between students (See Appendix 3).

In order that there is fair discrimination between students and minimal measurement error items and tests need to have certain statistical properties. An essential aspect of this is the use of a valid measurement model.

ACER test instruments are underpinned by a Rasch measurement approach (Rasch, G., 1960/1980). allowing the construction of a measurement scale which can be used to describe both characteristics of the assessment instrument (*Rasch difficulty*) and individual student performance on the test (*Rasch ability*). This approach allows test developers to construct test instruments which discriminate between students and effectively match the required ability range.

Test instrument development

In order to develop test instruments with effective constructs and best practice measurement models, test developers and psychometricians must follow a strict quality control process, including:

- Clear identification of assessment purpose and target student population
- Development of a test construct and specifications
- Design and development of test items matched to the test construct
- Trialling of items and initial data collection

- Construction of a Rasch measurement scale
- Selection of items targeted to student ability
- Design of test forms and reporting tools
- Continuous analysis and research to validate test instruments and constructs

Test administration

There are numerous practical and technical requirements for the effective and equitable administration of high stakes selection testing which is responsive to demands of transparency, equity and security throughout the program cycle.

A number of operational mechanisms must be employed to ensure consistency in test administration for high stakes selection testing. Clear and unambiguously articulated policies and procedures are required to be established in risk management, data security, confidentiality and disclosure, equity and fairness, security in test booklet printing, packaging and despatch, candidate identification and registration and reporting systems to manage results.

The management of test centres is an important component in the administration of high stakes selection assessments. Some of the key administrative requirements include the training and provision of test supervisors, accommodating regional, inter-state and overseas candidates, as well as making additional secure arrangements for candidates unable to sit on the designated test day.

Test administration procedures must provide an equitable test environment for candidates with special needs or requirements. Examples include the provision of altered test materials (i.e. Braille) and the allocation of extra time if appropriate. Comprehensive administrative procedures are required to provide assurance in this area including formal submission of candidate documentation relating to special circumstance and the establishment of dedicated review committees to evaluate all applications.

Conclusions and Recommendations

- Strategies for addressing the needs of academically able students can be properly targeted using effective tools to identify students with high academic potential.

- Identification of giftedness has been limited by inconsistent and problematic assessment. Objective assessment is an important factor in the identification of academically able students but such assessment should have appropriate research-proven validity, reliability, sensitivity and reference ranges.

- There are a range of practical and technical requirements to ensure effective and equitable administration of high stakes selection tests. Significant attention must be given to the management of assessment programs and the associated demands of transparency, equity, and security throughout the program cycle.

References

Andrich, D. (1988). *Rasch Models for Measurement.* SAGE University paper. CA: SAGE Publications

Alvino, J., McDonnel, T. C. and Richert, S. (1981). 'National survey of identification practices in gifted and talented education', *Exceptional Children*, 48, 124-132

Carrol, J. B. (1993). *Human cognitive abilities: a survey of factor analytic studies*, Cambridge University Press, New York

Deary, I, Strand, S. Smith, P and Fernandes, C. (2007). 'Intelligence and educational achievement', *Intelligence*, 35 (1), 13-21

Flynn, J. R. (1987). 'Massive IQ gains in 14 nations: What IQ tests really measure.', *Psychological Bulletin*, 101 (2)171-191

Flynn, J. R. (2009). *What is Intelligence: Beyond the Flynn effect*, Cambridge University Press, Cambridge

Frey, M. C. and Detterman, D. K., (2004), 'Scholastic Assessment', *Psychological Science*, 15 (6), 373-8

Gagné, F. (1993). Constructs and models pertaining to exceptional human abilities. In K. A. Heller, F. J. Mönks & A. H. Passow (Eds.), *International Handbook of Research and Development of Giftedness and Talen*t (pp. 63–85). Oxford: Pergamon Press.

Gardner, H. (1999). 'Intelligence Reframed: Multiple Intelligences for the 21st Century', Basic Books, New York.

Jarosewich, T., Pfeiffer, S. I. and Morris, J. (2002). 'Identifying Gifted Students Using Teacher Rating Scales: A Review of Existing Instruments', Journal of Psychoeducational Assessment, 20, 322-336

Mayer, J. Caruso, D and Salovey, P. (1999). Emotional Intelligence Meets Traditional Standards for an Intelligence. Intelligence, 27, (4), 267 – 298.

Mayer, J., Salovey, P. and Caruso, D. (2000). Models of Emotional Intelligence. In R. Sternberg (ed) Handbook of Intelligence. Cambridge: Cambridge University Press.

Pfeiffer, S. I. (2001). 'Professional psychology and the gifted: Emerging practice opportunities' *Professional Psychology: Research and Practice*, 32, 175-180

Pfeiffer, S. I., Petscher, Y. and Kumtepe, A. (2008). 'The Gifted Rating Scales-School Form: A validation study based on age, gender, and race', *Roeper Review*, 30, 140-146

Rasch, G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Education Research, 1960. (expanded edition, The University of Chicago Press: Chicago).

Robinson, N. M., Zigler, E. and Gallagher, J, (2000). 'Two tails of the normal curve', *American Psychologist*, 55, 1413-1424

Sternberg, R. J. (1997). A Triarchic View of Giftedness: Theory and Practice. In N. Coleangelo and G. A. Davis (Eds.), *Handbook of Gifted Education*(pp 43-53), Allyn and Bacon, Boston, MA.

Thomson, S. and De Bortoli, L. (2008) Exploring Scientific Literacy: How Australia measures up. The PISA 2006 survey of students' scientific, reading and mathematical literacy skills. Camberwell: ACER.

Thomson,S., Wernert,N., Underwood,C.& Nicholas,M.(2008). TIMSS 2007: Taking a closer look at mathematics and science in Australia. Camberwell:ACER.

Thomson, S., De Bortoli, L., Nicholas, M., Hillman, K., & Buckley, S. (2011) Challenges for Australian Education: Results from PISA 2009. Camberwell: ACER.

U.S. Department of Education, Office of Educational Research and Improvement (1993). *National Excellence: A case for developing America's talent*, Washington, DC

Non-ACER Tests

GATES - Gilliam, J. E., Carpenter, B. O., & Christensen, J. R. (1996). *Gifted and Talented Evaluation Scales*. Waco, TX: Prufrock Press

GES-2 – McCarney, S. B. and Anderson, P. D. (1989). *Gifted Evaluation Scale*, Columbia MO, Hawthorne Educational Services

GRS – Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales.* San Antonio, TX: The Psychological Corporation.

SRBCS - Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. L. (2002). *Scales for Rating the Behavior Characteristics of Superior Students: Revised edition.* Mansfield Center, CT: Creative Learning Press.

Raven Progressive Matrices Series – e.g. Raven, J., Raven, J.C., & Court, J.H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX: Pearson Assessment.

Stanford-Binet Series – e.g. Thorndike, E.P. Hagen, and J.M. Sattler. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*, Riverside Publishing Co.

Wechsler Series – e.g. Wechsler, D. (2004). The Wechsler intelligence scale for children—fourth edition. London: Pearson Assessment.

Appendix 1

ACER currently develops and/or administers assessment instruments used for identification and selection of academically able students at school level for a number of programs, including:

- **Assessment for entry into higher ability streams and programs**:
    - Higher Ability Selection Test (national and international)
    - Abstract Reasoning Test (national)
    - IGNITE (SA)
- **Selective school entry assessment**:
    - Selective High School Entrance Test (NSW)
    - Opportunity Classes Placement Test (NSW)
    - Queensland Academies Selection Testing (Qld)
- **Scholarship selection assessment**:
    - Co-operative Scholarship Testing Program (national)
    - Scholarship Selection Test (national)
    - Australian Co-operative Entry Program (SA, WA)
    - Primary Scholarship Program (national)

Three of the best known of these programs used to identify students with superior academic ability are the Higher Ability Selection Test (HAST), the Co-operative Scholarship Testing Program (CSTP) and the Selective High School Entrance Test (SHSET).

- HAST is an ability test used by secondary schools to identify academically gifted students for participation in accelerated learning and enhancement programs. The program delivers tests in reading comprehension, mathematical reasoning, abstract reasoning and written expression at all levels of secondary school including first level entry. The test material is presented in a wide range of contexts and includes some items that are highly verbal, some that require the ability to perceive concepts at an abstract level, others that require the ability to synthesize, extrapolate or make inferences and some that apply logical and strategic thinking to solving a problem.

  A version of HAST for use in primary school has been developed recently in response to requests from schools for a screening process that would assist them in identifying students for participation in gifted and talented programs at primary level.

- CSTP is a program that identifies academically superior students through the assessment of verbal and quantitative reasoning skills, and written expression. It operates mainly in the selection of scholarship students at the transition to secondary school (Level 1) but there are versions that operate at higher levels (Level 2 and 3). There is also a version (PSP) that operates at Year 4 for selection into Year 5.

- SHSET is the major assessment used to select students into NSW Selective High Schools. It consists of four components: Reading, Mathematical Reasoning, General Ability (Quantitative, Verbal and Abstract) and Writing. A similar test operates for selection into Years 5 and 6.

Appendix 2

Figure 2.1 below shows the proportion of students achieving at the highest levels of PISA 2009 (proficiency levels 5 and 6) in mathematical literacy. Sixteen per cent of Australian students, compared to 50 per cent of students in Shanghai and 36 per cent of those in Singapore, achieved these standards.

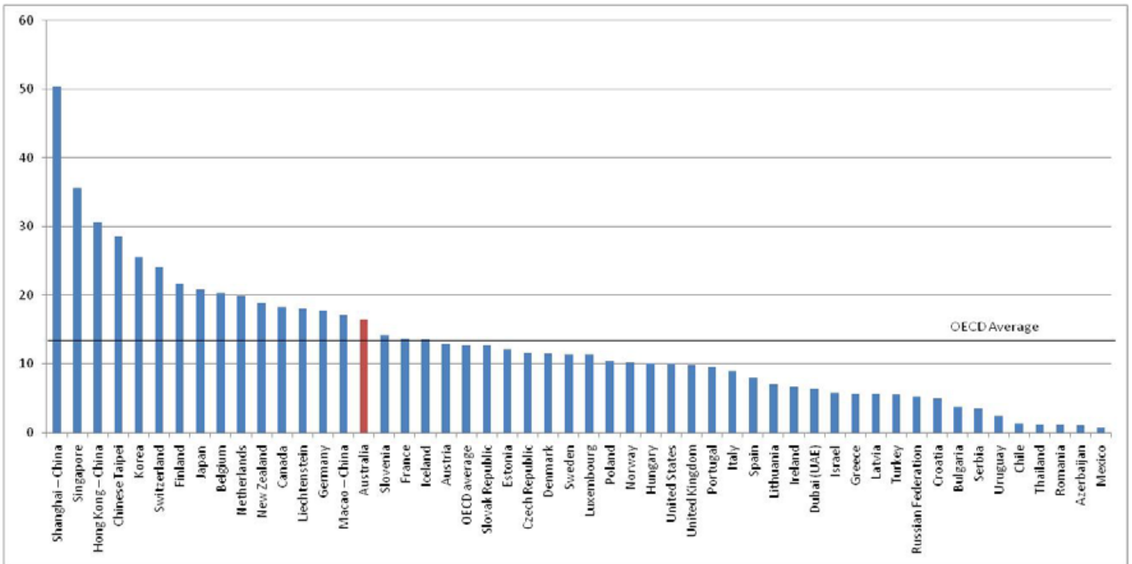

FIGURE 2.1: PROPORTION OF STUDENTS ACHIEVING BANDS 5 AND 6 IN MATHEMATICAL LITERACY, PISA 2009

Figure 2.2 below shows the proportion of students achieving at the highest levels of PISA 2009 (proficiency levels 5 and 6) in reading literacy. Thirteen per cent of Australian students, compared to 19 per cent of students in Shanghai and 16 per cent of those in New Zealand and Singapore, achieved these standards.

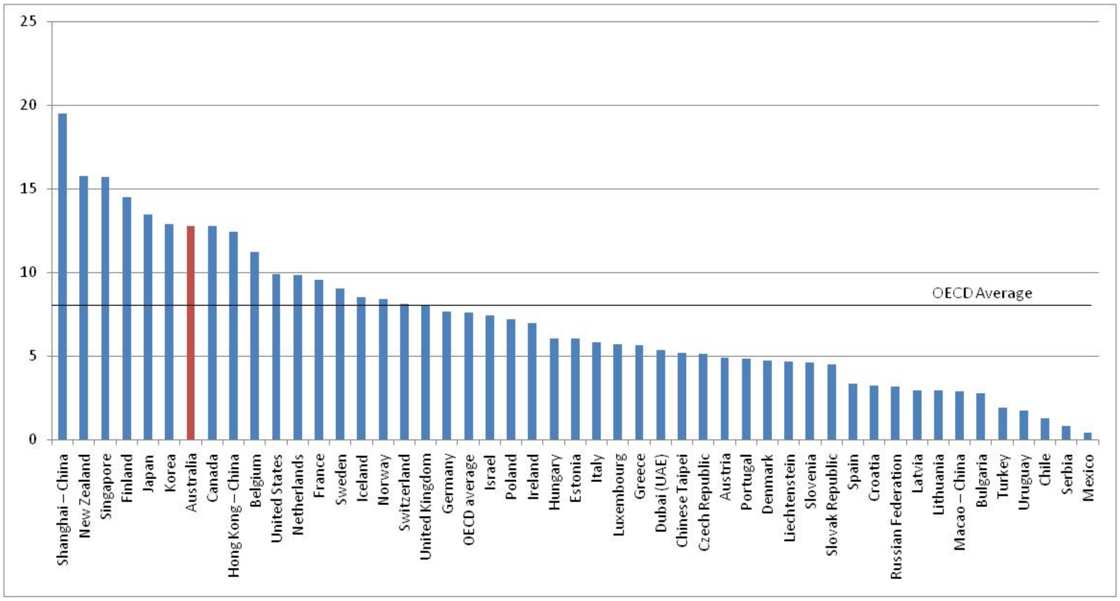

FIGURE 2.2: PROPORTION OF STUDENTS ACHIEVING BANDS 5 AND 6 IN READING LITERACY, PISA 2009

Appendix 3

Figure 3.1 below is a plot that shows the result when items were selected for a test without trial testing. Candidate ability is given by the distribution of Xs on the left and item difficulty by the item numbers on the right. The result is a mismatch between candidate ability and item difficulty, and mediocre reliability (0.5), with the test unable to discriminate properly between students of different ability, and particularly amongst those with higher abilities for whom there are few matching items. (Reliability also depends on the number of items, improving with more items. The reliability would have been worse with fewer items.)

```
           |
           |
          X|
          X|
          X|
         XX|
          X|
         XX|
         XX|
        XXX|
        XXX|
      XXXXX|
    XXXXXXX|28
      XXXXX|
      XXXXX|
     XXXXXX|14
   XXXXXXXX|27
   XXXXXXXX|
 XXXXXXXXXX|
   XXXXXXXX|31 46
   XXXXXXXX|26
   XXXXXXXX|17 20 30
  XXXXXXXXX|16
   XXXXXXXX|33
  XXXXXXXXX|5 87
   XXXXXXXX|85
     XXXXXX|
```

```
            XXXX|34 41
            XXXX|58 88
           XXXXX|
             XXX|
              XX|
              XX|4 8 59 73
               X|11
               X|45 86
               X|80
                |40 66 81 82 83
               X|
                |
                |10 52
                |7
                |6
                |44 71
                |74
                |2 15 55 76 77
                |21 57 84
                |
                |60
                |24 72
                |1 3 12 13 18 19 22 23 25 29 32 35 36 37 39 42 43 47 48 49 51 53 54 56 61
62 63 64 65 67 68 69 70 75 78 79
```

FIGURE 3.1: ITEM PLOT FOR TEST CONSTRUCTED WITHOUT ITEM TRIALLING

In contrast, the Figure 3.2 below shows a better match between candidate ability and item difficulty, and test reliability (0.8) is better than the previous case even though fewer items were used.

```
                               |
                          XX   |
                               |
                        XXXX   |      45  69
                               |
                         XXX   |      53
                    XXXXXXXX   |
                           X   |
                     XXXXXXX   |      54
                       XXXXX   |
                 XXXXXXXXXXX   |      52  61
                           X   |      55
              XXXXXXXXXXXXXX   |       4  33   60  62
         XXXXXXXXXXXXXXXXXXX   |      68
      XXXXXXXXXXXXXXXXXXXXXX   |
                   XXXXXXXXX   |      27  39
XXXXXXXXXXXXXXXXXXXXXXXXXXXX   |
          XXXXXXXXXXXXXXXXXX   |      25  47
             XXXXXXXXXXXXXXX   |      34
         XXXXXXXXXXXXXXXXXXX   |      10  46
                   XXXXXXXXX   |      26
              XXXXXXXXXXXXXX   |      28  38
              XXXXXXXXXXXXXX   |       9  23  24  37  44
              XXXXXXXXXXXXXX   |       3  17  29
              XXXXXXXXXXXXXX   |      16
             XXXXXXXXXXXXXXX   |      18
                      XXXXXX   |       1
                               |       2
                          XX   |      15
                     XXXXXXX   |
                           X   |
                          XX   |
                               |
                          XX   |
                               |
```

FIGURE 3.2: ITEM PLOT FOR TEST CONSTRUCTED WITH ITEM TRIALLING

Figure 3.3 below represents a test whose items are focussed on the top students. Although the previous test could provide reasonable discrimination for high ability students, this one may be even better.

```
                                        |
                                        |23
                                        |
                                        |
                                        |29
                                       X|35
                                       X|
                                       X|9
                                      XX|25 30 33
                                     XXX|26 27 28
                                   XXXXX|24 32
                                  XXXXXX|
                                  XXXXXX|31
                              XXXXXXXXXX|17 34
                               XXXXXXXXX|8 18
                            XXXXXXXXXXXX|16 21 22
                        XXXXXXXXXXXXXXXX|5 15
                       XXXXXXXXXXXXXXXXX|10
              XXXXXXXXXXXXXXXXXXXXXXXXXX|
                  XXXXXXXXXXXXXXXXXXXXXX|
             XXXXXXXXXXXXXXXXXXXXXXXXXXX|4 20
             XXXXXXXXXXXXXXXXXXXXXXXXXXX|
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|3 7 14 19
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|13
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|2
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|12
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|6
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|11
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
                XXXXXXXXXXXXXXXXXXXXXXXX|
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
                     XXXXXXXXXXXXXXXXXXX|1
                       XXXXXXXXXXXXXXXXX|
                   XXXXXXXXXXXXXXXXXXXXX|
                            XXXXXXXXXXXX|
                            XXXXXXXXXXXX|
                             XXXXXXXXXXX|
                                  XXXXXX|
                                    XXXX|
                                     XXX|
                                      XX|
                                      XX|
                                       X|
                                       X|
                                        |
                                       X|
```

FIGURE 3.3: ITEM PLOT FOR TEST TARGETED AT ACADEMICALLY ABLE STUDENTS

Only by trial testing and specifically selecting items for such a purpose and target population can the optimum test be produced.