Graduate Skills Assessment

Stage One Validity Study

Sam Hambur
Ken Rowe
Le Tu Luc
Australian Council for Educational Research

03/02

Evaluations and Investigations Programme

ISBN 0 642 77295 9 (Internet copy)

DETYA No. 6911.HERC02A

This report is funded under the Evaluations and Investigations Programme of the Department of Education, Science and Training.

Acknowledgments

The authors would like to thank the following people for their assistance and advice in this work. Don Anderson for providing advice, Adele Butler for co-ordination with the GSA testing project, Jennifer Bryce for conducting interviews with stakeholders, Neville Chiavaroli for collecting data from universities on student TER and GPA, Margaret Forster, Deirdre Jackson and Christian Monseur for reviewing aspects of the report, and Margaret McGregor for dealing with issues related to Writing. In addition, the authors thank Susan Bates for formatting the report, Kay Waters for proofreading, ACER staff involved in the development and publishing of the Graduate Skills Assessment test, staff at universities who provided information and those who evaluated various aspects of the test and its construct, including Alison Brown, Barry Fallon, Laurance Splitter and Margaret Wu.

Contents

Tables

Figures

Executive summary

The Graduate Skills Assessment (GSA) test has been designed to assess a set of valued and widely applicable generic skills that may be developed through the university experience, and which are relevant to university achievement and graduate work.

This GSA validity study was commissioned by the Commonwealth Department of Education, Training and Youth Affairs (now known as the Department of Education, Science and Training).

The study addresses the validity of the first two Graduate Skills Assessment (GSA) tests, GSA Exit 2000 and GSA Entry 2001 (Stage One). A total of 3663 students drawn from nine broad fields of study across 27 Australian universities were involved in one or other of these tests.

This summary provides:

- a description of the aims and scope of the study;
- key background information;
- a consideration of the study sample;
- key findings and conclusions;
- some recommendations for the future; and
- concluding remarks that reflect on the ongoing challenges of the GSA.

Aims and scope (Chapter 1)

The validity study has the following major aims:

1 To investigate the dimensional factor structure (discriminant validity) of the test;

2 To identify variables related to differential performance on the GSA;

3 To investigate the relationship between student performance on the GSA and other measures of student achievement;

4 To consider the suitability of current reference ranges; and

5 To evaluate the face/content validity of the GSA construct and items.

Background (Chapters 1, 2 and 3)

1 The GSA grew out of an increasing interest in generic skills related to the need for an adaptable workforce in modern economies. Both employers and universities have an interest in generic skills, though they do not necessarily value the same skills equally.

2 The GSA is based on an assumption that certain generic skills, though taught within a particular context, can be transferred to another context once there is sufficient familiarity with that context. It is expected that those with the highest levels of generic skills make such transfers most readily.

3 The skill domains chosen for assessment by the initial GSA are: Written Communication, Critical Thinking, Problem Solving and Interpersonal Understandings.

4 During test development, the focus of each domain was narrowed in a way that was expected to produce a test component that assesses a psychometrically coherent generic skill dimension.

5 The focus of the GSA is on cognitive skills since these are more amenable to assessment. The test does not assess directly those personality traits that may be related to putting into action the relevant skills/understandings. It is hoped that longitudinal studies will indicate an association between the skills/understandings and outcomes.

6 Each component of the GSA aims to present tasks that are generally meaningful, accessible and contextually appropriate, so that specialised knowledge is not required. Whereas Year 12 literacy and Year 9 numeracy is assumed, higher-level meta-strategic and meta-cognitive skills need to be applied.

7 The Critical Thinking (CT) component of GSA aims to assess some markers of the ability to think critically about viewpoints and arguments. Students are expected to use comprehension, analysis and synthesis to assimilate and evaluate viewpoints and arguments. Partly to distinguish CT psychometrically from Problem Solving, material is presented in text format.

8 The Problem Solving (PS) component of GSA aims to assess some markers of the ability to analyse and transform information as a basis for making decisions and progressing toward the solution of practical problems. Students are expected to show insight into the problem to identify and deal logically with key information. Analytical, logical and quantitative reasoning need to be applied. Partly to distinguish PS psychometrically from CT, the information is presented in low verbal and non-verbal formats.

9 The Interpersonal Understandings (IP) component aims to assess the ability of students to show insight into the feelings, motivation and behaviour of others, and into approaches related to helping or working with others, such as effective feedback and teamwork. The information is mostly presented as text but some pictorial material is used.

10 The Written Communication component aims to assess the ability of students to write effectively in two genres: Argument (ARG) and Report (REP). The Argument task requires students to develop a point of view about an issue and structure a clear, coherent and logical argument in support of that view. The Report task requires students to comprehend, select, organise and present clearly a summary report based on facts, figures and pictures presented in the stimulus.

Study sample (Chapter 4)

1 A total of 3663 students drawn from nine broad fields of study across 27 Australian universities were involved in the first two GSA tests.

2 Since the sample of students sitting GSA was largely self-selected, it is unlikely to be representative of the general university population. This is confirmed by observations of significant differences between the GSA and general university populations in terms of variables such as field of study composition and the proportion of students with English-speaking background.

3 Because it is unclear how the deviations of the GSA population from the composition of the general university population will affect the results of this study, particular caution needs to be taken in drawing conclusions.

4 It is expected that statistical methods based on linear relationships (such as correlation and linear regression) would not be greatly affected by the non-representativeness of the sample. Therefore, it is expected that general findings related to the factor structure of the test (Chapter 5), variables related to performance on the test (Chapter 6) and the relationship between performance on the GSA and other measures of achievement (Chapter 7) are likely to have significant validity.

Findings and conclusions

Factor structure and discriminant validity (Chapter 5)

1 In support of test validity, confirmatory factor analysis indicates that the test does measure five coherent and distinguishable (discriminant) dimensions in line with the test construct. Thus, there is no reason to collapse or combine dimension scales, unless students within a narrow field of study are considered (who tend to perform similarly on components other than Problem Solving).

2 A second-order factor is also observed on which each of the five dimension factors loads significantly. This second-order general factor may be related to a form of meta-cognitive general executive reasoning skill that can be applied to a range of tasks.

3 For appropriate measurements and comparisons to be made between years, it is essential that the factor structure of the test is monitored and maintained.

Variables related to student performance on GSA (Chapter 6)

1 Findings with respect to variables related to student performance on the test include the following:

- There are distinctive profiles of student performance on the GSA components related to field of study that seem meaningful on the basis of known strengths of field of study groups (e.g., humanities students do relatively well on Writing and Critical Thinking).
- When first-degree students are considered within fields of study, there is a statistically significant difference in GSA scores for all five components between first and third year students. This observation supports test validity but needs to be clarified by studies in which the same students are tested in first and third year.
- Multivariate, multilevel analysis indicates that field of study, year level and familiarity with English (i.e., English-speaking background – ESB) appear to be related to performance on all five GSA dimensions. Gender seems to be related to performance on Problem Solving (with males doing better) and Interpersonal (with females doing better). Age seems to be related to performance on Problem Solving (with younger students doing better) and Interpersonal (with mature age students doing better). Other variables may be relevant but need further investigation.
- The multivariate, multilevel models used (which consider field of study, English-speaking background, age, gender, school type and course year) explain about 30% of the variance in students' GSA scores, with field of study being the largest single contributor. However, the majority of the variance seems to be explained by other variables, including student-specific variables such as ability in relation to the skills assessed by the GSA and motivation.

2 Whether variables such as English-speaking background, age and gender are related to test performance inappropriately is not clear. Studies need to be done to monitor whether performance on the test with respect to these variables matches performance at university and in graduate work.

3 If student GSA achievement improves in a short period of time from first to third year, the GSA is likely to be assessing developing generic skills and not just a traditional fluid intelligence.

4 Obtained samples were inadequate to provide suitable 'value-added' estimates for either universities or fields of study within universities.

Relationship between performance on GSA and other measures of student achievement (Chapter 7)

1 Performance on the GSA should correlate with performance on similar tasks including those related to success at university and graduate work. An investigation was undertaken to examine relationships between GSA, tertiary entrance (TER) and grade point average (GPA) scores. Because universities

have different types of entrance scores and ways of predicting academic success in courses, such analyses were done at the university level. At this stage it is too early to investigate the relationship between students' GSA scores and their work performance.

2 In support of test validity, the data collected suggest that student performance on each GSA component is significantly correlated (statistically) both with TER and GPA performance for most university cohorts. In most cases, the GSA-GPA correlation was as good as or better than the TER-GPA correlation. For cases where performance on the GSA did not correlate significantly with GPA, neither did TER.

3 The predictiveness of the GSA components varied with the university cohorts, and this observation may be related to the field of study composition of the cohorts or other sample idiosyncrasies. Predictive validity studies that are focussed on individual fields of study could be informative.

4 The fact that performance on a short test of generic skills like GSA correlates significantly with measures like GPA and TER, which are related to a wide range of curriculum knowledge and skills, suggests the importance of generic skills in academic performance and supports GSA validity.

5 GSA-GPA correlations appear to be comparable to SAT[1]-GPA correlations seen in the USA.

6 Although data was only available for a handful of students, a small-scale study using a variant of the GSA (BMAT) tailored for entrance into a postgraduate business school found a statistically significant correlation between GSA Problem Solving performance and GMAT [2] performance. It would be desirable to expand this study.

7 Given the predictiveness of the GSA, it may be feasible for universities to use it as an additional predictor of performance for entry into undergraduate and post-graduate courses, perhaps weighting the GSA components to optimise the predictions (as is done with the Victorian General Achievement Test in another context). The GSA might also be used to provide university entrance score equivalents for students who do not have these.

Evaluation of GSA reference ranges (Chapter 8)

1 The GSA uses two main methods of indicating student performance on the Student Report Forms. One provides for comparison purposes the middle 60% of all student scores and the middle 60% of scores for students in similar fields of study. The other provides performance level descriptors, with student scores assigned to a level of performance.

---

1 The SAT (Scholastic Aptitude Test) is a widely used test of general academic ability in the USA.

2 The GMAT (Graduate Management Admissions Test) is used for selection into many postgraduate business courses in the USA and other countries.

2 Because GSA reference ranges are related to the sample that sat the first two tests, as discussed previously, there is doubt about how well the reference ranges apply to the whole university population.

3 As suggested by the TERs of participating students, reference ranges may be set high, with stronger students being over-represented.

4 Reference ranges are likely to be most problematic for the individual fields of study where few students have participated so far and where the field is composed of smaller sub-fields whose students differ markedly.

5 It should also be noted that, because insufficient data are available, current reference ranges for field of study groups do not take into consideration year level of students, and this is inappropriate.

6 In consultation with universities, more representative samples should be sought for the purpose of refining reference ranges.

7 In consultation with universities, further consideration could be given to the suitability of the described levels of performance.

8 The reliability of GSA multiple-choice components is likely to be satisfactory for many purposes though not for others. For example, it may be satisfactory for measuring relatively small changes in performance for groups of students between university entry and exit. However, it may not be satisfactory for determining small changes in an individual student's performance. The problem of test reliability would be most acute for the low and high ends of the reporting scale where there are few items to discriminate between students.

9 For assessments at the low and high ends of the scale, and for other purposes, specialised versions of the test might be used.

10 In order to improve test reliability for some purposes, it may be appropriate to reduce the number of multiple-choice components from three to two, one focusing on Analysis, Synthesis and Evaluation of information (addressing common elements of Problem Solving and Critical Thinking consistent with the initial stakeholder input) and the other on Interpersonal Understandings.

Review of test construct and items (Chapter 9)

1 Various stakeholders and content experts were asked to evaluate the GSA construct and a sample of items.

2 In general, the content experts in the various domains commented favourably on the face and content validity of the construct and items. However, they expressed general concerns about whether performance on the test would translate into university and workplace performance, and questioned the extent to which universities deliberately develop GSA-type skills (though such skills are mentioned in most university mission statements). In addition, there were specific concerns, such as those relating to the meaningfulness of some performance level descriptors.

3 In general, the graduate recruiters seemed to respond positively to the test, suggesting it was relevant. However, when asked for their preferences, they tended to emphasise most workplace skills such as applied interpersonal skills.

4 In general, the students responded positively to the test overall, suggesting it was measuring important skills and could give useful feedback to students and universities. Nevertheless, some expressed concerns about fairness for students whose first language was not English, as well as querying validity and reliability.

5 In the discussion with the group of other stakeholders, it was apparent that there were dramatically different views about aspects of the test, and to some extent these views were related to the background of the stakeholder (e.g. humanities academic vs engineering professional). In general, issues of concern for these stakeholders included: the possibility of league tables appearing; whether there are generic skills outside disciplines or work situations; privacy of results; whether universities actually teach such generic skills; limitations of multiple-choice items; relevance of interpersonal skills to researchers; audience specification and scaffolding for writing; relevance of the test to all university students; relevance to post-graduate work; cultural and ESL bias; and so forth.

Recommendations for the future development of the GSA (Chapter 10)

1 Continuing attempts should be made in association with universities to obtain representative student data.

2 The factor structure of the test should continue to be monitored to ensure that the test remains appropriately focussed.

3 Further investigations should be undertaken to confirm, clarify and more precisely quantify relationships between performance on the GSA and variables such as field of study and year level, and to investigate the appropriateness of differential performance on the basis of variables such as English-speaking background and gender. Investigations broadening the range of variables examined could be done.

4 Further investigations should be undertaken into the relationships between GSA performance and markers of achievement at university and work. Evidence could include reports on students and graduate workers by tutors and supervisors.

5 Consideration could be given to the use of the GSA for selection into university courses.

6 Reference ranges should be refined, including those for sub-groups, such as specific field of study and year level cohorts.

7 There should be further evaluations of whether test reliability and described levels of performance are suitable for the particular purposes for which the results are being used. If reliability is not sufficient for a particular purpose, consideration should be given to ways of improving it.

8 In consultation with stakeholders, consideration should be given to the refinement of face/content validity, and construct and level descriptions, where possible, these being based on a comprehensive and commonly accepted developmental model of generic skills.

9 The purpose(s) of the test should be clarified in consultation with stakeholders and, if appropriate, versions of the test tailored for specific stakeholder purposes could be produced that are linked statistically to the general test.

10 Assessment of validity should be ongoing as the test evolves and stakeholders should be involved in evaluation and research.

Concluding remarks

1 The challenge for test developers of producing an appropriate theory-based and empirically validated test of generic skills that satisfies a range of stakeholders with competing demands is a substantial one. In relation to this, more discussion with stakeholders about the purpose, design and value of the test, as well as more opportunity for stakeholder involvement in test design and research, may be useful.

2 Assessment of the validity of the GSA is a complex process. This study is a first step that provides evidence in favour of the validity of aspects of the GSA as it currently operates, but also raises some concerns. As the GSA evolves in response to feedback, ongoing assessment of validity will be required.

1. Aims, scope and methodology

1.1 Introduction

The Graduate Skills Assessment (GSA) is a new test with complex aims and is in the first stages of development and application. Principally, the test aims to assess a set of valued and widely applicable generic skills that may be developed by the university experience and which are relevant to university achievement and graduate work.

This publication reports on the validity of the first stage of the Graduate Skills Assessment test (GSA Stage One Validity Study), which covers the first two tests, Exit 2000 and Entry 2001. These tests involved the participation of 3663 students drawn from nine broad fields of study across 27 Australian universities. Details of these populations are given in the GSA Summary Reports (Hambur & Glickman, 2001, Hambur & Le, 2001). The study was commissioned by the Commonwealth Department of Education, Training and Youth Affairs (now known as the Department of Education, Science and Training).

To be valid the GSA should achieve the following:

- The test construct and items should have face/content validity for stakeholders and experts;
- The test should have a meaningful and discriminant dimensional/factor structure consistent with the described construct;
- Variables related to student performance should be meaningful and include number of years in a university course;
- Performance on the GSA should be related to performance on other assessments measuring similar skills, including performance at university and in employment;
- Performance on the GSA should not be affected inappropriately by variables such as gender; and
- The GSA should have suitable reference ranges and sufficient reliability for its purpose.

1.2 Aims

This validity study has the following major aims:

Aim 1 To investigate the dimensional factor structure (discriminant validity) of the test.

Aim 2 To identify variables related to differential performance on GSA.

Aim 3 To investigate the relationship between student performance on the GSA and other measures of student achievement.

Aim 4 To consider the suitability of current reference ranges.

Aim 5 To evaluate the face/content validity of the GSA construct and items.

1.3 Methodology

Sample analysis: In order to identify biases in the sample used for this study, characteristics of the students participating in GSA Exit 2000 and GSA Entry 2001 were compared with the general university population by means of relative percentage plots and standard t-tests of significance. (Chapter 4)

Aim 1: In order to confirm the intended five-factor/dimension structure of the test (i.e. its discriminant validity), correlation analysis and confirmatory factor analysis using LISREL 8.30[3] were used, these being based on score data from the GSA Entry 2001 test. (Chapter 5)

Aim 2: In order to identify variables related to performance on the test, mean score versus year level and field of study plots for each test dimension and multilevel, multivariate analysis[4] were used, these being based on score data from GSA Exit 2000 and Entry 2001. (Chapter 6)

Aim 3: In order to investigate the relationship between performance on the GSA and other measures of achievement, correlation analysis and regression ANOVA were carried out, using GSA Exit 2000 score data and tertiary entrance scores and grade point averages (or equivalents) for those students who allowed this information to be provided. (Chapter 7)

Aim 4: In order to consider the suitability of current reference ranges, results from the sample analysis (Chapter 4) were considered, together with score ranges for fields of study (based on data from GSA Exit 2000, GSA Entry 2001 and the trial test), and tertiary entrance scores for those students who allowed this information to be provided. (Chapter 8)

Aim 5: In order to evaluate the face/content validity of GSA, various stakeholders and content experts were asked to evaluate the GSA construct and a sample of items. These were:

- a professional with content expertise in each component;
- a group of recruiters of graduates;
- a group of students from one university who sat the test; and
- a variety of other stakeholders from university and the graduate workplace (Chapter 9).

[3] Jöreskog & Sörbom, 2000.

[4] *MLwiN* (Rashbash et al., 2000), LISREL 8.30 (Jöreskog & Sörbom, 2000).

2. Background to the GSA

2.1 General background

With increasing work complexity and job mobility in modern economies, there has been more need for people to develop skills that allow them to adapt to and operate in a variety of workplaces. As a result, there has been a growing interest in the development and assessment of generic skills (Kearns, 2001).

The Australian Council for Educational Research (ACER) was commissioned by the Department of Education, Training and Youth Affairs (DETYA) to consult with universities and other stakeholders to identify a set of valued generic skills that could be effectively assessed at university entry and exit level. The assessed skills needed to be transferable, and have broad relevance to academic work and graduate employment. Ideally, the skills would be developed by the university experience.

All Australian universities were invited to attend meetings at which representatives were asked to provide a list of skills they valued and would like to see assessed in their students. The result of this consultation is summarised in Table 2.1.

The resulting test is called the Graduate Skills Assessment test (GSA). The current format of the GSA is two hours of multiple-choice items and one hour of writing tasks.

2.1.1 Terms

Generic skills have been defined in various ways depending on the views and aims of the particular stakeholders (as discussed in section 2.2, Expectations of Stakeholders). Some stakeholders are interested in generic skills related to academic success or socio-cultural understanding; others are interested in generic skills more directly related to employability. In the latter context, the British National Skills Task Force (NSTF, 2000) has defined generic skills as *those transferable skills, essential for employability which are relevant at different levels for most* (people in the workforce).

With respect to the GSA, the term 'skill' is used to describe the crystallised ability of students to deal effectively with certain kinds of higher order generic reasoning task. The term '***generic skills***' refers to general, transferable skills of a kind that can be widely applied in academic work and graduate employment.

In this report, the following definitions are also used:

Domain – set of generic skills with a particular focus, such as Critical Thinking;

Dimension – psychometrically distinct aspect of a domain focusing on a single latent variable that can be used to construct a measurement scale; and

Component – set of items or task constructed by test developers to assess a dimension.

2.1.2 Transferability

Clanchy and Ballard (1995) point out that generic skills are learned in a context and that their form varies from discipline to discipline, and state that:

'…while such skills cannot be learned in vacuo, indeed they must be learned in the context of a specific discipline and body of knowledge, they do not – once learned – have to be learned totally anew in each context of learning. Some degree of transfer does occur, and the most effective learners are those who in fact most quickly recognise the relevance of previously learned skills to the new contexts and are most readily able to adapt them to those new contexts.'

The Graduate Skills Assessment (GSA) is based on premises that parallel these views.

Though expected to be transferable, the ability of a person to display a generic skill in a particular situation is dependent on a whole range of factors including familiarity with context and motivation.

Clearly, familiarity with context is important in the application of a generic skill. However, the emphasis in the GSA is on skills being applied across a range of accessible contexts. It is assumed that the more graduates are able to apply skills across the range of contexts provided, the more likely it is that they can apply them to other contexts with which they will become familiar.

In general, the skills assessed in GSA are meta-skills in that students need to identify, select and apply an appropriate repertoire of more specific knowledge and skills to deal effectively with the tasks. Such meta-skills are likely to be transferable.

2.1.3 Possible uses of the test

At this stage, the test aims to provide an indicator to universities of generic skills in their students at entry level and/or exit level.

At entry level, universities may use the test diagnostically to identify, for example, those who write poorly. Such students may be followed up and offered assistance. The GSA might also be used as additional information for student selection.

At exit level, results of the test may be used as an additional criterion for entry into post-graduate courses. Exit level information could be useful for employers.

Universities may be interested in profiles of student performance in different courses and changes in student skills over time.

Other uses of the test are possible and could evolve over time. For example, additional components, such as those related to *Basic Skills, Research Skills, Management Skills, IT Literacy* or *Personal Skills*, might be added to the battery at a later date.

The test might also be modified so students in each major study area deal with some material more specifically focussed on that area, and universities/departments may have the opportunity to add a set of items of specific interest to them.

Computer delivery of the test based on an item bank is possible and could enable selected test components to be delivered to selected students.

2.2 Expectations of stakeholders

2.2.1 Stakeholders

Stakeholders in the outcome of the development of such a test include the universities, employers and the government. There are tensions between the expectations of these stakeholders. For example, universities may wish to see a test that focuses on academic skills and personal qualities such ethical citizenship, which are described in university mission statements, whereas employers may wish the focus to be on practical employability skills (such as oral communication, problem solving, self-reliance and enterprise). Further, within universities, some departments may prefer to focus on discipline-specific skills rather than on more general skills.

To some extent, the different approach of stakeholders is reflected in the language they use. According to Curtis and McKenzie (in press), '*the lack of common understanding* (about which skills to assess with respect to employability and readiness for further learning) *is reflected in the language being used in different circles and forums'*. Their report reviews and comments on Australian and overseas views and experiences in assessing employability skills and sheds light on some issues relevant to the GSA, as does the report by Kearns (2001).

It is clear that stakeholders in the GSA have different expectations of such a test, which makes test development a particularly difficult task, requiring an approach that optimises its value to the range of stakeholders. An alternative approach is a set of tests, each focussed for a particular group of stakeholders.

2.2.2 Generic employability skills

Policy makers in several countries have considered and made recommendations about the generic employability skills that should be developed at school and university.

The Mayer competencies (Mayer Committee, 1992) have been much considered in Australia as a basis of employability skills that can be addressed by formal education**.** These are:

- Collecting, analysing and organising information;
- Communicating ideas and information;
- Planning and organising activities;
- Working with others and in teams;
- Using mathematical ideas and techniques;
- Solving problems;
- Using technology; and
- Cultural understandings.

Consideration is being given to the development of reliable and valid test instruments for these skills, but Kearns (2001) suggests that this set is limited (for example, by its omission of personal attributes and because the competencies are not based on a coherent theory of skill development).

ACNielsen (1998, 2000) report that Australian employers list the following as important graduate employability skills:

- Academic achievement;
- Literacy;
- Numeracy;
- Logical and orderly thinking;
- Computer skills;
- Time management skills;
- Written business communication;
- Oral communication;
- Creativity and flair;
- Interpersonal skills;
- Teamwork skills;
- Problem solving skills; and
- Comprehension of business processes.

These reports note that new graduates who have been employed are perceived to be most deficient (in comparison with employer expectations) in creativity and flair, problem solving skills, oral business communication skills and interpersonal skills, with some concerns also about numeracy, academic learning and logical and orderly thinking. However, there is variation between sectors and type of graduate.

In the UK, the NAB/UGC report (1984) stated that valued employability skills include:

'…the ability to analyse complex issues, to identify a core problem and the means of solving it, to synthesise and integrate disparate elements, to clarify values, to make effective use of numerical and other information, to work co-operatively and constructively with others, and above all perhaps, to communicate clearly both orally and in writing.'

Harvey and Green (1994) concluded that the following generic skills were valued by both employers and academics:

'… willingness to learn, team work, problem solving and a range of personal attributes including commitment, energy, self-motivation, self-management, reliability, cooperation, flexibility and adaptability, analytical ability, logical argument and adaptability to summarise key issues.'

The Association of Graduate Recruiters in the UK (1995) state that self-reliance skills are particularly important for graduates. These include the following: self-awareness, self-promotion, exploring and creating opportunities, action planning, networking, matching and decision making, negotiation, political awareness, coping with uncertainty, development focus, transfer skills, self-confidence.

In the United States, The Secretary's Commission on Achieving Necessary Skills (2000) has identified skill sets for many job types, and assessment instruments are being developed. A precursor to this was the work of Carnevale (1991) who identified a set of core skills as important to employability (Appendix 1).

*

Kearns (2001) points to two broad approaches to generic work skills, which are suggested in the preceding discussion and are summarised here.

Approach 1: *broader, more flexible, and more holistic set of generic skills, which include basic skills, personal attributes, values and ethics, learning to learn, as well as workplace competencies of the Mayer type*

Approach 2: *more narrowly focussed and instrumental set of skills competencies….personal attributes and values have been excluded*

Kearns states that Approach 1 is more common in the US (e.g. Carnevale and Pruitt, 1992), while Approach 2 is more common in England and Australia.

2.2.3 Generic academic skills

Typically, universities in Australia and overseas say in official statements of objectives that they wish to produce active, effective and ethical citizens who:

- are equipped and motivated for lifelong learning;
- communicate effectively;
- think critically, think independently and are open minded;
- solve problems;
- manage their behaviour effectively;
- work well and effectively with others, including leadership;
- understand their culture and those of others;
- have broad understandings;
- are expert in their chosen field;
- are IT literate; and
- are creative and entrepreneurial.

Generic academic skills tests

Generic academic skills tests are widely used, however they are much narrower in focus than the stated university objectives, usually focussing on scholastic aptitude for selection purposes. These are often referred to as generic or cross-curricular scholastic aptitude or achievement tests. Mc Donald et al (2001) have reviewed the use of such tests for university entrance internationally.

In Australia, such tests include the following ACER[5] tests: the *Australian Scaling Test* (used for UAI/TER scaling purposes in the ACT); the *Special Tertiary Admissions Test* (used for selection of mature age students into Australian universities); and the *General Achievement Test* (used for several purposes in Victoria, but in general as a comparison of academic ability against achievement).

[5] Australian Council for Educational Research, Melbourne, Australia

These three tests have a common three-component structure. The components are:

1 verbal comprehension/socio-cultural understandings based predominantly on humanities and social science contexts (multiple-choice);

2 logical and quantitative reasoning based on applied mathematics, science and social science contexts (multiple-choice); and

3 written communication.

In the US, generic academic skills tests are predominantly multiple-choice and include: ETS[6] tests such as the *Academic Profile* test (which assesses college students at entry and exit in reading/critical thinking, mathematical reasoning and writing, especially as these relate to the broad fields of Humanities, Social Science and Natural Science); the *Graduate Records Exam - General Test* (which selects students for post-graduate courses with components related to verbal ability, quantitative ability and logical/analytical reasoning); the *Scholastic Assessment Test 1* (which is similar to the GRE but pitched at a lower level and is used for selection into college); and the *Graduate Management Admissions Test* (which, again, is similar to the GRE and is used for selection into business colleges).

The US tests tend to be more related to a defined knowledge base (e.g. vocabulary or specified mathematical processes) than are the equivalent Australian tests, which are more often based in authentic contexts and are more focussed on cross-curricular skills.

2.2.4 Generic skills in relation to university and employer aims

As suggested above, employers in general are not averse to the skill sets that universities say they wish to develop in their students, or to the employability skill sets drawn up by education and government authorities. (See also Appendix 2.) However, employers tend to focus most on those skills that will help their organisation function, especially personal and interpersonal skills (such as self-reliance, self-management, effective oral communication, problem solving, logical and orderly thinking, creativity and flair in business, entrepreneurship, teamwork and leadership). On the other hand, universities often prefer to focus more on academic skills and student qualities related to national and international citizenship.

2.3 Views of Australian stakeholders in GSA

In preparation for the development of the GSA, ACER consulted with Australian universities and other stakeholders to identify a set of generic skills of relevance to academic performance and employability. The proviso was given that the skills needed to be practically assessable by means of two hours of multiple-choice items and one hour of writing.

[6] Educational Testing Services, Princeton, New Jersey, USA

Table 2.1 summarises the views of university representatives who participated in the consultation and other stakeholders (or their published official positions). The number of times a skill/attribute was mentioned by the stakeholders is indicated.

Table 2.1 Responses to consultation

Generic skills	Universities — specific response or general position	Other stakeholders (such as employers and careers councils)
Communication/structured written response	///// ///// ///// ///// /	///// /
Problem solving/applied reasoning/ strategic	///// ///// ///// /	///// /
Analytical skills	///// ////	/////
Critical thinking	///// ///// /////	//
Logical reasoning	///// ////	//
Ethics/citizenship/social responsibility/empathy	///// ///// /////	///
Creativity	///// ///	//
Interpersonal skills/teamwork/ leadership	///// ///// ///// ////	///// //
Sceptical but open-minded	///// ///	
Flexibility/tolerate uncertainty	///// /	//
Capacity for or commitment to lifelong/independent learning	///// ///// //	///
Numeracy/ability to quantify	///// /	//
Literacy	///	/
IT familiarity/IT use	///// ///// ///	///
Personal skills/self-management/ reflective/confidence/self-reliance/initiative	///// /	/////
Global/national/historical/cross-cultural perspective	///// //	//
Information literacy/ management/research skills	///// ///	

2.4 Selection of GSA generic skill domains

The development of a valid, reliable, time and cost effective instrument for the assessment of a set of generic skills of relevance both to academic work and graduate employment is a great challenge. In order to accommodate the divergence of views about what should be assessed, the GSA focuses on common elements of stakeholder views that have psychometric meaning and are measurable in the proscribed format. Ideally, the skills assessed will be enhanced by the university experience.

Of the skills suggested by stakeholders (Table 2.1), Written Communication, Critical Thinking, Problem Solving and Interpersonal Understandings were chosen for the initial test because they were 'popular', seemed to be essential elements of other skills (such as capacity for lifelong learning), and were likely to be transferable and readily measurable. These, or similar, skills are also mentioned frequently by universities, government agencies and employers in other countries.

Written Communication, Problem Solving, Critical Thinking and Interpersonal Understandings involve skills/approaches such as analysis, logical reasoning, creativity, functional literacy, functional numeracy, empathy and creativity, which are listed separately in the table. In addition, aspects of the capacity for lifelong learning and employability, such as an ability to identify, absorb and apply key information, reflect and logically organise one's thoughts and actions, would seem to be important for success in all the chosen skill domains.

The suggestions made in Table 2.1 can be divided into two groups, those focusing on cognitive skills and those focusing on non-cognitive attributes. The first versions of the GSA are predominantly cognitive in focus, because it was considered that this approach would be more reliable and valid in the first instance.

In the longer term it would be desirable to assess non-cognitive attributes as well.

2.5 Other issues

2.5.1 Test format

Multiple-choice questions have limitations. By necessity, they are relatively closed, providing options to choose from, rather than open-ended requiring the generation and application of a solution/view with limited prompting, as do most real-world tasks.

However, although GSA multiple-choice items provide options, the tasks are complex and students will normally need to generate solutions/views to match against the options if they are not simply to guess and thus attain a low score.

In contrast to multiple-choice items, the open writing tasks may be considered more like real-world tasks in that generation and application of solutions/views is required with minimum prompting. However, great care needs to be taken in selecting, briefing and monitoring markers of open-ended questions to avoid undue variability. Multiple-choice items have the advantage over open questions of avoiding variability in marker judgements.

Although it is believed the skills assessed in the GSA have real world relevance and transferability, this is a question that this study can only begin to answer.

At best, the GSA, like other tests, can only provide a glimpse of a student's skills. The value of that glimpse can only be gauged as validity work proceeds.

2.5.2 Generic skills and intelligence

Gottfredson (1997) states that although there is debate about the precise definition of intelligence, there is agreement that it reflects the ability to reason, solve problems, think abstractly and acquire knowledge. Intelligence affects the ability of people to recognise, acquire, organise, update and apply knowledge effectively. It is related to the ability of people to deal with complexity. She argues that such higher order thinking skills significantly affect the chance of success in work and life in general, though other factors such as motivation and personality are also important.

There is a massive literature on the topic of intelligence and there has been much debate about whether there is a general intelligence factor, *g*, or multiple intelligences. For example, Carroll (1993) lists scores of potential intelligence factors and Gardner (1993) talks about brain 'modules' dedicated to certain intelligences (e.g. verbal, mathematical/logical and interpersonal).

There is an ongoing debate about the extent to which intelligence is due to inherited or environmental factors. It is known that IQ scores, the traditional measures of intelligence, have increased, on average, over the past 50 years (Flynn, 1999). It has been hypothesised that this is a result of the increasingly stimulating and educating environment that people live in, but may also be related to factors such as nutrition.

Clearly, performance on GSA is affected by intelligence, whether due to genetics or environment, or both, and whether there is a general, executive intelligence, modular intelligences, or both. It will also be affected by other factors, including motivation and confidence.

Although intelligence is certainly a factor related to student performance on the GSA, what is important is that student performance on the GSA is affected by the university experience and is ultimately related to academic and work performance.

3. GSA construct

3.1 Component descriptions and rationales

Having selected the skill domains to assess in the GSA, it was necessary for test developers to define the domain constructs and test specifications in such a way that a valid and reliable test instrument could be developed. The test instrument had to measure psychometrically distinct skill sets (dimensions) that are relevant to industry and academic work. In order to measure psychometrically distinct test dimensions, it was necessary to narrow the focus of each of the broadly defined domains, describing them appropriately in the test construct.

In the development of the test construct, it was accepted that:

- the test would address five cognitive dimensions – Critical Thinking, Problem Solving, Interpersonal Understandings, Argument Writing and Report Writing. The first three of these would each be addressed by a 30-item multiple-choice test component and the last two (covering two forms of Written Communication) would each be addressed by a task requiring a written response;
- the generic skills to be assessed in the GSA should have a significant degree of transference to a new context once sufficient familiarity has been gained in that context (Assiter, 1995; Gibbs et al., 1994; Mumford et al., 1998); and
- the initial version of the GSA would focus on cognitive skills and understandings. The GSA would not attempt to assess directly personality traits that are related to putting the skills and understandings that are demonstrated on the test into action in the real world. It is hoped that validity studies will show a sufficient association between test performance and real-world outcomes.

In the following, a description is given of each of the five GSA test components, together with background information that informed the rationale for the GSA approach. Stakeholder evaluations of the face and content validity of the GSA are presented in Chapter 9.

3.2 Critical Thinking

3.2.1 Background considerations for Critical Thinking

Norris and Ennis (1989) have defined Critical Thinking as: '*reasonable and reflective thinking (concerned with what to do or believe)*'.

Jones and Ratcliffe (ERIC No. ED358772) state that: '*…items central to critical thinking are analysis, evaluation and inference; that critical thinking requires the use of cognitive abilities; that*

critical thinking includes meta-cognitive or self-monitoring skills; and that a student's thinking should meet certain criteria of good thinking or intellectual standards.'

Paul (1994) states that Critical Thinking's *most fundamental concern is excellence of thought and it is characterised by its responsiveness to intellectual standards, such as relevance, accuracy, precision, clarity, depth and breadth.* Critical thinkers can identify and make connections between elements of thought in order to produce predictable, well-reasoned answers to questions. They are aware of pitfalls in reasoning. Further, critical thinkers have intellectual traits such as intellectual integrity and humility. They are open-minded, and routinely self-assess to identify weaknesses in their own position and strengths in the position of others. Thus, they build logically consistent views of issues and problems but are aware that there may be more than one legitimate point of view. Good critical thinkers use their skills to identify the most appropriate intellectual position rather than to justify personal views.

Tests of Critical Thinking often focus on cognitive skills and have approaches much narrower than is Paul's. The Watson-Glaser Test of Critical Thinking (Watson & Glaser, 1980) has five sub-tests, namely: *inference, recognition of assumptions, deduction, interpretation of arguments, evaluation of arguments.*

The Smith-Whetton Critical Reasoning Test (Smith & Whetton, 1992) covers the following Critical Thinking skills:

- 'analysis – the careful analysis of available information, so as to identify the definitions used, claims being made, assumptions, arguments, logical implications, missing information;
- evaluation – objectively judging the validity of arguments and the strength and credibility of evidence; and
- planning – deciding on logical action sequence in order to achieve specified goals.'

The last of these, Planning, is more akin to Problem Solving in GSA.

Transferability of Critical Thinking skills

One of the major debates in the field of Critical Thinking has been about whether Critical Thinking is discipline-specific or transferable. Influential people in the field such as Paul and Ennis take the view that the most important Critical Thinking skills are transferable. Paul (1994), for example, says:

'As with any system, critical thinking is not just a random series of characteristics or components. All of its components – its elements, principles, standards and values – form an integrated, working network that can be applied effectively not only to academic learning, but in every dimension of living.'

Norris and Ennis (1989) make the following point in relation to the assessment of transferable Critical Thinking skills:

'...to evaluate students' critical thinking facility in general and thus estimate the likelihood they will think critically in new contexts, students should be presented with a wide variety of critical thinking tasks requiring background knowledge they already have.'

Pithers and Soden (2000) state that:

'It seems unlikely that broad forms of thinking, such as those involved in hypothesis generation and testing, have to be learned from scratch each time the graduate has to learn knowledge from another

discipline. Self-regulation of one's cognitive abilities is likely to be widely generalizable. Thus all the abilities and dispositions encompassed by the term 'critical thinking' are likely to facilitate the comparatively fast rate of assimilation required in academic study and in many occupations.'

Development of Critical Thinking skills

Little definitive work has been done on developmental models of Critical Thinking, though it is assumed that people become more sophisticated in their use of Critical Thinking as they go through educational experiences and mature. For example, they become more reflective, logical, able to identify appropriate intellectual criteria for making decisions and able to monitor their own thinking.

Kuhn (1999) describes a model for the development of Critical Thinking skills, focusing particularly on the meta-cognitive aspects, in which the emphasis is on the development of skills that enable one to generate appropriate criteria for belief and evaluate one's own thinking. For example, in early stages of the development of Critical Thinking, 'theory' is the basis of belief but at later stages 'evidence' is fundamental. In early stages, assertions can be taken as reality but in later stages assertions are evaluated against a set of intellectually rigorous standards. In early stages, people do not acknowledge the 'knower' as a constructor of knowledge, but in later stages this is accepted, as is an acceptance that humans can generate a multiplicity of valid representations of reality (valid with respect to appropriate intellectual criteria).

Kuhn points out that many adults do not reach higher stages of Critical Thinking development.

3.2.2 Critical Thinking in the GSA

Clearly, it is impossible to capture the breadth and depth of Critical Thinking in a 30-item multiple-choice test, and the difficulty is compounded by the unresolved debate about what it is and how it develops. (Critical Thinking is widely applicable and is assessed to some extent in other parts of the GSA.) However, GSA Critical Thinking aims to assess some key markers of the ability to think critically about views on issues and make decisions based on good intellectual standards.

At this stage, the GSA takes a fairly conservative approach to Critical Thinking, focusing mainly on the application of intellectual standards, and elements of thought and reasoning, in order to comprehend, analyse and evaluate viewpoints presented in text. However, intellectual traits of the thinker and meta-cognitive abilities are likely to be factors in success on the GSA.

To make GSA Critical Thinking distinct from GSA Problem Solving, which also has critical thinking aspects, it was decided that the viewpoints should be presented in text format, avoiding quantitative material (some data interpretation is included in the Problem Solving component).

The level of literacy that is assumed in GSA Critical Thinking is commensurate with that required for most beginning graduate level jobs. In fact, GSA vocabulary and sentence complexity should be within the grasp of most students successfully completing secondary school. However, because the focus is on thinking, bilingual dictionaries are allowed.

Since the ability to think critically depends on familiarity with the context, items used in the GSA tend to be generally accessible and avoid a need for specialised knowledge. The focus of Critical Thinking in GSA is reasoning in everyday contexts.

The items are multiple-choice in format and can be categorised as follows (though a single item may have facets of more than one category):

- *Comprehension* in order to identify explicit and implicit meaning;
- *Analysis and Inference* in order to identify definitions being applied, claims being made, points of view, key issues, lines of reasoning, evidence, conclusions, arguments, assumptions, logical flaws, logical implications, missing information, rhetorical devices, ambiguity, analogies etc; and
- *Synthesis and Evaluation* in order to judge the credibility and validity of evidence, lines of reasoning, conclusions and arguments.

Those with a wide-ranging and sophisticated repertoire of reasoning skills, including meta-cognitive Critical Thinking skills, and a basic understanding of issues that affect society, are expected to perform best on the test. It is expected that students doing well on GSA will be able to transfer their general Critical Thinking skills to work and academic situations once the specific details of the field have been mastered.

At this stage, the GSA probably is unable to distinguish between what Paul (1994) calls:

- the *sophist critical thinker* (who thinks skilfully but applies such thinking only insofar as it serves his/her views); and
- the *fair-minded critical thinker* (who assesses all views skilfully, including his/her own, according to the same rigorous intellectual criteria).

The scale used for describing student performance on the *first* GSA test is given in Figure 3.1 on page 25. Higher levels of performance are characterised by the ability to organise and control thinking effectively and subtly, including the ability to generate suitable evaluative criteria. The descriptions associated with the scale are likely to be refined as the test evolves.

3.3 Problem Solving

3.3.1 Background considerations for Problem Solving

Bransford and Stein (1993) state that:

'A problem exists where the present situation differs from the desired situation.'

Polya (1957), in his classic book *How to Solve It*, outlines the following approach for tackling mathematical problems, which can be more widely applied.

Understand the problem
Devise a plan
Carry out the plan
Look back to check the solution

Mumford et al., (1998) point to evidence that the ability to solve problems creatively is based on the following key processes:

- *definition and structuring of the problem situation (e.g., restatement of the problem);*
- *information acquisition or encoding in order to select and organise relevant information; and*
- *combination and reorganisation of knowledge to address the problem.*

As well as these general approaches, there are many specific approaches to problem solving. For example, de Bono (1977) emphasises lateral and creative thinking; Whimbey and Lochhead (1991) focus on comprehension in problem solving; Higgins (1994) discusses ways of analysing and recognising problems, particularly in business and administration contexts (techniques discussed include scenario analysis, story telling, fishbone diagrams, the use of analogies, brainstorming, making checklists); and Hoy and Tarter (1994) focus on administration problems, which are akin to some addressed in the Interpersonal Understandings component.

As this variety indicates, there are many types of problem and many approaches. However, the approaches generally have common elements, such as the following, that are relevant to the GSA:

- identification and analysis of the problem;
- selection and organisation of relevant information;
- representation of the problem and translation of relevant information in progressing toward a solution;
- identification of one or more strategies; and
- application and evaluation of strategies.

Problem Solving and transferability

Mumford et al., (1998) make the comment relevant to transferability that:

'*Assessment systems…that focus on general skills and that assess these skills by using general tasks intended to elicit these skills, may evidence transportabily across settings...*'

The comments made on the transferability of Critical Thinking skills are applicable here also, particularly since problem solving involves critical thinking (e.g. cognitive, meta-cognitive and meta-strategic kinds of critical thinking).

Further, insofar as problem-solving ability is related to general intelligence(s), which itself may be partly a function of experience, and intelligence is applicable and transferable between situations, general problem-solving ability is likely to be transferable.

3.3.2 Problem Solving in the GSA

Clearly, it is impossible to capture the breadth and depth of Problem Solving in a 30-item multiple-choice test. There is an enormous range of problem types and approaches. However, the GSA aims to assess some key markers of the ability to analyse and transform information in order to progress toward problem solution.

Psychometrically, it is likely that the form of the information provided and the type of problem are more important to solution than the general problem-solving approach. The specific content of a problem is likely to be a major determinant of its solution by a particular person with particular knowledge. It is also likely that only a few students will have been exposed to specialist problem-solving techniques.

Hence, the GSA approach has been to focus on generally applicable and accessible everyday practical problems that vary in complexity, and on the ability of students to identify, analyse, interpret, translate, reorganise, synthesise and appropriately apply problem-related information.

Students are expected to display a logical and organised approach in the analysis and application of relevant information. Analytical, logical, general quantitative and meta-strategic reasoning processes need to be applied.

Information is presented in low verbal and non-verbal form.

Although basic numeracy is assumed, including knowledge of simple arithmetic algorithms, all problems in GSA can be solved with lower secondary mathematical knowledge and calculators are allowed. The level of numeracy assumed is commensurate with that required for most citizens dealing with everyday problems and far below that required for graduate jobs in general.

Specialised mathematical, interpersonal and business/administration problems are not addressed.

Problem Solving in GSA does not attempt to assess directly personality traits that may be related to the ability to put problem solving into action.

GSA Problem Solving items can involve the following steps:

- Identify, comprehend, restate the problem;
- Identify and analyse information relevant to the problem;
- Represent features of the problem;
- Translate, reorganise, synthesise and apply information relevant to a problem;
- Identify or generate strategy for solution;
- Evaluate solution strategies and their outcomes;

and the following processes:

- Analysis, interpretation and evaluation of information for problem identification, understanding, restatement and representation;
- Categorisation, translation, reorganisation and synthesis of information in progressing toward problem solution;
- Application of reasoning skills to identify or generate a solution to a problem; and
- Evaluation of a solution.

It is expected that students doing well on GSA will be able to transfer their general problem-solving skills to work and academic situations once the specific details of the particular field have been mastered.

The scale used for describing student performance on the first GSA test is given in Figure 3.1 on page 25. Those with a wide-ranging and sophisticated repertoire of problem-solving skills are likely to perform best on the test. The descriptions associated with the scale are likely to be refined as the test evolves.

3.4 Interpersonal Understandings

3.4.1 Background considerations for Interpersonal Understandings

There is long history of a search for a factor that explains differences in how effectively people deal with others. Such ability is generally defined as how well one person understands others and can apply that understanding in social situations (Wechsler, 1958). In this context, concepts such as Social Intelligence (Legree, 1995), Interpersonal Intelligence (Gardner, 1993) and Emotional Intelligence (Mayer et al., 1999) have been theorised.

For Gardner (1993), Interpersonal Intelligence *'makes use of core capacities to recognise and make distinctions between others' feelings, beliefs and intentions'*. Gardner considers Interpersonal Intelligence largely distinct from other intelligences such as Verbal and Logical-mathematical.

According to Legree (1995), Social Intelligence, which is broadly similar in definition to Interpersonal Intelligence, loads on *g* (general intelligence), while also having unique elements. There is evidence that Social Intelligence is related to Verbal Intelligence. Studies by Rizzolatti & Arbib (1998) and Gallese & Goldman (1998) suggest that there may be an association between parts of the brain involved in identifying with the actions of others (empathy) and language/verbal skills. (This possibility is supported by observations in respect to GSA of a moderately strong correlation between student performance on Critical Thinking and Interpersonal Understandings, though, alternatively, it may be that there is a general reasoning/meta-cognitive skill that operates effectively on both.)

According to Mayer et al., (1999), '*Emotional Intelligence refers to the ability to recognise the meanings of emotions and their relationships, and to reason and problem-solve on the basis of them. Emotional Intelligence can be assessed most directly by asking a person to solve emotional problems ..*' These authors suggest a four component model, consisting of the abilities to reflectively regulate emotions, understand emotions, assimilate emotion in thought, and perceive and express emotion. In common with Gardner, they also suggest that Emotional Intelligence may be more distinct from traditional verbal intelligence than is Social Intelligence.

Such concepts are broad and it can be difficult to see how university courses, in general, can contribute specifically to their development.

In fact, as discussed in relation to Problem Solving, the GSA is not interested in measuring intelligence *per se*. It is most interested in assessing skills that can be deliberately developed by the university experience, including interpersonal skills.

These may in some way relate to a formal intelligence, but that is not the main focus of the test.

In descriptions of skills they wish to develop in their students, universities often describe skills related to working with others. Australian universities typically provide lists such as the following as guides to teamwork skills they wish to develop in their students:

- Work collaboratively and network to solve problems;
- Take responsibility and carry out agreed tasks;
- Take initiative and lead others;
- Operate in a range of supportive roles within teams;
- Negotiate, assert own values and respect values and contributions of others; and
- Evaluate team performance.

With respect to 'Work with others', the Canadian Employability skills (Conference Board of Canada, 2000) include the following guidelines:

- Understand and work within the dynamics of the group;
- Ensure your team's purpose and objectives are clear;
- Be flexible; respect, be open to and supportive of the thoughts and opinions of others;
- Recognise and respect people's diversity, individual differences and perspectives;
- Accept and provide feedback in a constructive and considerate manner;
- Contribute to a team by sharing information and expertise;
- Lead or support as appropriate, motivating for high performance;
- Understand the role of conflict in a group; and
- Manage and resolve conflict when appropriate.

A study in which employers and academics ranked valued generic skills of psychology graduates starting work reported in Appendix 2 (Australian College of Organisational Psychologists, 1999). In this study, *' establishing positive working relationships with people'* ranked highly for both employers and academics. The ability to establish such relationships is clearly related to interpersonal skills.

3.4.2 Interpersonal Understandings in the GSA

The Interpersonal Understandings component of GSA aims to assess some markers of the ability of candidates to show insight into aspects of interpersonal relationships that are relevant to the ability of people to work and live together effectively.

The stimulus for interpersonal items is usually presented as text, though pictorial material may be used. Verbal demand is usually moderate but some understandings require subtle descriptions that call on more sophisticated language. The material is not specialised in content and is readily accessible.

Some items are based on scenarios and require analysis and evaluation of the relationships described, and others on brief stimuli. Contexts may include situations drawn from work, education and life in general.

Because appropriate interpretations of and responses to situations may be a matter of opinion or changing fad, the focus is on principles and understandings that are generally applicable and relatively uncontroversial.

The items are multiple-choice in format and focus on the ability of students to:

- show insight into the feelings, motivation and behaviour of other people, and into issues related to helping or working with others; and
- recognise how such insight may be applied in order to effectively help or work with others, including effective feedback, listening, communication, negotiation, teamwork and leadership.

In more detail, aspects of interpersonal understandings that may be assessed include:

- identification and interpretation of roles and relationships;
- interpretation of feelings, attitudes, motives, values, personality, behaviour;
- identification and application of effective teamwork, leadership, negotiation, interpersonal communication and listening skills;
- identification, application and evaluation of approaches for optimising team performance or solving interpersonal problems in a work team with particular dynamics; and
- identification of individual differences, and application and evaluation of approaches for dealing with cultural diversity.

It is anticipated that students doing well on GSA will be able to transfer the interpersonal/social understandings that they have developed to work and academic situations once the specific details of the situations have been assimilated. (However, long-term studies will be needed to validate that success on the Interpersonal Understandings component of GSA translates into success in university and employment.)

The scale used for describing student performance on the first GSA test is given in Figure 3.1 on page 25. Those with a wide-ranging and sophisticated understanding of interpersonal matters, especially those related to working with others, and well-developed, socially focussed meta-cognitive skills, are likely to perform best on the test. The descriptions associated with the scale are likely to be refined as the test evolves.

3.5 Written Communication

3.5.1 Background considerations for Written Communication

The ability to communicate effectively in writing is considered an essential quality of university students and graduate workers. University faculties and employers stress the

importance of communicating information and ideas in writing and many tertiary institutions describe general communication and writing objectives for their graduates in broad and purposeful terms.

Australian universities typically provide lists such as the following as guides to the communication skills they wish to develop in their students:

- demonstrate oral, written, numerical and graphic communication;
- use the medium and form of communication appropriate for a given situation; and
- present well reasoned arguments.

These refer to communication in the broad sense. However, to be psychometrically coherent, it was necessary for the GSA to focus on certain kinds of communication, in particular on written communication.

Written communication is considered to be the purposeful, informed and effective control of the understanding, organisation and expression of ideas and information in writing.

3.5.2 Written Communication in the GSA

The Written Communication components of GSA were developed to reflect descriptions of the expectations of graduates, such as those given in the previous chapter.

It is important to use an appropriate writing form for the task at hand. This is reflected in the decision to design writing tasks focussed on two genres of writing that are considered to be valued in a range of faculties and workplaces, and appropriate for the maturity and expertise of tertiary students. These forms of writing draw upon aspects of generic employability skills, in particular collecting, analysing and organising information, communicating ideas and information, and planning. The genres selected were Argument and Report.

Report writing is a common form of substantive written communication in the university and graduate workplace. Skills in this form are commonly required and widely applicable. The Report task requires students to comprehend, select, organise and clearly present factual information.

Argument writing is commonly mentioned in descriptions of skills expected of graduates. The ability to present a clear, logical and soundly based argument is commonly required and widely applicable. The Argument task requires students to develop a point of view about an issue, and structure and present an argument in support of that view.

The stimulus material for each task provides students with a scaffold of data or information from which to build their piece. In addition, the guidelines (i.e. rubric) for each task include clearly expressed and differentiated instructions about the task and the qualities on which the writing will be judged. The stimulus material for the Argument task consists of a set of comments or opinions related to a social issue, while the stimulus material for the Report task consists of tabular and graphic data.

Criteria for assessment

Written Communication is assessed using a marking guide that describes three criteria, on a scale of 0 to 10, with 0 being the lowest score point and 10 being the highest. The criteria are:

- Quality of thoughts and ideas - differentiated for the Argument task and the Report task;
- Quality of structure and organisation; and
- Quality of language and expression.

The marking guide details the features of each criterion and describe the qualities of the writing related to each score point on the scale.

For "Quality of thought and ideas" (Argument task) there are four ordered points to describe each score point:

1. cognitive content;
2. range or breadth of ideas;
3. quality of development, commentary, evidence; and
4. evaluation, balance, opinion.

For "Quality of thought and ideas" (Report task) there are three ordered points to describe each score point:

1. cognitive content;
2. range or breadth of coverage; and
3. evidence of processing and integration of the material.

For "Quality of structure and organisation" (for both tasks) there are three ordered points to describe each score point:

1. logical structure;
2. development or building of ideas; and
3. knowledge of form or genre.

For "Quality of language and expression" (for both tasks) there are four ordered points to describe each score point:

1. communicativeness or expressiveness, tone or sense of audience;
2. sentence control and variety or syntax;
3. vocabulary; and
4. conventions (spelling, punctuation).

Application of the criteria

The tasks are sufficiently different to permit students possessing different communicative styles to demonstrate the extent of their skills. The characteristics of each genre are not unique and exclusive. Therefore they have not been strictly differentiated for each of the components. However, there are some underlying assumptions about those characteristics related to the purpose of the writing. In the Argument task the student who fails to recognise the underlying issue expressed in the

prompt is unlikely to be strongly rewarded on the thought and ideas criterion, although the structure and organisation of the material may be strongly appropriate for the presentation of a point of view. The Report task presupposes that the resource material presented will be synthesised and organised to succinctly communicate the ideas and information. This genre is not usually associated with broadly discursive or opinionative writing. In the case that such a piece of writing is presented, it is probable that the score for the structure and organisation criterion will be low.

Although the boundaries between criteria are necessarily somewhat blurred - for example, sometimes the demonstration of insightful understanding of the ideas is masked by difficulties in developing an effective structure for the piece - the markers consider each one as a separate quality. In this way, the student who is thoughtful, or who recognises the interrelationships between ideas in the stimulus, is not disadvantaged unreasonably by difficulties in organising or writing in English.

It is expected that students doing well on GSA will be able to transfer their general writing skills to work and academic situations once familiar with the specific requirements and conventions.

The scale used for describing student performance on the first GSA test is given in Figure 3.1 on page 25. Those with a mastery of writing skills, including those related to the analysis, organisation and clear presentation of information and viewpoint, and well-developed meta-cognitive skills that enable the effective application of these, are likely to perform best on the test. The descriptions associated with the scale are likely to be refined as the test evolves.

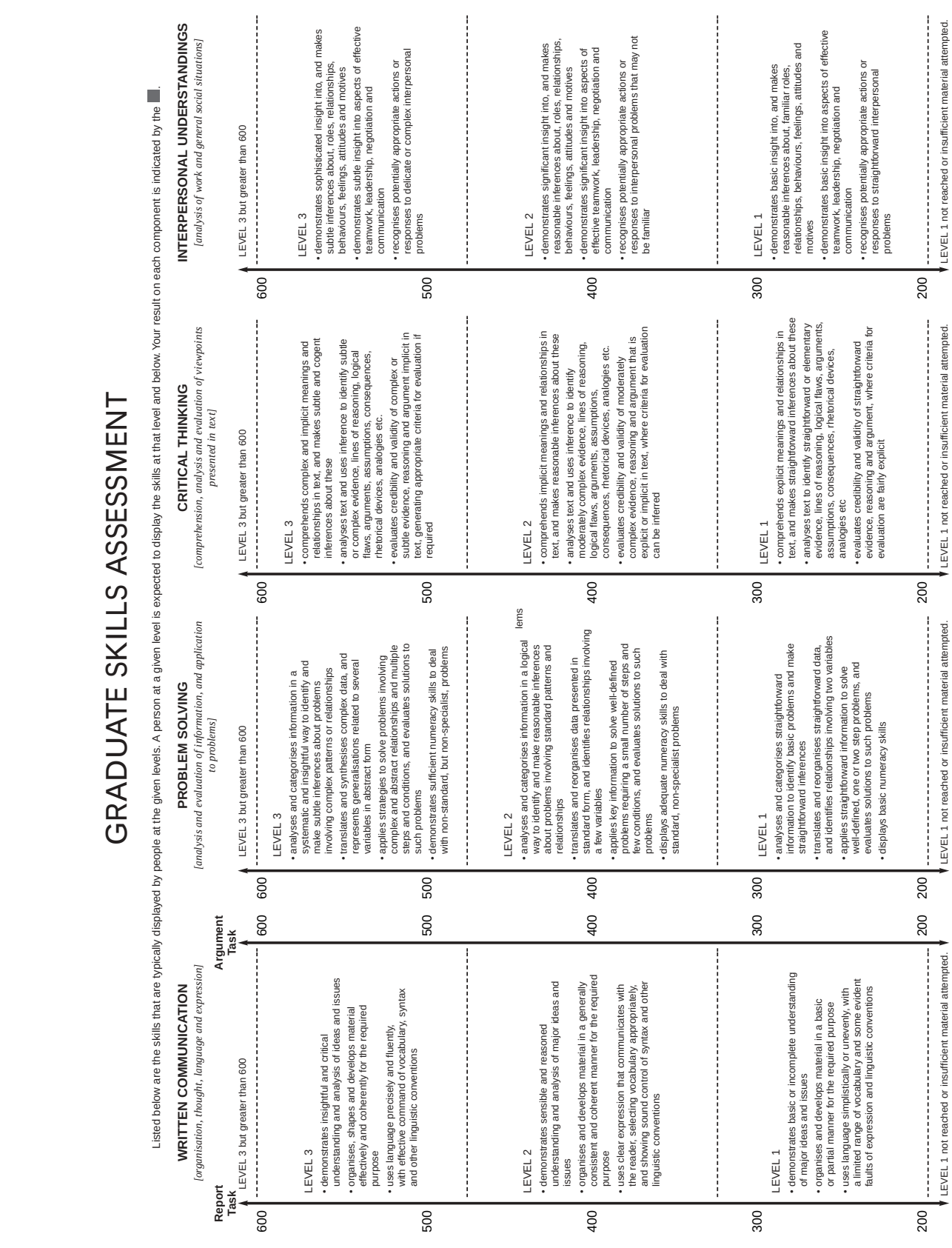


Figure 3.1 Student report for the first version of the GSA

4. The study sample

In order to produce suitable reference ranges, make legitimate comparisons between groups of students and address aspects of test validity, appropriate samples of students are required. This chapter describes characteristics of the sample of students that was used as a basis of the analyses performed for the validity study. It highlights some potential biases related to the sample, which is largely a self-selected group choosing to do the GSA, sometimes influenced in this choice by university departments. Some recommendations are provided on how to interpret the results presented in later chapters.

4.1 The GSA sample population

At present, the GSA is administered as parallel (linked) tests, one version in the first months following a student's admission to university (Entry test) and another in the final months prior to graduation (Exit test). The sample population used in this study was predominantly students who sat one or other of the first two GSA tests, Exit 2000 (mostly students in Year 3 or Year 4) and Entry 2001 (mostly students in Year 1).

Table 4.1 describes the sample of students in terms of field of study (FOS) and year level. Appendix 4 gives finer information about the field of study distribution of students sitting the first two tests. More information is provided later in the chapter, in the Summary Reports for the Exit 2000 and Entry 2001 tests (Hambur & Glickman 2001; Hambur & Le, 2001) and in Chapter 8.

Table 4.1 suggests that the field of study and year level composition of the sample may not be representative of the university population. This issue is examined in the following section.

Table 4.1 Number of students in field of study by year of course, exit 2000 and entry 2001 cohorts

	Year Level									
FOS	1	2	3	4	5	6	6+	NP*	All	%
Arts/ Humanities	152	7	180	95	25	7	3	14	483	(13.2)
Business/ Commerce	289	33	336	84	23	3	6	32	806	(22.0)
Computers/IT	175	12	80	32	9	1	1	12	322	(8.8)
Education/ Social	148	4	7	52	2	1	2	4	220	(6.0)
Engineering/ Architecture	179	5	7	52	26	12	8	4	293	(8.0)
Science/Maths	469	8	184	127	13	7	2	18	828	(22.6)
Law/Legal	19	3	8	9	14	0	1	1	55	(1.5)
Medicine/ Dentistry	315	0	1	1	12	7	0	5	341	(9.3)
Nursing	182	1	10	2	0	0	0	5	200	(5.5)
No Data/ Other	62	3	30	7	3	0	2	8	115	(3.1)
Total	1990	76	843	461	127	38	25	103	3663	(100.0)
%	(54.3)	(2.1)	(23.0)	(12.6)	(3.5)	(1.0)	(0.7)	(2.8)	(100.0)	

* NP means not provided.

4.2 GSA sample compared with the general university population

Students provided data by filling in a special section of the multiple-choice answer sheet. As suggested by Table 4.1, there was missing data. In addition, there is doubt that all students provided accurate data. Nevertheless, there appears to be sufficient appropriate data to make broad comparisons between the GSA sample and the general university population in order to identify potential biases.

4.2.1 Distribution graphs

Figures 4.1 to 4.7 compare broad characteristics of the combined GSA Exit 2000/Entry 2001 sample population with the total university population in 2000 (DETYA, 2001). For each figure comments are made comparing the populations. However, these need to be checked statistically (see 4.2.2).

Figure 4.1 compares the populations on the basis of age. The GSA population appears to be over-represented in the age groups younger than 25, and in the 30-39 age group, but under-represented in the 25-29 and 40 plus age groups.

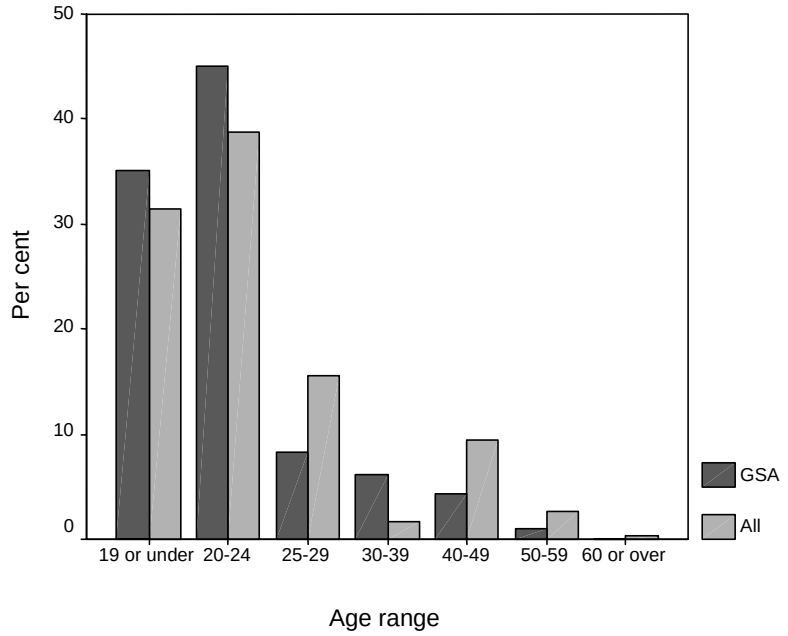


Figure 4.1 Age distributions of GSA and university populations

Figure 4.2 compares the populations on the basis of gender. Males appear to be slightly under-represented in the GSA population.

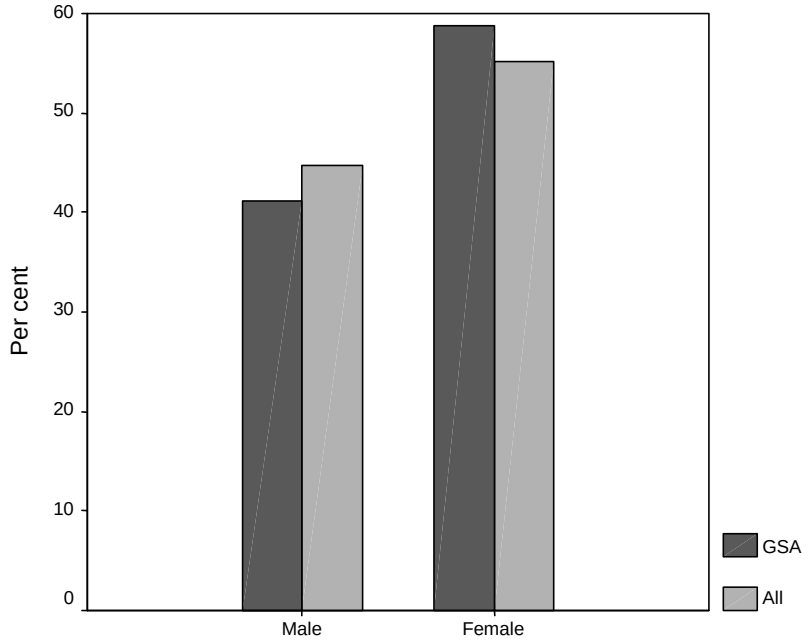


Figure 4.2 Gender distributions of GSA and university populations

Figure 4.3 suggests that full-time students are over-represented in the GSA population.

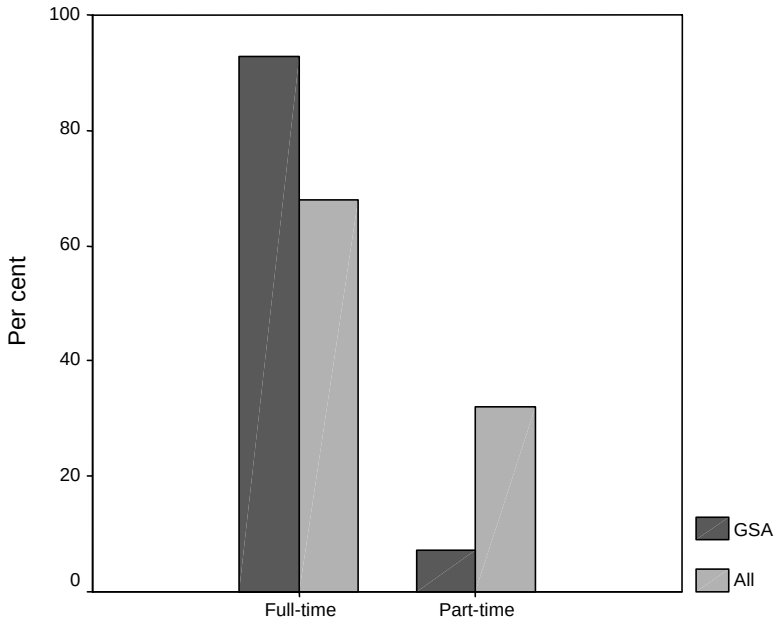


Figure 4.3 Full-time/part-time distributions of GSA and university populations

Figure 4.4 suggests that the proportions of undergraduates and post-graduates in the GSA samples is similar to that in the general university population, though post-graduates may be slightly over-represented in the GSA sample.

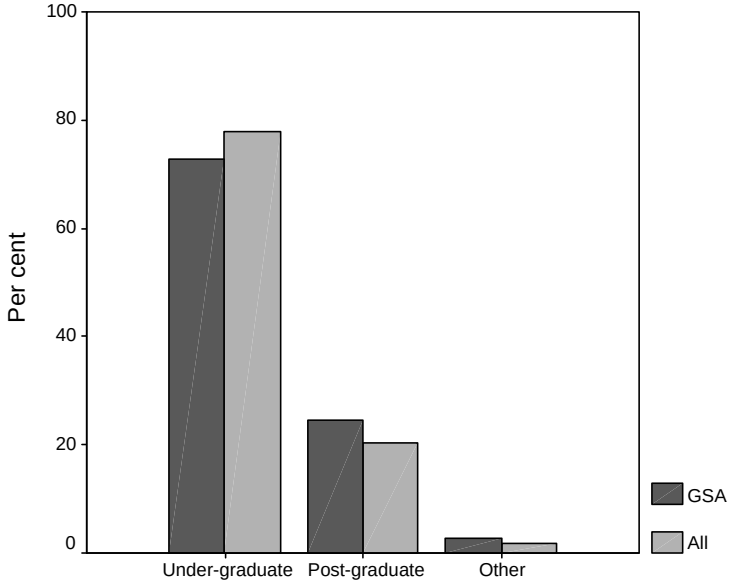


Figure 4.4 Undergraduate/post-graduate distributions of GSA and university populations

Figure 4.5 suggests that Science and Health students are over-represented in the GSA samples, while Arts/Humanities/Social Science, Law/Legal and Education students are under-represented. However, the way courses are combined to form GSA fields of study may vary to some extent from the method used for classifying the general university population.

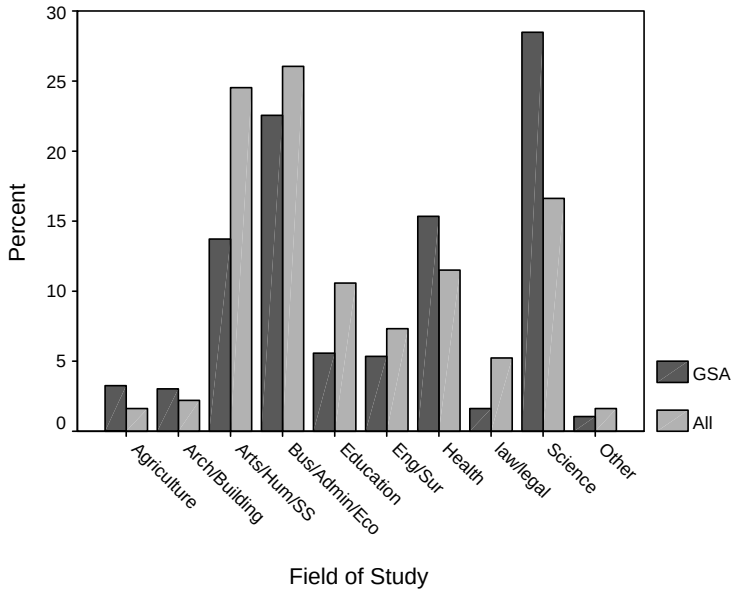


Figure 4.5 Field of study distributions of GSA and university populations

Figure 4.6 suggests that students from non-english speaking backgrounds (NESB) are over-represented in the GSA samples. (This observation may be consistent with the relatively high proportion of Science and Health students, and relatively low proportion of Arts/Humanities/Social Science and Education students sitting the GSA.)

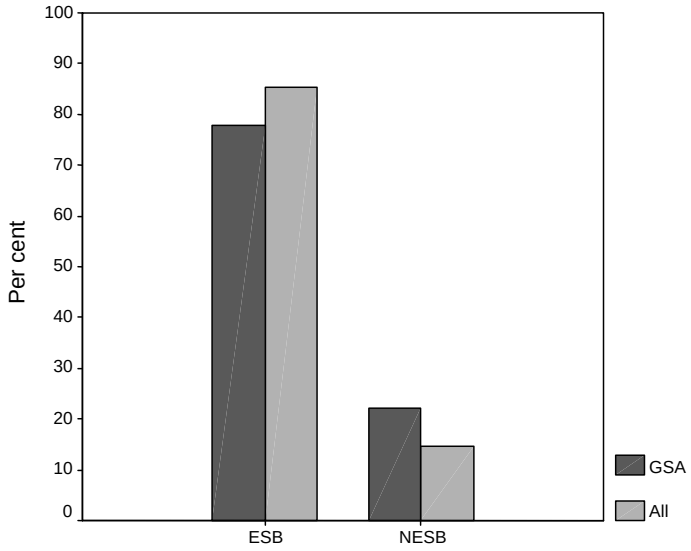


Figure 4.6 ESB/NESB distributions of GSA and university populations

Figure 4.7 suggests that relatively more students who reside in Australia sat the GSA as compared with the general university population. This is not surprising considering the way the GSA is currently administered.

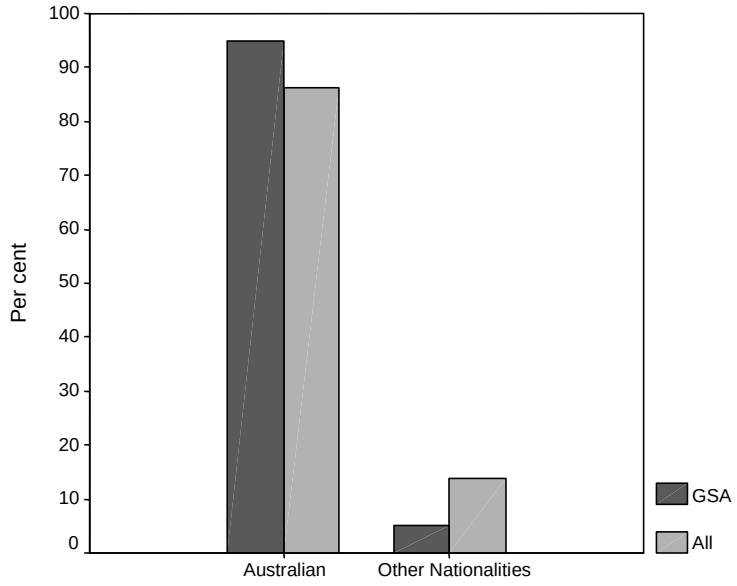


Figure 4.7 Nationality distributions of GSA and university populations

4.2.2 Statistical analyses

Although the preceding graphs suggest certain differences between the characteristics of the GSA and total university population, statistical analyses are required to confirm these.

Sampling theory allows computation of the likelihood of a difference between a population parameter and a sample estimate. Due to the hierarchical structure of the data, a classical method[7] should not be used to compute the uncertainty related to a sample statistic[8]. Indeed, it is important to account for the hierarchical structure of the data when estimating standard errors.[9]

Taking this approach, Table 4.2 provides results of the statistical tests of the comparisons presented in Figures 4.1 to 4.7. In this table, the various categories for a characteristic match those in the corresponding figures in order, and are represented by a number in the first column. For example, Category 1 for Age is 19 years or under.

[7] Software packages like SPSS or SAS provide standard errors on an estimate under the assumption of a simple and random sample. These standard errors are inappropriate for a cluster sample and usually underestimate the standard error.

[8] Indeed, two students within one university are more likely to be similar than two students from different universities. A way to express this uncertainty is related to the estimate of gender percentages in the GSA sample. Nursing students are mainly females in the GSA sample; there are about 200 nursing students who predominantly come from three centres. The exclusion of one of these centres would significantly affect the percentage of males and females in the sample.

[9] Two methods in this particular case are available: multi-level software like Mlwin or HLM and software packages based on replication methods.

Table 4.2 Statistical comparisons of GSA and university populations

Category		Est	SE(est)	Pop	T statistic	Significance
Age	1	35.19	4.599	31.44	-0.81539	0
	2	45.13	4.24	38.75	-1.50472	0
	3	8.29	0.869	15.64	8.457998	1
	4	6.1	0.84	1.73	-5.20238	1
	5	4.26	1.016	9.4	5.059055	1
	6	0.95	0.223	2.65	7.623318	1
	7	0.08	0.064	0.39	4.84375	1
Gender	1	41.2	2.745	44.8	1.311475	0
	2	58.8	2.745	55.2	-1.31148	0
Full/Part	1	91.64	1.456	67.97	-16.2569	1
	2	8.36	1.456	32.03	16.25687	1
Graduate	1	72.71	4.76	77.75	1.058824	0
	2	24.57	4.772	20.48	-0.85708	0
	3	2.72	0.566	1.77	-1.67845	0
FOS	1	3.23	1.138	1.50	-1.52292	0
	2	2.95	2.471	2.08	-0.35267	0
	3	13.59	2.456	22.88	3.784011	1
	4	22.42	4.291	24.26	0.429621	0
	5	5.57	2.732	9.90	1.586451	0
	6	5.29	1.618	6.83	0.949277	0
	7	15.27	5.587	10.72	-0.81483	0
	8	1.55	0.574	4.88	5.80779	1
	9	29.06	2.73	15.51	-4.96274	1
	10	1.07	0.324	1.43	1.124732	0
ESB	1	77.76	2.742	85.45	2.804522	1
	2	22.24	2.742	14.55	-2.80452	1
Nationality	1	95.26	0.834	86.25	-10.8034	1
	2	4.74	0.834	13.75	10.80336	1

In summary, the data in Table 4.2 indicate that:

1. there are no significant differences between the sample and university populations for students under 25, but older students are under-represented in groups aged 25 and over, except for the 30-39 group where they are under-represented;
2. there are no significant differences between the populations on the basis of gender.
3. part-time students are under-represented in the GSA sample;
4. there are no significant differences between the populations on the basis of undergraduate/post-graduate;

5. Arts/Humanities and Law/Legal are under-represented in the GSA sample but Science students are over-represented;
6. students with an English-speaking background are under-represented in the GSA sample; and
7. Australian students are over-represented in the GSA sample.

4.3 Concluding comments

Based on these results, it is clear that the GSA sample cannot be considered a random and representative sample and, therefore, appropriate care should be taken when interpreting the results of this study.

On the other hand, statistical methods based on linear relationships, such as correlation and linear regression, are less affected by such biases than other methods. Therefore, it is expected that general findings related to the factor structure of the test (Chapter 5), variables related to performance on the test (Chapter 6) and the relationship between performance on GSA and other measures of achievement (Chapter 7) will have validity

.

5. Factor structure and discriminant validity

5.1 Introduction

A key purpose of the GSA is to assess the ability of students to select appropriate skills from their repertoire and apply them to complex, authentic tasks related to five domains: Critical Thinking (CT), Problem Solving (PS), Interpersonal Understandings (IP), Argument (ARG) and Report Writing (REP). Interpreted broadly, there could be considerable overlap between the skills assessed by these domains. For example, if both Critical Thinking and Problem Solving focussed on verbal skills, performance on one may be equivalent to performance on the other.

Hence, the focus of the domains was narrowed as described in Chapter 3 in order to produce five psychometrically distinct test components, each addressing a distinct cognitive dimension (i.e. in order to have *discriminant* validity). If the test components were not sufficiently psychometrically distinct (i.e. there was evidence of excessive overlap between the skills addressed by two or more components), it could be argued that the components should be combined.

Hence, confirmatory factor analysis (CFA) was undertaken to assess the dimensional factor structure of the test in terms of the five target domains (i.e. its discriminant validity).

5.2 Inter-component correlations

At a gross level, the dimensional factor structure of a test may be suggested by component score correlations and their reliabilities. Table 5.1 records the correlations between component scores (Pearson product-moment) for the GSA Entry 2001 cohort, in which just over 2000 students from a wide variety of fields of study participated. These students were predominantly in their first year of study. Details of the student sample are indicated in Table 4.1 and, more specifically, in the Summary Report for this test (Hambur & Le, 2001).

Table 5.1 Component score correlations – entry 2001 cohort

	CT	PS	IP	REP	ARG
CT	1.00	-	-	-	-
PS	0.55	1.00	-	-	-
IP	0.65	0.50	1.00	-	-
REP	0.37	0.29	0.37	1.00	-
ARG	0.46	0.31	0.43	0.46	1.00

The magnitudes of these correlations are similar to those for the GSA Exit 2000 cohort, and, given the reliabilities reported in Table 5.2, suggest that the components differ significantly in psychometric focus and, therefore, address distinct cognitive dimensions.

Table 5.2 gives the reliability values (Cronbach's alpha) for the multiple-choice components of the test.

Table 5.2 Reliability coefficients for multiple-choice components – all students

	Reliability Exit 2000	Reliability Entry 2001
PS	0.83	0.82
IP	0.81	0.79
CT	0.81	0.78

It should be noted that reliability is partly related to the number of items/score points used for 'measuring' the dimension. Given the component length, the reliability values obtained are quite good, and suggest that the items in each component have an acceptable precision in terms of what is being measured. In addition, the reliability estimates are consistent between tests.

Given the measurement error inherent in such analyses, the correlation and reliability estimates nevertheless suggest that the GSA is made up of relatively consistent and distinct dimensions (i.e. has discriminant validity). This was checked by factor analysis as described in the next section.

5.3 Confirmatory factor analysis

To test the dimensional factor structure of the GSA suggested by the previous analysis, a confirmatory five-factor model was fitted to the data for the Entry 2001 test using LISREL 8.30 (Jöreskog & Sörbom, 2000).[10]

Analyses were undertaken at the unit level for the multiple-choice components (i.e. items were analysed in small sets or units, each based on a common introductory stimulus) or at the criterion level for the writing components (i.e. there were three criteria per component task). For brief descriptions of the units/criteria see Appendix 3.

5.3.1 Confirmation of the five dimension GSA construct

Table 5.3 presents the completely standardised CFA solution, showing loadings of the units/criteria on the five dimensions/factors. In the table, unit names are preceded by a prefix indicating the dimension that the unit/criterion was intended to assess. It can be seen that PS (Problem Solving) units load predominantly on PS, CT (Critical

[10]This method of analysis takes into account inter-item/component variances-covariances, as well as measurement error at the item and component levels.

Thinking) units on CT, and so on, indicating that the multiple-choice units and writing criteria are well targeted.

Although each component is unlikely to focus totally on a single, pure dimension/factor (because of the 'authentic' complexity of the tasks), the results indicate that the units/criteria used for the five dimensions consistently measure sufficiently distinct skills, suggesting no need to collapse or combine components.

Table 5.4 gives the Goodness of Fit Indices for the five-factor model, indicating that the model is an excellent fit to the data, with approximately 99% of the variances and co-variances in the data being accounted for by the fitted model.

Table 5.3 Five-factor completely standardised solution - GSA entry 2001

Unit/Criterion	PS	CT	IP	REP	ARG
PSshor	0.713	- -	- -	- -	- -
PSstaf	0.686	- -	- -	- -	- -
PStrav	0.660	- -	- -	- -	- -
PShous	0.655	- -	- -	- -	- -
PSchd3	0.624	- -	- -	- -	- -
PSbrid	0.555	- -	- -	- -	- -
PSsale	0.553	- -	- -	- -	- -
PSbatt	0.492	- -	- -	- -	- -
PSlaw	0.346	0.273	- -	- -	- -
PSchd2	0.326	0.355	- -	- -	- -
CThux	- -	0.741	- -	- -	- -
CTviol	- -	0.728	- -	- -	- -
CTsaw	- -	0.662	- -	- -	- -
CTstat	- -	0.630	- -	- -	- -
CTcorp	- -	0.604	- -	- -	- -
CTdrug	- -	0.594	- -	- -	- -
CTsued	- -	0.568	- -	- -	- -
CTshor	- -	0.538	- -	- -	- -
CTcit	- -	0.469	- -	- -	- -
IPdrm	- -	- -	0.641	- -	- -
IPt6	- -	- -	0.620	- -	- -
IPlist	- -	- -	0.614	- -	- -
IPshor	- -	- -	0.609	- -	- -
IPt4	- -	- -	0.605	- -	- -
IPdoct	- -	- -	0.572	- -	- -
IPt5	- -	- -	0.567	- -	- -
IPdrjo	- -	- -	0.493	- -	- -
IPemp	- -	- -	0.483	- -	- -
IPeng	- -	0.361	0.362	- -	- -
IPang	- -	- -	0.256	- -	- -
IPmar	- -	- -	0.218	- -	- -
WRRep2	- -	- -	- -	0.978	- -
WRRep1	- -	- -	- -	0.946	- -
WRRep3	- -	- -	- -	0.771	0.188
WRArg2	- -	- -	- -	- -	0.978
WRArg1	- -	- -	- -	- -	0.965
WRArg3	- -	- -	- -	- -	0.916

Notes on Table 5.3:

With respect to Table 5.3, it is notable that the Writing criteria (1 - Thought and Ideas, 2 - Structure and Organisation, and 3 - Language and Expression) load together on the individual writing task (Report or Argument), suggesting that assessed student performance on these elements tends to be global and task specific.

The results presented in Table 5.3 indicate that unit CThux is most strongly associated with the CT dimension, PSshor with PS, and IPdrm with IP. While loading significantly on their target dimensions, a few units/criteria also cross-load significantly with other dimensions. Test developers should consider the value of units that cross load.

Two units (IPang and IPmar) performed differently on the test compared with the trial and were excluded from the test-reporting phase by conventional criteria prior to factor analysis. Factor analysis shows them to have low loadings on their target dimensions, confirming the decision to drop these units.

Table 5.4 Model goodness of fit indices for the five factor model, GSA entry 2001

Goodness of Fit Index	0.995
Adjusted Goodness of Fit Index	0.994
Standardized RMR	0.079
Comparative Fit Index	0.996
Incremental Fit Index	0.996
Relative Fit Index	0.993

Thus, confirmatory factor analysis suggests that student performance on the GSA is consistent with the operation of five correlated dimension factors, with each of the five dimensions being related to a set of distinctive skills. This observation is significant evidence in favour of the discriminant validity of the five-dimension GSA test construct.

5.3.2 Second-order factor

A second-order CFA model was fitted to the data. In summary form, Figure 5.1 presents diagrammatically the second-order solution indicating the contribution of each of the five factors to a second-order general factor, *G*.

Table 5.5 provides the Goodness-of-Fit indices for the fitted second-order model (i.e. five first-order factors and one second-order general factor, *G*), indicating that the fit of this model to the data is excellent.

Table 5.5 Model goodness of fit indices – five first-order factors plus one second-order factor, GSA entry 2001

Goodness of Fit Index	0.995
Adjusted Goodness of Fit Index	0.994
Standardized RMR	0.089
Comparative Fit Index	0.996
Incremental Fit Index	0.996
Relative Fit Index	0.992

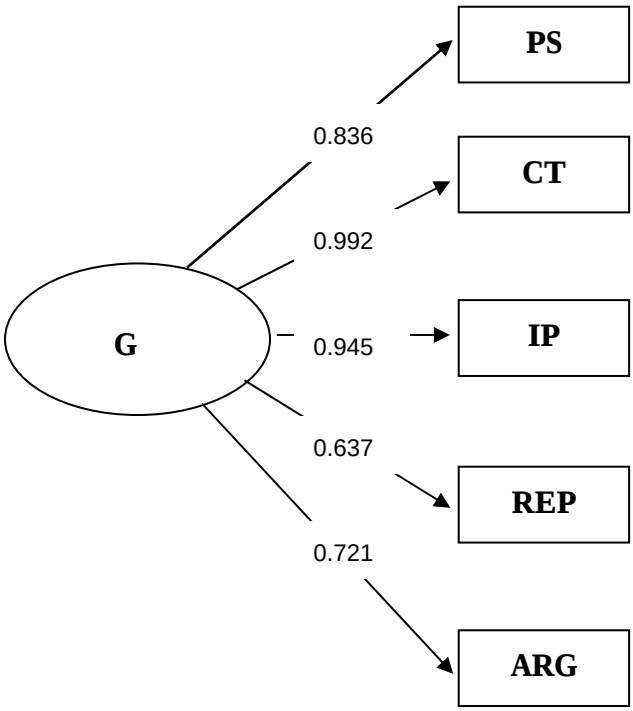


Figure 5.1 Second-order factor solution, GSA entry 2001

Notes on Figure 5.1:

Perhaps because it is most related to a kind of meta-cognitive general reasoning skill, Critical Thinking has the highest loading on the general factor (0.992).

Even though both CT and IP load highly on the general factor, and the correlation between student performances on these two components is around 0.7, confirmatory factor analysis was able to discriminate clearly between these two components. Further, some field of study groups perform quite differently on these components. For example, nurses generally perform poorly on CT compared with other students, but this is not the case for IP, where their performance is at least average (Chapter 6). Thus, there is a justification for maintaining CT and IP as separate components in keeping with test design, though test developers may consider whether and how to make them more differentiated.

Given the apparent improvement of performance of students on all GSA dimensions between Year 1 and Year 3 (Chapter 6), it might be hypothesised that the skills related to the general factor are being developed by the university experience, though other experiences may be involved. (Further, the apparent improved performance between Year 1 and Year 3 needs to be confirmed with observations on the same students between years or on properly matched samples, which are not yet available.)

It might be speculated that a separate score on the general factor could be calculated but this was not within the brief of GSA test development and there are theoretical objections and methodological problems related to such a procedure. Similarly, there are theoretical objections and methodological problems related to the removal of the effect of the general factor from the five dimensional factors.

Thus, confirmatory factor analysis suggests that student performance on the GSA is consistent with the operation of a general factor and five correlated dimension factors, with each of the five dimensions being related to a set of distinctive skills.

5.4 Maintenance of GSA dimensional structure

Although the discriminant validity of GSA appears to be satisfactory for GSA Entry 2001 in that five dimensional factors are observed, there is evidence that the fine structure and balance of test components can change without deliberate changes to the construct (Hambur 1997, 1998). This may result in different outcomes for students (e.g. with respect to relative performance on the basis of gender) with supposedly parallel tests.

This problem can occur with tests of higher-order complex skills such as GSA because, although each component is dominated by a single major factor, outcomes from such tests are also sensitive to minor factors. Unless carefully monitored, the focus with respect to the major and minor factors may not be consistent from test to test.

Test developers should monitor this phenomenon by routinely applying factor analysis and by the use of other checks (e.g. monitoring relative gender performance).

5.5 Inter-component correlations and field of study

Table 5.6 presents the component score correlations at the student level for GSA Exit 2000 and Entry 2001 combined, which are similar to those given in Table 5.1.[11]

Table 5.6 Student level correlation matrix - all students

Domain	CT	IP	PS	ARG	REP
CT	1.00	-	-	-	-
IP	0.66	1.00	-	-	-
PS	0.55	0.48	1.00	-	-
ARG	0.41	0.41	0.27	1.00	-
REP	0.35	0.36	0.27	0.43	1.00

Table 5.7 presents the correlations among the GSA component scores at the field of study level for the combined GSA Exit 2000 and Entry 2001 (i.e., correlations are observed within fields of study).

Table 5.7 Field of study level correlation matrix - all students

Domain	CT	IP	PS	ARG	REP
CT	1.00	-	-	-	-
IP	0.97	1.00	-	-	-
PS	0.70	0.57	1.00	-	-
ARG	0.97	0.99	0.52	1.00	-
REP	0.99	0.96	0.69	0.97	1.00

[11] This analysis was done following a multivariate, multilevel analysis of students' component scores (level-1) within students (level-2) and within fields of study (level-3).

The magnitudes of the correlations are notably larger at the field of study level. This finding is typical of such data because students within a given field of study tend to be more similar in terms of their abilities and skill sets than students from a range of fields.

The results presented in Tables 5.6 and 5.7 are consistent with the observation that field of study is an important variable with respect to GSA score (Chapter 6).

Table 5.7 suggests that the dimensional factor structure of the GSA may be effectively different for students within in a field of study. The GSA is designed for all students so the dimensional factor structure presented in Table 5.3 is most relevant but it may be of interest to analyse factor structure within fields of study in future.

In particular, high correlations at the field of study level related to Critical Thinking, Interpersonal Understandings and Writing suggest that students within a field have a similar set of generic skills and approaches relevant to these domains. The lowest correlations at the field of study level occur with Problem Solving, suggesting that students within a field vary most in Problem Solving skills and approaches.

It is interesting that in a small concurrent validity study with the Graduate Management Admissions Test (GMAT), Problem Solving was the only GSA component that correlated significantly with scores on GMAT (Chapter 7).

This finding may be associated with observations that a student's school performance in numeracy was more important than performance in literacy as a predictor of tertiary entrance score (Marks et al., 2001) and that student performance on the quantitative component of the Australian Scaling Test (actually logical, analytical and quantitative reasoning) was a better predictor of success in first year university than the verbal component, even for humanities students (Everett & Robins, 1991).

On the basis of such observations, it might be speculated that students within a field of study vary most in 'non-verbal' reasoning skills that appear to be related to academic success. However, such observations and suggestions require further investigation before firm conclusions can be drawn.

5.6 Concluding comments

Confirmatory factor analysis supports a five-dimension factor structure consistent with the intended test design. Although the analyses presented here support the GSA's discriminant validity, measures need to be taken to monitor and maintain the dimensional structure from test to test.

6. Variables related to performance on GSA

6.1 Introduction

This chapter examines variables that could be related to student performance on the five GSA dimensions, such as field of study, gender, year level, English-speaking background and age. It is important to carry out such an investigation because, if the GSA is operating properly, it would be expected, for example, that performance is related to year level and variables such as gender are not inappropriately related to performance. Data collected for the first two GSA tests, Exit 2000 and Entry 2001, were used for these analyses.

6.2 Relationship between GSA score, field of study and year level

Student performance on the GSA is expected to relate to variables such as field of study and year level. A relationship is expected with field of study because students in certain fields are expected to have certain generic skill strengths (for example, Humanities students would be expected to have relatively strong writing and verbal reasoning skills). A relationship is expected with year level because the GSA aims to assess generic skills that can be developed by the university experience and are related to performance at university. Observation of such relationships could provide evidence in favour of the validity of the GSA.

6.2.1 GSA score means by field of study

Figures 6.1 to 6.5 provide plots of point estimates of students' mean scores (bounded by 95% confidence intervals) for each GSA dimension. These are provided for students in each field of study who participated in one of the first two GSA tests (Exit 2000 and Entry 2001)[12]. Appendix 4 indicates how various courses were assigned to the broad fields of study used in reporting results.

[12] It is possible to undertake such analyses because the two tests have been equated, using common link items and score transformations, so that students undertaking different tests can be located on common dimension scales.

Separate plots are provided for each GSA dimension since the items relevant to each dimension were calibrated independently of those relevant to other dimensions. This provision also minimises the risk of invalid comparisons across dimensions within any given field of study.

The figures show distinctive profiles of performance related to field of study. Similar profiles are also evident when Entry 2001 (predominantly Year 1) or Exit 2000 (predominantly Years 3 and 4) cohorts are analysed separately.

The 'levels' shown on the plots are those described on the Student Report (see Figure 3.1). Each of the 9 fields of study groups considered had means within the Level 2 range, which covers approximately the middle 60% of all student scores.

Since the sampling characteristics of the two student cohorts are neither random nor perfectly representative of the university student population or the fields of study, it is important not to over-interpret the statistical 'significance' of confidence intervals presented in the figures. Moreover, since statistical significance is a function of sample size, the under-representation of students at both the university level and the field of study level is problematic.

Given this precaution, however, the distinctive profiles of performance related to the nine fields of study do seem meaningful. For example, Arts/Humanities students do relatively well on Critical Thinking, Interpersonal Understandings and Writing, but do not perform as well on Problem Solving. Engineering/Architecture students perform well on Problem Solving but not on Interpersonal Understandings. Nurses do not perform well on Critical Thinking or on Problem Solving, but are average on Interpersonal Understandings. Law and Medical students perform well on all dimensions.

The differential performances of students on the GSA dimensions across the fields of study may be partly explained by the relationship between academic ability, tertiary entrance score and course selection, to the extent that students with similar abilities, skills and cognitive styles may pursue similar courses and/or be selected into similar courses.

More work on representative samples (preferably using more closely defined groupings of students by field of study/course) is required to verify and elucidate generalisations and speculations such as these. However, the profile of performance on the basis of field of study observed with current data appears reasonable with respect to test validity.

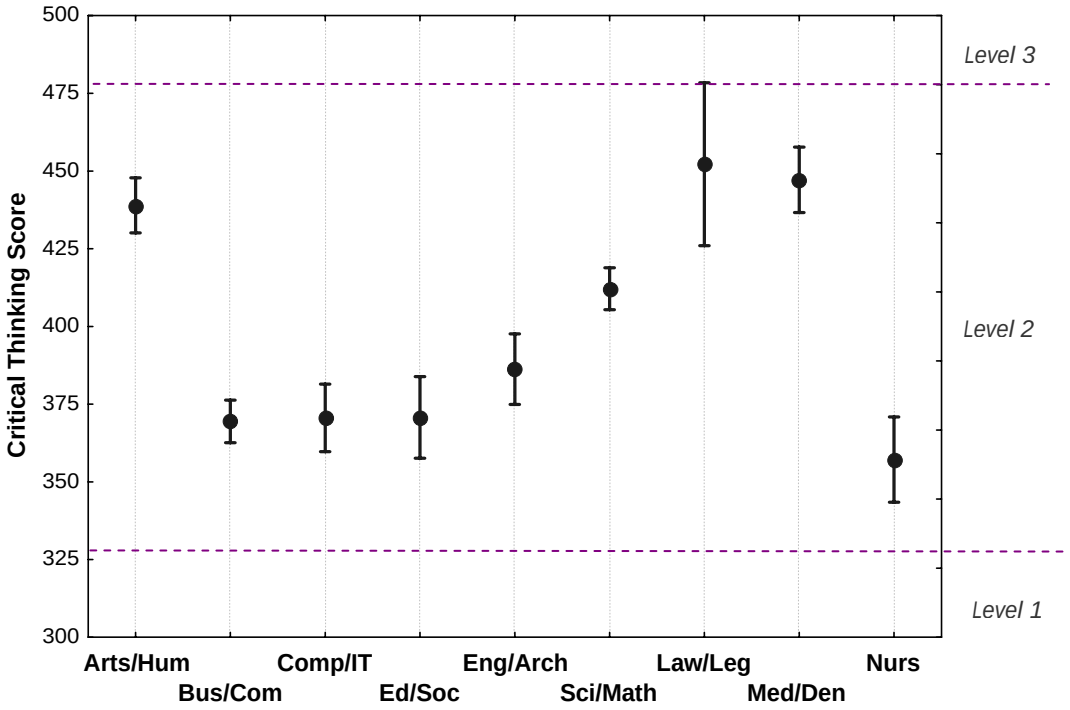


Figure 6.1 All students' means by field of study for *Critical Thinking* (CT), bounded by 95% confidence intervals

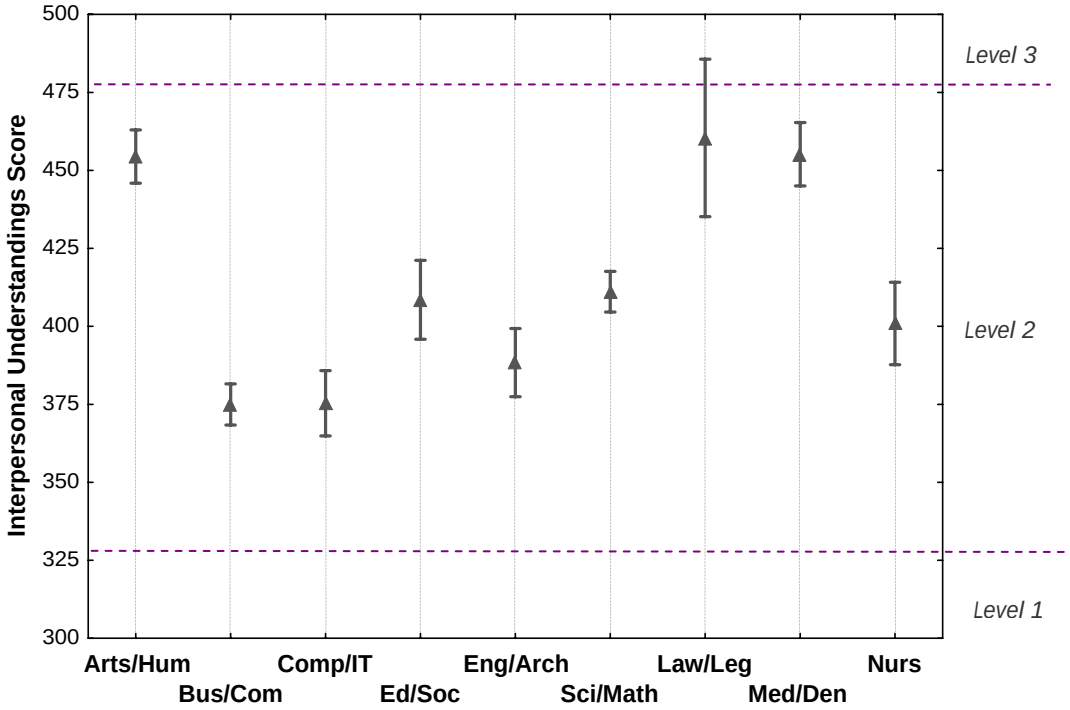


Figure 6.2 All students' means by field of study for *Interpersonal Understandings* (IP), bounded by 95% confidence intervals

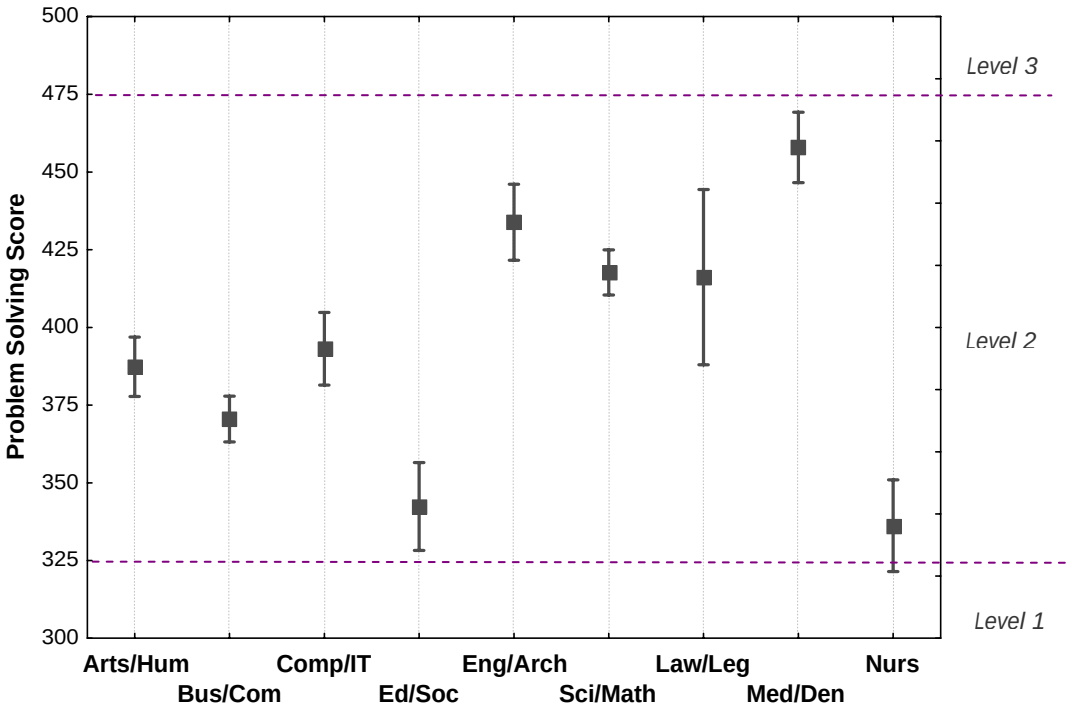


Figure 6.3 All students' means by field of study for *Problem Solving* (PS), bounded by 95% confidence intervals

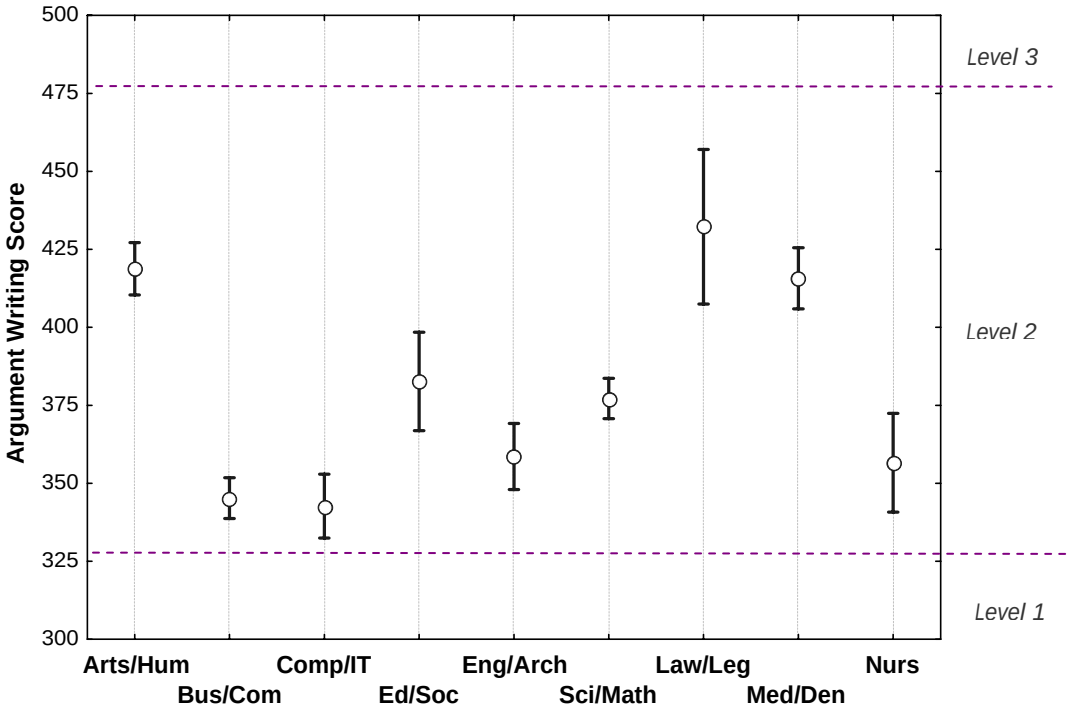


Figure 6.4 All students' means by field of study for *Argument Writing* (ARG), bounded by 95% confidence intervals

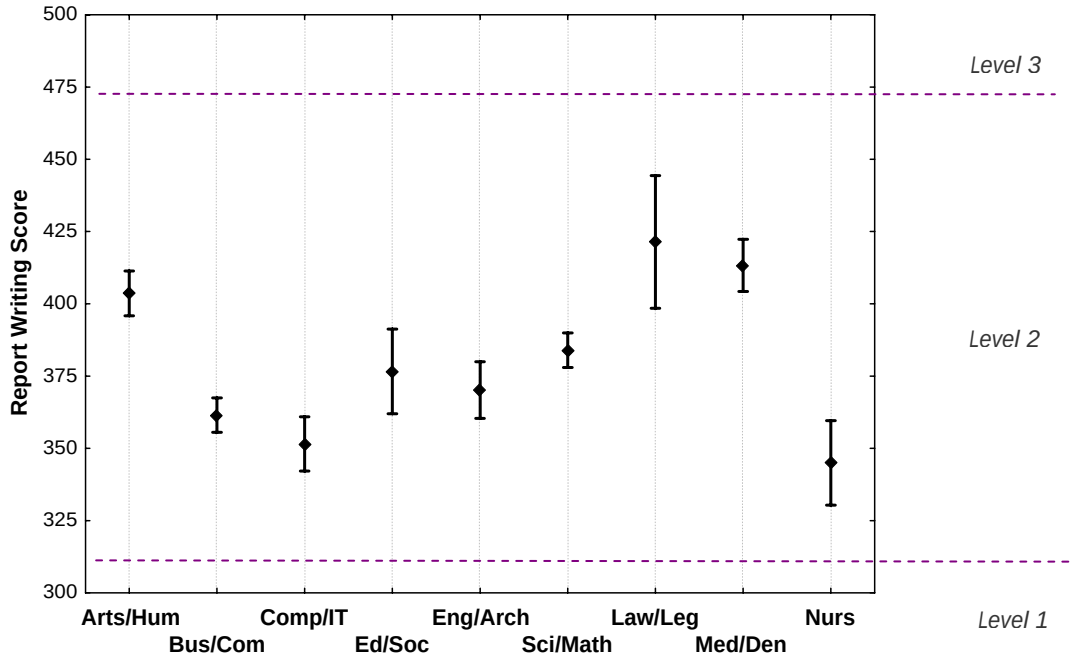


Figure 6.5 All students' means by field of study for *Report Writing* (REP), bounded by 95% confidence intervals

6.2.2 GSA score means by year level for selected fields of study

The data summarised in Figures 6.6 to 6.10 indicate GSA dimension score means and 95% confidence intervals just for those first and third year students who participated in the Exit 2000 or Entry 2001 tests who were known to be undertaking their first degree. Three fields of study (with the largest student representations) were selected for this analysis: Arts/Humanities (119 first year students, 91 third year students), Business/Commerce (216 first year students, 141 third year students), and Science/Maths (419 first year students, 106 third year students).

In the absence of repeat measures on individual students at the two year levels, the findings presented provide the best comparisons currently available with respect to score change with course year level, and are supported by the results of multivariate, multilevel analysis (Table 6.1). Although the data need to be interpreted carefully because first- and third-year groups are not matched in a systematic way, these results are consistent with significant improvement in student performance on the five GSA dimensions between Year 1 and Year 3. Alternatively, it may be that students with higher GSA skills tend to succeed and stay on at university. This issue needs to be investigated by a comparison of measures from the same students.

The difference between the scores of first- and third-year students suggests that the GSA may be measuring generic skills that may be subject to modification by experience in a relatively short time frame. A change in skill level of university students in such a short time seems more consistent with a change in learned generic skills than in a change in fluid intelligence.[13] However, this issue needs to be clarified.

[13] which is expected to remain reasonably constant in adults over two or three years.

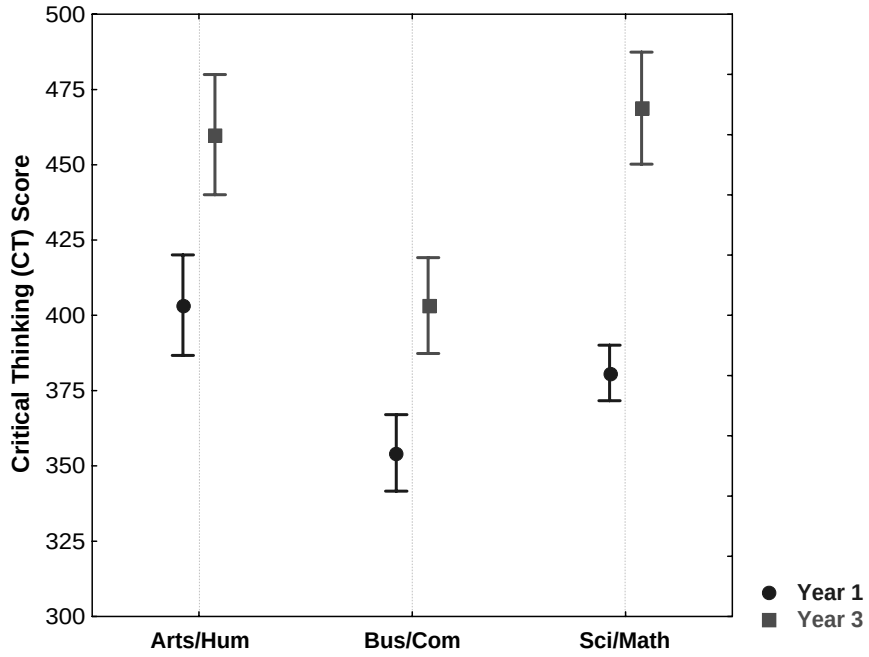


Figure 6.6 Mean scores for *Critical Thinking* (CT), bounded by 95% confidence intervals for first degree students (Year 1 & Year 3) in three fields of study

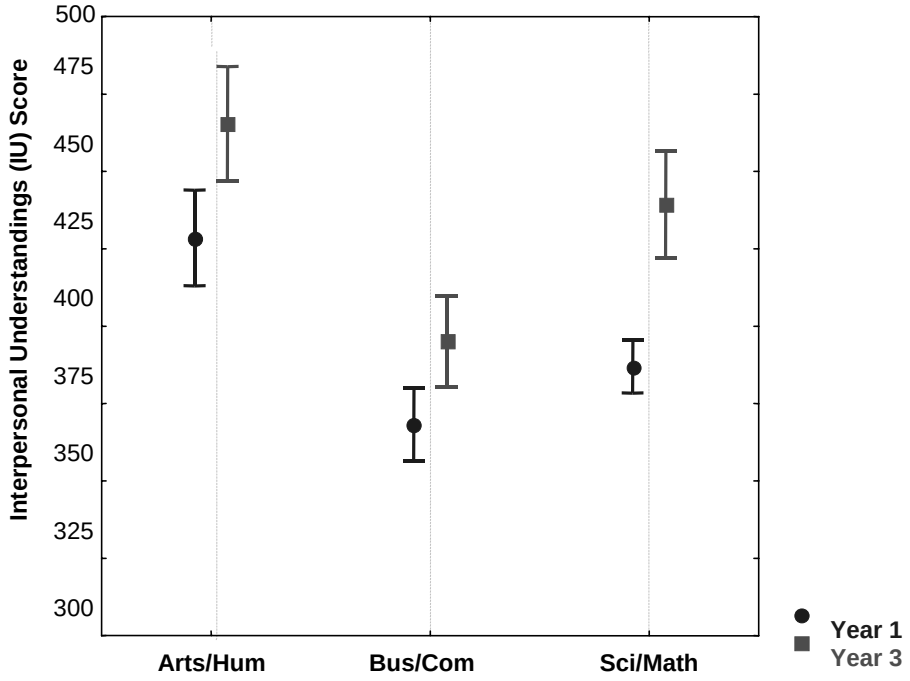


Figure 6.7 Mean scores for *Interpersonal Understandings* (IU), bounded by 95% confidence intervals for first degree students (Year 1 & Year 3) in three fields of study

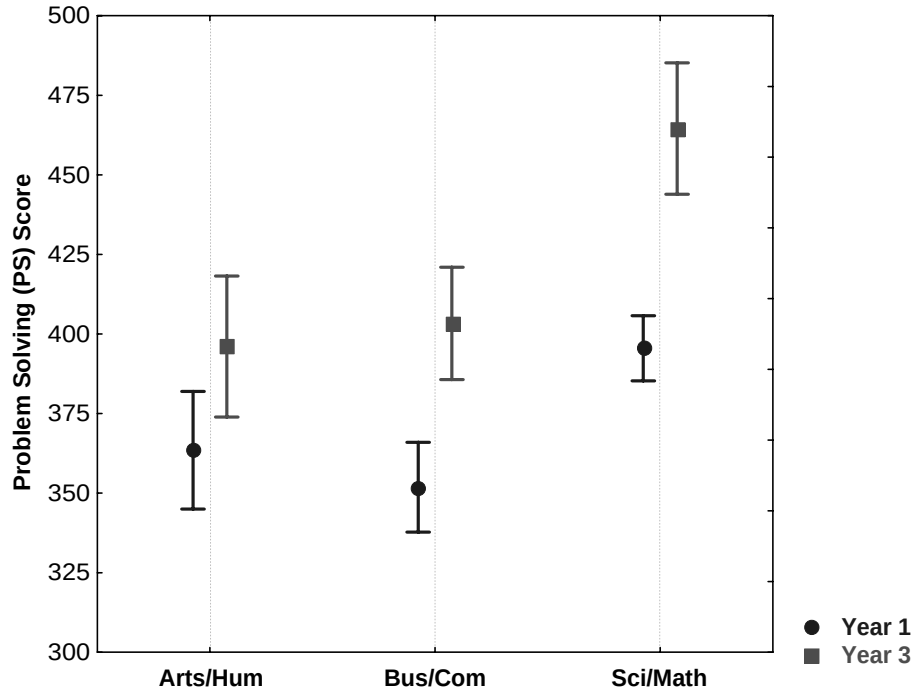


Figure 6.8 Mean scores for *Problem Solving* (PS), bounded by 95% confidence intervals for first degree students (Year 1 & Year 3) in three fields of study

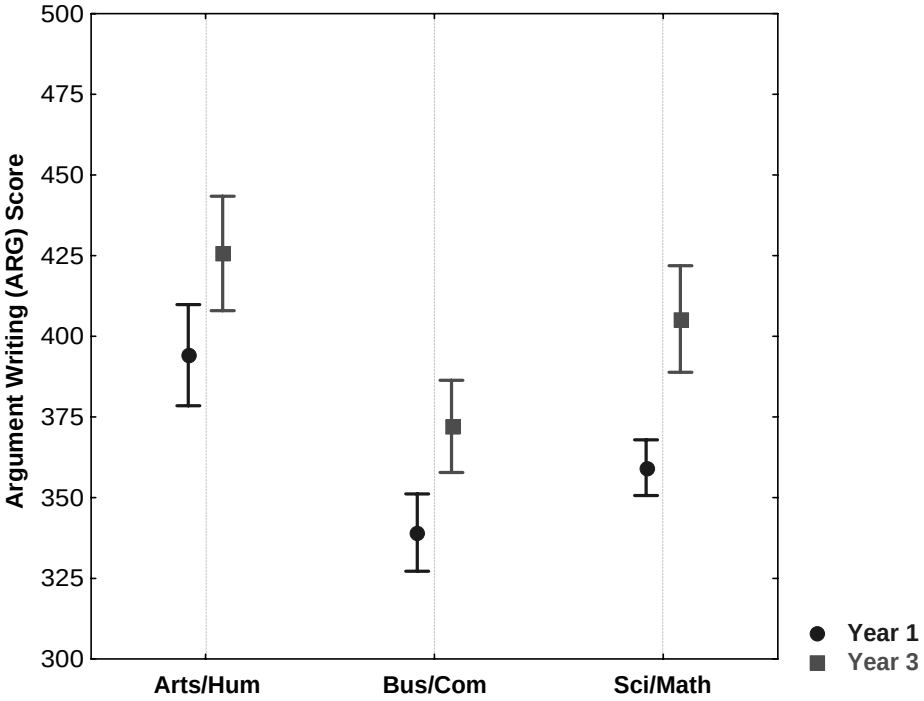


Figure 6.9 Mean scores for A*rgument Writing* (ARG), bounded by 95% confidence intervals for first degree students (Year 1 & Year 3) in three fields of study

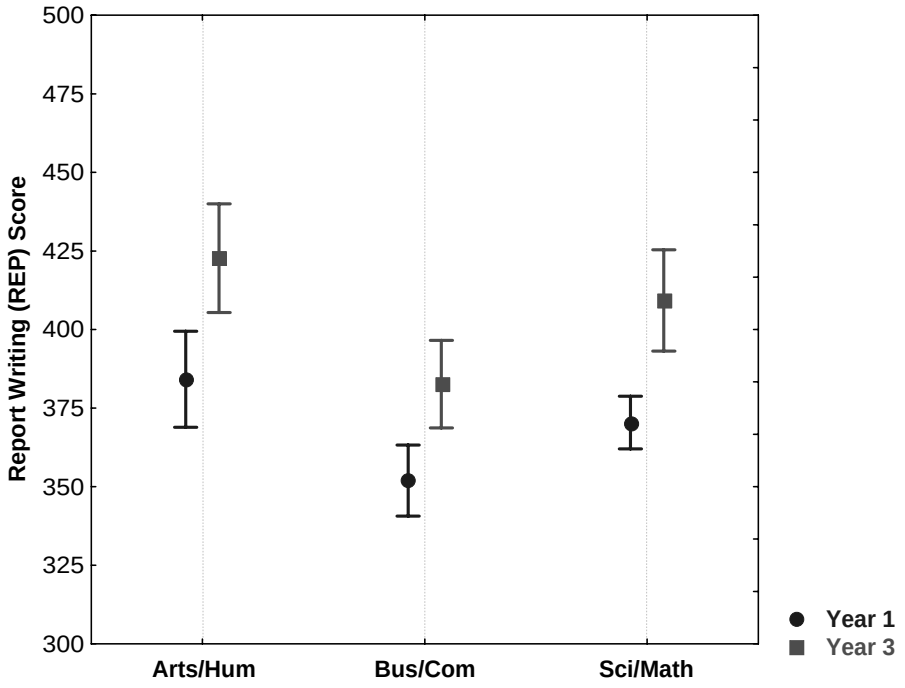


Figure 6.10 Mean scores for *Report Writing* (REP), bounded by 95% confidence intervals for first degree students (Year 1 & Year 3) in three fields of study

6.3 Fitting multilevel models to the data

The previous analyses examined the relationship between explanatory variables (field of study and year level) and GSA dimension scores, one at a time, and ignore the inherent hierarchical (multilevel) structure of the data. That is, these previous analyses are based on fitting single-level models. By contrast, in the following analyses multilevel models are fitted to the data and the effects of several variables can be considered together.[14]

6.3.1 Analysis using GSA total score

A three-level variance-components model was fitted to the data.[15] For convenience, total score was used as a basis of this analysis, though it is recognised that there are problems associated with simple addition of component scores.

The results indicate that 16.5% of residual variance in students' total GSA scores was at the field of study level, and that, when field of study was accounted for, residual variance at the university level was a statistically non-significant 6.5%. (Although there

[14] These models were fitted using both *MLwiN* (Rashbash et al., 2001) and LISREL 8 (Jöreskog & Sörbom, 2000).

[15] students (level-1) within fields of study (level-2) within universities (level-3).

appear to be significant differences between some universities, there is insufficient appropriate data to investigate this or the issue of 'value added' properly.)

This finding indicates that apart from the variation in total score at the student level (77%), most of the variation is at the field of study level.

Given that residual variance in students' total GSA scores is small at the university level when field of study is accounted for, a two-level regression model was fitted to the data (i.e., students within fields of study). In this analysis, using normalised scores, only field of study and several student-characteristic and background variables were included (i.e. *Course Year, ESB, Age, Sex, Course Delivery and School Type).*

In this case, the fitted explanatory variables account for a mere 5.5% of the variance in students' total GSA scores and the stable predictor variables were: *Course Year* (in favour of higher course year levels), *ESB* (in favour of students from English-speaking backgrounds) and *Age* (in favour of younger students). The proportion of residual variance at the field of study level was 11.1%.

6.3.2 Fitting a multivariate, multilevel model to the five GSA dimension scales (normalised scores)

Further analysis was done in which a multivariate three-level model was fitted to the GSA score data for the five components. In this analysis, field of study is accounted for and university is not considered (because it did not appear to be a major variable when field of study is accounted for – though this needs further investigation with appropriate samples). The results are presented Table 6.1.[16] The table indicates significant effects (beyond the 95% confidence level) on students' performances in bold type.

In summary, the results presented in Table 6.1 indicate that:

- English-speaking background (ESB) and Year Level (CRSYR) have significant effects on student performance on all five dimensions (in favour of ESB students and those in higher year levels). Sex is significant for Interpersonal Understandings (IP – in favour of females) and Problem Solving (PS – in favour of males). Age is significant for IP (in favour of mature age students), and for PS and REP (in favour of younger students). Course delivery (full-time/part-time) was significant only for ARG (in favour of full-time students), but could be related to the specific sample.
- School type (SCHTYPE – government, Catholic, independent) was not significant at the 95% confidence level for this sample.
- The fitted variables in the fixed part of the model account for a significant 17% of the multivariate variance in the five GSA dimensions.
- A significant 13% of the residual multivariate variance in students' CT, IP, PS, ARG and REP scores was due to variation between fields of study.

Whereas 30% of the multivariate variance in GSA scores was accounted for by field of study and the fitted variables, 70% of the variance in GSA scores was accounted for by

[16] In this case, scores on the five GSA dimensions (level-1) are clustered within students (level-2) within fields of study (level-3).

variables not fitted in the model. Such variables include student-specific variables such as motivation and student ability on the skill dimensions.

Table 6.1 Parameter estimates (Est.) and standard errors (S.E.) for variables affecting performance on the GSA domains (exit 2000 and entry 2001 students)*

Variable	Est	S.E.	Z-value	$p > \|Z\|$
SEX(CT)	0.016	0.035	0.476	0.6344
SEX(IP)	**0.279**	**0.034**	**8.153**	**0.0000**
SEX(PS)	**-0.215**	**0.035**	**-6.152**	**0.0000**
SEX(ARG)	-0.016	0.035	-0.459	0.6462
SEX(REP)	0.018	0.036	0.515	0.6067
CRSYR(CT)	**0.136**	**0.014**	**9.806**	**0.0000**
CRSYR(IP)	**0.076**	**0.014**	**5.522**	**0.0000**
CRSYR(PS)	**0.175**	**0.014**	**12.450**	**0.0000**
CRSYR(ARG)	**0.091**	**0.014**	**6.425**	**0.0000**
CRSYR(REP)	**0.105**	**0.014**	**7.247**	**0.0000**
CRSDEL(CT)	0.069	0.063	1.082	0.2793
CRSDEL(IP)	0.029	0.063	0.472	0.6369
CRSDEL(PS)	0.058	0.063	0.917	0.3590
CRSDEL(ARG)	**0.150**	**0.065**	**2.312**	**0.0208**
CRSDEL(REP)	0.012	0.066	0.176	0.8607
ESB(CT)	**0.576**	**0.038**	**15.228**	**0.0000**
ESB(IP)	**0.584**	**0.037**	**15.613**	**0.0000**
ESB(PS)	**0.404**	**0.038**	**10.619**	**0.0000**
ESB(ARG)	**0.542**	**0.039**	**14.027**	**0.0000**
ESB(REP)	**0.419**	**0.039**	**10.616**	**0.0000**
SCHTYPE(CT)	0.015	0.013	1.112	0.2663
SCHTYPE(IP)	0.025	0.013	1.930	0.0537
SCHTYPE(PS)	0.009	0.013	0.661	0.5089
SCHTYPE(ARG)	0.023	0.013	1.686	0.0919
SCHTYPE(REP)	0.018	0.014	1.296	0.1951
AGE(CT)	-0.034	0.022	-1.529	0.1262
AGE(IP)	**0.060**	**0.022**	**2.746**	**0.0060**
AGE(PS)	**-0.155**	**0.022**	**6.992**	**0.0000**
AGE(ARG)	0.025	0.022	1.093	0.2743
AGE(REP)	**-0.055**	**0.023**	**-2.421**	**0.0155**

* statistically significant variables are shown in bold. See text for an explanation.

6.4 Concluding comments

In terms of validity, it seems appropriate to conclude that:

- field of study is a significant variable (the profiles of performance related to field of study being meaningful); and
- year level is a significant variable (though the reason for this needs to be clarified with appropriate samples);
- student-specific variables such as motivation and student ability appear to account for much of the variance in scores.

However, whether performance on the dimensions of GSA is appropriately related to variables such as gender, age and English-speaking background is not clear. It is important to investigate this issue, including a study of relative university and post-graduate work achievements for these groups.

7. Relationship between performance on GSA and other measures of student achievement

7.1 Introduction

Given the purposes described for the GSA, it is expected that student performance on the instrument's dimensions would correlate with performance on similar tasks including those related to general academic achievement and graduate work. While it is too early to investigate the predictive validity of the GSA in terms of graduate outcomes, it is possible to undertake a preliminary investigation of the GSA's relationship with available measures of academic achievement.

7.2 GSA score, tertiary entrance score and grade-point average

7.2.1 Investigation of GSA scores, tertiary entrance scores and grade-point average

An investigation of the associations among GSA dimension scores, Tertiary Entrance Rank, TER (or tertiary entrance score, TES) and Grade Point Average, GPA (or weighted mean) was undertaken. For this investigation, the GSA Exit 2000 student sample was used (predominantly third and fourth year students).

Due to marked variations between institutions in methods used to calculate students' TER, TES, GPA and 'GPA-like' scores, summary estimates for the full GSA Exit 2000 sample are not comparable. Hence, separate within-institution analyses are reported. Note that the sample sizes for students may not match those reported elsewhere because TER/TES and GPA data were available only for a subset of students.

For example, Figure 7.1 presents three tables describing TER-GPA-GSA relationships for one university (University 1). (See Table 7.1 for N values.)

The first table in Figure 7.1 records the correlation coefficients among TER (or TES), GPA (or weighted mean), total GSA[17] score and individual GSA dimension scores. For University 1, total GSA-TER score correlation was 0.459, total GSA-GPA score

[17] Total GSA score was used in correlations for convenience. More refined indicators based on individual dimension scores, as discussed later would avoid technical problems related to addition of scores from separate scales into a total score.

correlation was 0.541 and TER-GPA score correlation was 0.399. All correlations, including those involving the individual GSA dimensions (CT, PS, IP, ARG and REP) were significant, except for the REPGSA-TER correlation. GPA had the largest correlation with the GSA's Critical Thinking (CT) dimension and the smallest with Report Writing (REP).

The other two tables in Figure 7.1 present information about the regression of students' GPA scores on their TER and total GSA scores in order to estimate the magnitude of TER and total GSA score predictive effect of GPA. According to the third table, total GSA was significantly predictive of GPA at the 95% confidence level, having a 0.34 SD (standard deviation) effect on GPA. Although TER had a 0.24 SD effect on GPA, this was not significant at the 95% confidence level.

University 1

	TER	GPA	CTGSA	IPGSA	PSGSA	ARGGSA	REPGSA	Total GSA
TER	1.000	.399**	.318**	.394**	.335**	.311**	.189	.459**
GPA	.399**	1.000	.542**	.476**	.374**	.383**	.285**	.541**
CTGSA	.318*	.542**	1.000	.665**	.569**	.424**	.355**	.801**
IPGSA	.394**	.476**	.665**	1.000	.611**	.511**	.518**	.874**
PSGSA	.335*	.374**	.569**	.611**	1.000	.365**	.443**	.796**
ARGGSA	.311*	.383**	.424**	.511**	.365**	1.000	.368**	.684**
REPGSA	.189	.285**	.355**	.518**	.443**	.368**	1.000	.668**
Total GSA	.459**	.541**	.801**	.874**	.796**	.684**	.668**	1.000

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Explaining the variance in GPA in terms of total GSA score and TER score:

Regression ANOVA Table[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	580.027	2	290.014	8.587	.001[a]
	Residual	1722.524	51	33.775		
	Total	2302.551	53			

a Predictors: (Constant), Total GSA, TER, Adjusted R Square = 0.223
b Dependent Variable: GPA

Coefficients for Explanatory Variables[a]

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	43.615	5.775		7.553	.000
	TER	.121	.068	.242	1.772	.082
	Total GSA	7.107E-03	.003	.343	2.513	.015

a Dependent Variable: GPA

Figure 7.1 GSA-TER-GPA correlations and regression for University 1

Table 7.1 presents a summary of similar data for the 11 universities that provided data on prior (TER/TES) and concurrent (GPA/weighted mean) measures of student

achievement. For reasons of confidentiality, university names are not given, nor do the university numbers match those given elsewhere.

In Table 7.1, the achieved sample sizes (*N*) given in the second column indicate that of the 99 students with GSA scores from University 1, 98 provided GPA scores, but only 54 provided TER scores.

The findings summarised in Table 7.1 indicate that GSA performance was a stable predictor of GPA for most institutions and the predictive value of GSA tended to be as good as TER, or better. For those institutions where GSA lacked predictive value, TER was generally not useful either. Further investigation of these observations is required.

Table 7.1 GSA-TER-GPA correlations and regression coefficients for 11 participating universities

Uni	N TER N GPA N GSA	GSA/ TER	GSA/ GPA	TER/ GPA	Coefficient (SD units)	
					TER	GSA
1	54 98 99	0.46*	0.54*	0.40*	0.24	0.34*
2	22 26 29	0.36	0.45*	0.45*	0.28	0.47*
3	160 179 207	0.34*	0.31*	0.34*	0.24*	0.28*
4	91 91 91	0.48*	0.35*	-0.15	0.0	0.35*
5	0 37 38	-	0.47*	-	-	0.47*
6	37 81 81	0.32	0.48*	0.60*	0.46*	0.42*
7	105 172 185	0.23*	0.03	-0.03	-0.04	0.06
8	150 175 176	0.47*	0.19*	0.23*	0.26*	0.0
9	33 57 56	0.12	0.32*	0.53*	0.49*	0.32*
10	14 14 14	0.55*	0.55*	0.39*	-	-
11	50 14 90	0.57*	0.62*	0.76*	-	-

* Indicates statistical significance beyond the $p < 0.05$ level.

- Indicates no or insufficient data.

It is interesting that the correlations of GSA with TER and GPA are moderately strong given that TER and GPA scores are both curriculum related and dependent on success across a wide range of differently focussed courses, while the GSA is a test of certain generic skills. This observation is consistent with a key purpose of the GSA, namely, to assess widely applicable underlying generic skills, and is evidence in favour of the validity of the test.

7.2.2 Analysis at field of study level

Detailed analyses of data for all 11 universities in this study indicate that all GSA dimensions had predictive value for GPA, though some were of more value for some student cohorts than others. This may be related to the field of study composition of the cohorts, student motivation, the modes of assessment at a particular institution, different GPA calculation methods used and standards applied by different departments and universities, and other characteristics of the samples provided.

Further, it might be expected that the more heterogeneous the university cohort with respect to field of study composition, the lower the correlation between GSA and GPA.

Also, cohorts with narrow ranges of GSA scores (e.g. mostly high scores) are likely to produce poor correlations.

Issues such as the differential predictiveness of the GSA dimensions for different cohorts merit investigation. Further analyses of the kind reported in this chapter should be undertaken at the field of study and university course level.

7.2.3 Comparison of GSA-GPA correlations with US equivalents

The GSA-GPA correlations appear to be comparable to SAT-GPA correlations reported in the USA (McDonald et al., 2001). Further, Bridgman et al., (2000) found that US High School GPA (similar to TER) accounted for 13% of the variance in university first year GPA, and adding the SAT to this, increased the prediction to nearly 20% across all students. Predictive value was seen to vary between various ethnic and other groups. The results reported here suggest that GSA may have similar predictive value (though this study focussed predominantly on third and fourth year exit students).

The TER-GPA correlations observed here are similar to school grade-GPA correlations reported by Power et al (1987). Indeed, McKenzie and Schweitzer (2001) report that university entry score accounted for 39% of variance in GPA for a sample of Australian Science and IT students at the end of their first semester of university study. They also report on the relationship of other academic, personal and psychological factors to GPA with some surprising results (e.g. study skills were not found to be a significant predictor of GPA). Such findings highlight the complexity of this area of investigation.

7.2.4 Use of GSA for tertiary selection

Given that both TER and GSA generally have significant predictive value with respect to GPA, it might be envisioned that a combination of the two could be used to more effectively select students into particular undergraduate and postgraduate courses. The TER score and the five GSA dimension scores might be weighted differentially to optimise predictive value for specific fields or courses (as is done for the Victorian General Achievement test in another context). Appropriate weightings for a particular field or course could be determined empirically.

Similarly, the GSA might be used to help select students without TER/TES scores or with scores from another state (where there may be some question about comparability).

7.3 Inter-correlations between GSA, GPA and GMAT for business studies group

A small-scale study was done using a version of GSA (BMAT) tailored for selection into post-graduate business school. Items of certain types and contexts were selected from the GSA multiple-choice item pool for this purpose.

As with GSA, there is a statistically significant correlation between total BMAT score and GPA score, with the total BMAT-GPA correlation being 0.589 (30 students). The scores on each GSA multiple-choice component (CT, IP and PS) also correlated significantly with GPA at the 95% confidence level.

In addition, though only five students had GMAT (Graduate Management Admissions Test) scores, the correlation between GMAT and BMAT Problem Solving scores was 0.933 and significant at the $p < 0.05$ confidence level, while the correlation between GPA and GMAT (0.659) was not significant at that confidence level. Other BMAT dimensions did not correlate significantly with GMAT.

Since the sample of students involved in this preliminary study was very small, this study should be expanded.

Nevertheless, these preliminary results suggest that useful specially focussed versions of GSA could be produced for different purposes.

7.4 Concluding comments

The statistically significant correlation between performance on GSA components and performance on traditional measures of student achievement is evidence in favour of the validity of the GSA. Since TER and GPA scores are both curriculum related and dependent on success across a wide range of differently focussed courses, while the GSA is a test of generic skills, this observation is consistent with a key purpose of the GSA, namely, to assess widely applicable underlying generic skills and is evidence in favour of the validity of the test. It is important to extend this study to investigate the predictive validity of the GSA with respect to individual courses and the workplace.

8. Evaluation of GSA reference ranges

8.1 Introduction

For the purpose of comparing the performance of a particular student with others in a field of study, a reference range is important. Since students sitting the GSA to date have been largely self-selected, and deviate from a representative sample as discussed in Chapter 4, there is doubt about the suitability of the current reference ranges. This chapter looks at current reference ranges in light of this problem.

8.2 GSA reference ranges and levels of performance

At this stage, a student's GSA scores are presented on a Student Report form (Figure 3.1) in relation to (i) the middle 60% of all students, (ii) the middle 60% of students in the same general field of study and (iii) broad described levels of achievement (see section 8.2.2 and Figure 3.1). See Appendix 5 information on how a student's GSA scores are obtained.

Middle 60% reference ranges applicable in mid-2001 are given in Table 8.1. These ranges are based on data from all students who completed one of the first two tests (Exit 2000 and Entry 2001) or the smaller trial test. (Data for the writing dimensions - Report, Argument - in Table 8.1, however, exclude the trial data where writing had a slightly different format.) These tests are linked by common items and raw score transformations, allowing student results on different versions of the GSA to be located on common dimension scales.

Table 8.1 Sample sizes, means and middle 60% ranges by field of study and for all students (GSA exit 2000, GSA entry 2001 and trial)

Field	N	Report		Argument		Problem Solving		Critical Thinking		Interpersonal Understandings	
		Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Arts/Hum	611	404	339-472	419	333-499	385	302-481	439	361-509	454	374-528
Bus/Com	992	361	286-434	345	264-427	369	277-462	368	281-448	373	286-455
Comp/IT	524	352	286-420	343	269-422	394	297-481	368	273-451	370	286-453
Ed/Social	262	377	313-445	383	313-451	347	277-425	372	299-439	406	322-476
Eng/Arch	355	370	286-454	359	285-423	444	343-530	391	299-479	389	304-468
Math/Sci	966	384	311-458	377	307-442	418	330-503	412	334-487	412	339-477
Law/Legal	113	421	366-482	432	377-503	437	320-526	460	395-529	465	387-545
Med/Dent	368	413	344-480	416	346-484	459	381-524	447	385-502	456	391-516
Nursing	270	345	285-415	357	273-433	336	258-417	361	290-439	400	339-468
ND/Other	142	361	277-431	355	281-448	314	216-425	326	216-439	318	210-436
All Students	4603	378	304-454	374	300-453	394	297-481	396	310-479	403	322-491

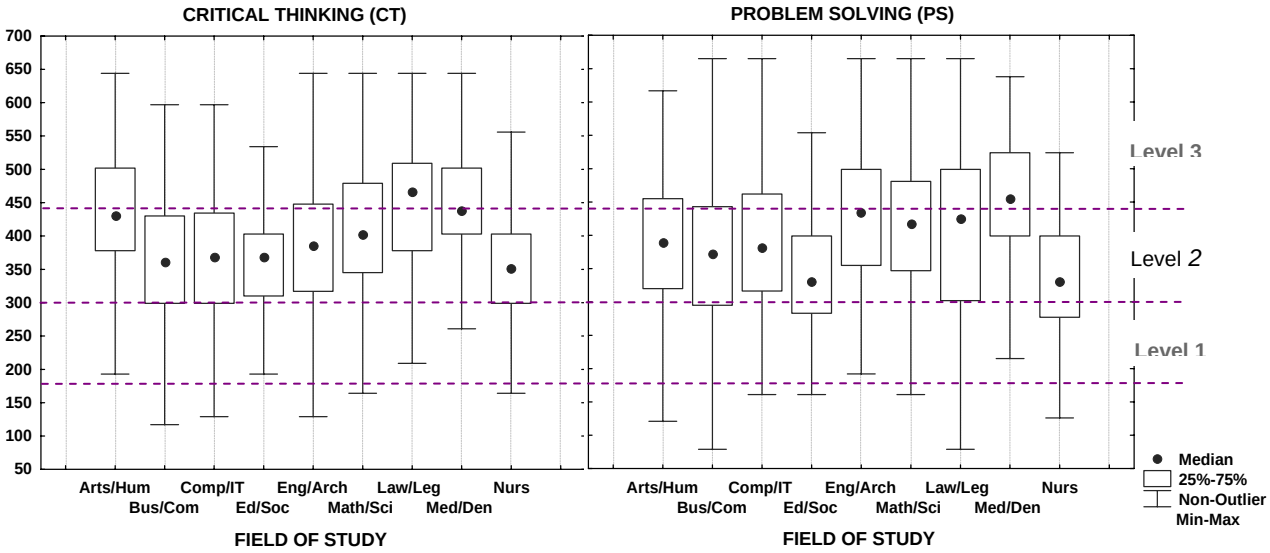
8.2.1 Inter-quartile ranges by field of study

To assist in comparisons, Figures 8.1 and 8.2 give the inter-quartile ranges for the five GSA dimensions by field of study in the form of conventional box-and-whisker plots.

The box for each field of study describes the inter-quartile range (25^{th} to 75^{th} percentiles). The 'dot' located within each box indicates the median value (50^{th} percentile, or the point above and below which 50% of the cases lie). The whiskers give the range of other scores up to a maximum and minimum of ± 1.5 inter-quartile distance. To assist readability, outlier and extreme values have not been included.

Figure 8.1 indicates that substantial proportions of Engineering/Architecture, Medicine/Dentistry and Law/Legal students are at Level 3 for Problem Solving, substantial proportions of Arts/Humanities, Law/Legal and Medicine/Dentistry students are at Level 3 for Critical Thinking and Interpersonal Understandings and substantial proportions of Arts/Humanities and Law/Legal students are at Level 3 for Argument Writing. (See Figure 3.1 and section 8.2.2 for descriptions of levels of performance.)

More work on representative samples (preferably using finer groupings of students by field of study/course) is required to verify and elucidate generalisations such as these.



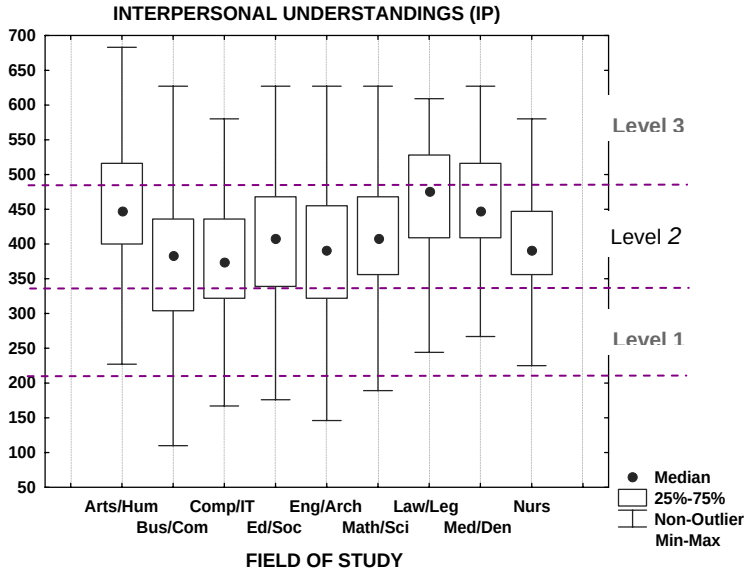
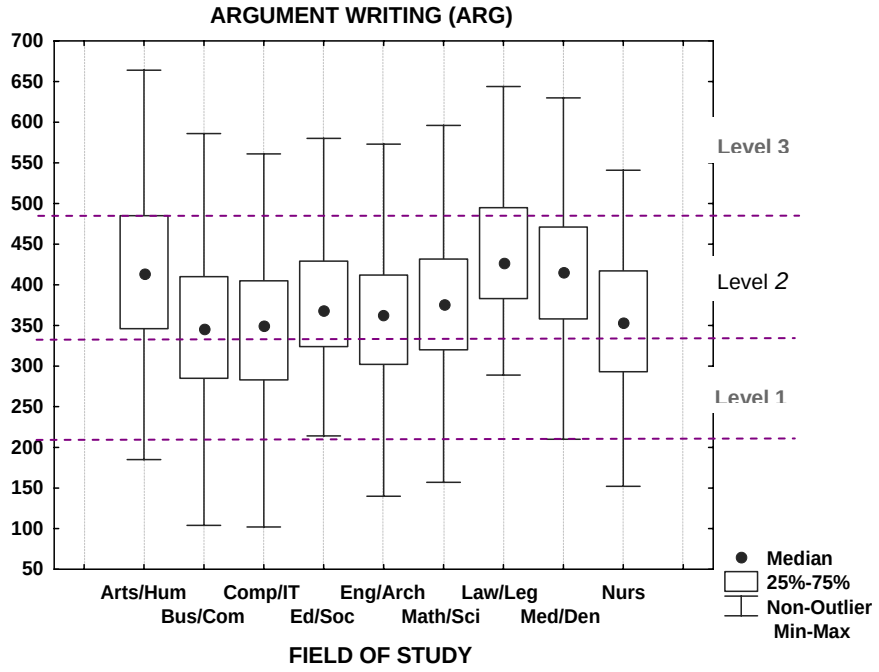


Figure 8.1 Box plots of three GSA dimension scores, showing inter-quartile ranges for nine fields of study (all students)



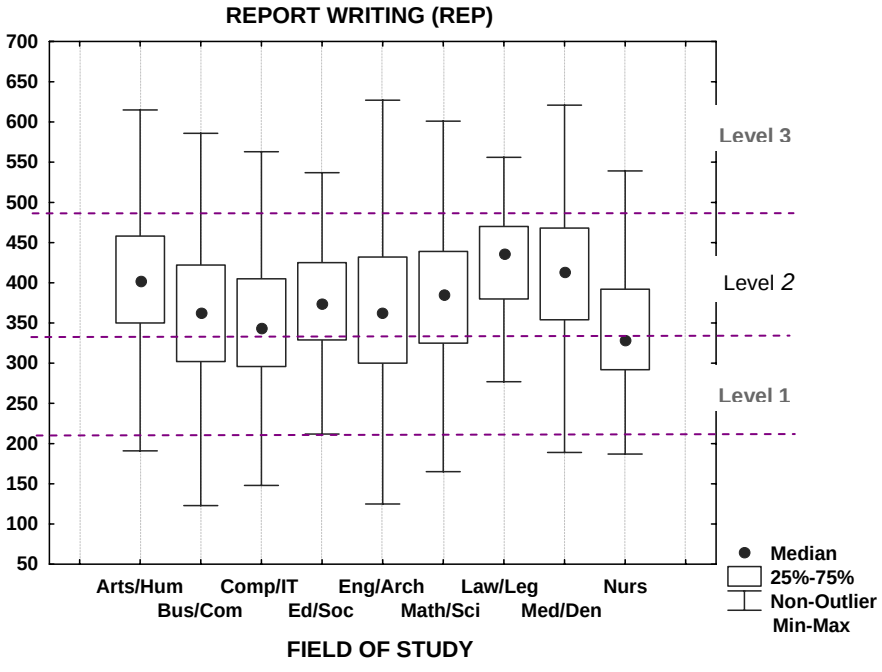


Figure 8.2 Box plots of the two GSA writing dimension scores, showing inter-quartile ranges for nine fields of study (all students)

8.2.2 Levels of performance

In Figures 8.1 and 8.2, and consistent with the Student Report form (Figure 3.1), three *levels* of performance are indicated. Note that Figure 3.1 also provides descriptions for these levels of performance. (Comments related to improving the descriptions are made in the next chapter.) These levels were determined both on the basis of preliminary judgements about appropriate student skills and on the actual performance of students in the trials, which, subsequently, were seen to match reasonably well with the performance of students in the first two tests.

Performance at Level 1 (between 200 and 325 GSA score points) suggests that a student has a basic or limited mastery of skills relevant to a particular GSA dimension. Performance at Level 2 (between 325 and 475 GSA score points) suggests that a student has a fairly solid mastery of skills relevant to a dimension, and performance at Level 3 (above 475 GSA score points) suggests that a student has a strong mastery of skills relevant to a dimension. In addition, GSA scores below 200 are reported as 'Level 1 not reached or insufficient material attempted' (though this may also be a result of inadequate language skills). Scores above 600 are reported as 'Level 3 but greater than 600' and suggest high-level mastery.

The data summarised in Table 8.1, and in Figures 8.1 and 8.2, suggest that most students perform at Level 2. Given that the levels are still based on preliminary judgements, though supported to some extent by student performance, it would be valuable to have more input from universities related to the 'appropriateness' of the levels and the associated descriptions of performance.

8.3 Relevance of the GSA sample to reference range validity

As described in Chapter 4, the sample of students on which GSA reference ranges are based deviates from that expected of a random and representative sample. Such deviations relate to age, field of study, full time/part time, English-speaking background and so forth.

Additional information about the GSA population may be provided by the distribution of GSA and TER scores for students in the GSA sample.

For example, Table 8.2 and Figure 8.3 give information about the GSA Report Writing score distribution for the Entry 2001 cohort. Although not normal, the distribution seems reasonably close to what would be expected from a random sample of students. The skewness in the scores for this writing task may be related to the relatively high proportion of NESB students.

Table 8.2 Descriptive statistics for GSA Report Writing scores, GSA entry 2001

Statistic	All Students	Males	Females
Mean	369.61	363.42	374.58
SD	88.17	90.27	86.24
SE	2.07	3.21	2.70
Skewness	0.10	0.17	0.05
Kurtosis	-0.12	-0.11	-0.11

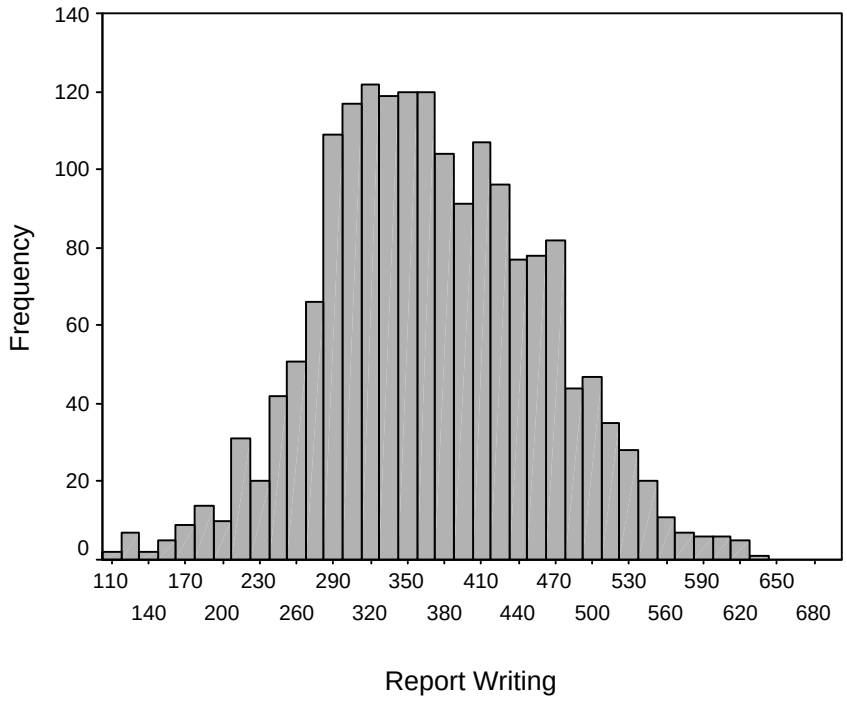


Figure 8.3 GSA Report Writing score frequencies, GSA entry 2001

Table 8.3 records the mean TER/TES scores for the 625 Exit 2000 students who agreed to supply these data. Taking into account the fact that these data were provided by different states using different methodologies, the mean TER values are, nevertheless, significantly higher than would be expected of the general university population. (The high mean TERs could result if students who expect to achieve low scores on the GSA avoid the test.)

It is not clear how the observed similarities and differences between the GSA and general university population reported here and in Chapter 4 affect the suitability of the current GSA sample in terms of providing suitable reference ranges and data for other analyses. Some effects of the deviations might be predicted. For example, it might be expected that the high proportion of students from a non-English speaking background would tend to lower mean scores for writing and high TERs in general would tend to raise GSA reference ranges. However, the sum of effects is impossible to predict.

Field of study reference ranges are likely to be particularly problematic for the individual fields of study where few students participated and where the field is composed of smaller sub-fields whose students differ markedly (e.g. Law/Legal Studies).

In addition, the current GSA reference ranges combine the results of students at all year levels, which is inappropriate.

Table 8.3 Means and standard deviations of TER scores by university (exit 2000 students)

	TER		
University	Mean	SD	*N*
University 1	77.80	13.14	54
University 2	84.14	8.87	22
University 3	75.99	12.35	160
University 4*	N/A	N/A	0
University 5*	N/A	N/A	0
University 6	71.07	15.56	37
University 7	83.65	12.44	105
University 8	91.07	5.29	150
University 9	86.07	9.14	33
University 10	96.61	3.13	14
University 11	74.93	15.02	50

* Tertiary Entrance Score not supplied or not comparable to other universities.

*

It is essential that representative samples are obtained in order that reliable reference ranges become available for fields of study and year levels.

Representative samples may not be necessary if students are simply compared with described levels of performance. What is important in that case is that the described levels are appropriate and the test outcomes match these levels. Described levels of performance should be reviewed in consultation with universities now that two tests have been administered.

8.4 Test reliability and sensitivity

The value of a reference range is related to the purpose of the testing. For the purpose of comparing the performance of a particular student with others in a field of study, the reference range is important. However, for simple ranking purposes such as for selection into a postgraduate course, score ranges may not be important provided that an appropriate cut-off score is chosen. Here test reliability and sensitivity become particularly important.

The GSA multiple-choice components have reliabilities around 0.8, which is considered suitable for a 30-item component. (It should be noted that GSA has about 30 items to be completed in 40 minutes for each multiple-choice component, compared with GMAT, which has about 40 items to be completed in 75 minutes for each multiple-choice component.) Nevertheless, whereas such reliabilities and their associated test sensitivities should be adequate to detect changes between first and third year groups of students, they may not be adequate to detect a relatively small change for a single student. This problem is particularly acute for students whose abilities are at the high or low ends of the reference range, where there are few items and test sensitivity is lowest.

If sensitive measures of student performance at the low or high end of the ability range are required, specific tests aimed at such groups should be used. Moreover, computer adaptive testing (CAT) could be employed, since CAT approaches tailor items to student ability. Alternatively, the GSA multiple-choice domains could be lengthened. However, there is resistance to making the test longer for pragmatic reasons such as student compliance. Alternatively, just two slightly longer multiple-choice components might be used instead of three, one based on Analysis/Synthesis/Evaluation of information, focussing on common elements of Critical Thinking and Problem Solving that are valued by stakeholders, and the other on Interpersonal Understandings.

Issues of reliability and sensitivity are relevant, similarly, to the writing tasks, though different solutions need to be sought for these.

8.5 Concluding comments

It is not clear how the observed similarities and differences between the GSA and general university populations reported here and in Chapter 4 affect the suitability of the current GSA reference ranges and the results of other analyses. It is important that, in consultation with universities, representative student samples are obtained so that appropriate reference ranges can be produced. Described levels of performance should be reviewed in consultation with universities. In addition, as the uses of the GSA evolve, test reliability should be reviewed and, if necessary, the test modified so that test reliability is appropriate for the given purpose.

9. Review of test construct and items

9.1 Introduction

In order to evaluate the face/content validity of GSA, various stakeholders and content experts were asked to comment on the GSA construct and a sample of items. These were:

- a professional with content expertise in each component;
- a group of recruiters of graduates;
- a group of students from one university who sat the test; and
- a variety of other stakeholders from university and the graduate workplace.

Since the test is secure, items cannot be provided here. A description of the units in GSA Entry 2001 is given in Appendix 3 and a small number of sample items is available on-line from the ACER GSA website, though, at this stage, these are not fully representative of the test.

9.2 Appraisals of the GSA by content experts

In this section are presented appraisals of the GSA construct and a sample of items by people with expertise in the each of the GSA components. Following each appraisal, test developers respond to some issues raised.

9.2.1 Interpersonal Understandings

Reviewer: Barry J. Fallon (BA [Hons], BD, MA, PhD)

Dr Barry Fallon is a Fellow of the Australian Psychological Society for which he served as President (1995/6–1996/7). He is a Registered Psychologist in the State of Victoria. He has over 20 years' experience of applied psychological research and tertiary teaching following his doctoral studies. He is the Foundation Professor of Psychology at the Australian Catholic University. Previously he was a member of staff in the Psychology Department at the University of Melbourne where he was most recently the convenor of the Post Graduate Program in Organisational and Industrial Psychology. Barry has experience in a wide range of applied psychological research that has been conducted at the Federal, State and local levels. He has been involved in course development and review at several Australian universities.

Review comments

The specific items appear to have considerable face validity with respect to the construct. It is appropriate to leave it to the psychometricians to decide on the empirical validity.

The variety of presentation formats of the scenarios is excellent.

There is an interesting juxtaposition regarding skills versus understandings. It is possible to argue that the understandings may be necessary but in and of themselves they are not sufficient evidence that the skills and behaviours will be in accord with the understandings. While it would be expected that there would be a positive association between understandings and behaviours within a particular domain in the present instance it remains an empirical question. The nature of the relationships between understanding and skills is something that requires further empirical investigation.

There is a statement that "It [the GSA] is interested in assessing skills that can be deliberately developed by the university experience". It could be argued that a considerable amount of the item content of Interpersonal Understandings has its basis in Social Psychology. How ".. the university experience" can be expected to contribute to the development of interpersonal understandings which have their basis in a particular sub discipline is difficult to understand. On the other hand (if) it could be argued that much of the content of Interpersonal Understandings is really just "good common sense" then it is distinctly possible that as a result of interactions with staff and students, experiences in tutorials and laboratories, working with other students and the like along with further developing maturity then "the university experience" may well contribute to improvement in Interpersonal Understandings. There is no doubt (that a) considerable amount of understandings of human behaviour which have their basis in general social psychological principles and theories has come to be part of our "every day general knowledge" and thus as students mature through their university experience they may further develop their understandings in this area.

Interpersonal Understanding is a construct that is very appropriate for the purpose of assessing skills of students relevant to success in graduate employment. Whatever field of employment a graduate enters there are interactions with others and hence an ability to interact successfully does require a certain level of Interpersonal Understandings. There is less certainty about the generality of the appropriateness of Interpersonal Understandings for success at university. The importance of Interpersonal Understandings for success at university will vary from course to course (or from discipline to discipline). While Interpersonal Understandings are of great importance for professional training in social work, education, medicine and other similar professions, it is difficult to see how relevant Interpersonal Understandings are for university success for physicists, geologists and disciplines like those.

The different descriptors for the levels for Interpersonal Understandings are so subtle that it does not really differentiate in a way that is particularly meaningful. "Sophisticated" is only meaningful because it applies to the top level when compared with "significant" for Level 2. Either term by itself does not provide particularly useful information. It may be worthwhile to reconsider the descriptors and to provide examples that would help in the articulations of the differences between the levels.

Test developer response

It is pleasing that Barry accepts the face validity of the items in terms of the construct and the range of scenarios. And it is agreed that empirical validity should be left to the psychometricians to evaluate as data are collected.

It is also agreed that the relationship between the demonstration of understandings in the test and the application of related skills and behaviours in practice needs to be investigated empirically.

As Barry allows, the understandings sought by the items of this component are just 'good common sense' about interpersonal interactions, some of which may have a basis in general social psychological principles and theories, and which could be enhanced by experiences with other students and staff (i.e. 'the university experience').

Barry's point about the Interpersonal Understandings construct being relevant to success in graduate employment in general, but possibly not being so relevant to success in university across the disciplines, is a good one. Preliminary data provided in previous chapters, however, does suggest a significant correlation between performance on Interpersonal Understandings and Grade Point Average, but more work will be needed to clarify such issues.

Also, it will be important to study the relationship between success on Interpersonal Understandings and success in graduate work.

It is agreed that descriptors for levels of performance need to be refined, perhaps with the aid of some examples.

9.2.2 Problem Solving

Reviewer: Margaret L. Wu (BSc [Hons], DipEd, MEd *Melbourne*, DipComStudies *RMIT*)

Ms Margaret Wu is a Senior Research Fellow at the Australian Council for Educational Research. Margaret has had a major role in the development of the problem-solving component of the OECD Programme for International Student Assessment (PISA) project. This work involved shaping the problem-solving framework and item development of the PISA test for 15-year-old students in about 35 countries.

Review comments

My overall impression of the Graduate Skills Assessment Problem Solving (GSA PS) component is that it captures very well the general definition of PS as found in most literature on the topic, and the majority of the items reflect the defined construct. In addition, I believe that the construct is appropriate for assessing generic skills of students relevant to success at university and employment, as the problem-solving processes covered are applicable to a wide range of problems and situations.

The term Problem Solving is widely used and the interpretation is broad, as shown in the general review of Problem Solving in the (construct) document. The GSA PS has a narrower focus, of course, as it is one of the four components of the GSA assessment, and the GSA has a specific targeted population and purpose. I wonder whether the section entitled "Problem Solving in GSA" (in the construct document) can be more explicit about the focus of the GSA PS component. For example, from the items, I get a sense that the PS component focuses more on logical reasoning and quantitative reasoning, and not so much on verbal reasoning (which is covered elsewhere). Could this be made more explicit (more than just a line noting that the information is presented in low verbal and non-verbal form)? That is, instead of stating what the GSA PS does not assess, state more about what the GSA PS does assess, and state more clearly how these differ from the other three components.

While the processes listed (identify, comprehend, analyse, represent, translate, re-organise, synthesise, generate strategy/solution) are all very relevant to the construct and items, these are more "procedural"

processes. Could there be some mention of the cognitive processes involved such as logical thinking, inductive, deductive, and analytical reasoning? I do believe the common strand underlying most of the items (and the procedural processes) are these reasoning skills.

The majority of the items reflect the PS construct very well. The only ones I have some trouble with are those on pages 28 and 29 (of the booklet provided). These don't seem to fit too well with the claim in the construct that "..focus on generally applicable and accessible everyday practical problems....". These are puzzle type mathematical items that are not very "everyday", although they clearly assess quantitative reasoning. These items have an appearance of "everyday" but the context is very contrived. In a way, I would prefer to see a totally fictitious setting (Indiana Jones) than contrived real-life contexts. In these items, the contexts do not really provide anything for the students to relate to.

While the descriptions for the three levels give a good summary of the tasks that students in these levels can typically handle, the descriptions themselves rely heavily on the use of words such as "straightforward", "standard", "well-defined", "basic", "complex", "non-standard". As people can have different understandings of these words, is it possible to give some examples along side the descriptions to demonstrate what is basic and what is complex, etc.?

Test developer response

It is pleasing that Margaret believes that GSA Problem Solving is consistent with common definitions of the area found in the literature and that the items generally reflect the defined construct, which takes a particular approach for psychometric and practical reasons.

It is also pleasing that she believes that the construct and range of items are appropriate for the assessment of generic skills relevant to success at university and in graduate work.

It is agreed that the description of the component in the construct could be made more explicit (indicating more clearly what is and is not addressed) and more emphasis could be put on the cognitive processes/reasoning skills required for the items.

In connection with this, the level descriptors need to be refined, perhaps with the aid of examples.

It is also agreed, that those items that appear contrived (e.g. puzzle type mathematics), as opposed to the authentic-looking items that the construct requires, should be removed from the test.

9.2.3 Critical Thinking

Reviewer: Laurance Splitter (BA [Hons] *Monash*, BPhil., DPhil. *Oxon*, MACE)

Dr Laurance Splitter has been a Principal Research Fellow and Director of the Centre of Philosophy for Children at the Australian Council for Educational Research. He has published widely both in Australia and overseas. His areas of professional expertise include: philosophy in schools, the teaching of thinking and reasoning, ethics and values education, inquiry-based pedagogy and meta-cognition. Laurance is currently a professor at the Department of Educational Foundations, Montclair State University, New Jersey, USA.

Review comments

It has been argued that critical thinking skills are highly discipline/domain specific, both in respect of their learnability and their application. If this were true, the kind of generic instrument proposed for the GSA would be virtually impossible to construct. However, many writers on critical thinking have rejected such a view. A reasonable counter claim is to distinguish, as the Construct does, between teaching such skills ab initio – which, arguably, is contextual with respect to some kind of disciplinary framework – and applying (which involves learning) such skills in new contexts and domains. Further, I support the view, reflected in the Construct and the range of items, that students who can apply skills across a range of contexts, are more likely to apply them in other contexts, perhaps not yet determined.

Like most writers in the area, I reject the view that Thinking (or Critical Thinking) is a discipline. It is, rather, a dimension of every discipline and of life itself. I think the introductory comments make it sufficiently clear that the reason for constructing such a test in its generic form, is that good thinkers should be able to deal with generally accessible/real world applications. Further, both the assumption of the transferability of Critical Thinking skills and the preliminary findings relating to the predictive value of the GSA, serve to justify the construction of a generic instrument.

Still, it does need to be said that a good critical thinking instrument can, at best, measure or record levels of competence and, perhaps, serve as a predictor of further achievement in university performance, etc.

I support the acknowledgement that a 30-item multiple choice test cannot capture the many different accounts and definitions of what it means to think critically (or, to be a critical thinker, which is not the same thing). It is, accordingly, unreasonable to expect the Construct to have analysed, compared and evaluated all, or even a large number, of these. In this context, it was reasonable to set out to "assess some key markers of the ability to think critically about views on issues and make decisions based on good intellectual standards."

<u>Definitions of Critical Thinking</u>

However, given the broad range of definitions of Critical Thinking, it is reasonable to expect the Construct to focus either on one such definition and then create and evaluate test items in line with this definition, or on a manageable range of such definitions. The latter seems to be the preferred course of action. For the sake of clarity, it would be desirable actually to articulate these common threads so that those involved in the GSA – including those who rely on the results – can be reasonably certain as to what it is that was tested. A clear summary of why the particular traits of comprehension, analysis/inference and evaluation are chosen would be appropriate.

One reason for not including Problem Solving as a form of Critical Thinking involves taking seriously the connection between critical thinking and meta-cognition. Critical Thinking is strongly meta-cognitive, where problem solving per se is not.

This is one distinctive feature of Critical Thinking, which is implicitly accepted in the Construct. Further, the reference to Kuhn suggests another, arguably more central feature:

> *Kuhn describes a model for the development of critical thinking skills, focusing particularly on the meta-cognitive aspects, in which the emphasis is on the development of skills that enable one to generate appropriate criteria for belief and evaluate one's own thinking.*

The importance of generating appropriate criteria for belief (or judgement) has been highlighted by Matthew Lipman, founder of the Philosophy for Children movement. Critical Thinking is precisely that mode of thinking, which results in judgement in virtue of its use of criteria, rules and standards.

Lipman's view is that there are several aspects to critical thinking, including (i) an inclination to correct our thinking when faced with good reasons for doing so, (ii) a sensitivity to context, and (iii) a concern with, indeed reliance upon, criteria. Of these, (iii) is the most important. I will comment on each of (i)-(iii) in the context of the Construct.

(i) It is not clear how an instrument such as the GSA could be expected to accommodate "self-correction" as a feature of critical thinking, mainly because judging that someone has changed their mind or thought again about an issue requires observations over time. However, it is the reference to inclination (disposition) with which I am particularly concerned. The problem, of course, is that dispositions, traits and inclinations are notoriously difficult to assess and measure, and would require an instrument of considerably greater sophistication – and, arguably, of considerably greater impracticality – than the GSA as presently constructed. Nevertheless, such a key ingredient of critical thinking should, at least, be acknowledged.

(ii) The idea that good Critical Thinking is sensitive to context is related to (i). Thinkers think in contextual situations, motivated by specific problems and tasks. Accordingly, a conception of critical thinking that focuses on specific individual skills, strategies and "moves", will not be adequate. We need to focus on "argument networks". Ideally, then, to measure someone's Critical Thinking capacity (or behaviour), we must at least assess their ability to engage in argumentation – to participate in a dialogue in which others are presenting opposing perspectives, etc.

I note the reference, in the Construct, to Paul's distinction between sophistical (or "weak") critical thinking, and fair-minded (or "strong") critical thinking, along with the acknowledgement – with which I concur – that the GSA, as constructed, does not seek to capture this distinction.

The quotation from Norris and Ennis, however, is more clearly supportive of the GSA structure, in so far as the test items do provide a "wide range of critical thinking tasks requiring background knowledge they already have". However, the range proposed – Comprehension, Analysis and Inference, and Evaluation" – is not all that wide.

(iii) Lipman, like Kuhn, is pointing to those meta-cognitive tools that facilitate the formation of judgement because they enable us to construct and apply criteria. Judgements, and the criteria underlying them, may be described as the essential ingredients of Critical Thinking. In slightly different terms, we could say that while much of our thinking is directed at forming judgements, not all of it is governed by the use of criteria. Critical Thinking is precisely that mode of thinking that results in judgement in virtue of its use of criteria, rules and standards.

It seems to me that these – admittedly contentious – features of Critical Thinking could be mentioned in the Construct, not just for reasons of comprehensiveness, but because it might serve as a marker in item writing and test construction in this conceptually tricky area.

Even if such features as sensitivity to context, fair-mindedness, willingness to self-correct, and other dispositional features of critical thinking cannot be incorporated into the test items, I suggest that the reference to criteria in forming judgements (as the conclusion of arguments) does need to be so incorporated. One way of achieving this is to include specific references to criteria in some items; for example: "In forming that particular conclusion, the writer is appealing to the criterion of (a), (b), (c) or (d)", or "Using the criterion of ___, which conclusion is the best one?"

<u>The Critical Thinking scale</u>

The Scale, as constructed, makes good sense to me, noting that it is likely to be refined as the test evolves. I shall assume that this kind of scale is commonly used in test analysis: it is not really within my expertise to be too critical. A few points might be considered:

- *Should the Construct be more explicit in explaining how the various Levels and their descriptors are to be applied to actual test scores? How are such distinctions as "reasonable inferences" (Level 2) and "subtle and cogent inferences" (Level 3) to be made? By contrast, the distinctions between "explicit" criteria (Level 1), "(readily) inferred criteria" (Level 2), and "generating appropriate criteria" (Level 3) make clear sense in light of the centrality of criteria in the definition of critical thinking (see above).*
- *While "inference" is listed with "analysis" in the second of the three categories of critical thinking strategies (pp. 8, 11), it is linked with both the first ("comprehension") and the second (analysis) in the Scale (Levels 2 and 3), and just with the first in Level 1. I am unclear as to whether there is any significant difference between "makes inferences" and "uses inference" that is meant to explain the dual categorization in the Scale.*
- *I note that the term "synthesis" has been added to "evaluation". Typically, where analysis is aligned with deductive thinking, synthesis is aligned with inductive and/or creative thinking. I do not see evidence that either the Construct or the test items make explicit use of this distinction, and it does not feature in the Scale. Nevertheless, I suggest that both deductive and inductive reasoning are involved in the specific items provided, so the issue becomes one of deciding whether or not to make the deductive/inductive distinction – hence the analytic/synthetic distinction – explicit.*

General comment on suitability of the items presented

I think that the components of Comprehension and Analysis/Inference are covered reasonably well - although I am not in a position to determine whether each and every component skill of these broad strategies is covered by the entire range of items. However, I am less confident that the Evaluation component has been adequately covered.

Final comment

In general terms, I regard the Construct, as presented, and the items as exemplified, as extremely worthwhile and timely.

Test developer response

It is heartening that Laurance has accepted several aspects of the construct. These include the view that Critical Thinking skills can be transferable and those who can apply such skills across a range of contexts presented in the test will be more likely to apply them to other contexts.

However, he does point out the limitations of evaluating Critical Thinking in a short multiple-choice test, many of which are admitted in the construct, so that the test can only assess some markers of a Critical Thinker. Least amenable to assessment by the current test are the dispositional characteristics of a Critical Thinker. The ability to produce an argument network is to some extent assessed in the Argument task.

Given the contentious and evolving nature of Critical Thinking, it might be worth considering renaming this component so that there is less expectation that it covers Critical Thinking in all its breadth. If not, better links should be made to commonly accepted definitions and broader aspects of Critical Thinking, as suggested by Laurance.

It is agreed that scale descriptors and some other aspects of the construct should be made more explicit or otherwise clarified (such as the analysis/synthesis distinction).

It is also agreed that Evaluation should be made more of a focus of the items.

9.2.4 Written Communication

Reviewer: Alison Brown (MA [Applied Linguistics])

Ms Alison Brown holds an MA in Applied Linguistics and is a Learning Skills Adviser at RMIT University. She is an experienced secondary ESOL/English teacher and a teacher of an accredited tertiary writing course.

Review comments

1 Appropriateness of construct of written communication in GSA

The performance level categories of the construct seem generally appropriate. There is a strong emphasis on synthesis and a grappling with complexity, as well as the integration of thought and writing, of structure and specific language control. Students who perform well on these quite demanding tasks are almost certainly going to manage the written demands of their courses and employers. However a number of students who perform poorly on these tasks (possibly at both entry and exit) may manage the content of their course to pass level and be successfully employed. In other words, my concern would be that the sweep may be too narrow.

As an assessment of students' generic written communication there may be too heavy a reliance on the creative generation of ideas. I think it would be possible to focus the tasks on the more basic skills of written English of the sort which mature age students, international students and those returning to postgraduate studies or entering the workforce really need. In my opinion these are: the locating of material in an appropriate context, in terms of form, audience and purpose; the grouping and sorting material (obviously there are varying levels of complexity and critical analysis here); the logical ordering of material and the maintenance of text cohesion both within and across paragraphs. More directed, scaffolded tasks may help to focus attention on these more "generic", transferable skills. The argument task as it stands is particularly abstract; the report task encouraging of "cut and paste" by weaker writers. (I will discuss this further in the following sections.)

The emphasis on independent generation of ideas may be too strong for a number of reasons:

1. *Many students (however much we might wish them to) do not mature into these independent critical skills until after their undergraduate study but may manage more explicit content-focussed tasks quite competently.*
2. *Many students from non-Western culture have little previous experience in the demands of independent critical analysis or the generation of "logical" argument as defined in a Western context (see Helen Fox: Listening to the World).*
3. *Few courses (in my opinion), particularly in areas such as Engineering, Applied Science and Business (the three largest faculties at my university) utilise the traditional argumentative essay as a genre, and rarely demand the kind of abstract socio-political thought required by the argument task.*
4. *Similarly, I suspect employers are more interested in their employees' ability to collate and present information in an appropriate manner than generate argument.*

Such factors could contribute to poor performance on tasks by otherwise competent students.

With regard to the performance level descriptors:

1. *Some language seems more emotive/judgemental than descriptive eg. "subtle, intelligent, obvious, banal, bland".*
2. *Level 4. The use of the term "subtle understanding". Should this read "understanding of the subtlety/ies"? Similarly "subtle processing"?*
3. *Is the intensive valuing of "subtlety" class/culture based? This may conflict with current emphasis on audience, clarity, and purpose.*
4. *The tendency to cut and paste could perhaps be reflected in more descriptive language such as "reproduces from text" rather than "obvious or predictable".*

2 Appropriateness of items and further suggestions

This type of open-ended argument task is not a genre commonly required by either university assignments or (I suspect) employers, and may privilege students from a literary/social science background. Even in subjects which use the traditional essay form, assignments will usually require a response to a specific question and be scaffolded with course notes, readings and discussion. Students are rarely asked to generate the arguments more or less "cold".

While some prompts are presented in the four boxes they offer few clues as to how to proceed with the task. The instructions are also a little obscure, requiring students to extrapolate from the "comments" to the "issues". This is quite a leap to make. The comments are also framed in a fairly conversational style eg. "just the modern form of…" and the contraction "they've read…" yet presumably a more formal response is required. Why not be more explicit? (Lack of explicit structural/stylistic direction is one of the major complaints of international and mature age students about their assignments).

I think this task could be better scaffolded and directed by providing a wider range of short statements about the topic, from relatively straightforward to more challenging (even some irrelevant) which would need to be sifted, sorted, aligned, expanded upon and then presented in a coherent way to a specified audience. This audience could be different to that of the report.

While the report represents a more familiar genre for students the lack of specified audience and purpose is a concern, since reports are almost always written in response to a particular request or to achieve a particular aim. This would also highlight students' ability to show stylistic flexibility in terms of structure and language choices. The task presupposes understanding of report format and this could be overcome with some suggested subheadings. It also requires students to bring knowledge of the subject from outside the material: for example, to make the link between body stressing and overuse injuries. I suspect that many students at the lower levels will recycle the wording from the source material in a "cut and paste" fashion. This is a legitimate skill in some contexts and is not here discouraged by the framing of the task. In fact being judged on "your… selection of material" may encourage it. The requirement of a recommendation section to the report may also add a further dimension of having to recast the material, rather than regurgitating it.

In summary, while the construct provides a challenging test of written communication, I think the tasks could be more clearly focussed to maximise the opportunities for students less confident in the area of generating written material outside the demands of their discipline area.

Test developer response

I have read Alison Brown's review of the GSA writing task and note that she makes some detailed suggestions for modifications to the tasks that appear to arise from her

special viewpoint as a Learning Skills advisor and expert in the teaching of English as a second language. However, the basic purpose of the writing tasks is to describe one strand of a range of generic skills for mainstream cohorts of tertiary students. It is not intended that the writing tasks focus on the basic skills of written English.

Response to the comments about the guidelines and descriptors

1 Performance level descriptors

I agree we could revisit some of the terms she mentions in the first and the final dot points and consider amending them.

Her point about *subtle understanding* in relation to *understanding subtleties* should be taken in.

2 Instructions for the argument task

It is of concern that the instructions appeared obscure and this has been recognised to the extent that, in the latest test guidelines, they have been reworded to clarify the relationship between *comment* and *issue*.

She makes some detailed suggestions for modifications to the tasks that appear to arise from her special viewpoint as a Learning Skills Advisor and expert in teaching English as a Second Language.

Response to comments about writing task modification

1 Reliance on independent generation of ideas

The tasks are developed to provide for the range of mainstream tertiary students. They are not intended to reflect any discipline area or to cater for specific student groups. They are designed to provide a stimulus to generally educated, informed and thoughtful students.

The 'sweep' of the tasks has been addressed by including two different kinds of writing task with different kinds of stimulus and ideas.

2 Argument task (appropriateness)

Although it is possible that the task instructions for the argument could have been clarified, the lack of structural guidance is one of the test characteristics and provides an indication about the capacity of the students to organise and express their ideas. Consideration was given to including a specified audience in the guidelines and rejected as unnecessarily constraining the task.

3 Report task (appropriateness)

The points about the support provided by the guidelines with respect to specifying purpose and audience are valid and accurate. However, the issue was addressed in recent test development by suggesting that students provide their own title for the piece rather than specifying one purpose and audience for all.

The tendency to bring in personal knowledge or to cut and paste or not are considered as indicators of the level of students' capacity for synthesis and higher levels of achievement. These indicators help assessors to discriminate between higher and lower achievement.

A few students include recommendations as part of their structure but it has not been considered as a requirement. They would be included within the suggested set of

scaffolding headings and sub-headings. Such a scaffold would either simplify and codify the task in unacceptable ways or unnecessarily constrain the task, especially for more able writers.

Summary

It is pleasing that Alison believes that the test provides a challenging test of written communication, even though she suggests it may not be accessible to special groups of lower achieving students. For those students who are less confident, the test may provide some indications of the areas of improvement required, both to individual students and also to their tertiary institutions.

9.3 Appraisals of the GSA by graduate recruiters

A meeting was held with ten representatives of recruitment and staff development officers of major firms and government departments. Material describing the GSA approach was presented at the meeting. Table 9.1 summarises their responses to a question about what skills they would like to see in graduate employees.

Although the meeting was informal, the responses reinforce some of the approaches preferred by employers as described in Chapter 1, in particular, a focus on Interpersonal/Teamwork and Communication skills (e.g., oral communication), though Problem Solving and Critical Thinking were commonly mentioned.

Table 9.1 Generic skills valued by graduate recruiters

Valued skill/attribute	Number of recruiters listing this skill (maximum possible is 10)
Interpersonal/emotional intelligence	10
Communication (oral or written)	6
Teamwork/collaboration	5
Problem solving	4
Critical thinking/reasoning	4
Adaptability/flexibility	3
Creativity/iInnovation	3
Initiative/Can-do attitude	2
Leadership	1
Ability to learn	1
Customer service	1
Computing skills	1

These responses are particularly meaningful when compared with the results of the ACNielson study (2000), the initial responses from stakeholders at the start of the GSA project (Table 2.1) and other input from stakeholders (Table 9.2).

The participants were also asked to rate the value of the GSA construct and a sample of items. In the ratings, '5' meant very important, '3' meant of moderate importance, and '1' meant of no importance.

In response to the question 'How do you rate the value of the generic skills assessed by the GSA to graduate employees in your organisation?', the typical rating was '4' (ranging from 3 to 5), which corresponds to 'important' on the scale provided.

In response to the question 'How do you rate the value of the generic skills assessed by the GSA to graduate employees in general?', the typical rating was 4 (ranging from 3 to 5), which corresponds to 'important' on the scale provided.

Thus, in general, despite the cognitive approach of the GSA, the respondents rated it as a useful tool by which to provide additional information of relevance on prospective employees.

9.4 Appraisals of the GSA by students

At one university, the 60 students doing the trial versions of the GSA were asked about various aspects of the test. Of these students, 19 responded. Their views are summarised below exactly as reported by staff from the university. The summary below is as reported, except that reference to the particular university has been deleted. Responses to the question 'Does the test measure anything important?' are probably most relevant.

What are advantages of doing the test?

The majority of respondents reacted positively to the test overall. Seven of the 19 respondents believed the test would be useful in giving them an edge in a highly competitive employment market, provided their results were good. The majority found the test challenging and reported enjoying it, although, as noted below, a significant number complained about its length.

Some saw considerable benefit in being able to compare their level of generic skill with the students in equivalent courses elsewhere. Others saw merit in being able to identify their specific strengths and weaknesses.

What are disadvantages of doing the test?

The majority of respondents believed that the test was too time consuming and that it should not be administered so close to the …exam period. One student suggested that, for graduating students, the ideal time might be in September as this is when graduating students are on the job market. Some students wondered what the impact would be on those students who received poor results and asked whether or not counselling would be available.

A number of students said that the test's intention remained unclear. They were unsure, for example, whether it was a psychological, vocational or basic skills test, or something else. A smaller group wondered whether it was really just an end of high school aptitude test.

A range of reliability issues was raised. These included the extent to which the language used in the test might disadvantage NESB students, and the extent to which communications' skills in particular can be reliably tapped using multiple choice or short answer questions.

One respondent was most concerned that GSA results, if aggregated, might be used as part of a new system of performance-based funding.

Five of the 19 respondents saw no disadvantages in undertaking the test, provided participation remained voluntary.

Does the test measure anything important?

17 respondents said yes, one said no and one was undecided. Those who gave a positive response said the test tapped skills which they believed were now central to effective professional practice, in particular critical thinking, problem solving, written communication, interpersonal skills, and comprehension. One respondent suggested that an item on interview skills might also be included. The team skill items were singled out by some respondents as being especially important. As noted earlier, this reaction generally aligns with the findings other studies.

The respondent who gave a negative response said that this was because the test did not accurately tap the above skills, that it was more like a high school test and that, as already noted, some areas (e.g. communicative skills) could never be reliably measured using multiple choice questions. Language difficulties for NESB students were also again raised.

Should the university participate in the GSA in future years?

14 respondents said yes, two said no, two were undecided and one did not respond.

Broadly, those who recommended future…participation said that this was because the test would help students self evaluate their relative areas of strength and weakness in key skill areas. They also said that, if the GSA was given in both year one and just before graduation, this would enable students to track their development in these areas. They suggested that using the GSA in this way could help … to position itself distinctively, especially if there was explicit employer endorsement of the test's validity. Another respondent said it was one way for the University to demonstrate how…assists young people to grow into more self-managed and capable adults.

Others said that the GSA could be used to compare…student performance in parallel courses at similar universities or that the data generated could help identify key areas for improvement in existing programs.

The student who said the University should not expand its use of the GSA repeated earlier points about the validity of the test, arguing that what it measures is not necessarily what makes a difference in the field. The current…'successful graduate' tracking project will help test this hypothesis.

Another student reiterated that using the GSA was fine, provided the results were given only to individual students and not aggregated for any external auditing purpose or used as a basis for performance-based funding.

The undecided respondents emphasised that their indecision came from a fear that the test might be made compulsory or used as part of the normal...assessment system.

Amongst other comments made, one relevant to validity was:

Respondents agreed that parallel versions of the test were needed to prevent cheating or collusion.

The responses of these students suggest that the GSA approach is generally, though not universally, seen as valid and useful. Some doubts about test validity expressed by students need to be addressed by appropriate longitudinal studies.

9.5 Appraisals of the GSA by some other stakeholders

Some other stakeholders were interviewed on general issues related to the test, the suitability of test components and the suitability of Interpersonal Understandings items (since this component was anticipated to be the most controversial). The stakeholders were:

1 A group from one university (dominated by Arts/Humanities academics);

2 A representative from an Engineering employer group;

3 A representative from an Architect employer group; and

4 A group from a private management education body.

A summary of the views expressed is given in Table 9.2. (Note that not all the items seen and discussed by the stakeholders are actually used in the test.)

In the discussion with these other stakeholders, it was apparent that there were dramatically different views about aspects of the test, and to some extent these views were related to the background of the stakeholder. Both positive and negative comments were made. For example, the Engineer employer and the Humanities academics had diametrically opposed views on certain Interpersonal units. Humanities academics tended to criticise the need for numeracy in Problem Solving while those from technical/science backgrounds tended to criticise the need to deal with information presented as text.

In general, issues of concern for these stakeholders (which often mirrored comments made by the experts and students) included issues such as: the possibility of league tables appearing, whether there are generic skills outside disciplines or work situations, privacy of results, whether universities actually teach such generic skills, limitations of multiple-choice items, relevance of interpersonal skills to researchers, audience specification and scaffolding for writing, relevance to university students, relevance to post-graduate work, cultural and ESL bias, and so forth.

Table 9.2 Summary of views of other stakeholders

Issue	Summary of key comments
General issues:	• Concerned about university 'league tables'. • Concerned about how results of the test may be used. • Could lead to a 'healthy' focus on generic skills. • Additional generic skills could be included in the future (e.g. research skills, commitment and drive). • Can suitable generic skills for employment be developed at university? • Different employers want different things. • Are the skills generic? Can you have a 'generic' report? • Length of test. • Privacy issue. • Wordy.
Construct: Problem Solving	• Too much text based, stimulus could be more graphical. • Too much numeric/mathematical. Lower secondary mathematics may be an inappropriate level for university students (implication that there could be more sophistication in thinking without requiring mathematical knowledge). • Suits engineers … it looks pretty good…comprehensive.
Construct: Critical Thinking	• Why mainly text based? • 'Critical Thinking' could be misinterpreted by international students (that it could mean criticising someone). Is 'analysis and synthesis' a better term? • Synthesis should be more prominent, as with Problem Solving. 'We said we wanted critical thinking whereas it may be more accurate to have Analysis and Synthesis' • Similarity of problem solving/critical reasoning: why not put them together?
Construct: Interpersonal Understandings	• Do you learn this at university? More likely to learn in part-time work? Still a good thing to assess, but depends on the purpose of the test. OK to use GSA interpersonal as a description of a person, but unfair to use it to reflect quality of institution. • Important, but must remember there are good researchers who are 'loners' (again, depends on purpose of test). • Would be better if spoken rather than written.
Construct: Written Communication	• Important to see that students can deal with different sources of information (as presented in the Report Task). • Task 1 (Report) particularly useful for Engineers but not task 2 (Argument). • For a graduate, information is often filtered higher up in the organisation first. • Too much visual material in Task 1 (Report). Too complex. • Very open ended, clearer framework desirable. Specify audience as in real life, there's no context. • Email is how people communicate nowadays. You also have to write letters.
Interpersonal Understandings Item Types	• Good to test how people deal with situations when they are equals, not in power. • Interpersonal items may not cover adequately ability to work in a team. • Very workplace related. How appropriate is this for students? Steer clear of public service type work place! (will students have experienced this? generic?) • Need to know how to extract information from people (not tested here). Use of the phone is important – don't know how you could assess it here. • Illustrations are a good idea but need to be better quality. • Importance of awareness of cultural issues in testing. • Quite a lot of reading.

9.6 Concluding comments

The challenge for test developers of producing an appropriate theory-based and empirically validated test of generic skills that satisfies a range of stakeholders with competing demands is a substantial one. Although overall there seems to be significant support for the GSA approach and its face/content validity, there are also concerns and differences of opinion.

A number of the concerns expressed may be addressed following empirical research, which may lead to validation of the current approach or improvements in the test design.

Some concerns can be addressed by refinement of the test items and level descriptors as suggested. Other concerns are related to policy issues (e.g. use of results) that cannot be addressed by test developers.

It may be that different versions of the test should be produced for different stakeholder purposes. Test developers and stakeholders need to consider this option. If a single GSA test is to be used, more opportunity for stakeholder involvement in test design and research could be useful as the test evolves.

10. Conclusions and recommendations

The first part of this chapter presents conclusions in relation to each of the five study aims. The second part of the chapter provides some recommendations for the future development of the GSA test.

10.1 Conclusions

Because of the inadequacy of the sample of students sitting the first two GSA tests with respect to its representativeness of the total university population (Chapter 4), caution needs to be used in drawing conclusions from the results of this study. Nevertheless, it is still expected that general findings related to the factor structure of the test, variables related to performance on the test and the relationship between performance on the GSA and other measures of achievement will have validity.

10.1.1 Aim 1: To investigate the dimensional factor structure (discriminant validity) of the test

As described in Chapters 1, 2 and 3, the GSA attempts to measure the generic skills of university students with respect to five cognitive dimensions. Confirmatory factor analysis was performed in order to validate this (Chapter 5).

Factor analysis indicates that the GSA does indeed consist of five psychometrically distinct dimensions, an observation that helps validate the test construct in terms of discriminant validity. A higher order factor is also apparent that could relate to a meta-cognitive generic reasoning skill of broad applicability. Although analysis indicates a suitable five dimension factor structure for GSA Entry 2001, it is important to monitor the factor structure of all GSA tests.

10.1.2 Aim 2: To identify variables related to differential performance on GSA

As described in Chapter 6, analyses were done in order to identify variables related to performance on the GSA.

The analysis identified variables that appear to be related to performance on GSA components in a way consistent with the design of the test, such as field of study and year level, but these observations need clarification in follow-up studies with appropriate samples. Whether the relationship with test performance of other variables, such as English-speaking background, age and gender, is appropriate also needs further investigation in relation to outcomes at university and in the work place. Most variance in student score appears to relate to variables such as individual student skill level and motivation, as would be expected.

10.1.3 Aim 3: To investigate the relationship between student performance on GSA and other measures of student achievement

Given the purpose of the GSA, student performance should correlate significantly with performance on measures of student achievement such as TER and GPA, as well as post-graduate performance. As reported in Chapter 7, an investigation of the relationship between GSA, TER (TES) and GPA (GPA-like) scores was done.

GSA performance correlates significantly (statistically) with both TER and GPA, despite performance on these two measures being related to subject knowledge in a wide range of curriculum subjects, which underlines the generic nature of skills addressed by the GSA and supports its validity. It seems feasible that GSA performance could be a useful predictor of performance in university courses, and, possibly, in postgraduate work, though further research is required to investigate this.

10.1.4 Aim 4: To consider the suitability of current reference ranges

In order to judge and compare student performance reliably and to draw valid conclusions to research questions, the GSA sample needs to represent the university population appropriately. Because of the way the sample was self-selected, there is doubt about the suitability of the sample's representativeness and Chapters 4 and 8 looked at this issue.

It appears that GSA reference ranges and reliability are likely to be limited for some purposes. Overall reference ranges may be set too high and particular problems may exist with reference ranges for fields of study where little data has yet been collected. More representative samples should be sought to check/improve reference ranges. In addition, described levels of performance should be validated in consultation with university and employer representatives. Reliability may be a problem when measuring small changes in the performance of a single student between years. Such issues need further investigation. Specially tailored, but statistically linked, tests may be appropriate for certain purposes.

10.1.5 Aim 5: To evaluate the face/content validity of the GSA construct and items

As described in Chapter 9, the views of content experts and stakeholders were sought in relation to the face/content validity of the GSA.

It appears that, although the GSA has considerable face/content validity for many content experts and stakeholders, several issues need to be addressed. Responses by test developers to some issues need to be based on more empirical evidence than is currently available (e.g. about the predictiveness of the test for graduate workers). Consultation with stakeholders on matters such as described levels of performance and the issue of the use of specialist versions of the test would be useful. Other issues may only be addressed by policy decisions.

10.2 Recommendations for the future development of the GSA

Given the preceding discussion, the following recommendations for the future development of the GSA are made:

1 Continuing attempts should be made in association with universities to obtain representative student data.

2 The factor structure of the test should continue to be monitored to ensure that the test remains appropriately focussed.

3 Further investigations should be undertaken to confirm , clarify, and more precisely quantify relationships between performance on the GSA and variables such as field of study and year level, and to investigate the appropriateness of differential performance on the basis of variables such as English-speaking background and gender. Investigations broadening the range of variables examined could be done.

4 Further investigations should be undertaken into the relationships between GSA performance and markers of achievement at university and work. Evidence could include reports on students and graduate workers by tutors and supervisors.

5 Consideration could be given to the use of the GSA for selection into university courses.

6 Reference ranges should be refined, including those for sub-groups, such as specific field of study and year level cohorts.

7 There should be further evaluations of whether test reliability and described levels of performance are suitable for the particular purposes for which the results are being used. If reliability is not sufficient for a particular purpose, consideration should be given to ways of improving it.

8 In consultation with stakeholders, consideration should be given to the refinement of face/content validity, and construct and level descriptions, where possible, these being based on a comprehensive and commonly accepted developmental model of generic skills.

9 The purpose(s) of the test should be clarified in consultation with stakeholders and, if appropriate, versions of the test tailored for specific stakeholder purposes could be produced, that are linked statistically to the general test.

10 Assessment of validity should be ongoing as the test evolves, and stakeholders should be involved in evaluation and research.

10.3 Concluding remarks

1 The challenge for test developers of producing an appropriate theory-based and empirically validated test of generic skills that satisfies a range of stakeholders with competing demands is a substantial one. In relation to this, more discussion with stakeholders about the purpose, design and value of the test, as well as more opportunity for stakeholder involvement in test design and research, may be useful.

2 Assessment of the validity of the GSA is a complex process. This study is a first step that provides evidence in favour of the validity of aspects of the GSA as it currently operates, but also raises some concerns. As the GSA evolves in response to feedback, ongoing assessment of validity will be required.

Bibliography

ACNielsen Research Services (1998). *Research on Employer Satisfaction with Graduate Skills Interim Report*. Evaluations and Investigations Programme Higher Education Division, DEETYA: Canberra.

ACNielsen Research Services (2000). *Employer Satisfaction with Graduate Skills: Research Report*. Evaluations and Investigations Programme Higher Education Division, DETYA: Canberra.

Adams, R.J. & Khoo, S.T. (1993). *Quest: The Interactive Test Analysis System.* Australian Council for Educational Research: Hawthorn, Victoria.

Assiter, A. (ed.) (1995). *Transferable Skills in Higher Education.* Kogan Page: London.

Australian College of Organisational Psychologists (1999). Comparison of Employers' and Academics' Views about the Importance of Generic Competencies for Psychology Graduates, *Interface.* 27, April 1999.

Bransford, J.D. & Stein, B.S. (1993). *The Ideal Problem Solver.* W.H. Freeman and Co.: New York.

Bridgman et al., (2000). Predictions of Freshman Grade-Point Average from Revised and Recentred SAT 1: Reasoning Test (College Board Research Report No. 2000-1/ETS Research Report 00-1). New York, NY: College Board.

Carnevale, P.J. (1991). *America and the New Economy.* Jossey Bass: San Francisco.

Carnevale, P.J. & Pruitt, D.G. (1992). Negotiation and mediation. *Annual Review of Psychology*, 43, 531–582.

Carrol, J.B. (1993). *Human Cognitive Abilities*. Cambridge University Press: Cambridge.

Clanchy, J. & Ballard, B. (1995). Generic Skills in the Context of Higher Education. *Higher Education Research and Development*, 14(2), 155–166.

Conference Board of Canada (2000). *Employabiliy Skills 2000+*. Conference Board of Ottawa: Ottawa.

Curtis, D. & McKenzie, P (in press). *Employability Skills for Australian Industry: Literature Review and Framework Development*. Australian Council for Educational Research: Melbourne.

de Bono, E., (1977). *Lateral Thinking*. Pelican: London.

DETYA (2001). *Students 2000, Selected Higher Education Statistics.* Commonwealth of Australia.

Everett, J.E. & Robins, J. (1991). Tertiary Entrance Predictors of First-Year University Performance. *Australian Journal of Education*, 35(1), 24-40.

Flynn, J.R. (1999). Searching for justice: the discovery of IQ gains over time. *American Psychologist*, 54, 5-20.

Gallese V. & Goldman A. (1998). Mirror neurons and simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493-501.

Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences.* Basic Books: NY.

Gibbs et al., (1994). *Developing Students' Transferable Skills.* The Oxford Centre for Staff Development: Oxford.

Gottfredson, L.S. (1997). 'Why *g* matters: The complexity of everyday life'. *Intelligence*, 24(1), 79–132.

Hambur, S. (1997). *Generic Factors and Curriculum, An Investigation of Test Balance in Cross-curricular Tests: 1. A Comparison of AST-C and AST-E* (unpublished internal monograph). Australian Council for Educational Research: Melbourne.

Hambur, S. (1998). *Generic Factors and Curriculum, A Review of Construct Validity and Factorial Balance in Cross-curricular Tests: 2. GAT and GAMSAT* (unpublished internal monograph). Australian Council for Educational Research: Melbourne.

Hambur, S. & Glickman, G. (2001). *Summary Report: GSA Exit 2000* (unpublished internal report). Australian Council for Educational Research: Melbourne.

Hambur, S. & Le, L. (2001). *Summary Report: GSA Entry 2001*, (unpublished internal report). Australian Council for Educational Research: Melbourne.

Harvey, L. & Green, D. (1994). *Employee Satisfaction Summary.* Quality in Higher Education Project: Birmingham.

Higgins, J.M. (1994). *101 Creative Problem Solving Techniques: The Handbook of New Ideas for Business.* The New Management Publishing Company. Winter Park: Florida.

Hoy, W.K. & Tarter, C.J. (1994). *Administrators Solving the Problems of Practice*. Allyn and Bacon: Boston.

Jones, E. & Ratcliffe, G. *Critical Thinking Skills for College Students.* ERIC-No.: ED358772.

Jöreskog, K.G., & Sörbom, D. (1999). *LISREL 8.30: Interactive LISREL for MS Windows 3.1+ Windows '95, Windows '98 and Windows NT*. Scientific Software International, Inc.: Chicago, IL.

Kearns, P. (2001). *Review of research: generic skills for the new economy*. NCVER: Leabrook, South Australia for the Australian National Training Authority.

Kuhn, D. (1999). A Developmental Model of Critical Thinking. *Educational Researcher*, 28(2), 16–26.

Legree, P.J. (1995). Evidence for an Oblique Social Intelligence Factor Established with a Likert-Based Testing Procedure. *Intelligence*, 21, 247–266.

Marks, G., McMillan, J. & Hillman, K (2001). *Tertiary Entrance Performance: The Role of Student Background and School Factors, Longitudinal Surveys of Australian Youth, Research Report Number 22*. Australian Council for Educational Research: Melbourne.

Mayer Committee (1992). *Putting General Education to Work: The Key Competencies Report*. AEC/MOVEET, Melbourne.

Mayer, J.D. et al., (1999). Emotional Intelligence Meets Traditional Standards for an Intelligence. *Intelligence*, 27(4), 267–298.

McDonald, AS. et al., (2001). *Aptitude Testing for University Entrance: A Literature Review.* The Sutton Trust.

McKenzie, K. & Schweitzer, R. (2001). Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development*, 20(1), 21–33.

Mumford, B. et al., (1998). Creative Thinking Skills, Chapter 7, in *Beyond Multiple-Choice: Evaluating Alternatives to Traditional Testing for Selection*, Hakel, M. D. (ed.). Lawrence Erlbaum Associates: New Jersey.

NAB/UGC (1984). *Higher Education and the Needs of Society*. National Advisory Board for Public Sector Education: London.

Norris, S.P. & Ennis, R.H. (1989). *Evaluating Critical Thinking*. Critical Thinking Press and Software. Pacific Grove: California.

NSTF (2000). Skills for all: Research report. DfEE: Sheffield.

Paul, R. (1994). *Critical Thinking*. Hawker Brownlow Education: Australia.

Pithers, R.T. & Soden, R. (2000). Critical Thinking in Education: A Review. *Educational Research*. 42(3), 237–249, Winter.

Polya, G. (1957). *How To Solve It*. 2nd edn. Doubleday Anchor: New York.

Power, C., Robertson, F. & Baker, M. (1987). *Success in Higher Education*. Australian Government Publishing Service: Canberra:

Rashbash, J., Browne, W., Healy, M., Cameron, B. & Charlton, C. (2000). *MLwiN (Version 1.10.006): Interactive software for multilevel analysis*. Multilevel Models Project, Institute of Education, University of London: London.

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

Rizzolatti, G. & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188-194.

Smith, P. & Whetton, C (1992). *Critical Reasoning Tests.* Australian Council for Educational Research: Melbourne, by arrangement with NFER-Nelson: Berkshire, England.

The Association of Graduate Recruiters (1995). *Skills for Graduates in the 21st Century*. The Association of Graduate Recruiters: Cambridge.

The Secretary's Commission on Achieving Necessary Skills (2000). *Skills and Tasks for Jobs: A SCANS Report for America 2000*. US Department of Labor.

Watson, G. & Glaser, E.M. (1980). *Watson-Glaser Critical Thinking Appraisal Manual*,. Psychological Corporation: New York.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th Ed). Williams and Wilkins: Baltimore.

Whimbey, A. & Lochhead, J. (1991). *Problem Solving and Comprehension*, Lawrence Erlbaum and Associates: New Jersey.

Wright, B.D. and Masters, G.N. (1982). *Rating Scale Analysis.* MESA Press: Chicago.

Wright, B.D. and Stone, M.H. (1979). *Best Test Design*. MESA Press: Chicago.

Wu, M.L., Adams, R.J. & Wilson, M.R. (1997). *Conquest: Generalised Item Modelling Software*. Australian Council for Educational Research: Hawthorn, Victoria.

Appendix 1: Carnevale's 16 job skills for the contemporary workforce

Learning to learn

1 Foundation skills: learning how to learn – how to collect, know and comprehend, how to give and receive feedback, and how to learn collaboratively.

Academic basics

2 Reading skills: basic literacy, reading in order to learn, reading in order to do.

3 Writing skills: preparing and organising information, writing, editing, revising.

4 Computational skills: quantification, computation, measurement and estimation, quantitative comprehension, quantitative problem solving.

Communication

5 Speaking skills: nonverbal skills, vocal skills, verbal skills.

6 Listening skills: assigning meaning to aural stimuli.

Adaptability

7 Problem-solving skills: the ability to bridge the gap between what is and what ought to be.

8 Creativity skills: the ability to produce a novel idea, and then turn it into a practical one.

Personal development

9 Self-esteem skills: the ability to maintain a realistic and positive self-image.

10 Motivation and goal-setting skills: the ability to translate work into an instrument for the development of self.

11 Personal and career development skills: the ability to adapt to changing work requirements to ensure employment security and to fulfil personal potential.

Group effectiveness

12 Interpersonal skills: the ability to judge appropriate behaviour, to absorb stress, to share responsibility, to deal with ambiguity.

13 Negotiation skills: the ability to overcome disagreements by compromising and accommodating.

14 Teamwork skills: the ability of groups to pool human resources to pursue common goals.

Influencing skills

15 Organisational effectiveness skills: the ability to work productively in the context of explicit and implicit organisational cultures and subcultures.

16 Leadership skills: the ability to influence others to serve the strategic purposes of an organisation or the developmental needs of an individual.

Figure A1.1 Carnevale's 16 job skills for the contemporary workforce

(derived from *Carnevale's America and the New Economy*, 1991, and adapted from Kearns, 2000)

Appendix 2: Expectations of employers and academics

Table A2.1 outlines the results of a study (Australian College of Organisational Psychologists, 1999) comparing the expectations of employers and academics of psychology graduates. Forty-one generic competency domains were rated from 1 (not at all necessary) to 10 (absolutely necessary). Mean ratings are given in the table.

Table A2.1 Comparison of employers' and academics' expectations

Competency Domain	Employer Ranking	Academic Ranking	Employer Mean	Academic Mean	Significant Difference ($p < .05$)
Establishing positive working relationships	1	4	8.74	8.14	.006*
Identifying information requirements	2	1	8.38	8.38	.985
Demonstrating self-management	3	7	8.34	7.89	.092
Managing workload efficiently and effectively	4	9	8.25	7.62	.093
Developing performance in response to self-reflection and feedback from others	5	8	8.19	7.88	.305
Working with others	6	10	8.16	7.43	.004*
Gathering and recording information	7	3	8.15	8.14	.939
Presenting information	8	5	8.14	8.12	.968
Demonstrating problem solving skills/abilities	9	6	8.10	7.90	.367
Managing own work functions and tasks	10	15	7.98	7.24	.049*
Demonstrating self-awareness	11	12	7.95	7.35	.077
Participating in team meetings	12	21	7.89	6.93	.016*
Utilising networks	13	33	7.86	6.51	.094
Demonstrating effective communication skills	14	14	7.84	7.25	.009*
Undertaking work activities according to a plan	15	22	7.83	6.87	.024*
Participating in training and development	16	19	7.80	7.07	.041*
Following instructions in the workplace	17	11	7.70	7.35	.427
Organising and maintaining own work performance	18	13	7.69	7.33	.201
Using feedback to review communication	19	23	7.65	6.85	.015*
Managing change in the workplace	20	18	7.65	7.17	.002*

Recognising and responding to the physical and psychological needs of oneself in the workplace	21	24	7.60	6.85	.110
Analysing and editing information	22	2	7.56	8.22	.018*
Using technology	23	32	7.54	6.51	.179
Managing interpersonal conflict in the workplace	24	29	7.47	6.60	.001*
Identifying purpose of work groups	25	35	7.39	6.50	.009*
Following workplace procedures for hazard identification and risk control	26	41	7.37	5.66	.155
Implementing organisation's processes in own work practice	27	34	7.31	6.51	.014*
Contributing to team commitment in work groups	28	25	7.29	6.84	.269
Developing and maintaining appropriate networks	29	39	7.17	6.29	.001*
Demonstrating understanding of, and ability to work with, differing roles and responsibilities of group members	30	30	7.16	6.59	.042*
Supporting group members of work teams	31	31	7.16	6.53	.233
Identifying the need for networks	32	38	7.16	6.30	.012*
Establishing relationships with work groups	33	26	7.13	6.70	.049*
Supporting group objectives when working in teams	34	37	7.09	6.38	.006*
Planning work activities	35	28	7.06	6.63	.217
Demonstrating ability to be self-promoting	36	16	7.05	7.19	.860
Promoting and disseminating inform-ation	37	27	6.89	6.63	.566
Planning and preparing information for communication	38	20	6.84	7.01	.494
Communicating in a range of contexts	39	36	6.81	6.41	.289
Demonstrating awareness of organisational structures, roles and goals	40	40	6.79	6.21	.264
Using mathematical ideas and techniques	41	17	5.87	7.18	.153

NB: Competency domains presented in order of employer preference; * p < = .05

There is quite a lot of agreement about the importance of the top dozen or so competencies.

Appendix 3: Description of units/criteria in GSA entry 2001

Table A3.1 briefly describes the units/criteria that form the GSA Entry 2001 test. Test items are secure and so are not available for presentation but were viewed by stakeholders responding in Chapter 9.

Table A3.1 Description of units/criteria in GSA entry 2001

Unit	Description
CTcorp	Analyse short passage to evaluate the strength of an argument, infer writer's intent and assumptions.
CTcit	Identify logical consequence of rules and relationships described in text, identify assumption necessary in order to logically draw a conclusion.
CTdrug	Evaluate statements' relevance to a proposition.
CThux	Comprehend writer's argument, make inferences, identify best counter-example.
CTsaw	Analyse passages to identify statements/evidence that most supports or counters claim/ argument or best summarises main argument, evaluate positions of author's of two passages.
CTshor	Identify claims and assumptions related to a line of questioning.
CTstat	Identify statement that most directly counters a proposition.
	Identify statement about a proposition that is most readily testable.
	Identify statement most consistent with an hypothesis.
	Evaluate reasoning implicit in an argument.
CTsued	Analyse passage to identify most appropriate response to an assertion and likely consequence, evaluate possible counters to a proposition and credibility/validity of a claim.
CTviol	Comprehend and reword key points in passage.
IPang	Identify attitude of character based on verbal and non-verbal cues.
IPdoct	Identify statements most consistent with view expressed by speaker.
IPdrjo	Identify most appropriate way of dealing with a late assignment.
IPdrm	From scenario, infer reason why a comment is inappropriate, level of awareness/understanding of another's attitude and type of response by one person that is most consistent with a second person's view of them.
IPemp	Infer most empathetic response in scenario.
IPeng	From scenario, infer most reasonable responses to dissatisfied client in order to have fair and reasonable negotiation.
IPlist	Evaluate listening effectiveness.
IPmar	Identify most appropriate way of dealing with an error in a joint assignment.

IPshor	Identify approach that is likely to minimise conflict in a meeting.
	Identify response most likely to indicate knowledge of effective teamwork.
	Identify advice most likely to encourage active participation in meeting.
IPt4	Identify probable underlying feelings of person making a comment.
	Identify behaviour least commensurate with effective learning from a teamwork error.
	Evaluate options for influencing selected groups in order to most effectively change organisational behaviour.
IPt5	Identify most appropriate listening behaviour.
	Identify statement that indicates greatest problem with teamwork.
IPt6	Identify behaviour most likely to cue person with problem you are not interested.
	Identify approach to negotiation that minimises conflict and maximises success.
	Identify most appropriate thing to say to help friend with problem.
PSbatt	Identify and apply appropriate tabular-numeric data in order to make a decision and evaluate conclusions
PSbrid	Evaluate possible outcomes by logically applying rules.
PSchd2	Apply appropriate tabular-numeric data to evaluate conclusions and reasons for treating data in a particular way.
PSchd3	Draw conclusions and make comparisons based on quantitative reasoning related to graph.
PShous	Interpret and apply rules to determine optimal solution to problem, and interpret, apply and evaluate suitability of proposed solutions presented in diagrammatic form with added constraints.
PSlaw	Apply rules to identify logical consequences of a decision.
PSsale	Identify key information about a problem presented in text. Translate/reorganise information into appropriate diagram format to identify relationships and optimum solution paths given constraints.
PSshor	Solve word problems requiring simple pseudo-algebraic, geometric and proportional reasoning.
PSstaf	Analyse text and tabular-numeric information and apply to staffing problem to determine optimum solution with and without constraints.
PStrav	Analyse and apply timetable information to identify worst and best case outcomes, and optimal assignment of vehicles.
WRRep1	Quality of thought and ideas.
WRRep2	Quality of structure and organisation.
WRRep3	Quality of language and expression.
WRArg1	Quality of thought and ideas.
WRArg2	Quality of structure and organisation.
WRArg3	Quality of language and expression.

Appendix 4: Assignment of students to fields of study

Figure A4.1 is the Field of Study Guide students used to match their courses to fields of study.

Field	Code
Architecture/Built Environment (e.g. Architecture, Construction, Drafting, Environmental Design, Landscape Architecture, Surveying)	010
Agriculture/Animal Husbandry/ Environmental Management (e.g. Agriculture, Animal Husbandry, Environmental Management, Environmental Science, Forestry, Parks and Wildlife, Plant Science, Soil Science, Resource Management, Rural Management)	020
Arts – Creative (e.g. Drama, Dance, Fine Arts, Graphic Design, Film, Media Studies, Music, Photography, Visual Arts)	031
Arts – Humanities (e.g. Asian Studies, Australian Studies, English History, Journalism, Library, Linguistics, Literature, Philosophy, Politics, Public Relations, Theology, Writing and Editing	032
Arts – Social Science (e.g. Anthropology, Geography, Government, Psychology, Sociology)	033
Arts – Languages	034
Business- Financial (e.g. Accounting, Actuarial Studies, Economics, Finance)	041
Business-Management/Marketing (e.g. Administration, Business Studies, Human Resource, Industrial Relations, International Business, Public Relations)	042
Computer-Programming/Design (e.g. Computer Studies, Multimedia, Programming, Systems Analysis)	051
Computer- Information Technology (e.g. Data Communication, Information Management)	052
Education – Primary/ Early Childhood	061
Education – Secondary	062
Engineering	070
Law	080
Legal Studies	081
Mathematics	090
Health – Medical (e.g. Dentistry, Medicine, Veterinary)	101
Health – Nursing	102
Health – Science (e.g. Nutrition, Medical Technology, Occupational Therapy, Optometry, Pharmacy, Physiotherapy, Podiatry, Speech Pathology)	103
Science – Applied (e.g. Animal Technology, Aviation, Biotechnology, Electronics, Food Technology, Marine Science, Sports Science)	111
Science – Biological (e.g. Behavioural Science, Biology, Biochemistry, Genetics, Microbiology, Pharmacology, Physiology, Psychology, Zoology)	112
Science – Physical (e.g. Chemistry, Earth Science, Meteorology, Physics)	113
Social Work/ Community Services	120
Tourism/Hospitality/Catering	130
Other	140

Figure A4.1 Field of study guide

Table A4.1 shows how smaller field of study groups were combined into larger ones for the purposes of reporting. Table A4.2 gives the field of study compositions of the Exit 2000 and Entry 2001 cohorts.

Table A4.1 Field of study groups

Field of Study Group	Field of Study Codes
Arts/Humanities	031, 032, 033, 034
Business/Commerce	041, 042, 130
Computers/IT	051, 052
Education/Social	061, 062, 120
Engineering/Architecture	010, 070
Science/Math	020, 103, 111, 112, 113, 090
Law/Legal	080, 081
Medicine/Dentistry	101
Nursing	102

Table A4.2 Field of study compositions of the exit 2000 and entry 2001 GSA cohorts

Field of Study	Exit 2000	Entry 2001	Combined
010	18	87	105
020	54	61	115
031	35	24	59
032	127	64	191
033	155	58	213
034	14	6	20
041	267	218	485
042	214	98	312
051	76	110	186
052	67	69	136
061	32	133	165
062	23	10	33
070	100	88	188
080	33	12	45
081	4	6	10
090	14	3	17
101	9	333	342
102	13	188	201
103	35	168	203
111	51	65	116
112	160	101	261
113	29	58	87
120	16	6	22
130	3	6	9
140	7	0	7
Missing	41	67	108
Total	1597	2039	3636

Appendix 5: Calculation of GSA scores

A student's GSA score on one of the five GSA dimension scales is obtained by converting the student's raw score into a *logit** score using an Item Response Theory model, namely the Rasch model (Rasch, 1960/1980). Quest (Adams & Khoo, 1993) and Conquest (Wu, Adams & Wilson, 1997) softwares are used for this purpose.

By using this model, both the item difficulty and student ability can be expressed in logits on the same latent scale. The estimates of item difficulties are independent of the abilities of the group of students who responded to them and similarly the estimates of students' abilities are independent of the difficulty of a test form (Wright & Stone, 1979; Wright and Masters, 1982). The difficulty of an item (relative to the other items) and the ability of a student are both expressed in logits on a scale ranging from minus infinity to plus infinity.

Common items are used to link various GSA forms so that all students and items can be put on the same scale, irrespective of which GSA form the student did.

Logit scores for each student are converted to GSA scores using linear formulae that, for convenience of interpretation, allow the described levels of the various components to be aligned on the report form.

** The logit is the natural logarithm of the odds of the event, where the odds of the event is defined as the ratio of the probability that the event will occur to the probability that the event will not occur. The logit scale is used because it is an interval scale. That is, if the difficulty of Item A is 1.0 logits greater than the difficulty of Item B, then the odds of a student responding correctly to Item B are 2.7 times the odds of the same student responding correctly to Item A, regardless of whether this student has high or low ability. Similarly, if the ability of Student A is 1.0 logits greater than the ability of Student B, then the odds of Student A responding correctly to an item are 2.7 times the odds of Student B responding correctly to the same item, regardless of how difficult the item is.*