Student Aptitude Test for Tertiary Admission (SATTA)
Pilot Program

Evaluation Report for the

Department of Education, Employment and Workplace Relations (DEEWR)

Hamish Coates, Daniel Edwards, Tim Friedman

March 2010

Australian Council for Educational Research
ACN: 004 398 145; ABN: 19 004 398 145

ISBN: 978-0-642-78041-6
# TABLE OF CONTENTS

## TABLES AND FIGURES

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>III</td>
</tr>
</tbody>
</table>

## ACKNOWLEDGEMENTS

IV

## EXECUTIVE SUMMARY

5

- Focus and scope of the evaluation
- Background and rationales
- Engagement with uniTEST
- uniTEST criterion validity
- A new admissions architecture

## 1 INTRODUCTION AND OVERVIEW

9

- Background and context
- Focus and scope of the evaluation
- An overview of uniTEST
- Overall research approach
- An overview of this report

## 2 FORMATIVE CONTEXTS

14

- Introduction
- Worldwide use of university admissions tests
- University admissions in Australia
- Institution’s experiences with aptitude tests
- Chapter summary

## 3 ENGAGING INSTITUTIONS AND APPLICANTS

31

- Introduction
- Characteristics of the applicant population
- Detailed analysis of the admitted population
- The independent impact of characteristics on uniTEST scores
- Chapter summary

## 4 VALIDITY ANALYSES

47

- Introduction
- The concept of validity
- Sample characteristics
- Analysis of concurrent validity
- Analysis of predictive validity
- Chapter summary

## 5 BOLSTERING GROWTH IN APTITUDE ASSESSMENT

64

- Taking stock on the added value
- A new admissions architecture
- An implementation approach
- Next steps

## REFERENCES

71

## APPENDIX 1: REGRESSION COEFFICIENTS DETAIL

75
TABLES AND FIGURES

List of tables
Table 1: Sample of discipline-specific university admissions tests used worldwide ........................................ 21
Table 2: Externally developed aptitude tests used for admission to Australian universities ........................... 27
Table 3: uniTEST applicant and admitted populations and samples by institution ........................................ 32
Table 4: uniTEST point difference for specified variables (unstandardised regression coefficients) ............ 45
Table 5: Year 12 ENTER point difference for specified variables (unstandardised regression coefficients) .... 45
Table 6: uniTEST evaluation sample characteristics .................................................................................... 50
Table 7: uniTEST and Year 12 correlations ..................................................................................................... 51
Table 8: uniTEST and Year 12 shared variance (per cent) .............................................................................. 51
Table 9: uniTEST and Year 12 correlations with academic performance at university ................................ 55
Table 10: Control and uniTEST group standardised regression estimates .................................................... 60
Table 11: Regression coefficients for GPA scores, Models 1 to 5, Semester 1 results .................................... 75
Table 12: Regression coefficients for GPA scores, Models 1 to 5, Semester 2 results .................................... 75
Table 13: Regression coefficients for GPA scores, Models 1 to 5, Semester 3 results .................................... 76
Table 14: Regression coefficients for GPA scores, Models 1 to 5, Semester 4 results .................................... 76

List of figures
Figure 1: Project schedule, stage 1 .................................................................................................................. 12
Figure 2: Project schedule, stage 2 ................................................................................................................ 12
Figure 3: Gender by admission type ............................................................................................................ 34
Figure 4: uniTEST item difficulty estimates by gender .................................................................................. 35
Figure 5: Age by admission type .................................................................................................................. 36
Figure 6: Proportion of university enrollees who enrolled directly after completing school by admission type .. 36
Figure 7: Language background by admission type ...................................................................................... 37
Figure 8: Socioeconomic status by admissions type ..................................................................................... 38
Figure 9: Parental education level by admission type .................................................................................... 39
Figure 10: Residential location by admission type ....................................................................................... 40
Figure 11: School sector by admission group ............................................................................................... 41
Figure 12: uniTEST entrants to university by Year 12 score ....................................................................... 42
Figure 13: Broad field of education by admission type ................................................................................. 43
Figure 14: Relative influence on uniTEST outcomes (standardised regression coefficients) ..................... 44
Figure 15: Impact of low socioeconomic status (as opposed to high socioeconomic status) on uniTEST and Year 12 outcomes, standardised regression coefficients .................................................. 46
Figure 16: Predictive validity analytical considerations ............................................................................... 54
Figure 17: Explained variance in GPAs from combinations of uniTEST scores and Year 12 marks ........... 56
Figure 18: Explained variance in GPAs from combinations of uniTEST scores and Year 12 marks .......... 57
Figure 19: Comparison of student participation over time for individuals accepted on the basis of their uniTEST performance and control students .................................................................................. 58
Figure 20: Mean GPAs for uniTEST and control students per semester with 95% confidence bands .......... 59
Figure 21: Control and uniTEST group engagement scale scores .................................................................. 61
Figure 22: Control and uniTEST group outcome scale scores ..................................................................... 62
ACKNOWLEDGEMENTS

The Australian Council for Educational Research (ACER) would like to warmly thank the Department of Education, Employment and Workplace Relations (DEEWR) for funding this evaluation.

Professor Sam Ball made a formative contribution to this report, and prepared material included in chapters two and four. Sadly, Sam died in December 2009. This report is dedicated to his lifelong interest in improving education.

Professor Richard James provided text on future contexts included in chapter five.

Professor Peter Hill offered formative thoughts and contributions on the most effective means of implementing an aptitude test with school leaver cohorts in Australia.

ACER colleagues who assisted with facets of this study include Ms Marita MacMahon Ball, Mr Luc Le, Ms Ali Radloff, Dr Gary Marks, Professor Geoff Masters, Ms Susan Nankervis and Ms Tanya Williams.

A large number of people and organisations contributed to this evaluation through various consultation processes. This includes all universities, many peak bodies, many government agencies, and many independent experts. We are very grateful for their input.

ACER is grateful to Professor Steven Schwarz, Professor Merran Evans and Professor David Andrich who gave early and formative feedback on the study’s methodology.

Finally, we are very grateful to the students who sat uniTEST and were willing to have their results used for research purposes.
EXECUTIVE SUMMARY

Focus and scope of the evaluation

Since 2007 the Australian Government Department of Education, Employment and Workplace Relations (DEEWR) has funded the Student Aptitude Test for Tertiary Admission (SATTA) pilot program. SATTA involves the supply, management and evaluation of uniTEST, and the evaluation of the Special Tertiary Admissions Test (STAT). This report documents the evaluation phase of the program, focusing in particular on various aspects of uniTEST. Seven recommendations are made.

Significant policy change is planned for Australian higher education over the next few years, with both government and institutions seeking new ways to make the system larger, more inclusive and more productive. In this context, it is vital to develop new transparent mechanisms for helping each student understand her or his potential and access the system. This report details how aptitude testing can play an important role.

The evaluation was conducted between late 2007 and early 2010. The evaluation involved background planning and review, data specification and collection, psychometric and statistical analysis, widescale consultation, and documentation and reporting. While many aspects of the data collection and analysis were difficult and complex, the project remained on schedule and has delivered a number of formative insights and findings on aptitude testing in Australian higher education.

The academic aptitude test, uniTEST, was developed jointly by ACER and Cambridge Assessment in the UK. Within the context of university selection, the purpose of uniTEST is to enhance the effectiveness of admissions processes as they attempt to select students with the ability to undertake tertiary education, despite discouraging or ambiguous achievement scores at the end of high school. uniTEST is administered by individual universities in association with ACER.

Evaluating the criterion validity of uniTEST was an important part of the national SATTA pilot. This has involved an analysis of concurrent validity – exploring how uniTEST relates to Year 12 achievement – and predictive validity, analysis of the extent to which test results predict future university performance. uniTEST is a high-stakes test that affects the future of test takers, therefore, its capacity to operate as an effective selection mechanism is critical. The criterion validity of the instrument provides an index of the extent to which this is the case.

Background and rationales

A university degree, and the higher order skills that it confers, is increasingly important for securing employment in the Australian labour market. In recent decades, growth in occupations that require a university degree is greater than growth in any other occupation type in Australia (Birrell, Edwards & Dobson, 2007; Birrell & Edwards, 2007). Student demand for university education is also high. As recent Federal Government policies have emphasised, the provision of university education is crucial to the health of the Australian economy (Australian Government, 2009).

In order to provide university education at a level of high quality and in fields that ensure graduates have successful labour market outcomes, and which facilitate growth in the economy, it is important that pathways into the system are well understood. Ensuring that all capable people are given the opportunity to study at university plays a major role in ensuring the productivity of Australian tertiary education.

Yet university admissions procedures in Australia have historically grown in ways that may not be most effectively servicing contemporary needs. Australia needs valid and efficient university admissions processes in order to optimise the equity and outcomes of higher education. Admissions procedures play a major role in the quality and productivity of our university education, but Australia lacks evidence about the comparability and efficacy of the various mechanisms currently used for selection.
This evaluation examines whether, through the provision of baseline and objective data, uniTEST offers a valid and efficient means for ensuring that people who are capable of success at university are able to gain admission. The study evaluates if these assessments enable the identification and inclusion of ‘latent talent’ that might otherwise be lost to educational development. As examined, to the extent that these assessments can function in this regard, they have the potential to play an increasingly significant role in helping ensure that the complexities of contemporary tertiary admissions processes do not threaten the validity or productivity of selection processes and outcomes.

The report analyses the role played by aptitude testing in comparative international systems and within Australian institutions. This research has revealed that the use of aptitude tests and multiple admissions criteria for selecting university candidates is common throughout the world. Such tests are used instead of, or to supplement, final-year school outcomes. There are tests used for general entrance as well as tests specific to particular disciplines and courses. Given the changes in the Australian system over the past half century, coupled with the recent higher education policy direction of the Australian Government, there appear to be cogent rationales for the wider and more transparent use of aptitude tests in selecting university candidates in Australia. This is not a radical shift, for aptitude tests are already used within the Australian higher education sector for student selection.

Importantly, the analysis, which involved a survey of institutional leaders within Australian universities, demonstrates that there appears to be an interest within the Australian system in using these tests – so long as there is evidence to suggest that they can be used effectively to aid student selection. Review of international, national and institutional contexts leads to the first recommendation:

**Recommendation 1:** Nationally coordinated implementation of uniTEST should be considered as a means of improving the transparency, efficiency and international relevance of university admissions in Australia.

**Engagement with uniTEST**

Chapter three reports an analysis of how institutions and individuals have engaged with uniTEST. uniTEST has been used by six institutions over the past few years to aid the selection process for admitting candidates to undergraduate degrees. While these institutions have used the test in a range of different ways, around 30 per cent of people who sat uniTEST were subsequently admitted to university.

Results from this analysis are important. They show that, in general, those who gain access to university via uniTEST have slightly different characteristics than are found in the general university population. This finding suggests that uniTEST has the potential to increase diversity within the university population, especially in terms of gender and socioeconomic status. Importantly, uniTEST scores – unlike Year 12 results – are not correlated with socioeconomic status.

University admissions in Australia are high stakes for individuals, institutions and Australia. With a view to continuous improvement, it is imperative that evidence-based approaches be used to enhance the efficiency and validity of how people are admitted to university – the second recommendation.

**Recommendation 2:** To ensure the most effective implementation, expansion of the use of aptitude tests with school student and leaver populations should be accompanied by ongoing analysis of the characteristics of the applicant and admitted populations.

**uniTEST criterion validity**

Current evidence on concurrent validity suggests a complex relationship between uniTEST and Year 12 scores. Most of these relationships are not statistically significant with the exception of those that pertain to the Quantitative Reasoning component. However the average shared variance between the measures is low. Overall, there appears to be a broadly divergent relationship between the measures, which suggest that they play a complementary role in the selection process. These results affirm that complementary role played by uniTEST in admitting school leavers to university. They underpin the third recommendation.
Recommendation 3: It is recommended that further work be undertaken to examine the extent to which aptitude (as measured by uniTEST) complements Year 12 achievement. With greater numbers and more information on school outcomes, examination in relation to Year 12 score bands and individual Year 12 subject scores to be examined.

Analyses of uniTEST predictive validity require comparison against suitable predictive measures. For this, grade point average (GPA) data was collected from participating universities. Unfortunately, while widely used, the empirical properties of the GPA metrics remain unknown. This introduces unexplainable variation into the results, and underpins the need to develop a robust generalisable measure of achievement for Australian higher education – the fourth recommendation.

Recommendation 4: Considerable value would be derived from developing a robust GPA for Australian higher education. Many GPA metrics already exist, but these are not well defined or validated, and are not implemented consistently.

Results from the analyses of predictive validity suggest that uniTEST results alone explain more variation in university GPAs as do Year 12 scores. From analysis of different combinations of uniTEST and Year 12 scores, it appears that for the populations under examination a combination of both measures offers a more powerful means of predicting first-year performance than either measure on its own. This is a major finding that affirms the value that an aptitude test can add to university admissions.

Recommendation 5: Predictive validity analyses demonstrate that aptitude test data adds to the power of admissions processes. To enhance the prognostic power of admissions processes, data on school achievement should be augmented with evidence from uniTEST.

Results from analysis of uniTEST and control group students suggest that uniTEST was able to facilitate the admission to university of students who otherwise would not have received a place, and that these students performed on par with their counterparts who gained entry through other means, most commonly through Year 12 scores. While the evidence is limited, both uniTEST and control group students appeared to report similar levels of academic engagement as well as learning and skill development. These findings must be hedged by the caveats that surround the current study, but nonetheless suggest there is a significant role that uniTEST can play in identifying individuals who have the potential to succeed at university, and enabling these people to be included in the system.

Together, analysis of the criterion validity of uniTEST affirms that it plays a valuable role in university admissions. This is not surprising given the widespread international use of aptitude tests, the need to grow and diversify admissions in Australia, and the extensive use of varying assessments by Australian institutions. This leads to the report’s most significant recommendation.

Recommendation 6: Based on evidence from the SATTA pilot it is recommended that uniTEST be implemented as a means of diversifying and complementing the data factored into the university admissions of school leavers in Australia.

By way of summary, early empirical indications highlight that the positive role to be played by uniTEST are favourable, and that it has the potential to identify ‘latent talent’ and facilitate the inclusion of able individuals in the system. But early empirical indications also show that many of the results are statistically inconclusive, due largely to the small and idiosyncratic nature of the available sample. Given the growing importance of assuring the validity of this assessment, there is an evident need for both larger and ongoing study. Ideally, the analysis of concurrent and predictive relationships should be woven into continuous quality improvement processes that underpin routine reflective practice.

A new admissions architecture

The opportunity now exists across Australia to develop new approaches to university selection that offer simplicity, consistency and transparency for prospective students and that maintain the benefits of
coordinated processes for application and selection for the majority of places. The final chapter in this report takes stock of university admissions in Australia, reviews emerging contexts and opportunities, and recommends an improved approach for national implementation uniTEST.

The report details an implementation process that involves:

- the assessment being promoted by key agencies as a credible alternative quantitative selection criteria to achievement tests;
- candidates sitting uniTEST during their senior secondary study, at some stage from the end of Year 10;
- the test being administered in a flexible mode, likely online, and in multiple sittings;
- informative reports being provided to assist students and institutions with their course choice and admissions decisions;

The test will vary in its relevance to institutions depending on factors such as selectivity, course characteristics and demographics, but it should be endorsed universally. As with current admissions practices, the process should be centrally coordinated – though not regulated – in a way that sustains institutional autonomy over selection decisions.

This process is highlighted in the study’s final recommendation.

| Recommendation 7: uniTEST should be implemented in a nationally coordinated way that is flexible, targeted at senior secondary students, and able to provide diagnostic information for both individuals and multiple institutions. |

This report closes by considering the options available to progress implementation of an aptitude test for use with school leavers seeking entry to Australian higher education:

- The first option involves a continuation of the past approach, which involves ACER working in a somewhat exploratory fashion with individual institutions.
- The second and recommended approach involves moving the assessment into schools, and could help to enhance student aspirations, inform subject and tertiary choice, enhance persistence, and provide a sound and complementary quantitative means of diversifying and perhaps compensating for the evidence used to admit school students into university.
- The third option involves factoring the implementation of an aptitude assessment into a much broader conversation about tertiary admissions. The reflection, consultation and evidence offered in this report highlights the significant dividends that may be yielded from this process. Admissions processes are a fundamental facet of university education in Australia, yet perhaps one of the least well researched and discussed. The private and competitive nature of the process may partly explain this state of affairs, yet it does not lessen the need for improvements that bring practice into line with contemporary system contexts and needs.

Of course, the third option given here may well emerge from the second, which has itself grown from the first. The third option does not necessarily (or at all) imply a radical revisioning of university admissions for school leavers in Australia. What it does advance is the need for ongoing research and development of this significant facet of Australian higher education. Indeed, this is the stance that underpins the ethos, approach and insights of this study, and which is imperative for ensuring that all school students who wish to study at university have the opportunity to demonstrate that they have the capacity to succeed.
1 INTRODUCTION AND OVERVIEW

Background and context

A university degree, and the higher order skills that it confers, is increasingly important for securing employment in the Australian labour market. In recent decades, growth in occupations that require a university degree is greater than growth in any other occupation type in Australia (Birrell, Edwards & Dobson, 2007; Birrell & Edwards, 2007). Student demand for university education is also high and will grow with current Australian Government plans for expanding the system (Australian Government, 2009). The provision of university education is crucial to the health of the Australian economy.

In order to provide university education at a level and quality and in fields that ensure graduates have successful labour market outcomes, and which facilitate growth in the economy, it is important that pathways into the system are well understood. Ensuring that all capable people are given the opportunity to study at university plays a major role in ensuring the productivity of Australian tertiary education. If able applicants are not given the chance to succeed, or if people are not able to advance their skill development, then the full potential of individuals and hence Australia’s skilled workforce remains unrealised.

Yet university admissions procedures in Australia have historically grown in ways that may not be most effectively servicing contemporary needs. Australia needs valid and efficient university admissions processes in order to optimise the equity and outcomes of higher education. Admissions procedures play a major role in the comparability and productivity of our university education, but Australia lacks evidence about the comparability and efficacy of the various mechanisms currently used for selection.

Today, tertiary admissions processes in Australia face a number of key challenges. By way of example:

- the participation in university of students for identified equity groups remains disappointingly low (CSHE, 2008; Edwards, 2008a; Coates & Krause, 2005), and in some cases has declined, raising questions about whether alternative entry mechanisms may help improve the participation of persistently under-represented individuals who are demonstrably able to succeed;
- much work has been done to build and clarify existing pathways between vocational and higher education (see, for example: MCEETYA, 2008; VRQA, 2008), but significant uncertainties remain that have the potential to hinder individual progression through the system;
- with expansion in the system, particularly over the last twenty years but increasingly into the future, ‘potential graduates’ have been accessing the system from increasingly diverse social, educational and professional backgrounds (DEEWR, 2008), putting pressure on processes that were developed to manage much less complicated student flows;
- while the achievement-oriented metrics on which much tertiary admission has historically been based can be influenced by demographic or educational factors, and may not provide sufficient or relevant evidence for making selection decisions, there is a lack of objective measures which can be used for cross-validation; and
- while the diverse schedules developed by multiple institutions and agencies to manage the complexities associated with student admissions are not inherently problematic, this complexity leads to a lack of transparency that is difficult for prospective students to follow and hinders the evaluation of effectiveness of admissions processes.

This formative evaluation examines whether, through the provision of baseline and objective data, an aptitude test offers a valid and efficient means for ensuring that school leavers who are capable of success at university are able to gain admission. As part of this, it considers whether aptitude testing enables the identification and inclusion of ‘latent talent’ that might otherwise be lost to educational development. As examined in the report, to the extent that such an assessment can function in this regard, it has the potential to play an increasingly significant role in helping ensure that the complexities of contemporary tertiary admissions processes do not threaten the validity or productivity of selection processes and outcomes.
Aptitude testing would appear to have an important role to play in a contemporary operating environment in which institutions and government are search for new forms of transparency, new pathways, and new ways of measuring performance and productivity. In principle, objective data on individuals’ aptitude for university study can enhance tertiary admissions processes in a number of ways. For instance, aptitude tests can:

- help identify students who independent of education background have the intellectual capacity to succeed at university;
- provide common and transparent inferential foundations for selection decisions;
- be designed and managed to ensure consistent performance across demographic subgroups;
- help manage competition for small numbers of highly prized university places;
- enhance the efficiency of a student’s flow through the tertiary system; and
- cross-validate information available through achievement metrics.

Of course, no assessment can address all problems or potentialities facing contemporary university admissions. Improving the access and participation of disadvantaged students hinges on implementing better methods of defining and measuring social disadvantage, strategies to build secondary students’ aspirations and address under-achievement well before the tertiary admission stage, and linkages between secondary and tertiary systems. Enhancing individual mobility between tertiary qualifications and providers also depends on promoting ongoing structural adjustments and alignments. Responding most effectively to individual and societal demands would likely involve increasingly sophisticated studies of economic and market trends, and possibly also new financing and regulatory models. While shaped by these considerations, this evaluation has a more modest focus on the technical benefits that the provision of valid objective data may confer on tertiary admissions in Australia.

**Focus and scope of the evaluation**

From 2007 the Australian Government Department of Education, Employment and Workplace Relations (DEEWR) implemented a pilot program of the Student Aptitude Test for Tertiary Admission (SATTA). The program involved the supply, management and evaluation of the aptitude assessment named uniTEST and, in 2008, the evaluation of the Special Tertiary Admissions Test (STAT). This report documents the evaluation phase of SATTA, looking in detail at various aspects of uniTEST. A report on STAT was provided to DEEWR in late 2008, and key results are published in Coates and Friedman (2010). This report does not provide further information on the STAT evaluation.

Establishing the criterion validity of uniTEST was the main focus of the evaluation. Criterion validity incorporates concurrent validity which involves review of the relationship between uniTEST results and those of other assessments undertaken simultaneously. Predictive validity is a further component of criterion validity, and requires determination of the extent to which test results predict future university performance. uniTEST is a high-stakes test that affect the future of test takers, and its capacity to operate as an effective selection mechanism is critical. The criterion validity of the instrument provides an index of the extent to which this is the case. Hence the primary question underpinning this evaluation was: Does uniTEST have suitable levels of criterion validity to support its use as a selection instrument?

The evaluation was expanded in late 2008 to look beyond criterion validity at various facets of uniTEST context and performance. Hence this report provides information on how comparison (largely OECD) countries use aptitude tests, on Australian institutions’ experience with such assessments, and on people’s engagement with the tests. A considerable amount of consultation has been conducted as part of the study, and by way of conclusion the report takes a wider look at how the pilot has progressed, and it offers suggestions for developing university admissions in Australia.

It is important to stress at the outset that uniTEST is not designed to predict achievement at university. The purpose of the assessment is to identify individuals with the capacity to undertake university study. As is emphasised throughout this report, the difference here is in the distinction between ‘ability’ and ‘achievement’, a difference commonly confused in informal conversation. While ability pertains to the capacity to perform, achievement refers to demonstrated performance, performance which is influenced by a
wide range of factors. Having said this, it is desirable that a positive relationship exists between measures which are used to admit students to university study and performance during the course.

Several objectives were beyond the scope of the current evaluation. The study did not seek to review individual university selection procedures in detail, although it recognised that the results may have indirect implications on the management of these processes. In order to make the analyses manageable, the study largely restricted its focus to first- and second-year students. Importantly, as is common in studies of predictive validity, the analyses did not include uniTEST candidates who were not successful in gaining entry to a university course. The study does not distinguish between first- and second-year results obtained using different forms of assessment of which, of course, there was a large and diverse number.

**An overview of uniTEST**

As the above remarks emphasise, much of this report focuses on uniTEST. It is useful to provide a few introductory remarks about this assessment.

uniTEST has been developed jointly by ACER and Cambridge Assessment of the UK. Its general aim is to assist universities with student selection. The purpose is to enhance the effectiveness of admissions processes as they attempt to select students with the ability to undertake tertiary education, despite discouraging or ambiguous achievement scores at the end of high school. uniTEST is administered by individual universities in association with ACER.

This report is an evaluation of the SATTA pilot program, and therefore of uniTEST. In reviewing the overall program, the report provides general analysis and recommendations about aptitude testing in Australian higher education.

Work on uniTEST in Australia began in 2007 after pilot work in 2006. Pilot administrations were run in 2007, 2008 and 2009. In total, just under 1,500 people have sat uniTEST, with around 400 of these gaining admission. Many of these people would not otherwise have been admitted to university based if their achievement at the end of high school had been the sole criterion. The expansion of opportunity that arises through diversification of admissions data is an important consideration explored in this report.

uniTEST focuses on academic skills deemed important in higher education and emphasises a student’s ability to reason in both familiar and unfamiliar areas of learning. There are three parts to uniTEST, each of which consists of 30 multiple choice questions:

- **Quantitative Reasoning** deals with interpreting mathematical and scientific information and problem solving;
- **Critical Reasoning** deals with making decisions on the basis of information provided and with the ability to analyse argument in a logical fashion – topics relevant to scientific, technical, business and humanities type subjects; and
- **Verbal and Plausible Reasoning**, which deals with interpretation of passages in a socio-cultural context, and is based on the type of reasoning typical in the arts, humanities and social sciences.

The content and construct validity of uniTEST is assured by ACER’s test production process. A large group of prospective items is developed by teams of experienced item writers at ACER. These items are then scrutinised by item editors who consider whether they elicit the kind of cognitive responses that are deemed to be valid measures of candidate skills. The items are also checked to ensure that ambiguities are not present and that the language is clear and fair from the perspective of gender and socioeconomic status. These latter considerations are checked statistically after the test items have been piloted using samples of students who are similar to the likely test candidates. The final content of the tests is decided having in mind the need for a range of topics, a highly reliable set of sub-scores and, again, the need to ensure fairness to candidates of various backgrounds.

It is imperative that aptitude tests are considered valid if they are to play a productive role in admissions processes. While the content and construct validation of uniTEST is incorporated into the development
process, it remains important to determine how the assessment operates in context, and what value it can add to university admissions. These latter considerations are the main focus of the SATTA evaluation and this report.

**Overall research approach**

This formative evaluation was conducted in two stages, with the first running between late 2007 and late 2008, and the second running from late 2008 to early 2010. Figure 1 shows the first stage of the evaluation. For this stage the overall project workflow was divided into four broad phases and each phase lasted around four to five months. While many aspects of the data collection and analysis were difficult and complex, the first stage remained on schedule and delivered a number of insights and findings on aptitude testing in Australian higher education.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Month (2007-08)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background planning and review</td>
<td>D J F M A M J J A S O</td>
</tr>
<tr>
<td>Data specification and collection</td>
<td></td>
</tr>
<tr>
<td>Psychometric and statistical analysis</td>
<td></td>
</tr>
<tr>
<td>Documentation and reporting</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1: Project schedule, stage 1**

In Figure 2 the stages and timeline of the second phase of the project is shown. Data collection for this phase incorporates collection of results from three semesters of study, as well as the introduction of an additional cohort of uniTEST entrants to those involved in the first stage of the project. Analysis and reporting of results was undertaken in stage two following the collection of the majority of the results data from universities.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Month (2009-10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data specification and collection</td>
<td>J F M A M J J A S O N D J F</td>
</tr>
<tr>
<td>Statistical and policy analysis</td>
<td></td>
</tr>
<tr>
<td>Documentation and reporting</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2: Project schedule, stage 2**

The evaluation commenced with a review of key issues and considerations relevant to the study. The background review examined aspects of uniTEST and its use, relevant aspects of university education and admissions, and contemporary methodological approaches for establishing the predictive and face validity of selection instruments. During this stage, contact regarding the evaluation was made with all Australian universities and Tertiary Admissions Centres (TACs). ACER provided further detail about the focus and approach of the evaluation, and developed institution-specific plans pertaining to the nature and provision of data.

The design of a detailed project methodology, including a data specification and analysis design, facilitated conversations with TACs and institutions on the project overall and data requirements in particular. As anticipated in the initial project scope, for a number of reasons these conversations were long and complex. It was necessary to secure approvals from agencies and institutions, and to arrange for appropriate work to be scheduled. A considerable amount of time was spent working with data providers to ensure a good understanding of requirements. Due to various technical and administrative complexities, an iterative process was required for such data collection, involving receipt, validation, re-specification and further provision. Once useable data was secured, a raw data file was built and refined into an analysis file. The analysis file was then validated against the source data and cross-checked to ensure its veracity and relevance to the study. Data collection and file preparation were completed in early 2010 as results became available from participating institutions.

Statistical analyses were conducted by ACER. The analytical approaches used for each component of the overall evaluation are detailed in relevant areas of this report.
The production of this report commenced in late July 2008. An interim report was delivered to DEEWR on 31 October 2008, a draft final report in December 2009, and a final report in early 2010. The report contains key results from a large number of analyses. These provide a basis for preparing a range of derivative reports for more specific audiences and purposes.

**An overview of this report**

This report continues in four chapters. Chapter two provides background on international, national and institutional contexts that shape the nature and use of aptitude assessments in university admissions. Chapter three examines how institutions and individuals engaged in the SATTA evaluation, and looks in detail of the characteristics of students admitted using uniTEST. Chapter four examines the criterion validity of uniTEST, looking in detail at its concurrent and predictive validity. Chapter five explores future contexts surrounding implementation of a national aptitude test for school leavers in Australia, and make suggestions for future development.
2 FORMATIVE CONTEXTS

Introduction
This chapter builds a picture of contexts that shape the use of aptitude tests in university admissions procedures. It is hierarchical in its approach, first looking at how aptitude tests are used globally in different systems of higher education. The international context is important given the growth of transnational student flows and the highly internationalised nature of higher education. Following this review of international practice, a brief historical overview is offered to help contextualise aptitude testing in the Australian context. Admissions are ultimately each institution’s responsibility; therefore the third part of this chapter reviews insights from a survey Australian universities about their admissions practices.

Worldwide use of university admissions tests

Introduction to the analysis
This analysis explores the use of admissions tests in the selection of students to university in a range of countries. Admissions tests are widely used worldwide and are employed in a variety of ways, including being the sole criteria for selection and being used in conjunction with other achievement measures (usually high school grades). There are system-wide general tests, institutional tests specific to one or a small group of universities, discipline-specific tests designed for entry into particular courses, and tests which help select students from non-traditional entry pathways (particularly mature age applicants). The first section of this international scoping review explores the use of generic tests used for entry into various courses. The second section examines the use of discipline-specific tests. This section is followed by a brief discussion of validity studies undertaken relating to some of these international tests. From the evidence examined in the literature, a multi-country exploration of university admissions tests such as this does not appear to have been previously undertaken.

Overall, the research presented here suggests that university admissions tests which cover a range of disciplines and fields are used quite widely across the world as either the sole selection method, or as an important component of admissions. In this sense, Australia appears to be somewhat behind other nations. In relation to discipline-specific admissions tests, practice in Australia in the area of health sciences in particular, is much more in-line with international processes.

Given the substantial scope that an analysis such as this could have, the discussion here focuses on a few key areas of university admissions process. In the main, the discussion is limited to entry to undergraduate level courses for domestic students. The primary focus has been on OECD member countries, although in some cases this focus has been broadened. In addition to these parameters, it is important to highlight that in exploring ‘admissions tests’ the discussion canvasses aptitude and content-specific tests designed for university entrance selection, but does not cover high school examinations, tests or certificate qualifications that may help to qualify candidates for university.

Exploring these selection tools worldwide is a difficult process. The nuances of university entrance criteria in each country are not always transparent and processes are often unclear to ‘outsiders’ and even to experts within a system. The authors are indebted to numerous international colleagues for assisting with this analysis. Even despite their formative input, in many cases specific details of a system were not entirely clear. Sometimes admissions processes are not centralised and used differently within systems. In cases where sufficient information was not available, no specific mention of the country or test that may be used is made here. Therefore, while the countries examined here are varied and numerous, this discussion does not necessarily cover every admissions process that exists within the scope of the analysis – a potentially enormous task. The country-specific discussions in this chapter vary in detail in relation to the amount of information available.

Overall, this exploration reveals that admissions tests are widely used throughout the world for selecting students into university education. There is a substantial range in the extent to which such tests are utilised and the types of tests that are used. Tests that examine discipline-specific content are common, particularly in the health sciences fields, but many countries have a broad test that is used across all areas of study and sat
by all potential candidates. In some cases the admissions test is the sole determinant for entrance to university, but in most a combination of test scores and school results are involved in the selection process.

It appears that the main driver for the inclusion of such tests in selection processes is the belief that they can offer a better (or at least supplementary) means for selecting the most appropriate students into university. In certain countries tests are administered because of a lack of consistency in senior high school assessment, while others have used tests in conjunction with other measures of achievement to strengthen selection methodology. However, identifying the specific policy drivers that sparked the implementation of these tests is difficult and undertaking a thorough analysis in this regard is a research project in itself. Where there is clear and readily available information and debate about the implementation of such tests for individual countries, this is included. This information tends to limited except in cases where the testing has only recently been implemented, for example in South Africa.

In the discussion of admissions tests in this chapter an attempt has been made to differentiate between those which are aptitude tests and those which are achievement tests. To further elaborate the earlier distinction, an aptitude test is a test designed to identify potential future achievement, given the opportunity to learn. An achievement test is a test measuring knowledge of concepts that a person has already learned. In general it is argued that achievement tests are more strongly influenced by environmental factors (such as the prevalence of books available to a child while growing up or the quality of tuition while in high school) and aptitude tests are designed to examine potential, net of environmental influences (Willerman, Horn & Loehlin, 1977). While these two separate definitions might seem clear, it can be difficult to differentiate between the two types of tests because both have the potential to predict future achievement (Gage & Berliner, 1998). In this chapter, which is intended as a scan of the admissions testing worldwide, it has been difficult in some instances to identify whether the admissions tests discussed are designed as aptitude or achievement tests. Where such a distinction is apparent, the type of test is noted in the discussion.

This review of international practice is split into two sections. The main discussion is contained in the first section which comprises an outline of the broad, generic (non discipline-specific) tests used across the world and exploration of the way in which they are used in choosing new university candidates. This section categorises three main uses of general admissions tests: system-wide tests used as the sole determinant of entry to university; system-wide tests used as a key measure, but supplemented with other achievement measures; and other uses of general admissions tests (i.e. tests that are not system-wide in usage and tests designed for mature age entry). The second section contains a brief overview of the use of admissions tests to determine entry into particular fields of education.

**General university admissions tests**

Admissions tests that are general or non discipline-specific in nature are used extensively by universities throughout the world. In many cases, the university admissions test is the key measure for which applicants are selected. Often this measure is supplemented by high school leaving achievement scores, or completion of prerequisite subjects in the senior levels of high school. In many countries, entry to university is heavily reliant on performance in a system-wide admissions test. Discussion of these tests is the focus of the next two sections. This is followed by an overview of general tests that are not necessarily system-wide or are used for smaller groups of university applicants. In some cases these tests are designed to examine student knowledge of the national high school curriculum (achievement tests) and in others the test is designed around items that identify potential to learn, but do not specifically test for content learned previously by the student (aptitude tests).

**Admissions tests that are the sole determinant of entry**

It appears relatively uncommon for an admissions test to be used as the only measure for university admission. In some cases, for example Portugal, the national admissions test is the only measure used to rank students for university selection, but certain prerequisite subjects at high school must also be completed for the student to be gain a place. This section includes scenarios such as the Portuguese case as well as others where a particular admissions test that serves as the sole criteria for university admission. The countries highlighted here are China, Portugal, South Korea and Greece.
The Chinese university entrance system, known as ‘gaokao’ is based on three key admissions tests, undertaken by prospective students following completion of their secondary schooling (Davey, De Lian & Higgins, 2007). The administration of the Chinese test is co-ordinated by the Ministry of Education, which also oversees the construction of the test. The test has run in a similar format within China since 1952, apart from a ten year period during the Cultural Revolution (Unger, 1980). Completion of this test is required for entry to all universities in the country and success determines not only whether a university place is gained, but also the type of university that a student is accepted into; with the prestigious institutions generally taking the highest scoring applicants (Davey et al., 2007).

There are numerous parts to the Chinese admissions test, which students sit over a two or three day period. These tests are achievement tests, designed to cover specific knowledge and theory across a range of disciplines learned by student during their schooling. However, there is a lack of practical or more aptitude based problem solving questions in the tests. This particular fact has been criticised as a limitation of the current testing process (Davey et al., 2007; Zhang, 1995).

Competition for places in China is very strong and many potential candidates do not succeed in gaining an offer from a university. Unsuccessful candidates have to wait another year before they can re-sit the test. As such, there is much emphasis placed on studying for the test from the early years of schooling (Davey et al., 2007; Zhao, 2007). The high stakes of this test also appear to open it to controversy surrounding bias towards the cultural and political elite, corruption, and cheating (Chunlin, 2005; Davey et al., 2007).

The admissions process for Portuguese universities is also based primarily on an entrance exam. Entry into publicly funded universities in Portugal is gained by sitting the Concorso Nacional, while for private institutions candidates sit the Concorso Local (European Education Directory, 2009a). These admissions tests are sat by all people under the age of 23 who wish to study at undergraduate level. The tests are designed to identify knowledge in particular subject areas, this suggests that they are purely achievement-related tests. However, the tests do not specifically cover content included in the senior school curriculum. In this sense, these tests could be considered as aptitude rather than achievement tests, although in this case it is difficult to determine. Candidates who sit the test are admitted to their selected courses on the basis of their achievement in the test and completion of specific prerequisite subjects in their final years of schooling.

In recent years, the outcomes of the national admissions tests in Portugal have assumed a greater level of importance. As of 2005, the minimum score for which a candidate can be admitted to university has been set at 95 (out of a possible 200). This change has meant that many candidates now miss out on available places, making the stakes of these tests higher than in previous years. From the perspective of policy makers and institutions, this change has been implemented to ensure that the country’s higher education standards are kept high in the policy setting of Bologna process.

As with the ‘gaokao’ in China, the South Korean national university admissions test, the College Scholastic Ability Test (CSAT) or ‘suneung’ is the sole determinant of whether students are admitted to university. The CSAT is developed and implemented by the Korea Institute for Curriculum and Evaluation (KICE) on behalf of the government (KICE, 2008). A huge amount of importance is placed on successful completion of this test, to the extent that students are coached for it years before they actually sit the exam. This test is an achievement (rather than aptitude test) and is based on the national school curriculum. All universities focus their admissions on CSAT results, with the most prestigious institutions taking those students who perform the strongest.

Entry from high school into Greek universities is also determined entirely by a national higher education entrance examination (Psacharopoulos & Tassoulas, 2004). Students in the final year of secondary schooling who wish to gain entry to university must sit this multi-disciplinary exam, which is administered centrally by the national government. Universities select students based on their exam result and the preferences for courses that they specify during the application process. As with many other countries which place a high value on the outcomes of one examination for determining entrance, this test has become part of the national psyche, with students and families investing substantial time and resources in studying and being tutored in the hope of increasing success in the national entry examination (Psacharopoulos & Tassoulas, 2004).
Admissions tests which are a key (but not sole) criteria for entry

There are a number of systems across the world where the admissions test is not the only criteria used to determine entry to university, but it still plays a dominant role. The systems discussed in this section have a general admissions test which most applicants for undergraduate courses sit as part of the admissions process. Turkey, Sweden, Japan, the USA and South Africa are included in the discussion here.

In Turkey, entry to university is based on the results of students on the ÖSS, a student selection examination that is based on two tests; a verbal test and a quantitative test. All those who wish to attend university must sit the ÖSS. The tests span a variety of disciplines including science, mathematics, Turkish and foreign languages, and social sciences (Karakaya & Tavsancil, 2008). The content and use of the ÖSS suggests that this test comprises both achievement and aptitude-related items. The ÖSS outcomes are combined with grade point averages from the Turkish school leaving certificate known as Lise Diplomasi (European Education Directory, 2009b). Applicants for university are selected based on these outcomes. Interestingly in Turkey, the criteria for selection and the actual selection itself is undertaken centrally by Yükseköğretim Kurulu (YÖK), a constitutional body for coordination, supervision and observation of the major activities of higher education institutions (Eurydice, 2008).

The Japanese university admissions process follows a similar process to that of Turkey and is not dissimilar to China, South Korea and Greece in the importance placed on the admissions test administered. For entry into the public universities and many private universities, applicants sit a test administered by the National Centre for University Entrance Examinations (NCUEE). Institutions individually decide which specific parts of the test devised by the National Centre they will include each year and candidates undertake the sections relevant to their courses of preference. The admissions process followed is different for each institution, but in many cases the test score is combined with applicant interviews or recommendation letters from teachers. However, Teichler (1997) notes that in general, the more prestigious the institution, the more likely it is to rely solely on the examination score for selecting candidates.

The Japanese National Centre Test for University Admissions has been administered since 1990, when it replaced a similar style test which had run since 1979 (NCUEE, 2009). Such tests were also run in Japan prior to 1979 in slightly different formats, but essentially this admissions system is well entrenched in the country. The current test is primarily designed to measure achievement in a range of disciplines for high school graduates. However, in its discussion of the test, NCUEE notes that it is also used by universities to judge aptitude of candidates (NCUEE, 2009).

Given the hierarchical nature of the Japanese education system, entry into the most prestigious higher education institutions in many cases provides a stepping-stone into successful careers and substantial cultural leverage. As such, the entrance test in Japan is considered very important. Most people enrol in special tutorials to prepare for the test and many of those who fail to gain entry on leaving school spend the subsequent year having further tuition (these candidates are known as ‘ronin’) in order to increase their chances of success for the next time they sit the test (Mori, 2002; Ono, 2007; Teichler, 1997).

It is not surprising that academic aptitude tests have been used prolifically in the USA. It must be recalled that in the USA there are some 12,000 local education authorities which is a consolidation from more than 30,000 just two generations ago. Thus, educational achievement is made heterogeneous by the variety of standards and curricula. Also, despite the attempts by state governments to help equalise educational inputs and budgets within their state, wide disparities with respect to educational resources exist among local education authorities. Note too that states themselves vary considerably with the amount of equity-based resources that they can provide.

Therefore, when tertiary institutions (colleges) in the USA make admissions decisions, there are no common metric achievement tests available for use. Even if there were state-wide achievement tests, they would be of limited use because many students travel interstate for their college years. For these reasons the USA adopted a strong emphasis on academic aptitude tests for college entry purposes and later for graduate school and professional school admissions.
Individual institutions in the USA again have the ability to make their own decisions about the student selection process. However, admissions tests are almost universally used as a key component of selection. While there are numerous general tests used throughout the USA, there are two specific tests that dominate the sector: the SAT and the ACT. These tests primarily assess student ability by testing their knowledge relating to subject areas and content included in the national school curriculum. In some jurisdictions in the USA, different admissions tests are used widely throughout the region, such as the Texas Higher Education Assessment (THEA).

The SAT, developed by the Educational Testing Service and administered by the College Board is a mixture of an achievement and an aptitude test. In terms of achievement, it measures knowledge based on the USA high school curriculum. Emphasising the aptitude facet of the test, the College Board (2009) states that the SAT assesses critical thinking and problems solving skills which are ‘attained in and outside the classroom’. The SAT can also be supplemented with SAT Subject Tests, which are specific to a number of disciplines and are purely achievement tests. The SAT Subject Tests are utilised by a number of institutions wanting additional information about student abilities in particular fields.

The ACT, also accepted widely by USA universities, has been running since 1959. It is strictly an achievement test, with all questions based specifically on the national high school curriculum (ACT, 2009). Substantial academic research and commentary about the SAT and ACT are further explored in the following section relating to validity studies and can also be found elsewhere (see for example: Clark, Rothstein & Schanzenbach, 2009; Geiser, 2009; Grove, Wasserman & Grodner, 2006; Simpson & Kadhi, 2009; Sternberg, 2006; Stringer, 2008; Thomas, 2004).

The Swedish tertiary entrance system is different to all those previously mentioned. However, it still has an admissions test which forms one key measure of entry to university. The Swedish Scholastic Aptitude Test (SweSAT) is an important tool used in the Swedish admissions process. Prior to 1991, the SweSAT was used only for mature aged applicants for university, but since this time, any applicant has been able to sit the test in order to have their results count towards their chance of admission to university (Berggren, 2006, 2007).

The SweSAT is not a mandatory test for all school leavers. However all universities accept students on the basis of SweSAT scores. Students gain access to university either through their high school Grade Point Average (GPA) or by their SweSAT score.

The Swedish National Agency for Higher Education, Högskoleverket, oversees the admissions processes within the country. It stipulates that of all new students admitted to each institution in a year, at least one third must be accepted on the basis of their SweSAT score, at least one third on their secondary school GPA and no more than one third on other forms of admission, including prior learning and experience, proficiency in specific areas and interviews (Högskoleverket, 2009). In a recent analysis of the Swedish admissions process, Berggren (2007) found that 43 per cent of the whole age cohort born in 1974 had sat the SweSAT.

The main reason behind the segmentation of admissions to university in Sweden (i.e. the stipulation that at least one third of new student must be admitted on the basis of their SweSAT score etc.) is to encourage a diversified university student body. By focussing on two distinct methods of selection – achievement in high school and an aptitude-focused test – it is hoped that selection of students from under-represented groups is more likely to occur. However, research by Berggren (2007) questions the extent to which broadening of participation occurs as a result of the current selection policies in Sweden.

In South Africa a new admissions system was implemented in 2009. A new national university admissions test known as the National Benchmark Test (NBT) has been rolled out with the aim of providing an improved selection tool for universities. Previously in South Africa the Standardised Assessment Test for Access and Placement (SATAP) was widely used by institutions as one measure for selecting undergraduate candidates (Scholtz & Allen-Ile, 2007). The new NBT is be used to supplement high school marks in determining which students are offered a place in the highly competitive university sector. The NBT is an achievement test, with items based on the National Senior Certificate (NSC) curriculum in the country.
At the time of writing, the pilot results from the NBT were being discussed within the South African Education Ministry (Parliamentary Monitoring Group, 2009). The early indication is that the first sitting of the NBT has attracted controversy in the country, mainly due to the fact that a large proportion of those who sat the test were identified as requiring additional learning support if they were expected to succeed at university (University of Cape Town, 2009). These issues with the test for selection of high school graduates in South Africa are made even more complex because in addition to the new selection test, the National Senior Certificate in high schools is also brand new. As a result, it has been difficult to identify whether the problems in the system lie with the National Benchmark Test or with the new NSC curriculum (University of Cape Town, 2009). While the South African system provides a recent example of the adoption of a new national selection regime, the parallel introduction of high school curricula and its setting in the developing (rather than developed) world make comparisons with the Australian system somewhat tenuous.

**Other systems using general admissions tests**

In many countries, the use of university admissions testing to supplement the selection process is used more sporadically than in those mentioned above – but such measures are widely used nonetheless. In some cases, such as Mexico, the UK and Australia, a sample of universities utilise achievement and aptitude tests for undergraduate entry. In other countries, admissions tests are used for particular cohorts of applicants – most commonly mature age students. Examples in relation to these two groups are explored in this section.

A number of universities in Australia and the UK utilise uniTEST as a supplementary tool for admissions. The use of this aptitude test is a substantial focus of other parts of this report and is therefore not explored further here. Overall four universities in Australia and seven in the UK use uniTEST in their selection process.

In addition to the use of uniTEST in the UK, some universities in the UK have other tests used for admission, for example Oxford and Cambridge universities both use specific versions of the Thinking Skills Assessment (TSA) to inform their admissions processes. However, in general there are no system-wide admissions tests that are implemented in the UK in the manner adopted by countries such as the USA, Japan and others mentioned in the section above.

In Mexico, there has been a trend towards the use of admissions tests to assist in the selection processes for universities, especially since the establishment of a national centre for the assessment of higher education (Centre Nacional de Evaluación de la Educación Superior) in the mid-1990s (Backhoff, Larrazolo & Rosas, 2000). However, currently there is no nationally coordinated selection test process in Mexico. Instead, individual institutions and groups of institutions in Mexico have been developing admissions tests that fulfil their requirements for entry. The EXHCOBA was developed at and is used by Universidad Nacional Autónoma de México Autonoma de Baja California and some other institutions. In addition a range of other large universities such as the Universidad Nacional Autonoma de Mexico and the Universidad Autonoma Metropolitana have developed their own specific admissions tests in recent times (Backhoff et al., 2000).

The Canadian higher education system currently operates relatively similarly to the Australian system. Admissions are generally coordinated by province-based university admissions centres (for example applications for universities in Ontario are facilitated through the Ontario Universities Application Centre) and the requirements for entry to courses are stipulated by individual institutions. Some use of testing is undertaken in specific subject disciplines (discussed in the section below) and there is ad hoc use of more general admissions tests, but domestic applicants are primarily selected on the basis of their achievement in the final years of secondary school.

In addition to these examples, many countries use admissions tests to aid the selection of mature-age entrants to university. In general these tests are administered in the absence of any high school mark, after applicants reach a certain age, or after a defined period out of schooling. As noted in this report, universities in both Australia and New Zealand use the STAT in this way. Other countries with tests used for similar purposes include Canada (Canadian Adult Achievement Test or CAAT), Austria (Berufsreifeprüfung), the Netherlands (Colloquium doctum), Portugal (Exame Extraordinário de Avaliação de Capacidade para
Acesso ao Ensino Superior), Switzerland and the USA (Graduate Record Examinations (GRE) is one of a number of such tests in the USA).

**Discipline-specific tests**

Australia appears to be much more in-line with global practice when it comes to the common acceptance and use of admissions tests and multiple selection mechanisms for admission into specific disciplinary fields. In particular, the use of aptitude testing in the health sciences, especially for entry to medical degrees, is universal across Australian universities is held in high esteem worldwide (Mullen, 2009).

There are tests for various disciplines used by many universities throughout the world. Table 1 provides a sample of the kinds of subject areas and countries in which discipline-specific tests are used. Given the complexity of higher education systems and university admissions processes across the world, the detail in Table 1 is likely to provide only the tip of the iceberg in terms of the kinds of tests being administered for entry into specific courses or for particular subject matter. The tests featured here are generally those that are well established and widely known. Rather than providing specific commentary about each of these tests, this table is designed to provide a broad overview and an indication of the extent to which the tests are utilised within these countries. Some of these tests are given more attention in the part which follows, relating to validity studies carried out on international tests.

As can be seen from the sample in Table 1, the health science fields, which include medicine and dentistry, are commonly linked with some kind of aptitude testing, while in most other fields there is less uniformity in terms of worldwide usage.

In addition to these tests, which have been devised with a specific subject-matter in mind, other tests that were discussed in the earlier section relating to broad or generic type admissions tests also include optional ‘add-ons’ that test specific disciplines. As discussed earlier, the SAT Subject Tests in the USA provide a notable example of this.
Table 1: Sample of discipline-specific university admissions tests used worldwide

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Country</th>
<th>Test name</th>
<th>Acronym</th>
<th>Use within Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical and health sciences</td>
<td>Australia</td>
<td>Undergraduate Medical Admissions test</td>
<td>UMAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>Belgium</td>
<td>Toelatingsexamen</td>
<td></td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>Dental Admissions Test</td>
<td>DAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>Ireland</td>
<td>Health Professions Admission Test - Ireland</td>
<td>HPAT-Ireland</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>Korea</td>
<td>Medical Education Eligibility Test</td>
<td>MEET</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>New Zealand</td>
<td>Undergraduate Medical Admissions test</td>
<td>UMAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>Nth Ireland</td>
<td>Health Professions Admission Test - Ulster</td>
<td>HPAT-Ulster</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>Medical School Admissions Test</td>
<td>MSAT</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>BioMedical Admissions Test</td>
<td>BMAT</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Dental Admissions Test</td>
<td>DAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Pharmacy College Assessment Test</td>
<td>PCAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Medical College Assessment Test</td>
<td>MCAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Optometry Assessment Test</td>
<td>OAT</td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Veterinary College Assessment Test</td>
<td>VCAT</td>
<td>Universal</td>
</tr>
<tr>
<td>Law</td>
<td>Australia</td>
<td>Australian Law Schools Entrance Test</td>
<td>ALSET</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>Korea</td>
<td></td>
<td></td>
<td>Universal</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>National Admissions Test for Law</td>
<td>LNAT</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>Cambridge Law Test</td>
<td></td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Law School Admissions Test</td>
<td>LSAT</td>
<td>Selected institutions</td>
</tr>
<tr>
<td>Education</td>
<td>Finland</td>
<td></td>
<td></td>
<td>Selected institutions</td>
</tr>
<tr>
<td>Engineering</td>
<td>Australia</td>
<td>ATN Engineering Selection Test</td>
<td>ATNEST</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
<td>ATN Engineering Selection Test</td>
<td>ATNEST</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
<td>Aptitude for Engineering Assessment</td>
<td>AEA</td>
<td>Selected institutions</td>
</tr>
<tr>
<td></td>
<td>Belgium</td>
<td></td>
<td></td>
<td>Selected institutions</td>
</tr>
<tr>
<td>Psychology</td>
<td>Finland</td>
<td></td>
<td></td>
<td>Universal</td>
</tr>
<tr>
<td>History</td>
<td>UK</td>
<td>History Aptitude Test</td>
<td>HAT</td>
<td>Selected institutions</td>
</tr>
<tr>
<td>Mathematics</td>
<td>UK</td>
<td>Sixth Term Examination Paper</td>
<td>STEP</td>
<td>Selected institutions</td>
</tr>
</tbody>
</table>

Criterion validiy analyses of international aptitude tests

While the focus of this current report is on the criterion validity of uniTEST, the short section below briefly explores some key research studies into the criterion validity of some of the larger tests mentioned above. Perhaps unsurprisingly, the most research in this regard have been in relation to tests administered in the USA. Therefore, tests from the USA are the primary subjects of this discussion.

The predictive validity of the SAT, for instance, has been extensively researched (see, for instance: Armstrong & Carty, 2003; Bridgeman, McCamley-Jenkins & Ervin, 2000; Geiser & Studley, 2001; Morgan, 1990). Bridgeman et al (2000) point out that the SAT predicts First-Year College GPA (FGPA) equally well across different ethnic groups. They also show that overall, men tended to get slightly lower grades than predicted and women get slightly higher grades than predicted. When SAT scores were correlated with FGPA the result, corrected for attenuation (restriction of scale), was 0.52. When SAT scores and high school GPA were combined and then correlated with FGPA the result rose to 0.61. At the high end of the SAT score range the combined correlation after correction was 0.69. Note that there was no correction for the unreliability of FGPA when calculating these correlations but an estimate of a revised correlation of 0.74 was suggested.

As well as the pervasive SATs for college entry, there is a plethora in the USA of aptitude type tests for more specific selection purposes. Major ones include the Graduate Record Examination (GRE) for admission into graduate school, the Law School Admissions Test (LSAT) and the Medical College Admissions Test (MCAT).

The GRE is developed by ETS as is the SAT. The GRE is not curriculum specific and assesses abstract thinking in quantitative and in verbal areas. There are also two writing tasks over 75 minutes. Since November, 2007 the types of items have begun to be revised in an evolutionary way. The multiple choice items are ‘computer adaptive’ meaning, in part, that the first item is of intermediate difficulty and success on that item leads the computer to administer a harder following item whereas failure leads to the administration
of easier items. The computer continues to adapt the test according to the applicant’s performance (ETS, 2008). The development and scoring of the items in the item bank are based on item response theory (IRT).

Most US graduate schools require the GRE as part of the admissions process to enable important practical improvements in admissions decision-making (Kuncel, Hezlett & Ones, 2001). This claim is backed by various studies that show that GRE plus undergraduate GPA predict first year graduate GPA more effectively than any other piece of information. Multiple correlations after correction for attenuation range from 0.55 to 0.60. (see, for instance: Powers, 2001; Schneider & Briel, 1990).

The LSAT, administered by the Law School Admission Council, provides law schools in North America with “a standard measure of acquired reading and verbal reasoning skills that law schools can use as one of several factors in assessing applicants” (Law School Admission Council, 2008). The LSAT is administered four times a year and, in the 2007-2008 cycle, more than 142,000 prospective students sat the test. The LSAT consists of just under three hours of multiple choice testing and a 35 minute writing task. The multiple choice section is divided into three parts, Logical Reasoning, Reading Comprehension and Analytic Reasoning. The LSAT results are used in the admissions process by almost all US law schools in addition to other information such as undergraduate GPA, letters of reference and applications statements. The two numerical scores (GPA and LSAT) are weighted idiosyncratically by each law school to obtain an admissions index.

A considerable amount of research (Law School Admission Council, 2008) has been conducted on the predictive validity of the LSAT alone and on the combination of the LSAT and the GPA (see, for example: Linn & Hastings, 1984; Powers, 1982; Stilwell, Dalessandro & Reese, 2005). The GPA is rarely used alone for admissions purposes because, as noted earlier in this chapter, there is a vast range of standards across states and across high schools within states.

The LSAT has a median predictive validity coefficient of about 0.4 when correlated with first year law school performance a year later (Stilwell, Dalessandro & Reese, 2005). It is slightly higher, about 0.5, when a multiple correlation is calculated using college GPA as the additional predictor (Stilwell, Dalessandro & Reese, 2005). It is also clear that the correlation increases when correction for attenuation and for the reliability of the predicted GPA is calculated. This is in line with the predictive results obtained for the GRE (see above) and for the Medical College Aptitude Test (MCAT).

The MCAT, previously known as the Professional School Aptitude Test, is designed to assess students wishing to enter medical studies in North America. It tests their problem solving, critical thinking, analysis and writing skills. There is also a section on scientific concepts and principles. Since January 2007, the MCAT has been administered by computer, and scores are provided on the use of cognitive skills in Physical Science, Biological Science and Verbal Reasoning. A score is also provided for writing skills which are assessed by two tasks over a period of about an hour. Overall, the MCAT test time is 4 hours and 20 minutes (AAMC, 2008).

A large number of studies of the predictive validity of MCAT have been published (see, for example: Brooks et al, 1981; Carlile et al, 1983; Friedman & Bakewell, 1980; Golmon & Berry, 1981; Jones & Thomas-Forges, 1984; McGuire, 1980; Veloski et al, 2000). Coates (2007a), after citing 24 research reports on the topic, suggests that the correlations between MCAT and subsequent academic results average about 0.3. Donnon, Paolucci and Violato (2007) conducted a meta-analysis of more than 20 studies and found small to medium validity indices for both medical school performance and Medical Board licensing measures. They cite specific results ranging from 0.39 to 0.60. Results published by the Association of American Medical Colleges (AAMC, 2002) gave correlations ranging from 0.28 to 0.81. More recently, Julian (2005) reported corrected correlations between MCAT and GPA in Medical School of about 0.55. The AAMC also reported that MCAT scores when used in combination with undergraduate GPA accounted for an additional 17 per cent of the variance in medical school performance in contrast with using undergraduate GPA alone.

As well as these major aptitude tests used mainly in North America there are a number of other aptitude tests used in the UK and in Australia. Such a test is the Graduate Australian Medical Schools Admission Test (GAMSAT). The GAMSAT is available to graduates of any discipline. It provides a standard national
objective assessment that adds value to other data when making admissions decisions. The test consists of three sections and takes a lengthy five and a half hours.

Donnelly (2006) reviewed the use of the GAMSAT in the United Kingdom and found that it predicted success in first year medical studies. He reported uncorrected (unadjusted) correlations of about 0.3. An ACER research report (Coates, 2007a, 2008a) looked at the criterion validity, including predictive validity, of GAMSAT. This study took into account the restricted range of the cohort under consideration and the reliability of the criterion measure.

The resulting average predictive validity of GAMSAT across the institutions studied, when related to first year medical studies, was 0.33. The best predictions occurred when both undergraduate GPA and GAMSAT scores were used in a multiple regression formula to predict first year medical results. The combined predictors explained an average 43 per cent of first year variance. These results are consistent with the MCAT results presented above.

Summary of international use

Overall, this analysis of admissions testing has shown that the use of aptitude or achievement tests is the norm in many developed countries around the world. In many systems, where a specific admissions test is a nationally instituted and almost universally undertaken test (for example in Japan, the USA, China, Portugal, Greece, Turkey, South Korea and Sweden), the function of the test is well embedded in the education landscape of the country. In other countries such as Australia, the UK, Mexico and New Zealand (among others), admissions tests exist but are administered in an ad hoc fashion with little continuity across the sector.

In those countries where admissions tests are an important and accepted part of the selection process, there are benefits in that students have a clear understanding of what the tests are designed for, when they will be required to sit them, how universities use their scores and what the testing process involved. However, there are also unintended consequences stemming from these tests, particularly the significant amount of pressure put on test candidates for success and the burgeoning private market for coaching on the national test(s). Those countries that use a nationally accepted test, but also balance these outcomes with school results seem to have fewer concerns with these problems.

When it comes to admission for particular courses, the patterns are slightly different. In particular, for entry into subjects in the health science fields (especially medicine) there is almost worldwide acceptance that the admissions process should include a core component that is an admissions test. Tests such as UMAT, HPAT-Ireland and Ulster, MEET, GAMSAT, MSAT and MCAT are all well established medical entrance tests in use throughout the developed world. Acceptance of entry tests for other disciplines is less universal than in the health sciences field, but as the examples in this chapter show, there are a number of widely used tests in a range of fields including law, engineering and education.

It appears from this scan of selection methods from across the world that the challenge for Australia in terms of acceptance of a generic and system-wide admissions test is to ensure that the complexities of such tests in other countries – that is the intense national focus and emphasis that the test is the only opportunity for gaining entry to university – are avoided by ensuring a balanced approach to selection based on a number of measures with no particular emphasis on any one of these selection tools. These ideas are explored further in the final chapter of this report.

University admissions in Australia

Introduction

Reviewing international practice is important, particularly in a higher education system that is as internationalised as Australia’s. Locating the evaluation in an international scope helps Australia learn from other countries and ensures that domestic approaches accord with what may be considered common or even ‘best’ practice.
Ultimately, however, countries adopt admissions practices which are seen to be valid and efficient in terms of the salient features of their domestic context. Thus the following section focuses on the Australian situation. It was suggested in the introduction to this report that uniTEST may be able to enhance tertiary admissions in Australia, and hence the overall effectiveness of the system. Beginning with an historical perspective, the following analysis investigates admissions practices in Australia, the more general role that aptitude tests can play, and then outlines the particular relevance of uniTEST. It closes with an overview of the need to study and enhance the validity of such tests, and hence of the need for the current evaluation.

**Historical background**

Traditionally in Australia, student admission into tertiary education has been mediated through the results of each state’s public examination of achievement in curriculum-specific courses. This tradition served tertiary education well enough in the twentieth century when Australian society was less complex than today.

In 1950, for instance:

- About 90 per cent of Australians were descended from people from the British Isles with a large majority having been born and educated in an English-speaking country.
- Fewer than 10 per cent of the school cohort finished high school and only about half of these went on to tertiary studies, while by 1975 (the next generation), 33 per cent finished high school (ABS, 2008) and by 1988, the retention rate had risen to 58 per cent (ABS, 2008).
- The population was not mobile and migration among states was relatively rare.
- Universities enrolled all students who ‘matriculated’ and students could then enter the faculty of their choice with few impediments. For example, in NSW matriculation was a pass mark of at least a B level in five academic subjects using a 3-point scale of A, B or F. The University of Sydney’s medical faculty enrolled many hundreds of matriculated students in its first year course but most of them were failed at the end of the year.

Thus, competition for university entry was very different from today with most of the culling then occurring during the mid-high school years (ages 14 to 16) or through heavy failure rates after tertiary entrance.

Towards the end of the first decade of the 21st century, the situation has changed dramatically:

- Australia is an ethnically diverse society, with only about 45 per cent of Australians having an ethnic background involving the British Isles (ABS, 2006).
- About 75 per cent (ABS, 2008) of young Australians finish secondary education and about 50 per cent (ABS, 2007) of young Australians will go on to some form of tertiary education within a few years of high school graduation. Higher education in Australia is moving from a mass towards a universal system (Australian Government, 2009).
- The population is more mobile and movement intra state and interstate is no longer rare. Furthermore, students are more likely to move interstate if they see an educational opportunity than they were two generations ago.
- Universities now attempt to control numbers entering courses and faculties. High attrition rates among admitted students are seen as economically wasteful and educationally distasteful.
- Competition for places in tertiary studies after high school graduation is now especially strong among prestige universities and courses, but it is also clearly present across the system. This will increase with policy for greater diversification less central planning of student numbers.
- Around a quarter of all tertiary students (DEST, 2006) are from overseas.

These changes have led to a questioning – with increasing vigour – of the continuing strong emphasis on achievement results as almost the sole criterion for tertiary admission (for a recent critique, see James, Bexley & Shearer, 2009). It is increasingly unclear that reliance on achievement scores alone provides a transparent and efficient means of ensuring that all talented students who would like to attend university are able to gain admission. Specialist courses have added other criteria such as art portfolios, aptitude testing and interviews but in the general perspective, these are relatively rare exceptions to the time-honoured tradition of achievement testing.
These achievement testing regimes have less supporting rationale in today’s society where many applicants for tertiary entry have sat their high school certificate years earlier, where potential applicants come from families where little of the important supporting mechanisms are present in home backgrounds, where schools in underprivileged areas are less able to provide the necessary intensive help their students need and where improved tertiary admissions processes are feasible, but usually ignored.

**Conceptual rationales relevant to Australia**

As noted, aptitude tests assess an individual’s potential for acquiring new knowledge and skills in an academic setting (Pellegrino, 1994). Of course there are many other kinds of aptitude (for example, musical, language, sporting, clerical) but here we are focused on academic or school-related aptitude. There are several principled reasons why aptitude tests would appear to be beneficial for Australian higher education.

First, a person has suffered from poor teaching or from being a member of a classroom where academic performance is derided as being for ‘uncool’ nerds. This person’s achievement level might not be at a level it would have been under more supportive circumstances. We might under-predict future achievement if we only consider their school achievement scores. This person may show comparatively better performance when curriculum specific items are missing but general items showing ability to solve problems or reasoning are emphasised. This would suggest that this person could do well in tertiary studies if an enriched educational environment that is supportive of learning were to be provided. In short, a major factor arguing for the use of aptitude tests is that it is a means of providing diagnostic information to students on what they are capable of achieving.

Second, in many tertiary programs there is a mix of new subjects not previously studied in school. For example, engineering, law, medicine, architecture, dentistry and a vast range of technical and vocational subjects are new subjects and there is no previous direct achievement background on which to call. Aptitude test scores become important decision factors in such circumstances. They can enable better matching between individuals and courses.

Third, from within each Australian state and among the states there are subjects with varying standards and demands, this is also the case internationally (see, for example Braun & Dwenger, 2008). In mathematics, Victoria and NSW have 3 and 5 levels of mathematics respectively. Did the demands of mathematics methods in Victoria equate with the demands of three-unit mathematics in NSW? Victorian students can take a course in psychology for their VCE but in NSW there is no such course. On the other hand NSW has a course called society and culture which does not exist in Victoria. If it is unfair to expect a 15 year old student to make up their mind about vocational choice, then should a student at the end of Year 12 be penalised for not having taken the ‘right’ subjects that allow them admission into the university course of their choice?

Fourth, many countries now have a major increase in immigrant numbers. Australia as an example now has considerably more than half a million immigrants who have settled in the past five years. This is an increase of some 2.5 per cent in the population. As well, the number of overseas students studying in Australia has risen from rapidly (Australian Education International, 2008). It would be a travesty to bar such students on the basis of Australian states’ achievement tests which assess state curricula. But equally it would be impractical to differentiate among high school graduates from dozens of different countries (for example, Singapore, Sri Lanka, Saudi Arabia, Sweden, South Africa and Switzerland) on the basis of achievement data from their own countries. Aptitude tests that all can sit for can be a useful common metric for admissions purposes.

Fifth, in many tertiary programs, such as medical degrees, a large number of students with very similar results compete for a limited number of places in a course. Although students’ previous academic results provide admission staff with some information on their expertise and experience, discriminating between these students on the basis of previous academic results alone can be extremely difficult (see, for example: Aldous, 2004; McManus et al, 2005; Nicholson, 2005). For this reason many universities base their admissions on non-academic criteria as well.
Sixth, if administered at an appropriate time, students can benefit from diagnostic information provided by aptitude assessments. Independent information can complement evidence provided by formative or summative achievement-oriented assessments to raise people’s awareness about what they’re capable of achieving in the final years of secondary school, and then at university. They can inform and lengthen the decision making choices of students. This information might also be used to assist universities with the allocation of advance placements or the provision additional support for people from disadvantaged backgrounds.

Seventh, aptitude tests can be developed to measure capacities that are not assessed by content-focused examinations. Increasingly it is recognised that academic and vocational performance hinges on the application of underlying intellectual capabilities that transcend particular subjects, industries and contexts. People’s capacity to function in certain situations can be shaped by these capabilities as much as specific forms of knowledge and skill. Assessing these, therefore, provides data that complements, reinforces and extends the information provided by assessments of a student’s demonstration of curriculum-specific competencies.

Eighth, as many of the above observations suggest, it is now very common for multiple forms of evidence to be factored into admissions decisions. Medical admissions in Australia rely on a combination of evidence from interview, of achievement and from aptitude tests. Many courses make use of portfolios or recognition of prior professional experience. Today, the over-reliance on a single measure appears unusual and even risky, particularly by international standards. It is a dubious proposition to rely solely on achievement scores when considering admissions to tertiary studies. However this would appear to be the prevailing approach with school leavers in Australia. Diversification that facilitates triangulation would appear to offer a means of enhancing the validity of the process. Aptitude test scores add extra dimensions in the important role of predicting future performance (see, for example: Beaton & Barone, 1981), and ensuring that all able students can participate in university study.

Ninth, a major benefit of aptitude tests is that they can enhance the transparency of admissions processes. In an era where people move between countries and school jurisdictions, and where competition for many places is high, aptitude assessments supply objective data for university admissions. The Graduate Australian Medical School Admissions Test (GAMSAT) illustrates this function, for it is the only generalisable data factored into admissions decisions which also reference data from interviews and prior achievement. There is room for aptitude data to enhance the transparency of the admission of school leavers to university in Australia. Numerous direct and indirect benefits flow from having a sound common data point as discussed and explored throughout this report.

In summary, given the changes in the system over the past half century and the recent higher education policy direction of the Australian Government, there appear to be cogent rationales for the wider and more transparent use of aptitude tests in selecting university candidates in Australia. Specifically, the above analysis suggests that aptitude assessments can add value to university admissions:

- when students come from different socioeconomic backgrounds that tend to depress achievement scores (the equity rationale);
- when future academic subjects are not simple continuations of previous subjects studied;
- when students come from various schools and states that emphasise or teach different curricula;
- when considering recent immigrants or applicants from and overseas student applicants;
- when there is a need to discriminate more finely between various levels of performance;
- when there is a case for giving people further insights on what they are capable of achieving;
- when information on general capability can complement that provided by curriculum competence;
- when there are good arguments that achievement scores need to be supplemented or replaced by academic aptitude tests; and
- when the addition of common and objective data can enhance transparency.
Institution’s experiences with aptitude tests

Introduction
Ultimately, universities are responsible for designing and managing their admissions processes. It is vital, therefore, to explore the use of aptitude tests within Australian universities and how institutions perceive aptitude assessment. For current purposes, this was done through a national survey along with a series of focused discussions about institutional practice. A brief online survey instrument was developed that probed university use and perceptions of aptitude assessments. This instrument was deployed to Deputy Vice Chancellors (Academic) and senior administrators (such as Registrars and Vice Principals) at all Australian universities. Completed surveys were received from 59 respondents representing 31 higher education institutions in Australia, giving an institutional response rate of 79 per cent.

Tests used by Australian universities
Aptitude tests have been widely used by Australian universities for many years. Table 2 details a sample of the aptitude tests in use at Australian institutions. This indicative list is derived from survey responses and is by no means exhaustive, as many aptitude tests are known to be in use that were not listed by respondents. As noted in terms of the international review, this largely stems from the devolved way in which aptitude tests are managed – typically at the faculty or even department level. Several aptitude tests have been developed by universities for internal use with specific cohort or course contexts. There are not discussed here due to the general lack of information about these tests.

Table 2: Externally developed aptitude tests used for admission to Australian universities

<table>
<thead>
<tr>
<th>Name</th>
<th>Acronym</th>
<th>Origin</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate Australia Medical School Entrance Test</td>
<td>GAMSAT</td>
<td>AUS</td>
<td>Graduate entry applicants to medicine courses</td>
</tr>
<tr>
<td>Graduate Management Admissions Test</td>
<td>GMAT</td>
<td>US</td>
<td>For international applicants for admission to business courses</td>
</tr>
<tr>
<td>Indigenous Student Intake Test</td>
<td>ISIT</td>
<td>AUS</td>
<td>For Indigenous applicants without formal education qualifications</td>
</tr>
<tr>
<td>Internal Selection Program</td>
<td>ISP</td>
<td>AUS</td>
<td>Course (including aptitude testing) for Indigenous applicants without formal educational qualifications</td>
</tr>
<tr>
<td>International Student Admissions Test</td>
<td>ISAT</td>
<td>AUS</td>
<td>International applicants</td>
</tr>
<tr>
<td>Medical College Admissions Test</td>
<td>MCAT</td>
<td>US</td>
<td>For international applications for admission to medical courses</td>
</tr>
<tr>
<td>SAT</td>
<td>US</td>
<td>For international applications where no recognised formal educational qualifications exist</td>
<td></td>
</tr>
<tr>
<td>Special Tertiary Admissions Test</td>
<td>STAT</td>
<td>AUS</td>
<td>Mature age students</td>
</tr>
<tr>
<td>Undergraduate Medical and Health Sciences Admissions Test</td>
<td>UMAT</td>
<td>AUS</td>
<td>Undergraduate entry to medicine or health science courses</td>
</tr>
<tr>
<td>uniTEST</td>
<td>uniTEST</td>
<td>AUS</td>
<td>Equity groups/students with Year 12 performance</td>
</tr>
</tbody>
</table>

As Table 2 shows, the most common aptitude tests reported to be used in Australia universities are relatively well known to the higher education sector. These include uniTEST, STAT, UMAT, GAMSAT and ISAT. Two tests, ISIT and ISP are designed to help Indigenous students without formal educational qualifications gain admission to university and include aptitude testing components.

Several universities reported using well known internationally developed tests as an alternative for international applicants seeking to gain entry. The SAT is used for applicants without formal qualifications who sought entry into a range of different courses, where as MCAT and GMAT (both also created in the US), were used for entry into specific Medical and Business related courses respectively at these universities.

That such a diverse suite of assessments is being used testifies to the weight that institutions place in aptitude tests. The diversity of practice also suggests that institutions have developed innovative ways of factoring aptitude test information into their selection procedures, quite likely in response to increasing numbers and diversification of incoming students. The pathways into university – as with study itself – are diverse, and it is inevitable that a wide range of measures will be required. In the interests of transparency and efficiency, however, there would appear to be value, where feasible, on reducing use to a smaller number of more documented assessments.
As noted, a number of aptitude tests are being used by institutions in Australia. These tests are used in varying ways, including:

- as a hurdle to gaining an interview for specific degrees;
- for applicants with an ENTER/UAI that is below a certain cut-off point;
- for applicants with an ENTER/UAI within a defined score range;
- for applicants for courses who missed out on a first round offer;
- as an additional admissions tool for applicants from specific regions or targeted schools; and
- alongside Year 12 scores and interview scores to provide a composite admissions score.

**Target populations for aptitude tests**

Respondents to the university survey indicated that aptitude tests are being used by universities for admissions for a range of different populations. There are several main populations:

- **Equity students**: Many universities see the benefits of using aptitude tests as a means of allowing greater numbers of equity students to gain admission. Theoretically the tests are designed to provide a fair way of assessing students on a level playing field. They are suitable for determining whether students from diverse backgrounds have the capabilities to undertake tertiary study. Universities indicated that the use of aptitude tests for selecting more candidates in the key equity groups is likely to rise over the coming years. Data from such assessments provides an authoritative quantitative foundation on which institutions can base assessments of students’ capacity to succeed.

- **Non-traditional populations**: Universities see aptitude assessments as particularly useful for students from ‘non-traditional’ populations, such as those who have not completed formal study or are not able to access university using common pathways. The tests allow these candidates to demonstrate that they have the capacity to successfully undertake university studies. In particular, the selection of mature age applicants using aptitude tests is well established in Australia through the existence of STAT.

- **Insufficient Year 12 scores**: Aptitude tests are also used for students whose Year 12 scores aren’t sufficiently high to gain admission to their chosen course. Such tests can provide information on these students that may not be reflected in their Year 12 performance. Additionally this is also true for cases of special consideration. High performance in an aptitude assessment can support claims that a Year 12 score does not reflect the ability of the student, possibly due to disadvantage, illness or disability. In many cases, students in this group also fit within one or more equity groups, thus making their inclusion in higher education increasingly important.

- **International applicants**: For overseas applicants, aptitude tests provide a means of assessing suitability for study for selected courses, particularly where data from senior secondary assessments from the home country are not available or informative. Several universities reported the use of tests of language proficiency in this regard.

- **As a requirement for particular courses**: In specific course areas such as medicine, where competition for places is highly competitive, aptitude tests offer extra information on applicants’ potential that assists universities in making selection decisions when there is little separating applicants on final year school results. Often such tests are specifically designed to provide differentiation between high performing candidates that is not obvious through final year results such as the ATAR. In general, discipline-specific tests ensure that students have a basic level of knowledge that would allow them to undertake their desired course and probe students to demonstrate higher capacity for knowledge in specific areas. It is also argued that the use of these subject-specific tests can help enable universities to select the candidates most suitable for their chosen profession.
• **Identifying personality traits:** One respondent noted that aptitude tests can be used to identify extreme personality or social interaction profiles that are associated with personality disorders, which can be important in academic and professional contexts where teamwork and specific orientations are essential.

• **Diagnostic for enrichment:** Another interesting suggestion for a use of aptitude testing was to develop an appropriate measure of aptitude that could determine if any kind of remedial/transitional work might be necessary before full admission is granted. Candidates identified as potentially suitably could be given a provisional offer or an offer that includes remedial work as a prerequisite to undertaking certain subjects. Clearly, in the later instance, inability to successfully complete the remedial work would need to be part of academic progression.

**Reported pros and cons of aptitude assessment**

That a diverse range of aptitude assessments are used with many different populations suggests that institutions see many advantages in such testing. Many respondents indicated that they see aptitude tests as offering a reliable means of measuring whether an applicant has the capacity to successfully undertake tertiary study. They are seen to provide an independent source of information on the abilities of an applicant which can be combined with Year 12 achievement scores for greater accuracy in estimations of future performance at the tertiary level.

Further, such tests are seen to provide a quick and relatively inexpensive way of assessing a large number of potential applicants in a short time. Having access to aptitude test scores supports a more automated approach to student selection that reduces the burden in selection offices. For example, at more than one university, respondents reported that aptitude tests are used to shortlist potential applicants for interview. One respondent reported that, “They provide additional information on which to base selection decisions in highly competitive programs”, and another that, “They ensure that the selection process for highly competitive medical/health courses is more rigorous by providing a second measure of applicants’ ability to undertake these courses”.

For potential applicants to universities who do not have a suitable final year school mark, aptitude tests are seen to open the possibilities for access to tertiary study where that may not be admitted otherwise. One respondent observed that, “As the Bradley review is rolled out aptitude tests could provide a means to identity students with the capacity for higher education studies but who lack conventional TER or other means of entry.” Another stated that, “Aptitude tests can play an important role in non traditional entrance, providing objective information which can be used to direct students into the right preparatory or mainstream programs. This becomes increasingly important in the context of widening participation” and “Aptitude tests could potentially be useful when considering applications from students from disadvantaged educational backgrounds”.

Aptitude tests can also be used by students to identify their own strengths and weaknesses in particular subject areas. They can assist students to test their readiness for university studies through self testing as well as potentially highlight a more successful pathway into the degree or career that suits their preferences.

Concerns about using aptitude tests relate to the lack of available evidence on the effectiveness of such tests. There are concerns that the tests currently being used may not be reliable and valid for the purposes for which they are being used. Without this evidence, institutions cannot be sure that the tests add to the admissions processes, and some consider that other pathways to admissions may be more beneficial.

Other concerns relate to the structure and content of the tests in that they may be culturally biased, which can be an issue when many of these tests are targeted towards specific groups. This is a legitimate issue given that a number of universities are interested in using aptitude tests for choosing students specifically to boost their equity enrolment numbers.

Another concern is related to the administration of aptitude tests. There are issues seen in both the practicalities of test administration and with the process of answering the questions in these tests. Others see
the cost of sitting some of the aptitude tests currently available as potentially being deterrent for some individuals. Linked to this concern is the issue that so far Australia has lacked a universally accepted and centrally administered undergraduate admissions test.

Chapter summary

This chapter has established the international and Australian contexts for the practice of using aptitude tests as an admissions tool for selecting university candidates. The use of aptitude tests for this purpose is common internationally both within developed and developing countries. In many cases, the common practice is to use such tests as one of a number of metrics for determining the capabilities of candidates to undertake university studies. In this regard, often the aptitude test is used to complement secondary school outcomes.

This is an area in which Australia can make further progress. The Australian system has not yet universally embraced the use of tools other than Year 12 scores for selection. In general, the current selection system has not fundamentally changed over the past few decades, despite substantial changes in the higher education system as a whole. Recent times have seen a boom in enrolment numbers and growth in the proportion of the population who attend university. If Australian Government attainment targets are to be met, this growth will have to continue into the future. With such growth and change, and the push towards wider representation in the system of students from historically underrepresented groups, it is imperative that some change in the current selection methodology for admissions is seriously explored.

Through review of Australian institutions’ experiences it is clear there is scope and, indeed, even appetite and willingness for such change. As the final section of this chapter emphasised, it is quite common for universities to use aptitude tests in their tapestry of admissions processes. However, apart from perhaps the use of UMAT for medical admissions and the STAT for mature-age entry, there is no sector-wide common adoption of such tests for school leavers in Australia. Aside from a few examples, the current use of such tools in admissions is relatively unknown publicly. Indeed, characteristics of use is, in many cases a mystery within institutions that have adopted such tests. There is a need to improve the transparency and clarity of practice in this area.

As emphasised in the individual sections of this chapter, therefore, evidence exists that Australia is falling behind world practice in university student selection. This facet of Australian higher education has not changed in step with significant growth and diversification of the system. However, institutions have clearly adopted such assessments and are clearly interested in making progress in this area. To enhance the efficiency and transparency of admissions practices there is room to make further progress on this front. Adopting a coordinated national approach would appear to be an important means of supporting people’s admission into a system which is increasingly universal in scope.

Recommendation 1: Nationally coordinated implementation of uniTEST should be considered as a means of improving the transparency, efficiency and international relevance of university admissions in Australia.

Having established the context and rationales for aptitude testing, the following two chapters explore the pilot implementation of uniTEST within a number of institutions in Australia. The first of these analytical chapters (chapter three) explores the take up of the test and the types of students who it has benefitted. Chapter four examines the criterion validity of uniTEST in order to establish the extent to which it is an effective tool for university admissions.
3 ENGAGING INSTITUTIONS AND APPLICANTS

Introduction
This chapter reviews the extent to which institutions and applicants engaged with the SATTA pilot. It begins with a review of the ‘applicant’ population – the population that sought to use uniTEST as a means of enhancing the admissions process. The chapter then examines the ‘admitted’ population in more depth. This is the population of people admitted to university study where it is reported that uniTEST played a role in the selection process. Together, these analyses highlight the productivity of uniTEST and, more broadly, the Student Aptitude Test for Tertiary Admission.

Characteristics of the applicant population
Australia has 41 Table A and Table B higher education providers – the population of institutions eligible to take part in the SATTA pilot program. Of these, six institutions have used uniTEST over the last three years. Four institutions took part in the 2007/2008 pilot administration of uniTEST. Four universities took part in the 2008/2009 pilot administration. Another institution, which had trialled uniTEST in 2006, also supplied data for the study. In addition to the data obtained from students sitting uniTEST as a means of obtaining entry into their chosen university, two institutions organised special sittings of current first-year students to sit uniTEST as a means of increasing student numbers for research purposes.

Overall, therefore, around 10 per cent of Australian institutions participated in the evaluation. An important explanation for this level of engagement in the study is that due to contracting processes for this evaluation, institutions were only invited to take part in the study in late 2007 and late 2008, well after most had finalised their admissions procedures for the following year. In addition, it must be stressed that these institutions self-selected into the evaluation. While they cover a range of jurisdictions and institutional types, they do not represent the national institutional population in any discernable way.

The selection of individuals within each institution was highly complex. Ideally, a study of predictive validity would likely involve administration of the assessment to a random sample (or complex equivalent) of at least 400 individuals at the relevant level of analysis – typically institution or field within institution. As is often the case, however, there is usually a gap between research requirements and institution practice.

In the current evaluation, the only aspect of the selection process that was common across institutions was that uniTEST was used as a secondary selection method in cases where Year 12 or other relevant scores were not seen as sufficient. Hence it was not used with the main student population, but with disadvantaged or academically marginal groups.

Within this general context, each institution deployed the assessment in different ways. As the use of uniTEST increases there would be value in further understanding the specific admissions practices used by institutions, and developing suggestions that help them enhance their practice. Such exploratory and benchmarking work is undertaken with more established aptitude tests and is seen to be useful for shaping enhancements in practice.

Table 3 provides a summary of each institution’s applicant and admitted (successful) student populations. It summarises the basis on which uniTEST was used, the number of students who took the test, the number admitted, and the number not admitted. Further to the numbers in Table 3, 154 first year students took part in a special sitting of uniTEST, designed to increase sample numbers for the evaluation. These 154 students had already been admitted to university on the bases of other criteria. In total, therefore, results for 1,594 individuals were available for the study.
work as vital for underpinning continuous quality improvement. admissions processes in Australia. This is in lin
undertaken to develop more comprehensive evidence more n
results for the specified and somewhat distinct populations within each institution. They cannot be
statistical estimates. The results themselves could still be conside
number that the technical ideal of 400 students.
the students involved in the study reflect a very small proportion of each institution’s overall intake, have
been sampled into the stu
drawn from the sample of 1,440 individuals who sat uniTEST for admissions purposes and for whom results were available for the current study:

- 1,047 (73%) were not admitted to university, regardless of the extent (unknown) to which uniTEST was or was not used;
- 165 (11%) sat uniTEST and received a place at university because of their achievement in uniTEST;
- 205 (14%) sat the test but did not have uniTEST included in admissions; and
- 23 (2%) sat uniTEST but it is unclear how the data was factored into admissions.

Therefore, while 27.3 per cent of the uniTEST applicant group subsequently gained access to university, at least 11 per cent of the cohort can be identified specifically as having been given an offer on the basis of their uniTEST result and gaining the opportunity to enrol when this may not have otherwise been the case based on their high school achievement scores. This is a basic but very significant result. It offers a broad affirmation of the capacity of uniTEST to provide a complementary basis for admitting people into university.

Consultation with institutions exposed several reasons why uniTEST data may not have been factored into selection procedures. For instance, for a cohort of students at one institution staff were simply not able to identify whether uniTEST was a factor in student admission. There are also examples of institutions that invited potential applicants to sit uniTEST before their Year 12 marks were available. In many of these cases, a uniTEST score was not needed as the Year 12 score ended up being adequate for admission. Conversely, there are also examples of other students for whom uniTEST performance may have been inadequate to gain admission to their preferred course. Some of these students may have already gained admission based upon their Year 12 performance, but sat uniTEST in order to gain admission to a different course. Not all of these students were successful. It should also be noted that because of time limitations, several institutions had not clearly established the target population for which uniTEST would be administered. Because of this, there would be students in the population who sat uniTEST, but would not eligible to gain entry based on their scores irrespective of their performance on the test. These students may have gained admission to the university via another means.

As already noted, and could be expected given the characteristics of deployment as part of the SATTA pilot, the students involved in the study reflect a very small proportion of each institution’s overall intake, have been sampled into the study in a range of non-random ways, and for many subgroups are much fewer in number that the technical ideal of 400 students. This has implications for the precision and generalisability of statistical estimates. The results themselves could still be considered informative given that they reflect results for the specified and somewhat distinct populations within each institution. They cannot be generalised beyond these select groups, however, to the wider national student population. In order to gain a more nationally representative understanding of the benefits of aptitude tests, further research should be undertaken to develop more comprehensive evidence of their role (both actual and potential) in university admissions processes in Australia. This is in line with contemporary international practice, which sees such work as vital for underpinning continuous quality improvement.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Condition</th>
<th>Applicant</th>
<th>Admitted</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Second round, ENTER&lt;60</td>
<td>414</td>
<td>130</td>
<td>31.4</td>
</tr>
<tr>
<td>B</td>
<td>Second round, ENTER&lt;65</td>
<td>101</td>
<td>44</td>
<td>43.6</td>
</tr>
<tr>
<td>C</td>
<td>Second round, targeted equity groups</td>
<td>614</td>
<td>99</td>
<td>16.1</td>
</tr>
<tr>
<td>D</td>
<td>Second round, all who miss first-round offer</td>
<td>182</td>
<td>84</td>
<td>46.2</td>
</tr>
<tr>
<td>E</td>
<td>Pre-second round, students at targeted schools</td>
<td>110</td>
<td>25</td>
<td>22.7</td>
</tr>
<tr>
<td>F</td>
<td>Second round, ENTER&gt;45</td>
<td>19</td>
<td>11</td>
<td>57.9</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>1,440</td>
<td>393</td>
<td>27.3</td>
</tr>
</tbody>
</table>

In line with the survey of institution experiences reported in the previous chapter, Table 3 shows that universities use uniTEST for different reasons. Further consultation illustrated that the data is also used in different ways. Specifically, of the 1,440 individuals who sat uniTEST for admissions purposes and for whom results were available for the current study:
Detailed analysis of the admitted population

Table 3 highlights the difference between what can be referred to as the ‘applicant’ and ‘admitted’ populations. The applicant population consists of 1,440 individuals who sat uniTEST in late 2006, 2007 or early 2008 seeking entry into university. Unfortunately, due directly to the selection process under investigation, not all of these students can be included in the analysis of criterion validity as most did not progress on to study at the target institution. Rather, the analyses here and in the following chapter focus on the 393 students in the admitted population who, due to whatever selection process, were successful in gaining entry into a university course in their first semester of study after completing uniTEST. This latter population thus excludes individuals who were not admitted into a university course, or who were admitted but deferred commencement until a later semester.

In any analyses of the validity and function of a test it is important to understand the cohort of students that the test has benefitted. To do this, the analyses in this chapter explore the cohort of students who specifically gained entry to university on the basis of a uniTEST result – a subset of the ‘admitted population’. It reviews the characteristics and outcomes of these students.

In total 165 university students have been identified as receiving a place at university because of their achievement in uniTEST. Certain candidates examined here are missing data for some of the characteristics examined, therefore the number of uniTEST entrants in some of the figures below is not always 165. This cohort is compared here primarily with all commencing bachelor degree students at the six institutions involved in this research across a range of variables relating to student characteristics. These benchmark figures are drawn from the DEEWR Higher Education Student Collection (DEEWR, 2007), from the Australasian Survey of Student Engagement (Coates, 2009), from ACER files, and from other national statistics publications.

In general, the findings show that there are some interesting differences between the uniTEST entrant group and the whole cohort of beginning students. These findings provide insight into those who gained access to university because of their uniTEST results. It must be remembered that given the small numbers of students in this group, the findings should be interpreted as indicative rather than conclusive.

When the distributions of these two cohorts are examined by gender, the results show that a greater proportion of males gained entry to university using uniTEST scores than females. This is interesting because among the full commencement cohorts at these universities, females made up a much larger proportion of the entrant group. Figure 3 shows that males comprised 56.4 per cent of the group who gained access to university because of a uniTEST score, yet made up only 41.1 per cent of all entrants to these universities.

This finding might be read as suggesting that uniTEST favours males. It could also be a result of the overall lower year 12 performance of males when compared with females (Edwards, Birrell, & Smith, 2005) and hence a product of the fact that a larger number of males are in the position where uniTEST becomes a factor needed in order to gain access to university.
To explore the possibility of gender bias, detailed analysis was conducted to determine whether individual uniTEST items have different levels of difficulty depending on the applicant’s gender. This analysis is based on the full cohort who undertook uniTEST in 2008 and 2009, so includes a wider spectrum of participants than those included in the analyses above.

Differential item functioning (DIF), colloquially referred to as ‘item bias’, involves investigation of the extent to which items provide consistent measurement across specific sub-groups of test takers. If an item is much easier to complete for one group of students than for another, then it may be said to be ‘biased’ against the first group. It is important to note that all assessment instruments are influenced to some degree by the contexts in which they are undertaken, the influences of various forms of measurement disturbance, and the individuals being assessed. What is important is that any bias does not reach unacceptable levels, and that the degree of bias is known.

Figure 4 provides a plot of item difficulty estimates for females against males, with 95 per cent confidence bands for the difference (the two curved lines). The bands were computed based on the standard errors of the item difficulty estimates for each of the two groups and, therefore, were affected by the corresponding sample sizes. They are narrow here because the sample sizes are large (which means that standard errors of the estimates are very small). Items plotted above the bands are relatively harder for females while those below the bands are relatively harder for males.

The results expose very little differential performance by gender. While there is a small amount of random variation between the two calibrations, the relationship is linear and there are no distinct outliers. This confirms that differences in male and female performance are a product of circumstance rather than the properties of the test.
There are also differences between the uniTEST cohort and all commencing students in relation to age. Figure 5 shows that the vast majority of uniTEST entrants 18 or 19 years old – the key ages for which the test is designed. By comparison, there is a greater overall spread into the older age groups for the overall commencement cohort at these institutions. On a similar theme, Figure 6 shows that students admitted on the basis of a uniTEST outcome were much more likely to enrol straight out of high school (87.5 per cent) than was the situation in general for commencing students at the selected universities in this study (56 per cent). These two figures highlight the strong use of uniTEST at this particular stage of a persons’ education. It demonstrates that for the pilot phase of this testing, the core group of students that would be targeted in any wider administration of the test were involved.
Figure 5: Age by admission type

Figure 6: Proportion of university enrolees who enrolled directly after completing school by admission type

Figure 7 shows that there was a greater concentration of people from English speaking backgrounds in the uniTEST entrant group in comparison to the overall commencing domestic students within the six universities involved in this study.
Unfortunately, due to the characteristics of available data, it is not possible to conduct the psychometric analyses presented above for gender that would be required to affirm whether the distribution of admitted students by language background is due to cohort, individual or test item differences. It is possible, for instance, that one of the reasons for this outcome relates to accessibility and knowledge about the availability of the test. It could be that those for whom English is not their first language were not sufficiently able to access information about uniTEST and its use in these institutions. If this is the case, then a more universal administration of the test (for example across each secondary school in the country) could have the potential to ameliorate this element of potential disadvantage.

Figure 7: Language background by admission type

It is well documented that the overall proportion of students from socioeconomically disadvantaged backgrounds is nationally low. Recent Australian Government policy has been set to increase the representation of this group in the university population (Australian Government, 2009). Figure 8 shows that the cohort who specifically gained access to university via a uniTEST score had a larger representation in the low socioeconomic status group than the overall entry cohort for these specific institutions. In total 15.3 per cent of those in the study who had gained entry as a result of a uniTEST score were from areas of low socioeconomic status, compared with 11.8 per cent of all domestic commencing bachelor degree students at the six institutions involved in this study.

This finding suggests that uniTEST does facilitate the entry of students from low socioeconomic backgrounds to a greater degree than do the usual practices of university admission in those institutions involved in the pilot. However, it must be noted that Figure 8 still shows that about half of all commencing students in these universities were from high socioeconomic backgrounds, regardless of whether they gained entry through uniTEST or traditional selection mechanisms. Therefore, these results should not be viewed as an indication that tests such as uniTEST is the solution to substantially increasing the representation of low socioeconomic status students within the sector.
Another measure that can be used as a proxy for socioeconomic status relates to the parental education levels of students. In the uniTEST cohort, a limited number (72) of students have this information recorded. In Figure 9 these uniTEST entrants are compared with all first year students in Australia who responded to the Australasian Survey of Student Engagement (AUSSE) in 2008 (Coates, 2009). The figure shows that uniTEST entrants were slightly more likely to have a parent with a university qualification and slightly less likely for a vocational qualification to be their parents highest qualification than was the case among the respondents to the AUSSE in 2008. This outcome is slightly contradictory to that shown in the socioeconomic status breakdown in Figure 8, although the exact comparability of the cohorts in Figure 9 is questionable given that this is a national comparison group rather than a direct comparison with the individual institutions involved in the uniTEST study.
Figure 9: Parental education level by admission type

Figure 10 shows that students who gained access to university via uniTEST were slightly more likely to come from a metropolitan area than the average commencing student across the institutions involved in this research. Just over one fifth (20.1%) of all commencing students at the six universities that form this research resided in a regional or remote area of Australia, while the comparative rate for uniTEST entrants was lower at 16.8 per cent. A key explanation for this slight difference is likely to be due to accessibility of uniTEST to geographically isolated students. Accessibility is a key issue in relation to any admissions measure that involves candidates being required to attend a central test site in order to sit the test and this appears to be the case for uniTEST in its current form. This supports the move towards the more flexible approach for the administration of such tests, as proposed in the last chapter of this report.

It is therefore likely that the outcome shown in Figure 10 is not a result of underlying bias against regional candidates in the test instrument itself (which it is not possible to specifically investigate due to data limitations), but rather a consequence of the implementation of the test during the pilot. Making uniTEST universally available to students, for example on site at their secondary schools, would ameliorate this element of accessibility disadvantage and open up the prospects of selection via this method to a greater range of students.

Given that regional and remote students are on average more likely to be from a low socioeconomic background, greater accessibility to uniTEST for such students has the potential to further increase the extent to which this test may be offering low socioeconomic status students additional opportunities for entry to university.
While numbers were very limited (only 49 uniTEST entrants had data available), the data on school sector show that the majority (57.1%) of students gaining access to university via uniTEST were from government schools. Figure 11 shows the share of uniTEST entrants by school sector, compared with the spread of students by sector across the country as measured in Year 12 (this comparative data was sourced from the ABS (2008) publication Schools Australia). For government school students, the rate of uniTEST entrants is shown to be representative of the whole Year 12 population. Catholic school students are slightly underrepresented, while independent school students are over represented.

Comprehensive data relating to the rate of actual university enrolments by school sector is not nationally available, however previous research (Edwards, 2005, 2007, 2008a, 2008b) has shown that in general, government school students are not represented in the university commencement cohorts at anywhere near the levels in which they are represented in the Year 12 student population. Therefore, this indicative outcome showing a representative share of those from government schools in the uniTEST entrant group is interesting in terms of boosting the participation of this large group of students. Again, such conclusions need to be interpreted carefully given the small numbers of uniTEST entrants for which school sector data was available.

Figure 10: Residential location by admission type
Figure 11: School sector by admission group

Figure 12 shows the distribution of Year 12 results among the group of students who gained entry to university on the basis of a uniTEST score. These results are based on the ENTER, UAI or similar entrance ranking provided to students. For the purpose of this comparison, such rankings can be interpreted as a percentile rank of the whole cohort. The figure shows that students from the full spectrum of Year 12 achievement have used uniTEST as a mechanism for gaining entry to their chosen university and course.

More than one third of this group (35.4%) had Year 12 results in the 70s. Another one third had a score of 80 or above (33%), while the final ‘third’ (31.7%) were those with scores below 70.
The final comparison in this section examines the field of education students enrolled in once they had entered university, comparing uniTEST entrants with the overall profile of commencers from the six universities involved in this research. The results displayed in Figure 13 show that uniTEST entrants were concentrated in a few core fields. Management and commerce is the most notable field in this regard, with 27.6 per cent of the uniTEST entrant cohort, but only 20.1 per cent of all commencing students. Society and culture (34.4% of uniTEST entrants but 30.3% of all students) and information technology (6.7% of uniTEST group but only 3.9% of all entrants) also revealed gaps in this regard.

At the other extreme, there was very low representation of uniTEST entrants in the health field (1.8% of this group) in comparison to the share in this field among all entrants (12.6%). The difference in the education field in this regard is also notable (4.9% of the uniTEST group compared with 9.8% of all entrants). The natural and physical sciences, and the creative arts fields showed the closest matches in relation to population share of these two groups.

These findings suggest that in the pilot phase of uniTEST, the test has had greater influence on some areas of study than on others. There are likely to be a number of factors in this regard. For example, the low relative numbers of uniTEST entrants in the health field is likely to be related to the fact that very high year 12 entrance scores are required to gain entry to many courses in this field. As shown in Figure 12 only small proportions of the uniTEST group have very high Year 12 scores. In addition to this, another aptitude test, UMAT, exist as a supplementary measure of entry into many of the health science fields, potentially making uniTEST redundant in the current implementation format.

**Figure 12: uniTEST entrants to university by Year 12 score**

The results displayed in Figure 13 show that uniTEST entrants were concentrated in a few core fields. Management and commerce is the most notable field in this regard, with 27.6 per cent of the uniTEST entrant cohort, but only 20.1 per cent of all commencing students. Society and culture (34.4% of uniTEST entrants but 30.3% of all students) and information technology (6.7% of uniTEST group but only 3.9% of all entrants) also revealed gaps in this regard.

At the other extreme, there was very low representation of uniTEST entrants in the health field (1.8% of this group) in comparison to the share in this field among all entrants (12.6%). The difference in the education field in this regard is also notable (4.9% of the uniTEST group compared with 9.8% of all entrants). The natural and physical sciences, and the creative arts fields showed the closest matches in relation to population share of these two groups.

These findings suggest that in the pilot phase of uniTEST, the test has had greater influence on some areas of study than on others. There are likely to be a number of factors in this regard. For example, the low relative numbers of uniTEST entrants in the health field is likely to be related to the fact that very high year 12 entrance scores are required to gain entry to many courses in this field. As shown in Figure 12 only small proportions of the uniTEST group have very high Year 12 scores. In addition to this, another aptitude test, UMAT, exist as a supplementary measure of entry into many of the health science fields, potentially making uniTEST redundant in the current implementation format.
Formative Evaluation of the Student Aptitude Test for Tertiary Admission

Evaluation Report

Figure 13: Broad field of education by admission type

Overall, this analysis comparing the group of students who successfully used uniTEST to gain entry to university with the wider cohort of commencing students has shown that there are some notable differences in characteristics of these groups. When compared with the wider student population, students gaining entry through uniTEST were more likely to be male, more likely to come from a low socioeconomic status background, more likely to live in a metropolitan area and less likely to be from a non-English speaking background. Each of these differences should be of interest to policy-makers and universities who are currently contemplating ways in which participation in university can be broadened.

Clearly the findings relating to socioeconomic status are of particular significance given that the Australian Government targets to increase low socioeconomic status enrolments in Australian universities. However, while these findings do indicate a higher representation of low socioeconomic status students among the uniTEST entrant group, it should not be simply assumed that this test is the answer to solving the low socioeconomic status participation rate issues.

The independent impact of characteristics on uniTEST scores

By way of summary, the analysis that follows explores the extent to which overall uniTEST scores are influenced by a range of variables. A number of regression models were constructed to explore the extent to which the measurable characteristics of students can explain the variation in uniTEST scores. This discussion is based on the most comprehensive of these models. The analysis is based on all uniTEST candidates for whom there is data available – therefore it is not restricted to only those students who gained access to university and are part of the predictive validity data collection. As with the item analysis above, using a wide cohort of candidates helps to increase our understanding of the way in which uniTEST works.

The model used enables exploration of the extent to which uniTEST results can be explained by a number of important variables. This analysis provides coefficients for each variable, while at the same time controlling for their influence on each other. In other words it provides an idea of the role that a number of important characteristics and factors play in explaining uniTEST scores of individuals. The variables used in this model
are Year 12 outcome (TER, ENTER or UAI), gender, language background, socioeconomic status and whether a uniTEST result was used successfully to gain entry to university (note that other candidates may have gained entry to university, but not on the basis of their uniTEST result).

The number of variables used in this analysis is limited by the availability of data relating to uniTEST candidates. However, the model used here explains 26.2 per cent of the total variance in uniTEST scores, which is a relatively good outcome given the number of variables and the nature of the analysis. The outcomes of this regression model are displayed in Figure 14 and Table 4 below. Figure 14 shows the standardised coefficients for the key variables, while Table 4 displays the unstandardised (uniTEST point metric) coefficients.

Figure 14 provides an indication of the different levels of influence – and the directionality of this influence – that certain variables have on uniTEST outcomes. It also enables the ability to compare the relative influence of these variables. As the figure shows, Year 12 scores are positively associated with uniTEST outcomes. In addition, those who gained entry to university with their uniTEST score are unsurprisingly likely to be more successful than other candidates, even when other characteristics are controlled for. The figure also shows that being female and from a non-English speaking background are negatively associated with uniTEST scores. In other words, females tend to perform worse than males when other factors are controlled for and those from a non-English speaking background have lower scores than those with English as their first language even when other variables are taken into account. The effect of being from a low socioeconomic status background (as opposed to a high socioeconomic status background) is shown in Figure 14 to be very small relative to the other variables and in this model was not statistically significant.

Table 4 provides the unstandardised coefficients of the model, which provide an indication of the actual difference in uniTEST points each of these characteristics has on uniTEST outcomes. This model includes 431 cases and explains 26 per cent of variation in the dependent variable.

The coefficients here show that for every one point rise in TER, uniTEST scores are estimated to grow by 0.26 points, even after accounting for sex, language, socioeconomic status and university offer. In more
simple terms, a person with a TER score of 75 would be expected to have a uniTEST score about 2.6 points higher than someone who had a TER of 65, regardless of other personal characteristics.

The other outcomes in this table estimate that females have uniTEST score 3.38 points lower than males, net of other influences and that candidates from non-English speaking backgrounds are estimated to perform 7.8 points lower than those from English speaking background, regardless of gender, ENTER, socioeconomic status and university entry success. Those who gained entry into university specifically on the basis of their uniTEST score are estimated to perform 8.7 points better than others even when controlling for other characteristics. The coefficients for the socioeconomic status variables are displayed in Table 4 but were not significant. This outcome shows that uniTEST outcomes are not affected by socioeconomic status and is an important finding.

Table 4: uniTEST point difference for specified variables (unstandardised regression coefficients)

<table>
<thead>
<tr>
<th>Variable</th>
<th>uniTEST points difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 12 ENTER</td>
<td>0.26*</td>
</tr>
<tr>
<td>Sex (female)</td>
<td>-3.38*</td>
</tr>
<tr>
<td>Language (NESB)</td>
<td>-7.80*</td>
</tr>
<tr>
<td>Socioeconomic status (low versus high)</td>
<td>-1.14</td>
</tr>
<tr>
<td>Socioeconomic status (middle versus high)</td>
<td>-0.86</td>
</tr>
<tr>
<td>Gained entry with uniTEST</td>
<td>8.70*</td>
</tr>
</tbody>
</table>

* p < 0.01

This outcome for socioeconomic status is contrary to that found for this variable when the regression analysis was instead focussed on Year 12 outcomes. An identical regression model, with Year 12 ENTER as the dependent variable rather than uniTEST was constructed to examine the influences of these variables with specific focus on socioeconomic status. The unstandardised regression coefficients (in the form of ENTER points) are displayed in Table 5. This analysing involving 431 cases explained 18 per cent of the variation in year 12 scores. This table shows that unlike uniTEST, the effect on sex and language background does not appear to be significant in estimating influence on ENTER scores. However, socioeconomic status is shown here to have a statistically significant impact on Year 12 outcomes.

The figures in Table 5 show that compared with candidates from high socioeconomic status areas, those from low socioeconomic status localities are estimated to have a ENTER 5.73 points lower after controlling for uniTEST outcome, sex, language and whether they gained entry to university as a result of their uniTEST result. The standardised coefficients show that the low socioeconomic status variable had almost as much impact on Year 12 outcome as the candidates’ uniTEST results. Those from middle socioeconomic status areas were also shown to have lower results than high socioeconomic status candidates, at a statistically significant level – by an estimated 4.28 ENTER points.

In many ways, the complementary aspects of uniTEST and Year 12 scores are revealed in Table 4 and Table 5. Where uniTEST shows some influence in relation to gender and language, Year 12 results do not and where Year 12 results show influence from socioeconomic status, uniTEST does not.

Table 5: Year 12 ENTER point difference for specified variables (unstandardised regression coefficients)

<table>
<thead>
<tr>
<th>Variable</th>
<th>TER points difference</th>
<th>Standardised coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>uniTEST</td>
<td>0.30*</td>
<td>0.29*</td>
</tr>
<tr>
<td>Sex (female)</td>
<td>0.74</td>
<td>0.03</td>
</tr>
<tr>
<td>Language (NESB)</td>
<td>2.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Socioeconomic status (low versus high)</td>
<td>-5.73*</td>
<td>-0.21*</td>
</tr>
<tr>
<td>Socioeconomic status (middle versus high)</td>
<td>-4.28*</td>
<td>-0.19*</td>
</tr>
<tr>
<td>Gained entry with uniTEST</td>
<td>4.52*</td>
<td>0.13*</td>
</tr>
</tbody>
</table>

These results in relation to the socioeconomic status variable are further highlighted in Figure 15, which compares the relative impact of low socioeconomic status on uniTEST and Year 12 scores (taken from the standardised coefficients of the above results). As is displayed in the graph, the impact of socioeconomic status on Year 12 scores among the group of students analyses here is much greater than its impact on
uniTEST outcomes. This outcome is net of the influence of other key variables on student outcomes and socioeconomic status is shown to be a significant predictor influencing Year 12 outcomes.

\[ \text{Figure 15: Impact of low socioeconomic status (as opposed to high socioeconomic status) on uniTEST and Year 12 outcomes, standardised regression coefficients} \]

**Chapter summary**

The findings of this analysis are important in providing a contextual basis for viewing uniTEST before entering into the criterion validity analyses that follow. They show that, in general, those who gain access to university via uniTEST have slightly different characteristics than are found in the general university population. This finding is important because it shows that uniTEST has the potential to increase diversity within the university population, especially in terms of gender and socioeconomic status.

Detailed analysis of the impact of certain important characteristics on uniTEST outcomes has shown that while sex and language do seem to have an influence on uniTEST outcomes, socioeconomic status does not have a significant impact on uniTEST scores. This finding was shown to be particularly important given that socioeconomic status does have a significant influence on Year 12 outcomes. The different influences on uniTEST and Year 12 outcomes in the final analyses in this chapter help to show that uniTEST and Year 12 should be viewed as complementary measures.

Currently the ability to explore these and other important facets of uniTEST are limited by the numbers of students who have been involved (and for who there is additional information). However, the potential for this type of test is to a large extent revealed in these analyses. It would appear that an impediment to greater use of the test by candidates and by institutions is the practicality of sitting the test (for students), and offering the test (for institutions). A system-wide approach to the running of uniTEST, creating greater opportunities for students to sit for the test, a centralised repository of results on which any institution can draw for selection purposes and a greater awareness of the test and its uses, would have the potential to substantially change the approaches in Australia towards university selection processes. These ideas are further discussed in the concluding chapter.

**Recommendation 2:** To ensure the most effective implementation, expansion of the use of aptitude tests with school student and leaver populations should be accompanied by ongoing analysis of the characteristics of the applicant and admitted populations.
4 VALIDITY ANALYSES

Introduction

As earlier chapters have suggested, this evaluation builds on a long tradition of examining the criterion validity of high-stakes tertiary selection tests. Criterion validity is seen as important because it helps to empirically situate selection tests within their broader contexts by, for instance, providing an assurance to educators and the public that test results are reliable and add value to selection decisions. As entry to university becomes more competitive, an increasing amount of development and research is conducted to produce evidence that can be used to optimise the validity and efficiency of selection processes.

Evaluating the criterion validity of uniTEST is an important part of the national SATTA evaluation. It reflects an important early analysis of this relatively new aptitude assessment. Earlier evaluation during the first stage of the SATTA evaluation determined the relationship between uniTEST and other measures used to admit students into university study, examined its capacity to operate as an effective selection mechanism, and demonstrated the ‘incremental validity’ (or ‘value added’) by uniTEST to admissions procedures in Australian higher education. This subsequent evaluation revisits these issues, with more data and after candidates have been at university for a longer period of time.

Specifically, this evaluation examines whether uniTEST has suitable levels of criterion validity to affirm its use as a university selection instrument. The primary question investigated in this chapter is: Does uniTEST have suitable levels of criterion validity to support its use as a selection instrument?

It is important to stress again that uniTEST is not designed to predict levels of achievement at university, but to reject candidates who are unlikely to succeed. The purpose of uniTEST is to identify individuals with the capacity to undertake university study. The difference here is in the distinction between ‘ability’ and ‘achievement’, a difference commonly confused in informal conversation. While ability pertains to the capacity to perform, achievement refers to demonstrated performance, performance which is influenced by a wide range of factors.

Having said this, it would seem desirable that a positive relationship exists between measures which are used to admit students to university study and performance during the course. This is an important point to make as it contextualises the size of the likely correlations between selection measures and first year marks, which typically lie in the range of 0.3 and 0.5. A much higher correlation would be anticipated if a follow-up measure of ability were taken in the first year of study.

As noted at the outset, several objectives were beyond the scope of this evaluation. The study did not seek to review individual university selection procedures, although it recognised that the results may have indirect implications on the management of these processes. Importantly, as is common in studies of predictive validity, the analyses did not include uniTEST candidates who were not successful in gaining entry to a university course. In addition, the study is not able to distinguish between university results obtained using different forms of assessment. This research focuses squarely on criterion validity, rather than on the other psychometric properties of uniTEST. While the results of the analysis may carry implications for the other aspects of the instrument, these are indirect rather than direct outcomes of the study.

The concept of validity

Validity is an essential characteristic of good assessment. A valid test (an assessment instrument) is one that successfully achieves the purpose for which it is being used. There are as many kinds of validity as there are purposes of testing.

The literature lists literally dozens of kinds of validity but they can be grouped into three general areas – content, construct and criterion-related. Content validity, as the name suggests, asks whether the test captures
the knowledge and skills it is meant to capture. For example, does a spelling test require students to spell a reasonable sample of the words that the students are expected to be able to spell. Content validity is determined when a group of experts consider the test and the subject area being tested and decide the extent to which the test is a fair representation of that area.

Construct validity is applicable when assessing an abstract quality. It was introduced into the literature by Cronbach and Meehl (1955). It asks whether the assessment gets at the essence of the quality being measured or assessed. For example, the quality being assessed might be ‘well-being’ or a ‘caring personality’. We need to first consider the kinds of behaviours expected of those with ‘well-being’ or with ‘caring personalities’. We then investigate whether those being assessed exhibit those behaviours which might be assessed by verbal responses to contrived situations or actual responses to structured, real-life situations, as when a rating scale is used by an expert observer. Presumably, high scorers will exhibit more of the construct than low scorers. Construct validity depends on both a rational approach involving expert opinion but also a statistical approach to ensure that the construct actually differentiates as expected.

The processes that go into the construction of a test should ensure that content validity is present and that the basis of construct validity is also present although extra statistical work will have to be carried out subsequently to verify this.

The current study is primarily concerned with criterion-related validity, the third kind of validity listed above. As noted, this kind of validity asks whether the results of a test are related to test results obtained by the same students on similar tests administered at about the same time (concurrent validity) or are related to the results of tests taken at some future time which the first test results were supposed to predict (predictive validity). Both concurrent validity and predictive validity are simply two forms of criterion-related validity differentiated by the time of the administration of the second test.

Criterion-related validity is usually expressed in terms of a correlation coefficient where the correlation is 1.0 if the two sets of test results are perfectly related such that the rank order of students of each test is exactly the same, or is -1.0 if the two sets of results are so unrelated that the first ranking on one test becomes the last ranking on the other and so on in exact reverse order. As noted and we shall see in the subsequent discussion, neither a correlation of 1.0 or of -1.0 is ever seen in educational measurement except perhaps in bizarre circumstances. Rather, the tendency is for concurrent or predictive validity indices to provide a moderate, positive correlation.

An example of concurrent validity would be the correlation between the results of a new group-administered test of physical fitness with the results of an old, established, individually-administered test of physical fitness. The new test would be relatively inexpensive to administer but the older, established test would be time consuming and therefore expensive to administer. If it can be shown that the new test correlates well with the old test (satisfactory concurrent validity) then strong consideration should be given to replacing the old with the new.

An example of the use of predictive validity would be provided when marks obtained by college entrants on a ten-point scale for a short essay on ‘Why I want to enrol at East Cupcake College’ are correlated with their results at the end of the first semester. If the predictive validity (the correlation) is negative, meaning that the better the essay mark the poorer the grade point average, then it would be wise to reject including the essay mark in the admissions decisions. If the predictive validity is positive and low to moderate (say, 0.2 to 0.6) consideration should be given to including the essay as a component of the admissions process. If, in the unlikely event that the correlation were very high, then consideration would have to be given to making the essay mark a major element in the admissions process.

To summarise, we note that content validity is established mainly through a rational/judgmental process involving expert groups critically considering whether the test items represent properly the specified domain of knowledge and skills. Construct validity also has this rational component but there are, as well, experimental and statistical methods which can be used to enhance the argument for construct validity. Criterion-related validity is dependent almost totally on empirical/statistical procedures that we have just alluded to.
Sample characteristics

The secured data reflect all eligible students from the participating universities. In total, useable data were obtained from 547 students who sat uniTEST. Note that this includes all students with a uniTEST score who were admitted to university including those from the special sitting. Further detail in this regard is provided at the beginning of Chapter 3.

For the purposes of the criterion validity analyses a further control group population was defined to provide a point of reference against which the group of students admitted using uniTEST could be compared. This group of 833 students was admitted to university using traditional entry methods, not an aptitude test. As much as possible, control group students were matched against successful uniTEST applicants in terms of field of education, qualification level, equity group status, Year 12 results, and student demographics.

Six universities (labelled A to F in Table 6 and subsequent analyses) were involved in the evaluation. One of these, Institution F, had very small numbers and therefore its results have not been separately displayed in subsequent analyses. However, the results of the students from this institution have been included in the analyses based on the combined data for all candidates.

Table 6 shows characteristics of these students. The percentage figures shown are the proportion for each subgroup within the sample. The figures show that the control group characteristics coincide roughly with those of the uniTEST group. Similar proportions of attendance type, disability and Indigenous status were found across both groups. The low percentage of students in the uniTEST group who had a disability, or identified themselves as being of Indigenous origin is perhaps surprising given the use of uniTEST in at least one institution as a means of providing those in disadvantaged groups an alternative entry into university.

The majority of students across both groups were studying in the management and commerce and society and culture fields of education. Similar proportions in each field of education were found across the uniTEST and control groups, and is a reflection of the use of broad field of education as a matching variable for selecting an appropriate control population. There was a similar age distribution for both groups, with the vast majority of students being aged between 18 and 20. This is to be expected, given that uniTEST is targeted towards those who have recently completed Year 12 study.

Slight mismatches in the two populations were present for gender, where the control group had a greater number of female participants and socioeconomic status, where the uniTEST cohort had higher representation in the low and the high groups.
Table 6: uniTEST evaluation sample characteristics

<table>
<thead>
<tr>
<th></th>
<th>uniTEST</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>547</td>
<td>100</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution A</td>
<td>130</td>
<td>23.8</td>
</tr>
<tr>
<td>Institution B</td>
<td>82</td>
<td>15.0</td>
</tr>
<tr>
<td>Institution C</td>
<td>99</td>
<td>18.1</td>
</tr>
<tr>
<td>Institution D</td>
<td>200</td>
<td>36.6</td>
</tr>
<tr>
<td>Institution E</td>
<td>25</td>
<td>4.6</td>
</tr>
<tr>
<td>Institution F</td>
<td>11</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>265</td>
<td>48.6</td>
</tr>
<tr>
<td>Male</td>
<td>280</td>
<td>51.4</td>
</tr>
<tr>
<td><strong>Attendance type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>465</td>
<td>94.1</td>
</tr>
<tr>
<td>Part-time</td>
<td>29</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Language background</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-English</td>
<td>105</td>
<td>19.7</td>
</tr>
<tr>
<td>English</td>
<td>428</td>
<td>80.3</td>
</tr>
<tr>
<td><strong>Field of education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural and physical sciences</td>
<td>77</td>
<td>15.6</td>
</tr>
<tr>
<td>Information technology</td>
<td>27</td>
<td>5.5</td>
</tr>
<tr>
<td>Engineering</td>
<td>11</td>
<td>2.2</td>
</tr>
<tr>
<td>Architecture and building</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Agriculture and environmental studies</td>
<td>6</td>
<td>1.2</td>
</tr>
<tr>
<td>Health</td>
<td>27</td>
<td>5.5</td>
</tr>
<tr>
<td>Education</td>
<td>18</td>
<td>3.6</td>
</tr>
<tr>
<td>Management and commerce</td>
<td>128</td>
<td>25.9</td>
</tr>
<tr>
<td>Society and culture</td>
<td>169</td>
<td>34.1</td>
</tr>
<tr>
<td>Creative arts</td>
<td>30</td>
<td>6.1</td>
</tr>
<tr>
<td><strong>Disability status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Disability</td>
<td>450</td>
<td>94.3</td>
</tr>
<tr>
<td>Disability</td>
<td>27</td>
<td>5.7</td>
</tr>
<tr>
<td><strong>Indigenous status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>520</td>
<td>99.6</td>
</tr>
<tr>
<td>Indigenous</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Age groups</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 to 20</td>
<td>477</td>
<td>87.4</td>
</tr>
<tr>
<td>21 to 24</td>
<td>52</td>
<td>9.5</td>
</tr>
<tr>
<td>25 to 29</td>
<td>13</td>
<td>2.4</td>
</tr>
<tr>
<td>30 to 39</td>
<td>3</td>
<td>0.5</td>
</tr>
<tr>
<td>40 plus</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Socio-Economic Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>51</td>
<td>13.0</td>
</tr>
<tr>
<td>Mid</td>
<td>121</td>
<td>30.9</td>
</tr>
<tr>
<td>High</td>
<td>219</td>
<td>56.0</td>
</tr>
</tbody>
</table>

Analysis of concurrent validity

Concurrent validity analyses were conducted to investigate the value added by uniTEST to the process of selecting applicants into university study. Normally, concurrent validation studies are conducted to determine the effectiveness of a particular treatment against a ‘gold standard’. They are often used, for instance, to determine whether a new drug is equally efficacious as a successful one which may already be in the market. In such contexts, for the new drug to be acceptable, it is desirable for its effects to correlate highly with the existing product. Appropriately high correlations provide evidence of the concurrent validity of the new drug.

The context of educational selection requires a different interpretation of concurrent validity. If uniTEST were being considered against an alternative objective assessment, then evidence of strong association would be desired. When considered against Year 12 results, however, it is likely that concurrent validity is affirmed by divergent rather than convergent relationships. That is, while some ‘redundancy’ or overlap may be required, uniTEST is validated as a useful component in a selection process to the extent that it provides relevant evidence that supplements Year 12 results. Therefore, positive correlations are desirable, but large correlation results are not because the aptitude test is being used in conjunction with, rather than instead of the Year 12 scores.

Concurrent validity in this evaluation was investigated by comparing uniTEST data sourced from ACER Assessment Services to Year 12 scores sourced from institutions. The uniTEST results included section
scores for questions assessing Verbal Reasoning (VR), Quantitative Reasoning (QR) and Critical Reasoning (CR), all of which are combined to form a Total Score (TL).

Table 7 presents correlations (scaled onto a 100-point metric to remove leading zeros and decimals) between uniTEST against Year 12 scores. Statistically significant correlations (assuming a conservative 5 per cent type one error rate) have been flagged by an asterisk. Note that the small number of observations within each institution means that the power of each correlation is relatively low, and hence significant correlations at the institution level are difficult to obtain.

The results show that there are large variations in the relationship between Year 12 and uniTEST scores across institutions. Of the 24 correlations, 8 are statistically significant which suggests in itself that there is a relationship. Results for the combined analysis range on the 100-point scale between 14 for the Verbal Reasoning score and 46 for the Quantitative Reasoning score. All combined uniTEST component scores and total scores were significantly correlated with Year 12 performance. Within institutions, the relationship varies between -8 for Verbal Reasoning and 41 for Quantitative Reasoning.

Note that the ‘combined’ correlation is not an average of all institution-specific correlations, but a correlation calculated for the combined data. The value appears somewhat anomalous on first inspection, but is interpretable given closer review of the cross-institutional distribution. Such variation affirms the importance of taking a cross-institutional perspective in these kinds of analyses.

Table 7: uniTEST and Year 12 correlations

<table>
<thead>
<tr>
<th>Institution</th>
<th>VR</th>
<th>QR</th>
<th>CR</th>
<th>TL</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9</td>
<td>22*</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>41*</td>
<td>24</td>
<td>29*</td>
</tr>
<tr>
<td>C</td>
<td>-8</td>
<td>15</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>D</td>
<td>-4</td>
<td>27*</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>E</td>
<td>-5</td>
<td>4</td>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>Combined</td>
<td>14*</td>
<td>46*</td>
<td>28*</td>
<td>36*</td>
</tr>
</tbody>
</table>

* p < 0.05

Table 8 shows the percentage of shared variance between the two measures. While this statistic is simply the square of the correlation coefficient, it provides a useful indication of the strength of relationship. For instance, within institution A only 0.8 per cent of the variance is shared between the uniTEST and Year 12 scores.

Table 8: uniTEST and Year 12 shared variance (per cent)

<table>
<thead>
<tr>
<th>Institution</th>
<th>VR</th>
<th>QR</th>
<th>CR</th>
<th>TL</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.8</td>
<td>4.8</td>
<td>0.0</td>
<td>1.7</td>
</tr>
<tr>
<td>B</td>
<td>0.1</td>
<td>16.8</td>
<td>5.8</td>
<td>8.4</td>
</tr>
<tr>
<td>C</td>
<td>0.6</td>
<td>2.3</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>D</td>
<td>0.2</td>
<td>7.3</td>
<td>0.2</td>
<td>2.0</td>
</tr>
<tr>
<td>E</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Combined</td>
<td>2.0</td>
<td>21.2</td>
<td>7.8</td>
<td>13.0</td>
</tr>
</tbody>
</table>

In sum, therefore, the current evidence on concurrent validity suggests a complex relationship between uniTEST and Year 12 scores. There is clearly some degree of overlap between uniTEST and Year 12 scores. Most institution level relationships are not statistically significant, however, and those that are range between being negative and positive in direction. Overall, there appears to be a broadly divergent relationship between the measures, which suggests that they play a complementary role in the selection process.

Given that uniTEST has been constructed with three specific components – Verbal Reasoning, Quantitative Reasoning and Critical Reasoning – the potential exists for it to be used in different ways to identify candidates for certain disciplines and courses. Ideally in analyses of concurrent validity, results relating to certain school subject outcomes would be matched to specific components of this test to examine specific correlations in this regard. Unfortunately, data relating to individual subject Year 12 results was not available for this study and therefore analyses that examine, for example, mathematics Year 12 scores with the
uniTEST Quantitative Reasoning component are is not possible. A discipline focus in the proceeding predictive validity analysis is also not possible due to the small numbers of participants when disaggregated by field of education.

While not explored here, however, such discipline-specific use of uniTEST is potentially very important. It is possible that institutions could weight different components of uniTEST depending on the courses that an applicant is applying to, thus adding a discipline-specific edge to the existing broad nature of the test. Further research into this aspect of the test will be possible with greater uptake of the test.

Recommendation 3: It is recommended that further work be undertaken to examine the extent to which aptitude (as measured by uniTEST) complements Year 12 achievement. With greater numbers and more information on school outcomes, examination in relation to Year 12 score bands and individual Year 12 subject scores to be examined.

Analysis of predictive validity

Data and analytical considerations

Studies of predictive validity are notoriously difficult due to inherent complexities associated with data characteristics and the nature of the selection process itself. The current study was no different. As such it is vital to take account of the data collected and of several analytical considerations that were factored into the analysis.

A number of data elements were required for this analysis. As anticipated given diversity in selection processes and management, the collection and aggregation of these data elements was complex and messy. Hence a considerable amount of consultation and checking was required to ensure the integrity, comparability and veracity of secured data.

uniTEST and Year 12 scores were secured from participating institutions, along with further data on course characteristics and student demographics. Evaluating predictive validity involves identifying appropriate measures of student performance. A wide range of elements were considered but the selection narrowed due to the complexities associated with provision and the small size of the uniTEST sample. Operationally, it was feasible to collect data on GPA and students’ perceptions of engagement and skill development.

Unit record GPA data from each student’s first four semesters (i.e. their first two years) of university was obtained from institutions and merged with the uniTEST results. For the cohort of students who commenced in 2009, only GPA data for each student’s first two semesters (i.e. their first year) of university was used in the study. Hence it should be noted that there are a greater number of cases for analyses of first and second semester GPAs compared to third and fourth semester GPAs. While the most appropriate measure of individual performance, it must be noted that due to various course, individual and assessment characteristics, GPAs are likely to provide only a very partial and inconclusive estimate of performance (Coates, 2007b, 2008b). One reason for this is that university assessments are particularly unstable in the early semesters of study. Another reason is that GPAs, particularly in the first year of study, are often scaled to fit specified distributions. Further, GPAs are not commonly used in Australian higher education and the empirical properties are not well established. For instance, the GPAs used in the current study were not reported on a standard metric.

To limit problems that may arise during statistical estimation, GPA data was transformed onto a standardised metric with a mean of zero and standard deviation of one. Even given this standard metric, however, it is important to stress that different institutions’ GPAs are not equated. That is, a GPA of 1.5 at institution A is likely to reflect a different standard of student achievement than a GPA of 1.5 at institution B, or perhaps even in another field at institution A. Unfortunately, due to the lack of moderation or calibration processes in Australian higher education, it was not feasible to psychometrically equate student assessment data and combine GPAs into a common variable. An indirect but important finding of this evaluation is the need to develop a robust national GPA for Australian higher education.
Recommendation 4: Considerable value would be derived from developing a robust GPA for Australian higher education. Many GPA metrics already exist, but these are not well defined or validated, and are not implemented consistently.

The stage 1 evaluation brief included analysis of students’ persistence, course selection and change, skill development, perceptions, and engagement. Data on skill development, perceptions and student engagement, and course selection and change was collected using a survey administered in August 2008. This collection was aligned with ACER’s Australasian Survey of Student Engagement (AUSSE) (Coates, 2008c; ACER, 2008), which in 2008 involved 29 institutions. This data was available for three of the five institutions that took part in this stage of the SATTA study. In line with the AUSSE, ACER’s Student Engagement Questionnaire (SEQ) was deployed to collect this information. The SEQ provides measurement of six engagement scales: academic challenge, active learning, student and staff interactions, enriching educational experiences, supportive learning environment and work integrated learning. The SEQ also measures six outcomes scales: higher order thinking, general learning outcomes, general development outcomes, average overall grade, retention intention and overall satisfaction.

Analyses of predictive validity are complicated by the selective nature of competitive admissions processes. While selection decisions are based on a range of considerations, it is likely that individuals with higher uniTEST scores are more likely to be admitted than others. As a result, the range of uniTEST scores for successful applicants is narrower than for all candidates who took the test. Further, criterion measures are only available for those students who were admitted into university. Such range restrictions result in attenuated estimates of the relationships between uniTEST and the criterion variables, leading to lower estimates of relationship and hence predictive validity.

A further complexity inherent in studies of predictive validity is associated with the reliability of the criterion measures (in this case university GPAs). Typically, these measures have reliabilities which are lower than the selection measures, or which are unknown. As discussed above, the reliability of the first and second year assessments which provide the criterion measures in the current study is not known. This places an obvious constraint on the study given that these provide the basis for assessing performance in at university and hence the predictive validity of uniTEST. The use of these measures is appropriate, however, given that they are the only data available and that they are used widely by individuals and institutions.

A final complexity worth noting relevant to the predictive validation is the partial role played by uniTEST in the selection process. Graduates can be accepted into university via a range of processes using various bases of admission. As documented above, these can vary between and even within institutions. Any relationship between uniTEST and achievement scores such as exam results will be influenced by these other measures. Of course, actual student performance will also be influenced by a range of educational, individual and other contextual factors which may or may not have been relevant when they took uniTEST.

Figure 16 attempts to summarise the situation which arises as a result of range restrictions, reliability uncertainties, and selection process variations. The distribution of applicant uniTEST scores is shown on the horizontal axis. The distribution of criterion measures, first and second year marks, is shown on the vertical axis, which is wobbly to indicate the unknown measurement properties of this data. The ‘theoretical joint distribution’ reflects the bivariate distribution which would arise if all uniTEST candidates rather than just successful applicants had first year marks. The three ‘observed joint distributions’ reflect observed distributions for different institutions. The distributions occupy different space due to variations in cut-off scores and the range of first year marks. Analysis of the relationship between uniTEST and first and second year marks must take account of these data contexts.
It is critical that several adjustments are applied to account for the attenuation of correlations due to range restrictions and the uncertain reliability of the criterion measures. Specific adjustment formulae for managing estimate bias caused by these factors have been developed (Givner & Hynes, 1979; Hynes & Givner, 1981; Julian, 2005; Nunnally & Bernstein, 1994; Muchinsky, 1996; Raju & Brand, 2003). These have been applied in this study where possible given data characteristics and model specifications.

A range of multivariate analyses were used to determine the relationship of uniTEST to measures of university success, and hence the predictive validity of the instrument. These analyses: sought to adjust, where possible and relevant, for demographic and education characteristics which may otherwise bias results; adjust for problems arising from selection biases and the unknown reliabilities of criterion measures; and identify the predictive validity of uniTEST (acknowledging that institutions use uniTEST in different ways).

Analyses were undertaken to investigate the ‘incremental validity’ of uniTEST. Incremental validity refers to the increase in predictive power associated with a measure. In the present study, it involves analysis of whether the use of uniTEST adds value to the selection process over and above the use of Year 12 results. For practical reasons, comparisons were only be made against uniTEST and Year 12 results.

**Analysis of different selection algorithms**

Predictive validity was evaluated by examining the relationship between uniTEST scores and GPAs for the four completed semesters of study. Correlations between these variables are presented in Table 9 for each institution and for all students combined. Correlations between Year 12 scores and individual semester GPAs are also provided, these being based on both uniTEST and data for control-group students, hence larger sample sizes and greater statistical significance.

Table 9 shows that uniTEST scores seem to have a predictive relationship with performance at university. Overall, there are significant correlations between the uniTEST component scores and GPA for all but the Semester 3 results. The Quantitative Reasoning scale has the strongest correlation with first- and fourth-semester GPAs. Interestingly, the correlation between GPA and Year 12 marks were lower than the total uniTEST score correlations for all but the third-semester in the overall comparison across the institutions involved.

---

1 In general, the semester three results in the analyses undertaken for this report do not seem to fit the pattern that appear for semesters one, two and four. This is difficult to explain given similar processes and procedures of data collection were carried out for all four semesters. When discussing trends in the following analyses, in some instances the semester three outcomes are ignored.
Among the specific institution-level data, Institution C, in particular, showed a strong relationship between uniTEST scores and subsequent performance across all four semesters that students were tracked. Institutions A and D also showed a relationship between uniTEST performance and performance in the first two semesters of study, but this was not the case in the third and fourth semesters. Institution B has low and non-significant correlations across all semesters for both uniTEST and Year 12 outcomes. Institution E displays a similar pattern to Institution B although it does have one significant Year 12 correlation for semester one.

Table 9: uniTEST and Year 12 correlations with academic performance at university

<table>
<thead>
<tr>
<th>Institution</th>
<th>VR</th>
<th>QR</th>
<th>CR</th>
<th>TL</th>
<th>YR12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>30*</td>
<td>27*</td>
<td>26*</td>
<td>33*</td>
<td>37*</td>
</tr>
<tr>
<td>Semester 2</td>
<td>31*</td>
<td>22</td>
<td>32*</td>
<td>33*</td>
<td>37*</td>
</tr>
<tr>
<td>Semester 3</td>
<td>-14</td>
<td>5</td>
<td>-10</td>
<td>-4</td>
<td>23*</td>
</tr>
<tr>
<td>Semester 4</td>
<td>13</td>
<td>31</td>
<td>22</td>
<td>25</td>
<td>24*</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>12</td>
<td>9</td>
<td>-3</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Semester 2</td>
<td>-17</td>
<td>1</td>
<td>-4</td>
<td>-8</td>
<td>6</td>
</tr>
<tr>
<td>Semester 3</td>
<td>2</td>
<td>-8</td>
<td>-12</td>
<td>-6</td>
<td>12</td>
</tr>
<tr>
<td>Semester 4</td>
<td>10</td>
<td>-24</td>
<td>29*</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>32*</td>
<td>30*</td>
<td>41*</td>
<td>43*</td>
<td>34*</td>
</tr>
<tr>
<td>Semester 2</td>
<td>30*</td>
<td>21*</td>
<td>31*</td>
<td>36*</td>
<td>34*</td>
</tr>
<tr>
<td>Semester 3</td>
<td>41*</td>
<td>52*</td>
<td>54*</td>
<td>55*</td>
<td>36*</td>
</tr>
<tr>
<td>Semester 4</td>
<td>34</td>
<td>46*</td>
<td>51*</td>
<td>50*</td>
<td>38*</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>10</td>
<td>40*</td>
<td>26*</td>
<td>32*</td>
<td>34*</td>
</tr>
<tr>
<td>Semester 2</td>
<td>18*</td>
<td>26*</td>
<td>20*</td>
<td>27*</td>
<td>24*</td>
</tr>
<tr>
<td>Semester 3</td>
<td>14</td>
<td>-3</td>
<td>-1</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>Semester 4</td>
<td>-18</td>
<td>9</td>
<td>-12</td>
<td>-10</td>
<td>22</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>-10</td>
<td>12</td>
<td>-17</td>
<td>-8</td>
<td>38*</td>
</tr>
<tr>
<td>Semester 2</td>
<td>14</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>Semester 3</td>
<td>-4</td>
<td>21</td>
<td>0</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Semester 4</td>
<td>21</td>
<td>33</td>
<td>21</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semester 1</td>
<td>18*</td>
<td>29*</td>
<td>24*</td>
<td>27*</td>
<td>21*</td>
</tr>
<tr>
<td>Semester 2</td>
<td>16*</td>
<td>15*</td>
<td>16*</td>
<td>20*</td>
<td>19*</td>
</tr>
<tr>
<td>Semester 3</td>
<td>9</td>
<td>13</td>
<td>10</td>
<td>14</td>
<td>18*</td>
</tr>
<tr>
<td>Semester 4</td>
<td>21*</td>
<td>35*</td>
<td>23*</td>
<td>32*</td>
<td>18*</td>
</tr>
</tbody>
</table>

* p < 0.05

Correlational analysis is useful for providing first insights into predictive relationships, but regression modelling provides a more nuanced view. Figure 17 presents the explained variance ($R^2$) for four regression models involving combinations of uniTEST and Year 12 scores. Data from each semester of university studies is reported separately. Each model regresses one or more explanatory variable onto GPAs for the relevant semester. Further statistical details relating to the results are provided in Appendix 1.

Comparison of Model 1 and Model 2 in Figure 17 show that uniTEST performance acts as a greater predictor than Year 12 performance for semesters one, two and four. For example, the model shows that Year 12 results explained 4.4 per cent of the variance in GPAs in the regression model, while the uniTEST
total score explained 6 per cent. Comparisons of Model 2 and Model 3 show there is a very marginal increase in explained variance from using a combination of uniTEST components, instead of the uniTEST total score.

Across the four semesters, exploration of Year 12 marks (Model 1) show that this metric explains the more variance in GPA outcomes in the first semester, than in later semesters. uniTEST (Model 2 and Model 3) also has the highest explanatory power in semester one. However, for the total score (Model 2) the predictive power is similar in semester two as it is in semester four and for the component score (Model 3) the explained variance increases between semester two and semester four.

In Model 4, Year 12 scores and uniTEST scores are combined in a regression model. While adding Year 12 scores to uniTEST performance does not greatly impact the explained variance for first and third semester marks, the combined models account for a much higher proportion of the variation in university achievement in the second and fourth semesters. Therefore, from this analysis it appears that for the populations under study a combination of both measures offers a more powerful means of predicting university performance and in particular over the longer term, a combined selection criteria can have greater predictive power than one selection component on its own.

In summary, a combination of Year 12 score and uniTEST components explains up to six per cent of the variance in GPAs and generally explains more variance than either measure on its own. This is a major finding of the evaluation – one that affirms the value added by an aptitude test such as uniTEST.

![Figure 17: Explained variance in GPAs from combinations of uniTEST scores and Year 12 marks](image)

Figure 18 examines whether various demographic variables influence the amount of variance of GPA outcomes of students included in the study. Model 4 (Year 12 score and uniTEST components) acts as the starting point. Model 5 incorporates demographic variables sex, language and age. Over the four semesters, the explanatory power of these variables increases slightly. Of the sex, language and age measures, age tends to be the strongest. Language background does not make much impact in the first two semesters, but is more important in the third and fourth, while gender has less impact in all but semester two.

Model 6 incorporates socioeconomic status. While this factor appears to have a negative or null effect on explained variance for GPAs in semester one and two, it has a large effect on marks in semester three and
four. Model 7 incorporates field of education of the student, while Model 8 accounts for the differences across institutions. As expected these last two variables account for a large amount of variance across students. Both course and institution differences account for a large increase in explained variance for first-year students, while course differences account for a large increase in explained variance for second-year students.

Figure 18: Explained variance in GPAs from combinations of uniTEST scores and Year 12 marks

These results are important, because they show that uniTEST results alone explain more of the variance in university achievement than Year 12 scores. These models show that the range of demographic and other characteristics of students (such as field of education and institution of enrolment) contribute to a notable amount of the variance in student outcomes at university. This helps to highlight the difficulties that compound selection processes and analyses of predictive validity of selection criteria. Recent evaluations conducted in the UK (Cambridge Assessment, 2008) found similar results to those reported here. This work, which involved six institutions, 1,589 applicants and 345 admitted students, found that results varied across institutions and that relations between uniTEST and first-year performance were difficult to assess due to variation in admissions policies and the statistically small number of students being assessed.

However, given the results displayed in Figure 17 in this analysis that show that a combination of the uniTEST component scores and Year 12 results generally yields greater explanatory power than either as an individual measure, it still appears that there is strong evidence to suggest that the complementary nature of these two measures offers the best selection scenario.
Comparison of uniTEST entrants and the control group

It is of interest to know whether uniTEST and control group students – those who were not admitted using uniTEST – perform equally well at university. If so, then it would appear that uniTEST is able to identify students who are capable of succeeding at university who would otherwise not have had the opportunity.

The most basic comparison that can be made of uniTEST versus control group students is the rate at which students in either group were still participating in university at the end of first, second, third and fourth semester. Figure 19 provides this comparison. It is important to note that the uniTEST group in this context is not all individuals with uniTEST scores, but only those for who uniTEST played a role in admission to university. The measure of student participation is defined as whether a mark was recorded by the university. It is expected that the majority of those students who did not receive a mark did not complete any assessment, or dropped out of the semester. It would be expected, however, that this variable may also be comprised of students who have legitimate reasons for not receiving a mark – such as deferring for a semester or changing university. Hence the results should be interpreted with caution.

The value of uniTEST as a means of admitting students to university would be supported if a similar proportion of these students were able to successfully complete their university studies as the comparison group. Figure 19 shows that, in all four semesters of study control students were more likely to receive a mark than the uniTEST group. However, the difference between these groups reduced over time and apart from the first semester this difference was not statistically significant. Therefore, with the exception of a small difference in the first semester, the results suggest that students admitted based on uniTEST have been retained at university over the first four semesters at an equivalent rate to their peers.

Figure 19: Comparison of student participation over time for individuals accepted on the basis of their uniTEST performance and control students

Figure 20 shows mean GPA scores for each semester for the 547 uniTEST and the 833 control group students. Control data for institution D was only collected for semester one Overall the differences in GPA between the control and uniTEST groups range between 0.23 to 0.04 of a standard deviation. The point estimates are given with 95 per cent confidence intervals. Overlap between confidence bands indicates that there is no statistically significant difference between the uniTEST and control group GPA distributions. The bands overlap in all data for all semesters and all institutions suggesting that there is no statistically significant difference in the academic outcomes of uniTEST and other students once admitted to university.
Despite being matched on key demographics, differences in key characteristics of uniTEST and control groups means that a direct comparison of straight arithmetic (marginal) means may not be the most pertinent analytical approach. The application of uniTEST to students in academically marginal positions meant that it was difficult to match the two groups in terms of prior achievement. While the mean Year 12 score for control group students was 75.4, the mean score for the uniTEST student group was 74.3. However, once students who completed uniTEST as part of a ‘special-sitting’ session at their university were factored out, the mean Year 12 score for the uniTEST student group dropped to 70.0. The positive correlation between Year 12 and achievement suggests, therefore, that it would be appropriate to partial prior achievement out of the comparison between uniTEST and control group first-year GPAs.

Table 10 presents results from regression modelling that compare uniTEST and control group semester GPA means by prior performance. In this ANCOVA-type specification, control group students are coded 0 while the uniTEST group are coded 1. The standardised parameter estimates for the group variable are small for all groups, and only one institution reached statistical significance at the 5 per cent error rate for the first semester. There was no significance between the groups at the institutional or overall level for semesters two, three or four. This evidence further affirms that after taking account of prior achievement uniTEST and control group students appear to perform equally in their first four semesters at university.
Table 10: Control and uniTEST group standardised regression estimates

<table>
<thead>
<tr>
<th>Semester</th>
<th>Institution</th>
<th>Group</th>
<th>B</th>
<th>SE(B)</th>
<th>(b)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Year 12</td>
<td>0.04</td>
<td>0.01</td>
<td>0.39</td>
<td>7.02</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.16</td>
<td>0.12</td>
<td>0.07</td>
<td>1.34</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Year 12</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.15</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.50</td>
<td>0.21</td>
<td>-0.21</td>
<td>-2.40</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Year 12</td>
<td>0.04</td>
<td>0.01</td>
<td>0.36</td>
<td>6.81</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.22</td>
<td>0.12</td>
<td>0.10</td>
<td>1.80</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Year 12</td>
<td>0.03</td>
<td>0.01</td>
<td>0.34</td>
<td>6.57</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.13</td>
<td>0.11</td>
<td>-0.07</td>
<td>-1.26</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Year 12</td>
<td>0.05</td>
<td>0.02</td>
<td>0.38</td>
<td>2.73</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.00</td>
<td>0.21</td>
<td>7.42</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.09</td>
<td>0.06</td>
<td>-0.04</td>
<td>-1.40</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.01</td>
<td>0.24</td>
<td>3.34</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.16</td>
<td>0.17</td>
<td>0.07</td>
<td>0.98</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.02</td>
<td>0.12</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Year 12</td>
<td>0.04</td>
<td>0.01</td>
<td>0.37</td>
<td>6.76</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.19</td>
<td>0.17</td>
<td>0.06</td>
<td>1.14</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Year 12</td>
<td>0.03</td>
<td>0.02</td>
<td>0.18</td>
<td>1.18</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.03</td>
<td>0.30</td>
<td>-0.01</td>
<td>-0.10</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Year 12</td>
<td>0.01</td>
<td>0.00</td>
<td>0.19</td>
<td>5.82</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.05</td>
<td>0.07</td>
<td>-0.02</td>
<td>-0.71</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.01</td>
<td>0.24</td>
<td>3.34</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.16</td>
<td>0.17</td>
<td>0.07</td>
<td>0.98</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.02</td>
<td>0.12</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Year 12</td>
<td>0.04</td>
<td>0.01</td>
<td>0.37</td>
<td>6.76</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.19</td>
<td>0.17</td>
<td>0.06</td>
<td>1.14</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Year 12</td>
<td>0.03</td>
<td>0.02</td>
<td>0.18</td>
<td>1.18</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.03</td>
<td>0.30</td>
<td>-0.01</td>
<td>-0.10</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Year 12</td>
<td>0.01</td>
<td>0.00</td>
<td>0.18</td>
<td>4.71</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.05</td>
<td>0.09</td>
<td>0.02</td>
<td>0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.01</td>
<td>0.25</td>
<td>3.34</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.13</td>
<td>0.18</td>
<td>0.05</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Year 12</td>
<td>0.01</td>
<td>0.02</td>
<td>0.06</td>
<td>0.60</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>-0.03</td>
<td>0.23</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Year 12</td>
<td>0.04</td>
<td>0.01</td>
<td>0.39</td>
<td>7.10</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.22</td>
<td>0.16</td>
<td>0.07</td>
<td>1.31</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Year 12</td>
<td>0.02</td>
<td>0.03</td>
<td>0.11</td>
<td>0.71</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.77</td>
<td>0.34</td>
<td>0.34</td>
<td>2.30</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>Year 12</td>
<td>0.01</td>
<td>0.00</td>
<td>0.18</td>
<td>4.71</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group</td>
<td>0.08</td>
<td>0.09</td>
<td>0.04</td>
<td>0.91</td>
<td>0.37</td>
</tr>
</tbody>
</table>

In principle, analysis of how the AUSSE engagement and outcomes scales vary across uniTEST and control groups would provide a basis for exploring the extent of skill development that takes place in the first year of study. As noted, however, the number of responses is very small, limiting the analyses that can be conducted and hence the conclusions that might be drawn. Indicative results are considered, however, to outline what may be possible in a larger study.
Figure 21 presents control and uniTEST scores (scaled onto a metric that runs from 0 to 100) for the six AUSSE engagement scales at institutions A, B and C. The results are shown with 95 per cent confidence intervals which for the uniTEST group are quite large due to the small number of observations. While there are variations across groups and institutions, none of the comparisons are statistically significant. The evidence base is very small, but this suggests that the engagement of uniTEST students in effective learning practices is on par with that of the control group.

![Figure 21: Control and uniTEST group engagement scale scores](image)

Figure 22 presents control and uniTEST scores (scaled onto a metric that runs from 0 to 100) for the six AUSSE outcomes scales at institutions A, B and C. The results are shown with 95 per cent confidence intervals which for the uniTEST group are quite large due to the small number of observations. These results, as per the engagement scales, suggest that the outcomes for uniTEST and control group students are similar.

![Figure 22: Control and uniTEST group outcomes scale scores](image)
Figure 22: Control and uniTEST group outcome scale scores

In general, therefore, results from analysis of uniTEST and control group students suggest that uniTEST was able to facilitate the admission to university of students who otherwise would not have received a place, and that these students engage with study to the same extent as other students, and that key outcomes may be roughly similar. While this finding must be hedged by the technical caveats that surround the current study, the results suggest that uniTEST can play a significant role in identifying individuals who have the potential to succeed at university, and enabling these people to be included in the system.

Chapter summary

This chapter has investigated the criterion validity of uniTEST to explore the role it plays in university selection processes, how it relates to other selection methods, and its capacity to predict success in the first two years of study.

The evaluation established and confirmed a cross-institutional methodology for undertaking such an evaluation in Australian higher education, and identified many of the inherent complexities that are involved. This is a notable outcome of the study in itself, because it is important that such work is undertaken on a routine basis for the purposes of evidence-based quality improvement, both to inform instrument development and institutional selection processes.

The current evaluation was limited in scope by the small number of institutions and individuals involved, and by the complex and varied nature of each institution’s student cohorts and selection approaches. All findings must be read with this context in mind. They cannot, for instance, be generalised to the broader population of Australian students. Nonetheless, the study did produce indicative findings which provide insight into various relationships between uniTEST, Year 12 and university marks.

The current evidence on concurrent validity suggests a complex relationship between uniTEST and Year 12 scores. Most of these relationships are not statistically significant with the exception of those that pertain to the Quantitative Reasoning component. However the average shared variance between the measures is low. Overall, there appears to be a broadly divergent relationship between the measures, which suggest that they play a complementary role in the selection process.

From the analysis of predictive validity, which are based on the small amount of data available for the specific subgroups being analysed, the results suggest that uniTEST results alone explain more variation in university GPAs than do Year 12 scores. From analysis of different combinations of uniTEST and Year 12 scores, it appears that for the populations under examination a combination of both measures offers a more powerful means of predicting university performance. This is a major finding that affirms the value that an aptitude test can add to university admissions.

Results from analysis of uniTEST and control group students suggest that uniTEST was able to facilitate the admission to university of students who otherwise would not have received a place, and that these students performed on par with their counterparts who gained entry through other means, most commonly through Year 12 scores. While the evidence is limited, both uniTEST and control group students appeared to report similar levels of academic outcomes and engagement, as well as learning and skill development. These findings are constrained by the caveats that surround the current study, but there is a significant role that uniTEST can play in identifying individuals who have the potential to succeed at university, and enabling these people to be included in the system.

By way of summary, empirical indications here highlight a positive role can be played by uniTEST and that this test has the potential to identify ‘latent talent’ and facilitate the inclusion of able individuals in the system. But the empirical indications also show that many of the results are statistically inconclusive, due largely to the small and idiosyncratic nature of the available sample. Given the growing importance of assuring the validity of this assessment, there is an evident need for both a larger and an ongoing study. Ideally, the analysis of concurrent and predictive relationships should be woven into continuous quality improvement processes that underpin routine reflective practice.
Recommendation 5: Predictive validity analyses demonstrate that aptitude test data adds to the power of admissions processes. To enhance the prognostic power of admissions processes, data on school achievement should be augmented with evidence from uniTEST.
5 BOLSTERING GROWTH IN APTITUDE ASSESSMENT

Taking stock on the added value

Together, insights in previous chapters suggest that evidence from an aptitude assessment provides a transparent and useful foundation for university admission in Australia. Aptitude tests are widely used around the world, both as a primary data source as well as a means of complementing information on school-level achievement. There are many principled rationales for expanding the use of aptitude assessments with school leavers in Australia, not least that it provides a means of coordinating advances in national practice.

The current results suggest that uniTEST has been able to facilitate the admission to university of at least 165 people who might otherwise have not had the opportunity to participate. Scores appear to be particularly helpful for students from historically underrepresented backgrounds, and have been shown to be less influenced by important characteristics like socioeconomic status. uniTEST scores complement rather than replace Year 12 achievement scores. Further evidence of the value they add is that, used in combination with achievement scores, they provide an improved predictor of GPA over the first two years of university. It appears that uniTEST offers an independent and complementary data source for improving university admissions in Australia.

Recommendation 6: Based on evidence from the SATTA pilot it is recommended that uniTEST be implemented as a means of diversifying and complementing the data factored into the university admissions of school leavers in Australia.

Given this, this chapter examines how an aptitude assessment – specifically uniTEST – might be incorporated into university admissions process, and the likely benefits that would result. It begins by reviewing broad contexts of relevance to the national implementation of an aptitude assessment, and continues through analysis of specific strategies. It concludes by documenting what would appear to be the most significant next steps.

A new admissions architecture

As suggested throughout this report, a richer information architecture is required to make university selection in Australia a more informed and sophisticated process. Aptitude assessment is an important element in this architecture. It is possible to imagine a scenario in which aptitude assessment is routinely undertaken on a voluntary basis by school students towards the end of their senior secondary study. The timing point would offer careers advisors information with which to guide students and would provide students themselves with objective insights into their strengths.

The opportunity now exists across Australia to develop new approaches to university selection that offer simplicity, consistency and transparency for prospective students and that maintain the benefits of coordinated processes for application and selection for the majority of places. For this to occur, the university sector must undertake the consultation necessary for a common university application framework and for a new range of agreed selection criteria.

One of the principal objectives for introducing a common and agreed approach to the use of aptitude assessment is to expand and diversify the information available for university recruitment and selection. A broader information architecture will be one precursor for building a stronger, more responsive, more accessible and more diverse university education sector.

As proposed in chapter two, aptitude assessment has the potential to provide valuable new information for student advising, recruitment and selection. Importantly, as results from the concurrent validity analyses show, information from aptitude assessment is distinctive and is not associated with school achievement. Hence aptitude assessment opens up the potential for a wider set of recruitment and selection criteria. It will allow prospective students to demonstrate abilities and areas of skill not commonly identifiable through ranked aggregate measures of school achievement. A richer information base for selection and recruitment
into tertiary education would be likely to support greater diversity in institutional practices and widen the opportunities open to students to progress to tertiary education.

Achieving these goals requires considered upscaling of implementation. The principal purpose of aptitude assessment should be to identify the potential or disposition for particular fields of study. For some students, aptitude assessment will identify potential in students that the school system has not identified and in this regard is likely to make a modest but discernible contribution to equity. Aptitude assessment will also open up wider opportunities for mature-age students who are presently somewhat disadvantaged by not having recent school achievement results.

Aptitude assessment should not simply be utilised for the purposes of competition, but used to demonstrate a positive rationale for the selection of students for particular programs of study. It will be important to create a context in which aptitude assessment is primarily diagnostic and informative rather than summative and judgemental. There are of course fine balances to be achieved here, for these purposes are rarely entirely separable in practice.

Establishing new selection criteria and practices is now a high priority. The Review of Australian Higher Education (Bradley, Noonan, Nugent & Scales, 2008) and the subsequent Australian Government higher education policy (Australian Government, 2009) has set Australia on a course towards universal higher education participation, a higher education system in which it is possible that more than half the population will go to university at some stage in their lives.

A universal participation system requires reconsideration and renewal of many aspects, including entry pathways, curricula, patterns of student participation and course delivery, and the structure of the tertiary sector itself. A more highly differentiated tertiary sector is likely to be shaped in the next decade. New types of public and private tertiary institutions may emerge that bridge universities, VET institutions and secondary schools. Private providers will grow in number. The boundaries between higher education and vocational education and training are likely to be increasingly blurred. New forms of bridging awards at the secondary-tertiary interface may also emerge. The nexus between undergraduate and graduate studies and research will also change.

Overall, entry to tertiary education will be more ‘open’. For some institutions and courses the concept of selection will have reduced meaning, for pathways can be foreseen in which partner institutions offer student guarantees for seamless student transitions that offer curriculum coherence and continuity.

In the likely tertiary education context of the future it is highly advantageous for the sake of clarity and transparency to maintain common frameworks for application and selection as far as possible — the alternative — processes administered by individual institutions — is potentially costly, complex and confusing for prospective students. A transparent information architecture offers relative simplicity from the applicant’s point of view, transparency and standards.

Aptitude assessment should be located within agreed student selection and recruitment frameworks alongside other common selection criteria, which would include the Australian Tertiary Admission Rank. The distinctive role of aptitude assessment would be to provide a valid and reliable measure for particular fields of study. Aptitude assessment would not replace or preclude the need for interviews and portfolios, these will continue to play an important role for particular courses. It will be desirable to establish protocols for quality assurance for the use of interviews and portfolios in the interests of transparency, fairness and equity.

Institutions must offer highly transparent statements of the precise ways in which they use various selection criteria for particular courses. This is vital as the tertiary system diversifies and pathways and criteria multiply. A ‘one-stop’ guide that offers precise information on the criteria and application and selection processes for all courses is desirable. In addition, institutions might be expected to make their selection criteria explicit on their websites as is already commonplace.

Given that institutions have autonomy over selection, and this is paramount, it is not possible to predict with precision how institutions might use information from aptitude assessment alongside other information for
particular courses. Hypothetical possibilities can be sketched, however, to illustrate the diversity that is possible:

- 50 per cent of places are allocated on the basis of ATAR and 50 per cent allocated on aptitude assessment for a number of its courses. This strategy is designed to diversify the student intake and to offer prospective students alternative ways of demonstrating their suitability for admission.
- 90 per cent of places are allocated on ATAR and 10 per cent on results from aptitude assessment plus Statements of Recommendation.
- Applicants must achieve a threshold band on the aptitude assessment for the field of study for which they are applying, following which selection is based on ATAR. This strategy is used to ensure suitability for particular careers.
- Applicants must achieve a threshold ATAR, following which selection is based on aptitude assessment. This strategy is also used to ensure suitability for particular careers.
- Students must achieve a threshold band score on aptitude assessment in order to be later judged on the basis of their design portfolio. This approach places no emphasis on school achievement but looks instead to demonstrated talent in creative areas.
- Students with an ATAR above 80 or an aptitude assessment above a certain band gain entry to a ballot for the allocation of places. A strategy such as this might be used for high demand courses in which ATAR and aptitude assessment results are both known to be predictors of academic success in tertiary education.
- Students with an ATAR above 80 are admitted. Students with ATAR in the 70-80 region may be selected on the basis of aptitude assessment.
- Applicants are selected on the basis of ATAR set at a certain level and an aptitude test score that has been weighted to reflect the course in which the candidate has applied (for example a science applicant might have the quantitative reasoning section double-weighted).
- Entry is entirely based on portfolios.

This brief sketch does not seek to trivialise the complexities inherent in selection – some reflect current approaches – but to illustrate the permutations that might emerge. Institutions might choose to use aptitude assessment as the basis for ‘conditional’ offers to students prior to Year 12 completion. For example, institutions might make conditional offers to students from partner schools towards the end of Year 11, based on aptitude assessment. Students might be then required to successfully complete their VCE studies to pass level in order to confirm their place.

**An implementation approach**

Since its inception in Australia, uniTEST has been administered by the universities who have chosen to consider uniTEST results as part of their selection criteria. This has meant that the registration of candidates has been conducted by ACER or the user university, and all arrangements for the test venues and invigilators have been the responsibility of the user universities. The outcome of this is that testing has principally been conducted only at the user universities and on a single day at each university, although for 2010 entry ACER instituted additional ACER managed test sessions.

To manage the test sessions ACER developed and distributed manuals for the management of uniTEST sittings. Test papers and response sheets for the recording of answers were printed by ACER and couriered to a nominated person at each centre, where they were stored securely prior to the test day. Following testing, the response sheets and all used and unused test papers are couriered to ACER for scoring. Both candidates and the user universities receive uniTEST results following their scoring and analysis.

While this system has allowed for flexibility, it is burdensome for institutions and does not allow easy access to the test for interested candidates. The most recent sittings of uniTEST have highlighted just how cumbersome and time consuming the process is if uniTEST is to be used for multiple applications.

To address the above issues ACER has investigated a number of delivery options and has formulated a streamlined process that should allow for flexible delivery and easier access to uniTEST. ACER’s extensive
test development and test management experience ensures that it has the knowledge, capacity and experience to successfully manage the transition to a new delivery model.

This section considers characteristics of a feasible cross-institutional approach for implementing an aptitude test for use with school leavers seeking admission to Australian higher education. The approach has been formed by drawing on ACER’s extensive experience administering aptitude tests in higher education both in Australia and abroad, consultation with peak bodies and experts, participation in round table meetings, analysis of several options, consultation with computer-based or online test delivery providers, and the reviews of international, national and institutional practice given in this report. The approach advanced here is not explored in sufficient detail to enable immediate implementation. Rather, the following analysis documents broad characteristics that have arisen during the current evaluation.

The core principles underpinning the approach are that:

- candidates are able to sit uniTEST during their senior secondary study;
- candidates are only able to sit uniTEST once in any six month period;
- uniTEST will be a computer-based test;
- a detailed list of test venues and test dates will be made available well in advance to candidates, with venues including schools, universities and, if an independent testing organisation is contracted to manage the testing, their secure test venues will also be utilised;
- the cost for the first sitting of uniTEST is the responsibility of the federal and/or state government/s, while the cost for subsequent sittings is to be borne by the candidate;
- should the Commonwealth Heads of Government (COAG) proposed ‘national student identifier’ be instituted it will be used to monitor uniTEST candidates;
- uniTEST registrations and reporting of results on a national scale will be managed by ACER, or its appointed secure testing company; and
- uniTEST results will be provided to candidates and delivered directly to the TACs, unless otherwise required, for inclusion in the various course algorithms.

The consultations highlighted that the main options for administering the aptitude assessment include:

- around the end of Year 10 or start of Year 11;
- during Year 12, likely during the middle of the year;
- after the completion of secondary studies; or
- on a fixed date with varying applicant cohorts.

There are pros and cons with each approach. After detailed consideration of educational and practical factors, ACER recommends the option listed first – that the assessment be conducted around the end of Year 10 or start of Year 11. The assessment should be offered to students on a voluntary basis.

There appear to be several rationales for conducting the assessment at the end of Year 10 or start of Year 11:

- students who are ‘at risk’ of leaving school before completing Years 11 or 12 and who are intellectually capable of success at university can receive diagnostic information that encourages retention through the senior secondary years – it enhances their educational literacy;
- the process extends the time available to individuals, families and institutions for making decisions about course choice, thereby adding breathing space to what is currently a very tight process;
- data can be made available to universities via a carefully designed means to assist them identify able students or offer additional supports and advanced placements; and
- examination and admissions agencies might access data to assist with calibration and validation activities.

While it is proposed that the assessment is administered in specific window of time, it would be preferable if the assessment was administered at a time of an individual or school’s choosing. As an alternative, testing could be conducted at several point throughout the year. This stagewise approach carries the advantage of
enabling students to sit the test when they feel most prepared and able to receive maximum diagnostic return. If they chose to do so, students could resit the assessment after a nominated period of time.

It follows from the above remarks that the assessment should have the capacity to enhance individual decision-making and advising by providing information on people’s capacity to succeed at university.

It is preferable that the assessment is deployed online. The following outline provides an opportunity for Australian governments and tertiary institutions to easily employ the use of uniTEST for applicants to their respective institutions. It envisages a structure that offers scalability as the use of uniTEST increases. It also allows for a candidate and institution friendly approach to the collection of information about the capacity of students to succeed in the tertiary education sector, and recognises Australia’s capacity to embrace sound assessment principles and utilise current technology. Online administration supports flexible administration, enables the use of a greater number of item formats, and facilitates efficient individual and system-level reporting of results. Hence, it is proposed that:

- from June 2010 uniTEST will be available as a computer-based assessment;
- during 2010 and 2011 uniTEST will also be available as a paper-based assessment (paper-based testing will only be available where computer-based testing is not feasible);
- uniTEST will be available by provision of secure USBs or remote desktop technology – ACER has explored a variety of secure computer-based testing delivery options and is confident that uniTEST can potentially be delivered in this manner; and
- ACER or the relevant education departments will be responsible for organising test centre supervisors and invigilators, and by 2012 all uniTEST Chief Supervisors must be accredited by ACER via online training and assessment modules.

The assessment should be promoted by key agencies as a credible alternative quantitative selection criteria to achievement tests. The test will vary in its relevance to institutions depending on factors such as selectivity, course characteristics and demographics, but it should be endorsed universally.

Management of the philosophical shift in the approach to university admissions needs to include a number of stakeholders. ACER is just one of those stakeholders. Others include federal and state governments, school systems, admissions and assessment agencies, universities, and private and public vocational and technical institutions. How to harness this collective group to embrace the use of uniTEST is the challenge ahead.

It is important that key stakeholders – particularly universities and TACs – need to develop ownership over the aptitude assessment. While coordinated by TACs, admissions processes are largely deregulated, even within institutions, and building this sense of ownership will require a considerable amount of consultation and technical development.

As with current admissions practices, the process should be centrally coordinated – though not regulated – in a way that sustains institutional autonomy over selection decisions. A collective approach reduces inefficiencies arising from overlapping processes and, importantly, confusion with applicants. It also provides a forum for institutions to develop and benchmark their approaches within a collaborative structure. Hence people who have undertaken aptitude assessment would need to indicate if they wish test results to be made available to institutions via the TACs. Similarly, institutions might request aptitude assessment information from TACs.

The implementation approach must be cross-institutional in nature. While institutions shape selection processes that suit their unique missions, the proliferation of testing on an individual institution basis is to be avoided. The term ‘cross-institutional’ is used deliberately, for individual institutions are responsible for admissions, and a single ‘national’ approach cannot be mandated. Of course, there are often compelling reasons why institutions choose to adopt a mechanism to such an extent that it may be considered a ‘national standard’. It may be considered, for instance, that adopting a uniform approach to assessing aptitude is considered useful because it:
enhances the transparency of the admissions process for students, which is vital in times of increasing diversification;
provides a further calibrated and streamlined source of data for admissions agencies;
supplies a sound metric against which institutions can monitor and benchmark their performance;
facilitates economies of scale in the production and administration of the assessment; and
provides a sound foundation for validation and evaluation activities conducted for the purpose of continuous improvement.

Recommendation 7: uniTEST should be implemented in a nationally coordinated way that is flexible, targeted at senior secondary students, and able to provide diagnostic information for both individuals and multiple institutions.

Next steps
The major proposition advanced by this report is that significant value would be derived from implementing a common aptitude assessment for school students seeking admission to Australian higher education. This proposition forms this evaluation’s major recommendation given above. This is a bold recommendation, yet it is one that, if analyses given in this report are accepted, is likely to add significant value to university admissions in Australia.

Aptitude tests are already commonly used in Australian higher education. A relatively large number of tests have been implemented for specific demographic and educational contexts. Considered in this regard, many of the analyses delivered in this report might be considered informative but relatively trivial from a policy perspective.

To date, however, the uptake of uniTEST in Australia has been limited. This in part is confirmation of the reality that the adoption of changed education processes is guarded and generally embraced in a measured fashion. Nevertheless confidence in the use of Year 12 achievement data as the almost sole criteria for tertiary admission is waning. In this climate it is expected that the use of an aptitude test to complement academic results is more likely to be given credence. It is also believed that a system that facilitates the use of uniTEST, such as that described above, will assist with increasing the engagement and use.

On reflection, there would appear to be three alternative ways in which Australia could advance the use of an aptitude test as part of university admissions. The first involves a continuation of the past approach. As described earlier in this report, this involves ACER working in a somewhat exploratory fashion with individual institutions. This approach has facilitated access to the system for students who may not otherwise have had the chance to participate. Importantly, it may have provided an independent and complementary metric that is less influenced by an applicant’s socioeconomic background. This approach appears to be limited, however, inasmuch as it is reactive to specific institution’s inclinations and is not promoting cross-institutional synergies and the most widespread distribution of the benefits that an aptitude assessment can provide.

The second approach – that recommended above – involves changing the focus and scope of the implementation. This is doubtless considered to be a more significant development but, in light of the evidence given in this report, would appear to generate the most advantages for students, institutions and hence the country as a whole. Moving the assessment into schools would appear to carry advantages for enhancing student aspirations, informing subject and tertiary choice, possibly enhancing persistence, and providing a sound and complementary quantitative means of diversifying and perhaps compensating for the evidence used to admit school students into university. This approach is cross-institutional in scope, which promotes synergies and efficiencies not realised by implementations confined to a limited number of institutions.

The third option for future implementation involves factoring the implementation of an aptitude assessment into a much broader conversation about tertiary admissions. The reflection, consultation and evidence offered in this report highlights the significant dividends that may be yielded from this process. Admissions processes are a fundamental facet of university education in Australia, yet perhaps one of the least well
researched and discussed. The private and competitive nature of the process may partly explain this state of affairs, yet it does not lessen the need for improvements that bring practice into line with contemporary system contexts and needs.

In many respects, as this report has charted, an aptitude test provides a critical agent to prompt this broader process. Change on this scale needs to be designed and implemented in a considered and highly consultative way. One of the first steps, for instance, would be to form a advisory group to oversee the review of university admissions and the ongoing implementation uniTEST. There would also appear to be considerable value in holding a national summit on tertiary admissions. This meeting would bring together stakeholders to work through the complexities associated with national and institutional implementation. Issues might include, for instance, managing coaching, sustaining institutional engagement, assisting institutions with assessment decisions, how best to tune aptitude data with specific fields, the nature and extent of policy engagement that is needed, and links between admissions processes and quality assurance.

Of course, the third option given here may well emerge from the second, which has itself grown from the first. The third option does not necessarily (or at all) imply a radical revisioning of university admissions for school leavers in Australia. What it does advance is the need for ongoing research and development of this significant facet of Australian higher education. Indeed, this is the stance that underpins the ethos, approach and insights of this study, and which is imperative for ensuring that all school students who wish to study at university are able to demonstrate that they have the capacity to succeed.
REFERENCES


## APPENDIX 1: REGRESSION COEFFICIENTS DETAIL

### Table 11: Regression coefficients for GPA scores, Models 1 to 5, Semester 1 results

<table>
<thead>
<tr>
<th>Institution</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YR12</td>
<td>R²</td>
<td>TL</td>
<td>R²</td>
<td>VR</td>
</tr>
<tr>
<td>A</td>
<td>0.368</td>
<td>0.136</td>
<td>0.277</td>
<td>0.077</td>
<td>0.204</td>
</tr>
<tr>
<td>B</td>
<td>0.023</td>
<td>0.001</td>
<td>0.072</td>
<td>0.005</td>
<td>0.205</td>
</tr>
<tr>
<td>C</td>
<td>0.337</td>
<td>0.113</td>
<td>0.361</td>
<td>0.130</td>
<td>0.081</td>
</tr>
<tr>
<td>D</td>
<td>0.340</td>
<td>0.116</td>
<td>0.267</td>
<td>0.071</td>
<td>-0.165</td>
</tr>
<tr>
<td>E</td>
<td>0.377</td>
<td>0.142</td>
<td>-0.063</td>
<td>0.004</td>
<td>0.070</td>
</tr>
<tr>
<td>Combined</td>
<td>0.210</td>
<td>0.044</td>
<td>0.245</td>
<td>0.060</td>
<td>0.006</td>
</tr>
</tbody>
</table>

### Table 12: Regression coefficients for GPA scores, Models 1 to 5, Semester 2 results

<table>
<thead>
<tr>
<th>Institution</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YR12</td>
<td>R²</td>
<td>TL</td>
<td>R²</td>
<td>VR</td>
</tr>
<tr>
<td>A</td>
<td>0.369</td>
<td>0.136</td>
<td>0.165</td>
<td>0.027</td>
<td>0.243</td>
</tr>
<tr>
<td>B</td>
<td>0.058</td>
<td>0.003</td>
<td>-0.066</td>
<td>0.004</td>
<td>-0.267</td>
</tr>
<tr>
<td>C</td>
<td>0.341</td>
<td>0.116</td>
<td>0.300</td>
<td>0.090</td>
<td>0.175</td>
</tr>
<tr>
<td>D</td>
<td>0.239</td>
<td>0.057</td>
<td>0.226</td>
<td>0.051</td>
<td>0.064</td>
</tr>
<tr>
<td>E</td>
<td>0.241</td>
<td>0.058</td>
<td>0.114</td>
<td>0.013</td>
<td>0.129</td>
</tr>
<tr>
<td>Combined</td>
<td>0.190</td>
<td>0.036</td>
<td>0.200</td>
<td>0.040</td>
<td>0.094</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL</td>
<td>YR12</td>
</tr>
<tr>
<td>A</td>
<td>0.051</td>
</tr>
<tr>
<td>B</td>
<td>0.036</td>
</tr>
<tr>
<td>C</td>
<td>0.282</td>
</tr>
<tr>
<td>D</td>
<td>0.127</td>
</tr>
<tr>
<td>E</td>
<td>0.119</td>
</tr>
<tr>
<td>Combined</td>
<td>0.126</td>
</tr>
</tbody>
</table>
### Table 13: Regression coefficients for GPA scores, Models 1 to 5, Semester 3 results

<table>
<thead>
<tr>
<th>Institution</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YR12</td>
<td>TL</td>
<td>R²</td>
</tr>
<tr>
<td>A</td>
<td>0.230</td>
<td>-0.032</td>
<td>0.001</td>
</tr>
<tr>
<td>B</td>
<td>0.120</td>
<td>-0.053</td>
<td>0.003</td>
</tr>
<tr>
<td>C</td>
<td>0.364</td>
<td>0.461</td>
<td>0.213</td>
</tr>
<tr>
<td>D</td>
<td>0.217</td>
<td>0.028</td>
<td>0.001</td>
</tr>
<tr>
<td>E</td>
<td>0.180</td>
<td>0.044</td>
<td>0.002</td>
</tr>
<tr>
<td>Combined</td>
<td><strong>0.177</strong></td>
<td>0.115</td>
<td>0.013</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>YR12</td>
<td>R²</td>
</tr>
<tr>
<td>A</td>
<td>-0.123</td>
</tr>
<tr>
<td>B</td>
<td>0.013</td>
</tr>
<tr>
<td>C</td>
<td>0.462</td>
</tr>
<tr>
<td>D</td>
<td>-0.167</td>
</tr>
<tr>
<td>E</td>
<td>0.061</td>
</tr>
<tr>
<td>Combined</td>
<td>0.070</td>
</tr>
</tbody>
</table>

### Table 14: Regression coefficients for GPA scores, Models 1 to 5, Semester 4 results

<table>
<thead>
<tr>
<th>Institution</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YR12</td>
<td>TL</td>
<td>R²</td>
</tr>
<tr>
<td>A</td>
<td>0.244</td>
<td>0.211</td>
<td>0.045</td>
</tr>
<tr>
<td>B</td>
<td>0.063</td>
<td>0.046</td>
<td>0.002</td>
</tr>
<tr>
<td>C</td>
<td>0.377</td>
<td>0.419</td>
<td>0.176</td>
</tr>
<tr>
<td>D</td>
<td>0.224</td>
<td>-0.079</td>
<td>0.006</td>
</tr>
<tr>
<td>E</td>
<td>0.138</td>
<td>0.270</td>
<td>0.073</td>
</tr>
<tr>
<td>Combined</td>
<td><strong>0.180</strong></td>
<td>0.197</td>
<td>0.039</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>YR12</td>
<td>R²</td>
</tr>
<tr>
<td>A</td>
<td>0.264</td>
</tr>
<tr>
<td>B</td>
<td>0.095</td>
</tr>
<tr>
<td>C</td>
<td><strong>0.424</strong></td>
</tr>
<tr>
<td>D</td>
<td>-0.312</td>
</tr>
<tr>
<td>E</td>
<td>0.276</td>
</tr>
<tr>
<td>Combined</td>
<td>0.187</td>
</tr>
</tbody>
</table>