

A Review of Aspects of the Queensland Core Skills Test

A review of the Queensland Core Skills (QCS) Test to ascertain the ongoing relevance of the test and the capability of the test to act as a statistical scaling device in the calculation of Overall Positions (OPs) and Field Positions (FPs) for tertiary selection

**Produced by Enterprising Minds Pty Ltd
Tony McLeod and Michelle Davidson**

**in association with
John Bennett
George Cooney
Bob Edwards**

**and in consultation with
Jim Tognolini**

February 2012

Contents

Executive Summary.....	3
1. Introduction.....	7
2. Review of the test procedures.....	10
3. Consideration of test-specific issues related to the QCS Test.....	20
4. Other issues identified during the stakeholder consultation.....	29
5. The future.....	37
6. Concluding comments.....	40
Appendices	
A. Documents considered by the Review Team.....	41
B. QSA officers interviewed by the Review Team.....	42
C. Meetings with supervisors of marking, immersers/supervisors and markers..	43
D. School visits, other meetings and phone interviews.....	44
E. Psychometric analysis of the QCS Test.....	46
F. The scaling procedure.....	58

Executive Summary

The Queensland Studies Authority (QSA) commissioned this review of the Queensland Core Skills (QCS) Test *'to ascertain the ongoing relevance of the QCS Test and the capability of the test to act as a statistical scaling device in the calculation of Overall Positions (OPs) and Field Positions (FPs) for tertiary selection'*.

The QSA specified that the review should consider all the test processes, investigate certain specific issues identified by the QSA, and provide advice on future contextual issues.

The QCS Test has been conducted for 20 years. It was previously externally reviewed in 1992 and 1996. Both of these reviews were positive about the test and its associated processes and practices.

In addition to reviewing operational manuals and reports, and analysing data provided by the QSA, the Review Team met with many external and internal stakeholders to ascertain their views on a wide range of issues related to the test. The Review Team is grateful for the cooperation and assistance of all involved, and their willingness to share their experiences and opinions of the test is very much appreciated.

The Review Team is of the view that the QCS Test is a sound and well-conducted program. The test itself is of high quality due to its design criteria, the care and expertise of those involved and the sound procedures employed in its development. The administrative procedures are well established and run smoothly and effectively. The marking processes are thorough, carefully applied and well monitored. The Review Team believes that the processes applied to the developing and reviewing items, to assembling the test, to administering the test, and to marking student responses represent 'best practice'. Considerable effort is put into ensuring that the test will engage students and provide for a range of student interests and abilities. The QSA and its employees are to be commended for their commitment and success in delivering a quality program.

The purpose of the QCS Test is primarily to act as a scaling instrument in the calculation of OPs and FPs used for tertiary selection. However, because of its close connection to the Queensland senior secondary curriculum through the Common Curriculum Elements (CCEs), there is scope for the QSA to provide additional information on student performance. This additional information would enable schools to improve their teaching and learning programs.

The CCEs are an important feature of the curriculum as well as being the foundation for the QCS Test so it is important that the QSA continues to ensure that the CCEs remain current and that their place in any revised or new curriculum documents is made explicit. This is especially true in relation to the curriculum documents being produced as part of the Australian curriculum.

During the review it was evident that there was strong support for the QCS Test and the school-based assessment program, although it was evident that many people, including many teachers, did not fully understand its role and purposes other than at a fairly superficial level. There is considerable scope for conducting professional development programs for teachers addressing a number of aspects of the test, so that their understanding of the test and its processes is enhanced. The Review Team felt that some further materials would also be useful to help parents understand the test better.

An analysis of some performance data available to the Review Team has identified differences in the performances of males and females in co-educational schools as measured by their QCS Test results compared to their Within School Measure (WSM). These differences need to be investigated in greater depth by the QSA to identify the underlying causes and find strategies to minimise them. This is necessary to ensure that the QCS Test continues to operate to its full potential as a scaling test.

A Review of Aspects of the Queensland Core Skills Test - February 2012

More use of information and communications technology (ICT) could offer significant benefits in the delivery and marking of the QCS Test. While good use is already made of ICT in many aspects of the test, the introduction of 'onscreen' marking could be expected to provide significant efficiencies and improved marking practices. The QSA should also begin to investigate the place ICT can play in the delivery of the test in the future.

In addition to its specific recommendations below, the Review Team has noted, in the body of the report, other findings and suggestions for consideration by the QSA. These comments and suggestions are written in italics and serve as a summary of the discussion of an issue or as an indication that the QSA might consider the issue in more depth.

Overall, the Review Team believes that the QCS Test continues to perform well the functions for which it was designed and introduced. The recommendations made by the team are intended to further improve the program and test instrument, and extend its relevance and effectiveness well into the future.

Recommendation 1

that the QSA considers revising and re-using each year a relatively small number of items that have proved to be effective from the tests administered over the past 20 years (Section 2.1)

Recommendation 2

that the QSA conducts a risk assessment of the Marking Centre security arrangements to determine whether any additional or revised security measures are warranted (Section 2.7)

Recommendation 3

that each year after the QCS Test has been administered and analysed, the QSA systematically compares the item-level data obtained from the trialling program with that obtained from the test, in order to ascertain the extent to which the test developers' revision of items, which is based on the data from the trialling, is resulting in improved item performance in the test (Section 3.4)

Recommendation 4

that the QSA conducts further research to determine the reasons for the gender-related difference in students' rankings based on QCS Test scores compared to their rankings based on WSM (Section 3.6)

Recommendation 5

that the QSA conducts further research to determine the extent of any dependencies that occur, the effects of such dependencies on the outcomes of the scaling process, the potential causes of the dependencies and, if necessary, strategies for minimising the dependencies and their effects (Section 3.6)

Recommendation 6

that, unless items in the QCS Test are specifically designed to assess students' expertise in vocabulary and/or language, processes be put in place during the development of the test to ensure that the language used is as clear and simple as possible (Section 4.2)

Recommendation 7

that the QSA considers whether those students following OP-eligible programs who do not wish to receive an OP should be required to sit for the QCS Test (Section 4.2)

Recommendation 8

that the QSA addresses the concern about the perceived negative impact of the scores of OP-eligible non-English speaking background students on the scaling process by providing public advice on how this matter is addressed (Section 4.2)

Recommendation 9

that the QSA Governing Body considers whether any change to the current scheduling of the QCS Test should be made (Section 4.3)

Recommendation 10

that the QSA undertakes a review the CCEs to ensure that those on the existing list are still current both in their relevance and expression, and to include any elements that meet the criteria to be a CCE that have emerged recently (Section 4.4)

Recommendation 11

that the QSA updates the documentation explaining the role and importance of the CCEs and their place in each course (Section 4.4)

Recommendation 12

that the QSA works to ensure that the CCEs are appropriately captured in the curriculum documents being produced by the Australian Curriculum Assessment and Reporting Authority (ACARA) (Section 4.4)

Recommendation 13

that information showing the alignment of the CCEs with the Australian curriculum documents is provided to teachers before the implementation of those curriculum documents (Section 4.4)

Recommendation 14

that the QSA establishes a secure service that enables principals to access the QCS test results of their students and related information for any year at any time (Section 4.5)

Recommendation 15

that the QSA holds talks with principals to ascertain what additional information on the performance of individual students in the QCS Test would be useful to them and how it could be presented in a helpful and user-friendly manner (Section 4.5)

Recommendation 16

that the QSA develops software that will assist principals to analyse and present the performance data of their students in ways that will support the evaluation of the school's teaching programs (Section 4.5)

Recommendation 17

that the QSA investigates ways in which some additional information on each student's performance in the QCS Test could be made available to the student at the time their results are released (Section 4.6)

Recommendation 18

that the QSA holds discussions with QTAC and the universities to determine the usefulness of information currently provided to universities for course selection processes, and to make changes to the information if warranted (Section 4.7)

Recommendation 19

that the QSA prepares a document to be provided in the early part of each year to every parent of a student sitting for the QCS Test that explains the nature and function of the test (Section 4.8)

Recommendation 20

that the QSA holds discussions with QCS coordinators and Community Representatives to identify ways in which additional support could be provided effectively and efficiently to support less experienced personnel (Section 4.9)

Recommendation 21

that the QSA examines the range of professional development activities it provides to teachers related to the QCS Test and identifies additional activities that can be effectively and efficiently incorporated into its programs, including how information on student and school performance in the test can be used to improve teaching and learning (Section 4.10)

Recommendation 22

that the QSA continues to monitor proposals and developments concerning the proposed Australian Baccalaureate and assesses any likely impacts it might have on the QCS Test (Section 5.1)

Recommendation 23

that a high priority is given to the project of investigating, and developing a plan for, the introduction of onscreen marking of the Writing Task and Short Response papers of the QCS Test (Section 5.2)

Recommendation 24

that the QSA establishes a project team to investigate the issue of having students complete part of, or the whole, QCS Test by computer and develop feasible options and timelines for further consideration (Section 5.3)

A Review of Aspects of the QCS Test

1 Introduction

1.1 Outline of the task

The Queensland Studies Authority (QSA) commissioned this review of the Queensland Core Skills (QCS) Test 'to ascertain the ongoing relevance of the QCS test and the capability of the test to act as a statistical scaling device in the calculation of Overall Positions and Field Positions for tertiary selection'.

The members of the Review Team decided to approach this task by dividing it into three broad areas: test processes, test-specific issues, and future contextual issues.

Each of these broad areas has been further broken down as follows:

i. Test processes

These include:

- development and review of test items
- trialling of the items
- test construction
- printing of the test
- the delivery and handling of tests
- administration of the test and the collection of the completed test papers
- procedures at the Marking Centre
- analysis of the test results
- the reporting of results.

ii. Test-specific issues

These include:

- the relevance of the design criteria according to current research into quality testing
- the functioning of the test and whether it favours any particular group of students
- student engagement with the test and whether it is at a level that ensures maximum participation by students (because of the contribution of this factor to producing reliable results)
- the need for additional criteria or processes to be added to the testing process
- the extent to which the current tests meet the original, and any additional, design criteria
- the viability of the test as a scaling instrument
- the extent to which the administration of the test has the confidence of the stakeholders in regard to security and integrity
- the fairness and reliability of the marking of the test.

iii. Future contextual issues

These include :

- the future design of the test with particular reference to the mode of delivery
- the relationship of the test to the future introduction of the Australian curriculum
- administrative issues surrounding the security of the test.

The Review Team was also required to consult with a cross-section of external stakeholder groups including students, school administrators, tertiary institutions, the Queensland Tertiary Admissions Centre (QTAC), participants involved in the trialling of test items in another state, and internal stakeholder groups including members of the Analysis & Reporting Branch, and the Queensland Core Skills Unit.

1.2 Background of the Queensland Core Skills (QCS) Test

The QCS Test is based on 49 Common Curriculum Elements (CCEs) identified within the Queensland senior secondary syllabuses. For an element to be included on this list all students need to have had the opportunity to encounter and engage with it during the course of their studies, and it needs to be able to be assessed by means of pen and paper. An initial list of possible CCEs was reviewed and modified by meetings of teachers across the state before the development of the first QCS Test.

Following extensive planning and preparation for the test, a pilot study was conducted in 1991 at the conclusion of the 1991 Common Scaling Test (CST). The information gathered from this pilot study was used to inform and 'fine-tune' the QCS Test and underpin the test development processes.

The QCS Test was first introduced in September 1992.

1.3 The purpose and structure of the QCS Test

The QCS Test is an achievement test that is aligned to the Queensland senior curriculum and assesses samples of the 49 Common Curriculum Elements (CCEs). It is open to all Queensland Year 12 students, regardless of their subject selections. It is compulsory for those students who are undertaking an OP-eligible program of courses for the Queensland Certificate of Education (QCE).

The purpose of the QCS Test is to provide:

- group results for comparing the achievement of students in different subject groups and different schools in order to calculate Overall Positions (OPs)
- group results for calculating Field Positions (FPs) – up to five per student, depending upon subject choice
- individual results reported on a five-point scale.

The QCS Test comprises:

- One Writing Task (WT) paper
- One Short Response (SR) paper
- Two Multiple Choice (MC) papers.

The WT paper gives students two hours to write about 600 words of prose. They are given a variety of stimulus material (visual and written) grouped around a theme/topic/concept and they are advised to plan their response, prepare a draft, edit their draft and then produce their final response.

The SR paper gives students two hours to respond to a number of items that relate to stimulus material and cover many CCEs. The responses to items can vary and might consist of a mathematical or a visual expression, a sentence, a paragraph, or even a longer prose piece.

The two MC papers contain 50 items each. The items have a stem and four response options. Students are given 90 minutes to complete each paper. The items are presented individually or in units based on common stimulus material. The material is drawn from many disciplines

A Review of Aspects of the Queensland Core Skills Test - February 2012

including language, literature, philosophy, history, the physical and life sciences, the social sciences, art and mathematics. The MC papers test the CCEs embodied in the senior curriculum rather than the content that defines the subject.

The QCS Test is administered on the Tuesday and Wednesday of Week 8 in Term 3.

1.4 Previous external reviews

The QCS Test has been externally reviewed twice previously.

In May 1992, Gunter Trost, Director of the Institute for Test Development and Talent Research conducted a review of the planning, preparation and trialling of the QCS Test prior to its first administration in May 1992. This review, *The Queensland Core Skills Test: Evaluation of Design Criteria and Process*, considered the design criteria, the design process, the extent to which the test seemed to meet the design criteria, the face validity of the test, the reaction of stakeholders, and the Board's evaluation.

Trost reported that:

The Queensland Core Skills Test presents itself as a high-quality product that is well suited to fulfill its task.

It has many innovative features and meets strict standards by virtue of the thoroughness of the specifications for the design process, the manifold checks and revisions, the empirical trialling of multiple-choice and short response items, the differentiated marking schemes and last, but not least, a most sophisticated marking process.

The design criteria are very demanding, and they seem to be met by the test to a large extent.

In August 1996 Trost conducted a second review, *The Queensland Core Skills Test: A Follow-up Evaluation of the Design Criteria and Process*.

On this occasion he concluded that:

The high quality of the QCS Test that I was able to confirm in the year 1992 has been maintained, if not further increased, in the subsequent years. The content of the test is absolutely convincing. The marking process has been further refined. Some new analyses yielded additional and quite favourable results. Thus it is not surprising that the 'Queensland Model' has found keen interest and most respectful recognition in many other countries.

1.5 Methodology used for this review

The Review Team used several different strategies for collecting and assessing the necessary information. The strategies included the following:

- reviewing documentation such as operational manuals, annual reports, past test papers, reports of earlier reviews
- analysing data provided by the QSA to obtain evidence of the effectiveness of the test and its match to requirements, including addressing the key issues nominated by the QSA
- meeting with members of a wide range of stakeholder groups, including those nominated by the QSA, to seek their views on a range of issues about the test, its administration and uses, and to obtain further evidence about the effectiveness of the test and its match to requirements.

1.6 Acknowledgements

The Review Team would like to acknowledge the assistance and cooperation of everyone involved in this review.

The QSA officers involved in all aspects of the QCS Test willingly provided advice and materials, and arranged meetings with stakeholders. This greatly assisted the Review Team in fulfilling its objectives. The markers and clerical staff at the marking centre were very helpful and accommodating. The principals, teachers, students, parents and community representatives involved in the review's structured interviews were open, supportive and cooperative and provided much useful information.

2 Review of the test processes

2.1 The development and review of test items

When writing items for the QCS Test, writers are obliged to ensure that:

- each item assesses at least one of the 49 CCEs
- the items are cross-curriculum in nature and do not examine the content of any particular syllabus
- the overall test is targeted to the ability levels of 17-year-old Queensland students irrespective of whether they are eligible for an OP
- the stimulus material is self-contained, ie there is enough information in the material for the students to answer the items without calling on substantive syllabus content
- the material used to write items does not come from texts and publications that are routinely used as sources of stimulus for QSA syllabuses
- enough items are written to account for the fact that a significant proportion of items will fail at internal or external review or as a consequence of malfunctioning during field trials.

The item-writing teams abide by the following security protocols when developing items.

- All working documents are shredded after they are finished with.
- Item writers and test developers are prohibited from discussing any aspects of the instruments or draft items with work colleagues outside the item development team.
- Protocols are clearly articulated and enforced when items are panelled during the test development process.

The following key elements were identified by the Review Team as being indicative of the item development team's commitment to quality.

- The QSA employs a group of highly experienced item writers from a range of backgrounds as full-time employees of the QSA.
- The item development team involves the use of external panels and stakeholders in the process of developing and refining high quality test items.
- The items developed by the item development team are trialled on relatively large samples of students with similar characteristics (ie comparable) to the Year 12 Queensland secondary student population, and the subsequent analytical data are used to revise and improve the items before their inclusion in the final test.
- Detailed guidelines are prepared and used in the test design process.

A Review of Aspects of the Queensland Core Skills Test - February 2012

- Each and every SR item that is prepared for inclusion in the test is accompanied by a detailed marking scheme that is reviewed as part of the item review process.
- Clear marking criteria and rubrics are prepared at the time that a WT is prepared and reviewed as part of the development process for the WT.

The QCS Test is underpinned by a design process rather than by the mechanical application of a set of rules. Part of the process requires the panelling of items by review panels consisting of subject matter experts and equity stakeholders.

After the item writers have developed units, an initial panelling session is conducted internally by QSA staff. The panels provide thorough feedback to the item writers who then respond to the feedback. The items are then modified where necessary and resubmitted to the panels.

After consideration by the panels the items are then placed in trial test booklets. These trial test booklets sometimes also contain items that did not perform well when first trialled and have been re-worked as a result of the initial trial. These trial tests are administered to samples of students to ascertain how the items function with groups of students with similar characteristics to the students who will eventually sit the QCS Test.

The responses from the students are then analysed and the review panels consider the resulting items and test analysis data. The panels also evaluate the suitability of marking schemes including the extent to which the marking schemes cover the whole range of actual student responses, and they also examine any differences in the functioning of items that occurs in population subgroups.

On the basis of the information obtained from the trialling, the items either proceed without alteration, are revised or are rejected. Items that are revised are re-trialled before they are used in an actual test to ensure they function well.

Item writers observe the trialling process and obtain first-hand data about which items engage and interest students.

When the penultimate version of the test paper is produced, an expert external scrutiny panel is convened to work the paper to consider difficulty, timing, inclusivity and engagement. This panel has academic, pedagogic and equity expertise.

The Review Team noted that considerable attention is paid to every step of the development of test items as part of the process: the construction of new items, establishment of the relationship between the Common Curriculum Elements (CCEs) and the item content, the development of a differentiated marking scheme for each item, the multi-looped panelling process, the repeated revisions, the trialling and the editing processes. The Review Team believes that the processes established to develop and review test items are sound and represent 'best practice'.

In his 1992 report, Trost recommended that 'the Minister of Education considers not releasing each of the MC and the SR papers every year – say from 1994 on.'

His reasons were as follows:

1. The saving of time, effort and cost would be enormous if single items, after long intervals and in an unsystematic way, could be re-used.
2. An individual's reservoir of good item ideas is limited; after a certain time, item writers tend to invent exotic contents or to modify old ideas. This process could be counteracted by re-using good items from time to time.
3. Not releasing test papers after use would allow for trialling new items under real conditions on the original population by interspersing small numbers of trial items into the actual test paper. A model for such a procedure has been developed and is operational in Germany.

The Review Team understands that there are sound reasons for releasing the test papers every year, which makes it impossible to re-use items on a large scale.

However, in the 20 years since the QCS Test was introduced some 2000 multiple-choice items and more than 300 SR items have been used. While these are in the public domain, it would still be feasible to revise and re-use each year some items that had been used in the past. Items that had worked well in the past could be re-worked or revised so that they are superficially different, but are essentially the same in all other respects. Using these 'shadow items' has the advantage that statistics on their difficulty and discrimination when they were previously used will be known.

Adopting this strategy would enable the advantages outlined by Trost to be realised, at least to some degree.

Recommendation 1

that the QSA considers revising and re-using each year a relatively small number of items that have proved to be effective from the tests administered over the past 20 years

2.2 Trialling of the items

The test items are trialled annually in another Australian jurisdiction. The sample used for the trial includes students who are likely to obtain an Australian Tertiary Admissions Rank (ATAR) in that year, as well as some students who are not eligible for an ATAR. The test is managed under the same administrative and security procedures as those applied in Queensland as part of the QCS Test administration.

In 2010 the trialling involved approximately 2400 students of whom 2100 went on to achieve an ATAR. This is believed to be a good representative sample of the ATAR population in this jurisdiction. The ATAR population in this jurisdiction was 65% of the Year 12 Certificate population. In Queensland the OP population was 58% of the Year 12 Certificate population.

The Review Team was advised that the trial tests are regarded very highly and taken seriously in this jurisdiction. The results are considered to be a good indicator of the likely performance of each school group as a whole in the jurisdiction's own assessment regime.

All students involved in the trialling of the QCS Test receive a certificate with two grades, one summarising their performance on a quantitative scale and the other summarising their performance on a verbal scale. Participating schools also receive a report. This report shows students' performances in the test, and contains:

- a table showing grades and performance indicators for verbal and quantitative items, as well as an overall performance indicator for each student
- boxplots of the above data for the school and system broken down by gender
- histograms showing the distributions of grades on verbal and quantitative items.

The writing task is not trialled. However, considerable effort goes into ensuring that it will be suitable for the candidature, including checking and revising, if necessary, marking criteria and rubrics. The rubric and criteria remain relatively constant over time.

The program for trialling test items is a necessary component of the test development process. When items are trialled in another jurisdiction it would be expected that there will be some differences between item-level data obtained from the trialling and data obtained from the actual QCS Test. However, as the items have been trialled in the same jurisdiction over a

number of years the panels should be aware of any significant differences in the curriculums and take them into account during the review process.

The Review Team is confident that the trialling program runs smoothly and provides important information to the QSA as part of the test development process. Further discussion on increasing the effectiveness of the information from the trialling program, based on an analysis of data from the 2009 and 2010 tests, can be found in Section 3.4 and Appendix E.

2.3 Test construction

The QCS Test is thoroughly grounded in the Queensland senior curriculum through the testing of the curriculum's CCEs. The test assumes that students possess an elementary level of general knowledge and a level of vocabulary and mathematical competence consistent with the level of sophistication associated with a sound general Year 12 education. Any substantive vocabulary whose meaning cannot be determined from the context is explained. Mathematical operations and concepts such as percentage, ratio, area and angle are assumed knowledge.

The rest of this section describes some of the key features of the various papers.

2.3.1 The multiple choice (MC) papers

Members of the Review Team attempted the 2010 and 2011 versions of the MC papers as part of the review process. In the opinion of the Review Team the papers meet the design criteria and the papers are indicative of international best practice. The items are interesting and appropriately demanding. They display great variety in terms of content and represent a sound balance across the curriculum areas.

Some item types are highly creative. The phrasing of items is precise and appears to show no signs of bias. The papers are clear of typographical and editorial errors. The layout of the paper is effective, and there is no evidence of any distracting or inappropriate elements. Page layout is clear, and there is no need for students to needlessly turn back and forth across pages while answering item sets. There is appropriate space for student work. The directions are easy to understand and follow.

A more quantitative analysis of the functioning of the items has been carried out as part of the review and is presented in Appendix E.

2.3.2 The short response (SR) papers

The SR paper is divided into a number of items arranged within units. A unit comprises a common piece of stimulus material and a set of items. The units may be undertaken in any order, and the physical construction of the tabbed material supports students in approaching the paper in this way.

Each short response item is constructed in a standard way. The stimulus material may be verbal, numerical, visual images, tables, graphs or a combination thereof. Additionally, the items may include other information and guidance to students on the elements required to answer a particular question. This information includes a 'star rating' of items, which provides students with important guidance on the relative difficulty or requirements of an item, so students may direct their efforts to start with the easier or the more demanding items.

A key positive aspect of the SR papers is the provision of fold-out elements of the booklet that provide for additional engagement with stimulus material (eg 2011, Unit 6), and spare pages for additional response areas where students may wish to provide re-worked solutions (eg 2011, Units 5 and 7).

A Review of Aspects of the Queensland Core Skills Test - February 2012

Significant attention is paid to every step of the design process. This includes establishing the relationship between the CCEs and the item content, developing a marking scheme for each item, the panelling and revision processes, and the editing of the test paper.

These features ensure that the paper has high content validity and that the various formal dimensions are balanced and widely covered.

Quantitative information regarding the functioning of these items within the QCS Test is provided in Appendix E.

2.3.3 The writing task paper (WT)

The WT is a two-hour paper involving the development of one sustained extended piece of prose, of approximately 600 words. A new WT is administered each year. It presents students with written and visual material on a theme that allows students to demonstrate their writing abilities across a range of perspectives. The richness of the material, around a common theme, is of interest to Year 12 students and allows students to demonstrate their ability to produce expressive tasks involving extended pieces of continuous writing.

The WT paper has a number of highly appealing features:

- It contains a diversity of content, sources, and style of presentation in the stimulus material.
- It contains a wide range of options offered to the students for responding to the material.
- It complements the other elements of the QCS Test in that it allows for assessment of imagination, creativity and expertise with language.

The tasks are developed with the holistic nature of the writing task in mind. The writing task developers consciously work towards developing materials that provide a number of entry points for candidates. There is a deliberate attempt to develop papers that afford students the ability to respond in a range of genres, as opposed to some other tests which focus unduly on the genre of argument or persuasion, to the exclusion of other writing styles. The only genre that is expressly excluded is poetry.

The test development team chooses prompts that can provide an appropriate stimulus for a range of students regardless of specific subject selection in the senior years (ie to include students who have undertaken a mathematics and science oriented program of study, as well as students who have undertaken a humanities focused program).

In addition, the item development team attempts to ensure a mix of possible subjects and stimuli that have the capacity to appeal to and interest candidates from both regional and urban areas, and from different cultural backgrounds.

The development of each writing task involves a scrutiny panel and the Manager, QCS Test. The panel includes item writers from a variety of curriculum backgrounds. There is also an external representative on the panel who, though linked with school education, may come from a tertiary education context. A graphic artist is involved in panel deliberations because of the high priority given to visually engaging materials. There is a deliberate attempt to avoid stimuli and images that are in any way clichéd.

The Review Team is of the opinion that considerable care is taken during every stage of the QCS Test construction process and that this results in a test that provides for a range of student interests and abilities, does not contain observable bias, places appropriate demands on candidates, and has high face validity. The Review Team considers that the QCS Test is an instrument of high quality and rigour. Those involved in its construction are commended.

2.4 Printing of the test

The printing of the QCS Test and related materials is well organised and secure. Multiple printers are used and QSA staff monitor the printing process very closely to ensure it is conducted in accordance with the strict requirements established by the QSA.

Once the printing is complete the test papers and materials for each school are packed in an off-site location under tight security. A high level of supervision is applied to every step.

2.5 Delivery and security of test materials within schools

There are two key roles related to the handling of test materials and the conduct of the test in each school. One is the role of QCS Coordinator; the other is the role of the Community Representatives.

The QCS Coordinator is a senior teacher at the school who is responsible, on behalf of the principal, for the administrative arrangements related to the test and its administration. This person is also responsible for ensuring that an appropriate number of teachers are scheduled to supervise the test and that they know what is expected of them.

The Community Representatives are members of the community appointed to see that the QSA's procedures for preparing for and administering the test and for dispatching completed test responses to the QSA are followed. The 'home' school nominates one Community Representative. A second Community Representative is nominated by an 'outside' school.

The test papers and related materials are packed to take account of the number of candidates in the school and an additional loading is also included. They are sealed in envelopes with a label showing the details of the contents of each package. The packages for each school are locked in one or more metal boxes.

Two different keys are needed to open the boxes. One is issued to each of the Community Representatives.

On the day the metal boxes are delivered to the school by courier the two Community Representatives meet with the QCS Coordinator at the school, open the boxes and check the labels to ensure that the contents are in accord with expectations. The boxes are then locked again, and placed under lock and key in a secure area of the school. The Community Representatives retain their keys. This process ensures that neither of the Community Representatives nor any person at the school can access the test papers before the test.

The QCS Coordinators, principals and Community Representatives interviewed by the Review Team were confident that the procedure was very secure and that a breach of security during this phase of the process would be difficult, if not impossible.

The Review Team is of the view that the current procedure is working well and provides an appropriate level of security during this phase of the operation.

2.6 Administration of the test and collection of the completed test papers

On the days of the test the Community Representatives attend the school. The metal boxes containing the materials for the relevant sub-test are taken from the secure room. The Community Representatives, along with the QCS Coordinator, open the boxes and remove the required materials.

During the test the Community Representatives observe the process and ensure that the supervisors are undertaking their duties as expected. The Community Representatives complete a report on the conduct of the test for the QSA.

At the completion of the test the supervisors collect and count the number of scripts. This is checked by the Community Representatives.

The responses are then sealed in the appropriate bags, locked in the appropriate metal box and collected by the designated courier.

During the days of the test QSA officers visit a sample of schools to check that the required procedures are being followed and proper arrangements have been made.

The QCS coordinators and Community Representatives who were interviewed by the Review Team indicated that the procedures for the administration of the test and collection of student responses are thorough and that the process is sound and secure.

The Review Team also is of the opinion that the process is well planned and implemented.

2.7 Procedures at the Marking Centre

The Review Team visited the Marking Centre on the first two days of its operation. During these visits discussions were held with a wide range of permanent and casual officers from the QSA who play a role in the marking operation. In addition, meetings were held with coordinators and immersers/supervisors, and with markers appointed to mark both the WT and the SR papers. Members of the Review Team also attended some marker training sessions for both the WT and SR papers. Some of the comments and suggestions made during these meetings are discussed in Section 4 of this report.

The QSA staff members who were interviewed and observed demonstrated commitment and pride in their work. The senior markers and markers interviewed also showed commitment to the process and determination to do the job well.

The procedures and methods currently used to mark the QCS Test are sound, well documented and well administered. The training of markers and the ongoing monitoring of the marking process are thorough and effective.

Each unit in the SR paper is marked at least twice with an item-specific marking scheme. For each item, markers are given a range of material to assist them to achieve consistency and comparability: an item-specific marking scheme, the performance domain(s) being tested by the item, descriptions of the standards for each of the available grades, and an exemplar with the characteristics of an A-grade response. Referee marking takes place whenever there is a significant discrepancy between markers on the same response.

The marking used for the WT is criteria-based holistic marking. Each student response receives two grades awarded independently by two markers on each of the following criteria: responsiveness (R); central idea (CI); grammar, punctuation and spelling (GPS); structure and sequencing (SS); and vocabulary (V). There is also an overall judgement made as to the length (L).

A Review of Aspects of the Queensland Core Skills Test - February 2012

Grades awarded in these criteria reflect the contribution of each aspect to the holistic worth of the script. At least three markers will mark each script, with each marker awarding grades in four of the six criteria. For the first five criteria the marking scale has six levels ranging from 1 (highest) to 6 with +, 0, - qualifiers for each level. In relation to length, the judgement is made in one of five categories ranging from far too short, to far too long. If the two grades awarded for a criterion differ by more than a pre-determined amount, a third marking of that criterion will occur and the two grades in closest agreement out of the three used.

The approaches used to monitor the reliability of marking and the procedures for re-briefing aberrant markers are sound and well conducted, and appear to operate effectively. They are discussed further in Section 3.

The Review Team is of the view that the current marking procedures are relevant and appropriate, and meet the needs of the program. The Review Team also believes that there are some efficiencies and benefits that can be obtained through greater use of ICT in the marking process. This latter point is discussed in Section 5.

Breaches of security at a marking centre can be of two basic types. The first type is where someone seeks to obtain information or material either to alter a student's marks, or simply to find out a student's marks. The second type of security breach is where someone, either deliberately or accidentally, gains access to a secure area, whether or not they cause any further damage.

There is a focus on security at the QSA Marking Centre during the day. Marker identity cards are used and QSA permanent and casual staff members are vigilant and careful about who has access to the Centre. After marking is completed for the day, the Marking Centre is locked and a security service is engaged. While such arrangements may not completely eliminate the possibility of a security breach, in the opinion of the Review Team they are adequate to deal with most security threats.

The Review Team is of the view that the QSA should conduct a risk assessment of the Marking Centre to determine the adequacy of current security procedures and whether any additional or revised procedures are warranted.

Recommendation 2

that the QSA conducts a risk assessment of the Marking Centre security arrangements to determine whether any additional or revised security measures are warranted

2.8 Analysis

The QCS Test results are analysed by the QSA as part of their quality control procedures.

Both Classical Test Theory (CTT) and Item Response Theory (IRT) techniques are used in the analysis of test items. The results of the item analyses are provided in the QCS Yearbook.

Dimensionality of the test is examined using CTT and IRT procedures. The IRT analyses include both Infit and Outfit statistics for each item to assess fit to the model.

Reliability coefficients are also calculated for the multiple choice and short response sections, and for the total test.

The QSA also checks data for subject groups and for whole-school groups to determine whether the mean QCS Test performances are consistent with the overall school performances as measured by the school assessments. For each school a polyscore is generated for each student.

A Review of Aspects of the Queensland Core Skills Test - February 2012

A polyscore provides an estimate of the overall achievement independently of the procedures used for determining OPs. These estimates are based on treating each level of achievement in each subject as equivalent.

School groups that have large negative mean residual polyscores are selected for further examination. A large negative mean residual polyscore suggests that students in this group tend to have an Overall Achievement Index (OAI) much lower than their estimated overall achievement. Similarly, school groups with a much larger polyscore spread than their OAI spread are selected for further analysis. In these cases, the students well above the school mean may on average have OAI's much lower than their estimated overall achievement.

School groups with noteworthy inconsistency between the QCS Test and level-of-achievement information are referred to the QSA's Scaling Anomalies Committee for consideration of special-case calculations to remove the inconsistency.

Schools may request that the QSA look at their data if they believe that extraneous circumstances might have had a negative effect on their students' performance on the QCS Test.

The Review Team is of the view that the procedures used to analyse and verify the results from the QCS Test are sound and statistically defensible.

2.9 Reporting of results

The QSA provides a wealth of information to the community, schools and students on the performance of students on the QCS Test. This includes the following reports:

- the yearly *Data summary – QCS attendance and grades* which provides the State distribution (both overall and within genders) of Overall Positions (OPs) and Field Positions (FPs), the QCS attendance in terms of total, OP-eligible who sat, were exempt, were ineligible, and did not sit, and grades awarded by eligible and ineligible cohorts
- the yearly *QCS Test – OP-eligible QCS candidates – Data for the State*. This lists the QCS Test results by subject groupings.

During the end of year processing, after receiving assessment data from schools, the QSA makes available QCS Test summary and QCS Test versus Within School Measure (WSM) plots for 'non-visa' schools. The WSM is calculated within a school among students based on the 'wins' and 'losses' in terms of Subject Achievement Indicators (SAIs). It is the difference between these two measures that is used to weight the QCS Test results when calculating the OPs. (See Appendix F for details.)

Further data are uploaded to the website early in the following year. These data provide details for a school of the difference between components of the QCS Test performance, the groupings of CCEs and the state mean for each component. They also provide information about the performance of a school's Year 12 students, over a length of time. In addition they show trends in student performance, and also show the trend in the polyscore measure for a school.

While students may have little use for the results beyond feedback on their performance, teachers can use the results from the test to inform their teaching with a view to improving the performance of subsequent Year 12 students. For this reason the Review Team strongly believes that the QSA should provide teachers and schools with more 'fine-grained' item-student data that teachers can use to improve the performance of students in future cohorts. This issue is also raised later in the report.

Raw scores are converted to letter grades (A to E) and are reported on the Senior Statement.

There are two steps in the process of converting raw scores to grades.

A Review of Aspects of the Queensland Core Skills Test - February 2012

First, the top and bottom grade cut-offs are determined by a bookmarking standard-setting method whereby items are ranked according to their IRT difficulty. Each item is then assessed against verbal descriptors of the A/B and D/E cut-offs and the corresponding difficulties determined for each basket (ie criterion).

The A/B and D/E cut-offs for the test then are determined by aggregating these cut-offs across the five criteria*. The number of students awarded A and E grades are then determined. Table 1 shows the 2010 QCS Test grade cut-offs according to basket (criterion).

Second, the numbers of students awarded grades of B, C and D are generated using a 4:5:3 ratio for grades B, C and D respectively. The 2010 grade distribution is shown in Table 2.

TABLE 1
2010 QCS Test basket cut-offs

	α	β	θ	π	φ
A/B	57	34	57	88	34
D/E	20	10	18	34.8	5

TABLE 2
Distribution of 2010 QCS Test grades

QCS grade	A	B	C	D	E
Number	4,434	7,785	9,013	3,953	35
Percentage	17.6%	30.9%	35.7%	15.7%	0.1%

Having a mix of methods means that the descriptors for the top and bottom grades are fixed, but the descriptors of the remaining grades are determined by student performance.

The Review Team suggests that the QSA should consider providing more detailed student-level information, which could be used to help teachers improve the performance of subsequent cohorts of students. The fact that every question is linked to an aspect of the curriculum means that the information would be relevant to the teaching and learning within schools.

In addition, the Review Team believes that the QSA should consider providing student scores, grouped by subject on the criteria of the QCS Test, to each school and also show a link with the Overall Achievement Indicators.

The next section of this review addresses some of the test-specific issues identified by the QSA for consideration by the Review Team.

* The five criteria are:
 α : Comprehension and Collect
 β : Structure and Sequence
 θ : Analyse, Assess and Conclude
 π : Create and Present
 φ : Apply Techniques and Procedures

3 Consideration of test-specific issues related to the QCS Test

The Review Team conducted an analysis of the test-level and item-level performance of the QCS Test using results from the 2009 and 2010 QCS Test administrations. The results from this analysis are presented in Appendices E and F of this report. The information from the analysis is used, where appropriate, to answer the test-specific issues identified for consideration by the QSA.

3.1 Relevance of the design criteria

The Review Team believes that the design criteria identified by Matters in 1991 are just as important now as they were then. That is, the QCS Test must:

1. be an integrated whole
2. be balanced and include a suitable range of items
3. be suitably difficult
4. discriminate among students
5. be without bias
6. make appropriate demands on the test-takers
7. have face and content validity
8. present items within each paper in such a way that their sequence assists students to demonstrate their optimum performance.

3.2 Functioning of the test and whether it favours any particular groups

Table E7 of Appendix E shows the QCS Test performance data for various groups of those OP-eligible students who were in Year 12 and completed all four papers of the QCS Tests in 2007, 2008, 2009 and 2010.

It can be seen from Table E7 that the mean for the total group was 117.8 out of a possible 235 (50.1%) in 2009, and 109.3 out of a possible 237 (46.1%) in 2010. These scores suggest that the tests are of average difficulty (which is the requirement for a scaling test) for the population of students. The spread of scores is relatively large which means that it is discriminating across the range of scores.

A second point worth noting from Table E7 is that males outperform females in all categories in the four years, except for the WT and the related π (Create and Present) criterion. These differences are not statistically significant, but they are persistent and consistent and probably warrant further investigation. The most likely cause of these test statistics is that the ability distributions of males and females sitting for the test are quite different. There are significantly fewer males meeting the OP-eligibility criteria than females. This would suggest that there is some self-selection occurring in the total population whereby the less well performing males are opting out of the OP-eligible program. For example, in 2010 males accounted for 43.8% of the OP-eligible cohort who sat the QCS Test.

The Review Team has noted that a Differential Item Function (DIF) analysis has been conducted as part of the statistical analysis of the 2010 QCS Test and that there appears to be no overall DIF favouring either group.

Given the critical importance of the QCS Test in the scaling of school-based assessment, the Review Team suggests that the QSA continue to monitor this issue and collect further evidence to show that the test is not functioning differently for males and females.

3.3 Whether student engagement with the test ensures maximum participation and commitment by students

Students who undertake an OP-eligible program of courses must sit the QCS Test. Students who are ineligible for an OP may sit the test if they wish. Over the past decade the proportion of those students receiving a Senior Statement who follow an OP-eligible program has declined from 72% in 2001 to 58% in 2010. However, the actual number of OP-eligible students has not changed significantly. Over the same period the proportion of certificated students ineligible for an OP sitting the test has remained relatively stable at around 7%, although the proportion of students ineligible for an OP who sat the test to those with an OP has increased slightly in recent years.

The Review Team was told by stakeholders that the reasons for the decline in the proportion of the Year 12 cohort sitting for the QCS Test include the increasing availability of jobs and hence attractiveness of VET pathways, and competition between schools because of comparisons of OP numbers in the press. Some stakeholders told the Review Team that some schools encourage lower performing students to not undertake an OP-eligible pattern of courses. Whatever the reason for the decline in the candidature, it does not seem to be directly related to the QCS Test.

The proposition that there is an increasing proportion of students who are following a pattern of study that makes them ineligible for an OP, and who elect to sit for the QCS Test as an alternative pathway to university, does not seem to be supported by the data. The proportion of students in this category has remained relatively steady at about 7% over time. (See Table 3 and Figure 1.)

TABLE 3
Numbers of students with a Certificate, sitting the QCS Test, and their OP status

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Senior Certificate	37837	38821	38721	38471	38953	39579	40886	41152	43196	44652
OP-eligible	27303	27749	27973	27235	27027	26233	26185	25030	25306	25704
Sat QCS	29976	30534	30297	29233	28886	28183	28267	26262	28302	28420
OP ineligible who sat	3008	3178	2743	2436	2344	2408	2530	2795	3393	3169
% OP/Cert	72.24	70.79	72.24	70.79	69.38	66.28	64.04	60.82	58.58	57.57
%QCS/Cert	78.24	75.99	78.24	75.99	74.16	71.21	69.14	63.82	65.52	63.65
%Ineligible/Cert	7.08	6.33	7.08	6.33	6.02	6.08	6.19	6.79	7.85	7.10
%Ineligible/OP	11.02	11.45	9.81	8.94	8.67	9.18	9.66	11.17	13.41	12.33

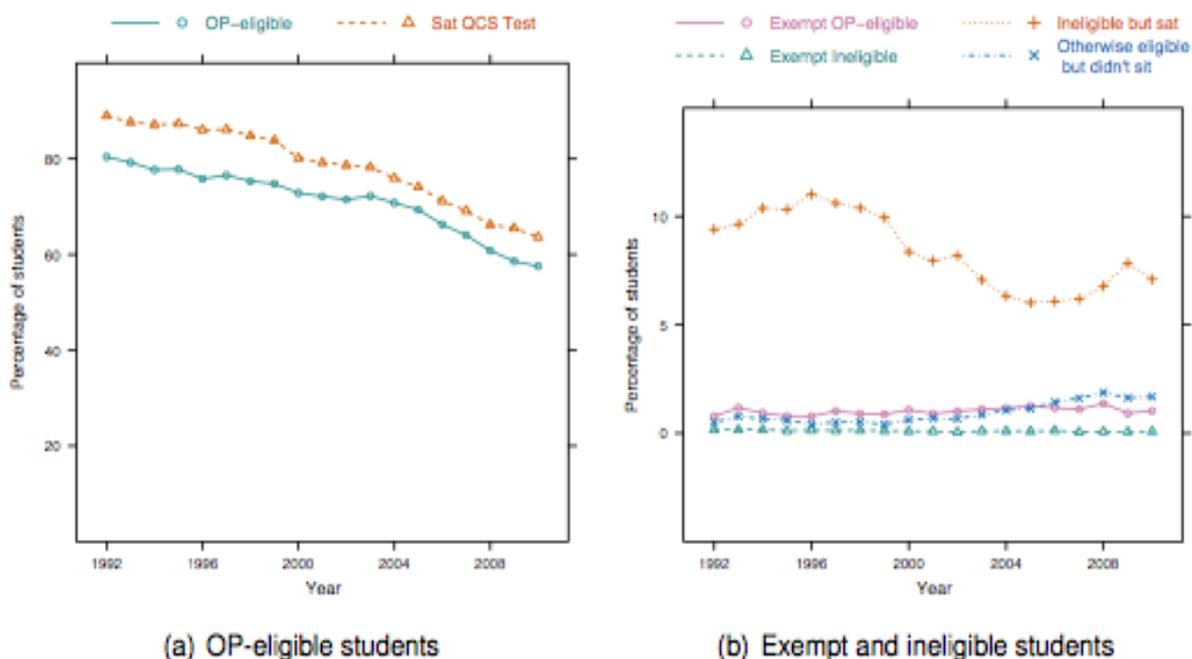


Figure 1 Trends in percentages of OP-eligible and OP-ineligible students
(Source: QSA 2010 Data Summary – QCS attendance and grades)

Another aspect of student engagement is the extent to which the students connect with the content of the QCS Test. This generally can be measured by reviewing the omit rates for multiple choice and short response items, evaluating the stimulus and the source of the items to see what item-reviewers thought about their reactivity with students, and asking the students what they thought about the test.

The Review Team was given no statistics on a question-by-question basis that showed the omission rates, but students interviewed indicated that they believed it was important for them to do well and so they tried hard during the test.

The QSA staff involved in the development of the test emphasised the fact that considerable effort is put into making the test items interesting and engaging for students. Members of the Review Team who sat some of the papers also felt that the tests contained material that would engage students and that it was presented in engaging ways.

During the consultation, groups of students were asked directly 'What did you like about the test?' The responses covered the range from 'the free breakfast organised by the school' to 'the star values for the SR questions helped'. When this issue was explored further most students felt that the test was easy to follow and was clearly set out, and had something for everybody. Most students enjoyed the WT and felt that it gave them plenty of scope to plan and produce a response. Students felt that the star values helped them in planning their pathway through the SR paper. Most students were of the view that the test was targeted at an appropriate level of difficulty with the SR paper being more difficult than the other three.

Furthermore, when asked 'How important do you think it is for you to do your best in the test?' the students interviewed were all of the opinion that it was extremely important that they do their best, not only for their individual benefit, but also for the benefit of their school group as a whole.

A Review of Aspects of the Queensland Core Skills Test - February 2012

The engagement with the overall test as measured by distributional statistics appears to be high, as indicated by the lack of skew of the raw score distributions reported by the QSA. In 2010 the skewness of the test is slightly positive (0.129), as are each of the subtests.

The Review Team is convinced that the QCS Test is engaging the students and consequently the students are committed to doing their best in the test, and the participation rates have remained relatively high.

3.4 Whether any additional processes are needed

The Review Team has some more comments about obtaining the maximum benefit from the trialling process. This issue is raised because, as discussed below, a number of items in the QCS Test do not discriminate well.

The purpose of the test is to place all students' scores onto one scale so that the scores can be added as the basis for the calculation of an OP. The scaling test should effectively spread the scores along the measurement scale. In order to maximise spread, the intention is to have an average facility of 0.5 for multiple-choice items as this maximises the spread of the cohort across the scale. In reality, there will be a range of item facilities with the majority of the items having facility values close to 0.5. The QCS Test achieves this quite well.

It is also important that the items have a high discriminating power. The point biserial correlation is used to give an indication of the item discrimination. Items with a point biserial value less than 0.2 would be considered to be of dubious value in a test that requires the items to discriminate so that the resulting distribution is appropriate for scaling. While the point biserial values cannot be determined until after the test, the items that are used in the test are trialled and any item that is seen to be problematic in terms of its discriminating power is reviewed with a view to improving its discrimination power before it goes into the final QCS Test.

Table E8 in Appendix E shows the item facilities (difficulties), point biserial correlations and criterion measure for each of the multiple-choice items in the 2009 and 2010 QCS Tests. In 2009 there were 18 multiple-choice items (18 percent of MC items) with point biserial values less than 0.2, and in 2010 there were 25 multiple-choice items (25 percent of MC items) with relatively low point biserial correlations. In 2010 there was one item with a negative point biserial correlation (item 49).

The Review Team understands that the test developers use the results from the trialling program to review and revise items before they are included in the final test. This being the case, it would be worthwhile for the QSA to see whether the point biserial values that are obtained in trialling are providing the test developers with sufficient pertinent information to identify items that fail to discriminate effectively. If such items are not being identified because of differences between the trialling population and the Test-taking population, the value of trialling is limited. The Review Team notes that if the QSA makes changes to MC items that have not performed well in the trial, then these revised items are re-trialled.

The information contained in Table E9 in Appendix E suggests that there was a relatively significant increase in the difficulty of the SR paper from 2009 to 2010. Hopefully the Test Development Team became aware of this increase in difficulty during the trialling process. If this is not the case, there is further concern for the Review Team regarding the comparability between the characteristics and motivation of the trial group and that of the QCS Test population.

The Review Team is of the view that these comments and the information provided in Appendix E should be considered in the context of seeking to gain maximum benefit from the trialling program.

Recommendation 3

that each year after the QCS Test has been administered and analysed, the QSA systematically compares the item-level data obtained from the trialling program with that obtained from the test, in order to ascertain the extent to which the test developers' revision of items, which is based on the data from the trialling, is resulting in improved item performance in the test

3.5 The extent to which the test meets the original, and any additional, design criteria

It can be seen from Table E6 in Appendix E that the composition of the test has remained relatively stable over time. The criterion with the highest content weighting every year has been π (Create and Present). The second most frequently assessed criterion most years, particularly over the past 12 years, is θ (Analyse, Assess and Conclude). The criterion that has the least weighting in most years is Φ (Apply Techniques and Procedures).

These data strongly suggest that there has been no significant shift in the mix of criteria that constitute the variable. This supports the claim that the test is basically assessing the same construct as it was when it was first developed 20 years ago. The criteria that comprise the construct are generic, so they are likely to remain embedded in the curriculum even when it changes.

The data also suggest that the items included in the test cover a suitable range of the CCEs and they represent a comprehensive range of item types.

Since its inception 20 years ago, there has been no widespread criticism, from the public or users of the test, of the construct being assessed. It would seem, therefore, that the QCS Test has a high level of face validity.

The Review Team is confident that the test was well balanced when it was introduced, that it has retained this balance over time, and that the construct being measured by the test is well supported by the various jurisdictions as being appropriate for the purpose.

As a consequence of discussions with stakeholders, the Review Team notes that there is a need to make sure that the CCEs are appropriately captured in the curriculum documents being produced by the Australian Curriculum Assessment and Reporting Authority (ACARA).

In addition, the Team believes that the QSA should consider updating the documentation explaining the role and importance of the CCEs and their place in each course.

These views are captured in Recommendations 11, 12 and 13 (in Section 4.4) of this report.

3.6 Viability of the test as a scaling instrument

As an instrument for placing all subject assessments across all schools on the same scale the QCS Test must first provide valid estimates of the relative performance and spread of the scores of students from the various subject groups within a school.

Individual data were available but subject groups were not identified, so the Review Team carried out analyses at school level.

The QSA calculates three within-school measures of overall achievement: the scaled Overall Achievement Index (OAI), the scaled Within School Measure (WSM) and the scaled polyscore. The QCS Test scores play no part in the calculation of the WSM and polyscore except for the final step where they are scaled to place them on the same scale as the QCS Test.

Neither polyscores nor OPs were available for individual students so analyses were carried out on the relationship between the WSM and the QCS Test scores.

3.6.1 Relationship between WSM and QCS test scores

If the QCS Test is a valid scaling test, the relationship between the WSM and QCS Test will be independent of gender, school type, sector, location and size.

The following analysis examines the relationship between WSM and QCS Test scores and gender. Table 4 shows the means of WSM and QCS Test scores for male and female students in single-sex and co-educational schools.

TABLE 4
WSM and QCS Test mean scores by gender and type of school

School type	School WSM		QCS Test		Correlation between WSM and QCS Test scores	
	Male	Female	Male	Female	Male	Female
Co-educational	170.8	175.5	177.4	170.4	0.637	0.679
Males only	181.7		181.6		0.679	
Females only		183.5		183.5		0.717

Table 4 shows that students in single-sex schools have had higher mean QCS Test scores than students in co-educational schools.

It is interesting that in co-educational schools males have outperformed females in the QCS Test, but have lower mean scaled WSM scores.

As would be expected, the scaled WSM and QCS Test mean scores are the same in the single-sex schools, and for the co-educational schools when males and females are aggregated. However, when males and females are treated separately, there are obvious differences in the means. For males the QCS Test mean exceeds the scaled WSM mean, whereas for females the converse is true.

The correlation between the within school ranking, scaled WSM, and the QCS Test scores is higher in the single-sex schools than in the co-educational schools. In both types of schools the correlation is higher for females than for males.

Taken together, the evidence indicates that the relationship between the WSM and QCS Test scores is different for males and females. To investigate this issue further a series of regressions was carried out, with WSM regressed on QCS Test scores for males and females within co-educational schools, and for single-sex schools.

In co-educational schools a significant QCS Test score by gender interaction was found so separate regressions were calculated. The two regression lines differed significantly in slope.

$$\text{Females in co-educational schools: } WSM = 58.012 + 0.689 * QCS$$

$$\text{Males in co-educational schools: } WSM = 62.550 + 0.610 * QCS$$

The two regression lines are shown in Figure 2. For students with similar QCS Test scores the predicted scaled WSM scores are higher for females than males for similar QCS Test scores, and the difference increases as the QCS Test score increases.



Figure 2: Regressions of scaled WSM on mean QCS Test score for males and females in co-educational schools.

A different picture is seen, however, when the regressions for males and females in single-sex schools are calculated.

<i>Males in single-sex schools:</i>	$WSM = 60.14 + 0.673 * QCS$
<i>Females in single-sex schools:</i>	$WSM = 53.53 + 0.706 * QCS$

Figure 3 shows the two regression lines. The slopes are not significantly different. And, unlike the co-educational schools, the lines intersect within the range of scores. Students with similar QCS Test scores, whether male or female, have similar predicted scaled WSM scores.

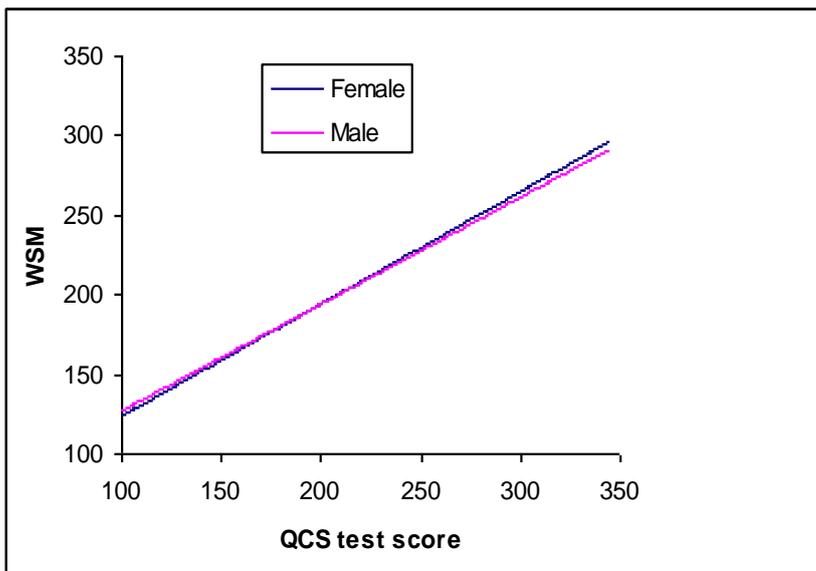


Figure 3: Regressions of WSM on QCS for males and females in single-sex schools.

Two further comparisons were carried out. First, the regression for females in co-educational schools was compared to that for females in single-sex schools. The small difference in slopes

A Review of Aspects of the Queensland Core Skills Test - February 2012

was not significant but there was a small but significant difference of 1.53 in the intercepts. This means that females in co-educational schools had slightly higher predicted WSM scores than females in single-sex schools with similar QCS Test scores.

The second comparison was between the regression for males in single-sex schools with that for males in co-educational schools. This showed that males in co-educational schools had lower predicted WSM scores than males in single-sex schools for the same QCS Test scores.

The results of the regression analyses were supported by an analysis of the relationship between OPs and QCS Test scores. Mean QCS Test scores were available for each OP band for both males and females. A two-way analysis of variance (OP band by gender) showed that the gender effect was significant.

The average mean QCS Test score for male students was 170.0 compared to 162.7 for females. On average, females received higher OPs than suggested by their QCS Test scores.

Further analyses were carried out to investigate the effects of school-type, sector, location and size on both the relationship between WSM and QCS Test scores and between OP and QCS Test scores. Small differences in relation to school type, location and sector were observed. To a large extent, these differences could be explained by the spread of single-sex and co-educational schools across sectors and locations.

Differences between co-educational and single-sex schools can be readily accounted for by the gender differences observed earlier. As the majority of government schools were co-educational and all remote schools were co-educational, the observed sector and location differences can be attributed to gender differences.

3.6.2 Implications

These analyses show that the relationship between WSM and QCS Test scores is not independent of gender in co-educational schools: males performed better than females in the QCS Test, but not in the school assessments; and on average, females received higher OPs than suggested by their QCS Test scores.

There are two implicit requirements of a test which is going to be used to transform the scores of tests (or measures) to ensure that no difference results from the subjects (or measures) the candidates take or from the school they attend.

The first requirement is that the tests (measures) to be used are measuring the same construct (variable). There is no doubt that the QCS Test is measuring the CCEs that are part of the same curriculum that is being measured by school-based assessments. Therefore, it is generally measuring the same construct as the school-based assessments.

The second requirement is that the result of the scaling does not depend on factors other than achievement in the QCS Test. The above analyses show that this is true for students enrolled in single-sex schools, but for students enrolled in co-educational schools the results of scaling are affected by student gender.

Gender has also been identified as having an effect on the scaling of within school measures. However, it is possible that the differences associated with gender are also attributable to some other factors. If the population is subdivided by other criteria, there will probably be differences in the groups' performances in the test and their school-based assessments. For example, dividing the population according to mathematics/science students and others could also give an indication of inconsistency between the scaling measure (predictor) and the criterion.

The Review Team believes that this outcome warrants further analysis.

Recommendation 4

that the QSA conducts further research to determine the reasons for the gender-related difference in students' rankings based on QCS Test scores compared to their rankings based on WSM

Recommendation 5

that the QSA conducts further research to determine the extent of any dependencies that occur, the effects of such dependencies on the outcomes of the scaling process, the potential causes of any dependencies and, if necessary, strategies for minimising the dependencies and their effects.

3.6.3 Further comments

The QSA is to be commended for the extensive quality control measures that are in place to locate and rectify anomalous results at individual and school levels, and the provisions for small schools. These provisions are seen to be necessary because the calculation of OPs, FPs and the reported grades are based on school-based assessments.

On the other hand, the extensive nature of these provisions means that the scaling process (see Appendix F) is very complex and therefore not always transparent. For example, some of the feedback to schools is provided in terms of the polyscore – a construct used for quality control but not for actual calculation of the OAI.

The Review Team strongly believes that the QSA should work with schools and teachers to identify ways in which reporting of QCS Test scores can be used to enhance the transparency of the scaling process. The QSA might also wish to examine the scaling procedures it uses to see whether a simpler approach would meet its needs and the needs of tertiary institutions.

3.7 Extent to which the administration of the test has the confidence of the stakeholders with respect to security and integrity

Interviews with stakeholders convinced the Review Team that the stakeholders have the greatest confidence in the security and integrity of the QCS Test. The fact that there have been no known breaches of security since the inception of the test gives further support to the view that the processes used in the administration of the QCS Test have integrity and are secure. This is addressed further in Section 4.

As discussed in Section 2.7, the QSA should conduct a risk assessment of the security measures at the Marking Centre to ascertain whether any changes are warranted. This should be regarded as standard management practice and, over time, could be applied to other aspects of the QCS Test program as well.

3.8 Fairness and reliability in the marking of the test

Considerable attention is given to ensuring that the marking is fair and reliable. The procedures for marking the SR and WT papers are outlined in Section 2.

Marker judgements are monitored throughout the marking operation. The monitoring is to identify discrepant markers; identify aspects of the marking schemes, criteria or standards that might need adjustment; and identify pairs of grades awarded to particular student responses that are sufficiently discrepant to warrant an additional marking.

The QSA uses two main quality control processes that focus on the marking of individual markers. These are referred to as 'marker monitoring' and 'check marking'. Marker monitoring identifies those scripts that have received significantly different grades from different markers, and through this identifies those markers who are regularly in disagreement with other markers or who demonstrate other inconsistencies. The check marking process involves the unit manager, immerser or marking supervisor marking a group of scripts already marked by a particular marker to ascertain whether they agree with the grades awarded. The check marking may be 'routine' (ie when a folder of student responses is checked at random) or 'targeted' (ie when the work of a particular marker is selected due to some concerns about the reliability of their marking).

Those markers who are identified as being 'discrepant' from other markers and are not applying the marking scheme correctly are taken aside for further training or 'refocusing'.

The statistics provided by the QSA reporting the level of absolute agreement between markers during the initial marking give the impression that the reliability of the SR paper is reasonable on the whole, but that the reliability of the marking of the WT paper is low. For example, the Krippendorff alpha statistics calculated by the QSA for the SR items typically range between 0.5 and 0.8, whereas the values for the WT typically range from 0.05 to 0.07. When interpreting these statistics it must be remembered that they are based on the grades and categories awarded by the two markers initially and are based on achieving complete agreement between the two grades awarded. Given that each of the first five criteria associated with the WT is marked on an 18-point scale, it is not surprising that these values are low.

The Review Team is confident, given the quality assurance procedures that are in place, that the marking of the QCS Test is reliable and fair. The Review Team would suggest, however, that a more suitable and useful measure of marker reliability for the WT be considered; one that better takes account of the particular situation and gives more realistic measures of the reliability that is being achieved.

4 Other issues identified during the stakeholder consultation

4.1 Public support for the QCS Test

In the interviews with stakeholder groups conducted by the Review Team very strong support was expressed for the QCS Test and its place in the senior secondary assessment program. Principals, teachers, students and parents were strong advocates for school-based subject assessments, the way they are moderated, and the QCS Test and its role in determining the Overall Positions (OPs) used for tertiary entry.

A few individual teachers advocated the introduction of external curriculum-based examinations or criticised aspects of the current test. In addition, an internet chat site that was used by students and others to criticise one particular multiple-choice item included in the 2011 test contained a number of comments that claimed that external examinations should be introduced. Such criticisms or calls for change are not considered to be widespread or substantial.

The Review Team concluded that the QCS Test has strong support across the community.

4.2 Suitability for all students

An aspect of suitability to do with apparent differences in the performances of males and females in the test is addressed in sections 3.6.1 and 3.6.2. This will not be discussed further here.

Some principals and teachers raised another issue regarding the suitability of the test for all students. They expressed the view that students from non-English speaking backgrounds find the test very difficult.

There are two different groups of non-English speaking background students. The first group consists of students who have come to Queensland to complete the final year of schooling and gain an OP equivalent. These students (usually referred to as 'international' or 'visa' students) do not usually undertake a full program of study required for the award of the QCE. The second group consists of students who live in Queensland, but come from a home where English is not the first language. Some of these students have difficulty in understanding English.

Some schools have large numbers of 'visa' students. Although they are required to sit the test if they want an OP equivalent, the results of these students are not necessarily included in the scaling exercise for their school group. This ensures that the results of this group of students will not adversely impact on the results of other students in the school group.

Some principals and teachers expressed concerns about the second group of non-English speaking background students who are following an OP-eligible program. (All students who undertake an OP-eligible program of study are required to sit the QCS Test.)

When this issue was discussed during the interviews several different views were expressed. Some felt that those students from a non-English speaking background should be given extra time and other assistance, such as access to dictionaries, to ensure they are able to adequately attempt the test. The point was made that some schools currently give non-English background students these extra provisions for assessment tasks administered as part of the school-based assessment and in order to be consistent and fair, these students should be given the same consideration when attempting the QCS Test.

A counter argument put was that the QCS Test is used in the process of determining tertiary entrance and strong skills in English are required when students go to tertiary study. Hence, accommodations should not be made in this regard because other students might be disadvantaged.

Some of those interviewed suggested that non-English speaking background OP-eligible students should not be compelled to sit for the QCS Test if they do not intend to go directly to tertiary study. Those who saw this non-participation as a solution quoted anecdotal evidence to the effect that some schools already encourage their non-English speaking background students not to sit for the QCS Test.

The QSA already has procedures for identifying those students whose QCS Test results are significantly below what might be expected from other evidence. The QSA then adjusts the scaling procedure to ensure that those who perform well below or above expectation on the test will not adversely affect the results of other students.

The Review Team is of the view that there is sufficient concern in the community to warrant the QSA addressing the issue by assuring the community that the OPs of other students within a school are not affected by having OP-eligible non-English speaking background students sit for the QCS Test.

After considering the views expressed by the various groups the Review Team makes the following recommendations:

Recommendation 6

that unless items in the QCS Test are specifically designed to assess students' expertise in vocabulary and/or language, processes be put in place during the development of the test to ensure that the language used is as clear and simple as possible

Recommendation 7

that the QSA considers whether those students following OP-eligible programs who do not wish to receive an OP should be required to sit for the QCS Test

Recommendation 8

that the QSA addresses the concern about the perceived negative impact of the scores of OP-eligible non-English speaking background students on the scaling process by providing public advice on how this matter is addressed

4.3 The timing of the administration of the test

The QCS Test is scheduled for the Tuesday and Wednesday of the third last week of Term 3. This enables the marking to occur during the school holidays between Term 3 and Term 4.

Some students and teachers who were interviewed indicated that this was a time when most schools also ran a major part of their internal assessments. The students claimed that they felt stressed and exhausted by the end of this period.

When asked for suggestions for an alternative time to administer the test some stated that the end of Term 2 would be better while others nominated the third week of Term 3.

Other groups of principals and teachers, however, expressed the strong view that the tests should continue to be administered at the current time.

It is possible that in the future the greater use of technology, both in administering the test and marking, will present some further opportunities in relation to the scheduling of the test. This matter is discussed in some detail later in the report.

At this point the Review Team makes the following recommendation:

Recommendation 9

that the QSA Governing Body considers whether any change to the current scheduling of the QCS Test should be made

4.4 The role and prominence of the Common Curriculum Elements

The items in the QCS Test are based on 49 Common Curriculum Elements (CCEs). These CCEs were established by identifying those skills that are developed through the study of courses from the available curriculum. To be included among the 49 CCEs an element has to be present in a number of courses. This is necessary to ensure that every student has a chance to develop these skills irrespective of the pattern of courses he or she takes. Furthermore, these CCEs have to be able to be tested by way of a pen-and-paper test.

Principals and teachers interviewed during this review understood the significance of the CCEs in relation to the QCS Test. They also understood the importance of ensuring that a school's

teaching programs for the senior secondary courses ensure students have opportunities to develop these skills.

A number of principals and senior teachers also indicated that they ensure that their teachers take account of the CCEs when preparing and delivering teaching programs in the junior school. The Review Team believes that this is a sound approach. Although the focus of education is broader than the CCEs, they nevertheless should be regarded as important when teaching the various courses.

Some experienced teachers made the observation that when the QCS Test had first been introduced there was much useful information on the CCEs and their place in each senior secondary course. They indicated that over the years some of this documentation has not been kept up to date, particularly when new syllabuses have been introduced. There is a danger that the role and importance of the CCEs will be diminished over time if they are not widely documented and presented, particularly to new teachers.

In recent years the Australian Curriculum Assessment and Reporting Authority (ACARA) has been developing and rolling out courses that will form part of the curriculum for students in all states and territories. Although the curriculum documents released so far only relate to Foundation to Year 10, it is important that the documents provide students with opportunities to develop the knowledge, skills and understanding covered by the CCEs. This will become even more critical when ACARA develops curriculum documents for the senior secondary years.

The QCS Test links directly to the curriculum through the CCEs. Because of this, it is not a test of the specific courses in the curriculum but it is an achievement test based on the curriculum. The school assessment program is based on the curriculum, so the validity of the QCS Test as a scaling device is dependent on how well these two aspects of assessment assess performance on the CCEs. If the focus on the CCEs in teaching is diminished, it is likely that the effectiveness of the test as a scaling instrument will also suffer.

After considering the views expressed to it during the review, the Review Team makes the following recommendations:

Recommendation 10

that the QSA undertakes a review the CCEs to ensure that those on the existing list are still current both in their relevance and expression, and to include any elements that meet the criteria to be a CCE that have emerged recently

Recommendation 11

that the QSA updates the documentation explaining the role and importance of the CCEs and their place in each course

Recommendation 12

that the QSA works to ensure that the CCEs are appropriately captured in the curriculum documents being produced by the Australian Curriculum Assessment and Reporting Authority (ACARA)

Recommendation 13

that information showing the alignment of the CCEs with the Australian curriculum documents is provided to teachers before the implementation of those curriculum documents

4.5 Feedback to schools on the performance of their students

The QSA provides a range of useful summary information to principals on the performance of their students on the QCS Test. This information is generally used by schools to monitor the performance of their students over time and to identify any general trends in the performances of the students that may warrant a review of the teaching programs.

Several observations were made to the Review Team regarding this information.

One observation was that the reports provided by the QSA were only available for a relatively short period of time. This requires schools to download their information and store it on the school's system so that it can be accessed after the QSA removed the reports from their system.

Another observation was that there was very little feedback to schools regarding the relative strengths and weaknesses of the performance of individual students. Interviewees making this point were drawing attention to the fact that the QSA is missing an opportunity to make the test even more significant by taking advantage of the fact that it is directly linked to the curriculum and, as such, the information could be used to improve learning. This link to the curriculum is a feature of the scaling test, which is unusual. Generic tests that are used for scaling are usually of limited value in informing teaching and learning.

While it is undoubtedly true that the group performance is the most significant feature of a scaling test, the Review Team believes that a more detailed breakdown of the performances of the individual students would be helpful to a school in better understanding the effectiveness of its teaching programs.

The Review Team also believes that the QSA could investigate developing and providing some software to schools that would produce helpful and meaningful reports. Discussions with some other jurisdictions that have developed a similar system would assist in exploring what would be the best approach for the QCS Test and Queensland schools.

After considering the views expressed to it during the review, the Review Team makes the following recommendations:

Recommendation 14

that the QSA establishes a secure service that enables principals to access the QCS test results of their students and related information for any year at any time

Recommendation 15

that the QSA holds talks with principals to ascertain what additional information on the performance of individual students in the QCS Test would be useful to them and how it could be presented in a helpful and user-friendly manner

Recommendation 16

that the QSA develops software that will assist principals to analyse and present the performance data of their students in ways that will support the evaluation of the school's teaching programs

4.6 Feedback to students on their performance in the test

The performance of students in the QCS Test is currently reported on their Queensland Certificate of Education (QCE) using an A to E scale. Some of the students interviewed indicated that they would appreciate more detailed information on how they had performed. This was no doubt due to the fact that students generally receive helpful information from their teachers about their performances in school-based assessment tasks.

Currently, some schools engage the services of a commercial company to administer, mark, analyse and report on the students' performances in a trial QCS Test. The results from these trial tests are then presented to the school and student in a form that makes explicit the strengths and weaknesses in their performance.

There is an argument that as the students who have finished the test will not be in a position to use the information to improve their performances; the diagnostic component is of limited value. However, it may be possible for the QSA to make some additional information available to individual students to assist them in better understanding their performance.

The Review Team is of the view that the QCS Test is unique as a scaling test in that it assesses a substantial number of key elements that are directly embedded in the curriculum. As stated in Section 4.5, a greater emphasis on feedback on the performances of individual students would emphasise this unique attribute and strengthen the links between teaching, curriculum and assessment to improve learning.

The Review Team makes the following recommendation:

Recommendation 17

that the QSA investigates ways in which some additional information on each student's performance in the QCS test could be made available to the student at the time their results are released

4.7 The information provided to the universities about student performance

The QSA provides the Queensland Tertiary Admissions Centre (QTAC) with student Overall Positions (OPs) and Field Positions (FPs). The FPs are additional rank orders that supplement an OP. The term "field" refers to areas of emphasis in the senior curriculum. Tertiary institutions can use FPs when greater differentiation between students in an OP band is required, for example, when the number of eligible applicants exceeds the number of places for a course.

A student may receive up to 5 FPs, depending on subject choices. FPs are reported in 10 bands, from 1 (the highest) to 10 (the lowest) in the following fields:

- Field A – extended written expression involving complex analysis and synthesis of ideas
- Field B – short written communication involving reading, comprehension and expression in English or a foreign language
- Field C – basic numeracy involving simple calculations and graphical and tabular interpretation
- Field D – solving complex problems involving mathematical symbols and abstractions
- Field E – substantial practical performance involving physical or creative arts or expressive skills.

When calculating FPs, the QSA uses the Subject Achievement Index (SAIs) for the subjects that contribute to a particular field. The extent to which a subject contributes to each FP depends on the weighting of that subject. Although subjects do not contribute equally to FPs, most subjects contribute at least some weight to most fields.

QTAC reports that in 2011, just 13 (0.8%) of the 1654 courses available in tertiary institutions used FPs in the decision-making process. Considering the considerable work of the QSA in calculating FPs, the Review Team makes the following recommendation:

Recommendation 18

that the QSA holds discussions with QTAC and the universities to determine the usefulness of information currently provided to universities for course selection processes, and to make changes to the information if warranted

QTAC uses the QCS Test results when processing applications for tertiary places from those students without an OP such as Queensland Year 12 students who are ineligible for an OP and students who have taken the International Baccalaureate (IB). This is a reason for non-OP-eligible students to sit the QCS Test. It was noted in Section 3.3 that the proportion of these students has remained steady at around 7% of certified students.

4.8 Communication with students and parents about the test

In addition to the range of information provided by the QSA on the nature and purpose of the QCS Test, senior teachers at each school ensure that students understand the structure, nature and focus of the test. All the schools visited by the Review Team had programs in place to ensure students were familiar with the test. Programs included, but were not limited to, giving students the opportunity to sit for past tests and receiving feedback on their performances. Schools also provided appropriate advice to students on effective strategies for undertaking the test, such as managing their time effectively and checking their responses.

While there is no doubt that students are well briefed as to the nature and purpose of the test, there is scope to provide further information to parents. Those parents who had read the material provided by the QSA to their son or daughter, and those who had had other children sit for the test in the past, had a better understanding than those parents whose children were experiencing the QCS Test for the first time.

The Review Team concluded that there would be value in the QSA producing a document particularly designed for parents explaining the nature of the QCS Test and its function. This document could be organised around the common questions that parents ask and provide answers to those questions in clear and simple language. Involving a number of parents in the preparation of the document would help keep the focus on their needs.

The Review Team makes the following recommendation:

Recommendation 19

that the QSA prepares a document to be provided in the early part of each year to every parent of a student sitting for the QCS Test that explains the nature and function of the test

4.9 The materials provided to schools about the administration of the test

The QSA provides detailed manuals and other material to schools to assist them in preparing for and administering the QCS Test.

The Review Team discussed the nature of the administrative materials with a number of QCS Coordinators and Community Representatives. Those coordinators and representatives who had fulfilled their roles for a number of years felt that the information was very thorough and comprehensive. Some did express the view that when they were performing the tasks for the first time they found the manuals somewhat daunting at first and a little hard to follow. Nevertheless they did agree that all the information they needed was provided. The QCS Coordinators and Community Representatives interviewed indicated that when they needed to contact the QSA to clarify matters or discuss issues, the QSA officers were helpful and supportive.

In order to make the tasks of new QCS Coordinators and Community Representatives easier the QSA could hold a meeting with some experienced and inexperienced coordinators and hold

a similar meeting with community representatives, to discuss what further direction and support could be provided to the people carrying out these roles. This support might take the form of additional explanatory materials, some briefing sessions, an animated DVD or pairing new representatives with experienced colleagues.

The Review Team makes the following recommendation:

Recommendation 20

that the QSA holds discussions with QCS coordinators and Community Representatives to identify ways in which additional support could be provided effectively and efficiently to support less experienced personnel

4.10 Professional development for teachers

A number of principals and senior teachers commented in interviews that when the QCS Test was first introduced the QSA supported it with a strong program of professional development for teachers. They said that over time, as the program became established and common practice, this form of professional development has largely disappeared.

There is no doubt that overall, the QCS Test program runs very smoothly. However, some less experienced teachers said they would benefit from having a more detailed understanding of the test and the place of the CCEs in the education of their students.

Markers who were interviewed said they had gained greater understanding of the QCS Test by being involved in marking the scripts. They felt they had developed knowledge they could take back to their schools and share with colleagues.

The QSA makes a point of employing a significant number of markers from country areas. This is a sound approach as it means that those country teachers can assist in disseminating information on marking practices to their colleagues. This practice is to be commended.

To further support teachers in understanding the QCS Test, its connection to the curriculum, and its practices, the QSA could look at the related professional development activities it conducts and identify any additional needs and opportunities.

The Review Team makes the following recommendation:

Recommendation 21

that the QSA examines the range of professional development activities it provides to teachers related to the QCS Test and identifies additional activities that can be effectively and efficiently incorporated into its programs, including how information on student and school performance in the test can be used to improve teaching and learning

4.11 Comments from QSA staff involved in the test development process

QSA staff members involved in the test development process expressed a view during discussions with the Review Team that the quality of the QCS Test was largely due to the rigour and robustness of the panelling process.

Some members of the item development team expressed concerns about what they saw as the 'unfounded and pervasive myths about the QCS Test' – in particular that either 'humanities' orientated students are disadvantaged or that 'quantitatively' orientated students are advantaged. Those who raised this issue believed that engaging a broader range of teaching staff in the development and marking processes would attest to the 'fairness' of the test and they would become ambassadors for it. This was supported by anecdotal evidence that

teachers who have been involved in the marking process are more likely to discredit the myths about the QCS Test.

The QSA Test Development Team also felt that:

- broadening the pool of markers would lead to a greater awareness of the procedures for constructing the papers and this would provide more support for the validity of the QCS Test amongst teachers and the wider community
- increasing the number of teachers from regional and remote areas in the panelling and marking activities would be worthwhile for a number of reasons
- recruiting a small number of additional item writers to create multiple 'voices' and more diversity within the papers would be advantageous to the test development process. They felt that this was becoming harder to achieve with the current circumscribed number of item writers
- there was a tension between short-term secondments to promote diversity and the need for item developers to work within the unit for some time before they 'hit their stride'. Linked with this is the need to actively consider succession planning, given the unique and constrained nature of the item development team.

The Review Team is of the view that these suggestions from committed and experienced QSA staff members have merit. However, as the Review Team did not explore the issues surrounding these comments, it believes that they should be considered by the QSA senior management team in the context of the total program. In particular, the Review Team believes that the quality of the QCS Test items is due to a large degree to the highly skilled item writers. The Review Team is of the view that the QSA needs to give consideration to ensuring it has an ongoing pool of writers capable of producing high-quality items.

5 The future

5.1 The impact of the introduction of the Australian Curriculum

Some initial discussion of this issue can be found in Section 4.4 above and the comments here should be considered in conjunction with the points made earlier. In addition, Recommendations 12 and 13 address aspects of this issue.

The CCEs are an important aspect of the Queensland curriculum. Not only are they the basis for the QCS Test, but also they help to ensure that schools' teaching programs appropriately and adequately enable these important elements to be addressed.

It is important that the QSA reviews draft curriculum materials produced by ACARA to ensure that they adequately cover the CCEs that would be expected to be present in the particular courses.

Once the curriculum documents have been finalised and approved, the QSA should produce and disseminate documentation that shows the CCEs covered by that course. This process should be undertaken for all secondary school courses introduced as part of the Australian curriculum.

At this stage there is not sufficient information available about the proposed Australian Baccalaureate for the Review Team to evaluate the likely impact of its implementation on the

QCS Test and the QCE. If clear and detailed information is provided in this regard in the future, further analyses of the impact of such a change can be made.

The Review Team makes the following recommendation:

Recommendation 22

that the QSA continues to monitor proposals and developments concerning the proposed Australian Baccalaureate and assesses any likely impacts it might have on the QCS Test

5.2 Greater use of ICT in the marking process

The QSA already uses ICT to support aspects of the marking operation. The discussion that follows focuses on how further use of technology could deliver efficiencies and enhancements in the marking program. The approach discussed here is generally referred to as 'onscreen marking'.

Onscreen marking for the QCS Test would consist of having students complete the test as they do at present, scanning their responses and distributing the SR and the WT responses to markers' computers. The markers then record a mark or grade using their keyboard.

The software that controls the onscreen marking is written to contain whatever processes and features the testing authority requires. For instance, all steps currently incorporated into the marking operation could be included if desired.

A major advantage of such an approach is that once the electronic images of a student's responses are captured they can be distributed to markers and their supervisors far more flexibly and efficiently than can be achieved with manual processes. Delays due to manual handling of responses, delays in getting quality feedback on marker reliability, delays in loading marks onto the computer and delays in resolving discrepant marks can be significantly reduced. Some states already use such approaches as part of the marking of parts of their senior secondary examinations. A similar approach is used in Queensland by the QSA in marking the extended responses of the students sitting for the NAPLAN Tests.

In addition, such a marking operation can either be conducted completely at a single centralised marking centre, or after some initial centralised training in the marking scheme and in using the software, markers can undertake the marking task elsewhere. This would open the marking operation to more country markers. It could also open it up to some potential markers in the metropolitan area whose personal circumstances currently prevent them from marking. In addition, it would have the potential to enable the number of markers to be expanded if the QSA wishes to do this.

The different groups of markers and their supervisors who were interviewed at the Marking Centre raised the issue of onscreen marking. Those associated with the marking of the Writing Task, particularly those who had marked NAPLAN, were very open to the idea of using an onscreen approach.

Those associated with the marking of the SR papers were less enthusiastic about the introduction of onscreen marking. The main concerns they raised were the belief that online marking would require restrictions on the way items were constructed and that it might be difficult to mark certain items due to the need to consult or manipulate additional materials.

It is important that the test items used should not be unreasonably limited by what the technology can accommodate. If some items would prove difficult to mark using an onscreen approach then they could continue to be marked in the traditional manner. It would be expected, however, that these would be the exception.

The Review Team is of the view that the QSA should investigate the introduction of onscreen marking as a matter of priority. As some other states are currently using this approach, QSA

officers could visit these jurisdictions, hold discussions with key personnel and observe a demonstration of their systems.

The Review Team understands that the QSA has established a project team to investigate this matter and develop an implementation plan for approval. The project team will need to determine how best to proceed and whether to stage the implementation over two or more years. In this regard, however, it might well be possible to implement onscreen marking to some extent for the 2012 QCS Test. If this does not prove to be feasible, some pilot studies and testing could be undertaken using the 2012 test materials with a view to widespread implementation in 2013.

The Review Team makes the following recommendation:

Recommendation 23

that a high priority is given to the project of investigating, and developing a plan for, the introduction of onscreen marking of the Writing Task and Short Response papers of the QCS Test

5.3 The use of ICT by students in completing the test

The issue of changing the QCS Test so that students would complete it by computer was discussed at each of the schools visited and at the Marking Centre.

Some of the teachers interviewed felt that the use of ICT by students was getting to a point where consideration should be given either to changing the test so that students would complete it using a computer, or to incorporating some use of a computer in part of the test. They argued that if the CCEs were revised, skills in the use of ICT would be included in the list of CCEs. One suggestion was that the writing task be turned into a task to be completed by students online. A further suggestion was that an additional online writing task could be added to the test.

While the great majority of principals and teachers were of the opinion that it was essential that students be given opportunities to develop their skills in the use of ICT, most were not in favour of moving to any form of online test at this time due to the variability in capacity and reliability of computer infrastructure in schools.

A number of groups indicated that, while the rollout of computers in schools was proceeding, it was not yet at a point where the QSA could be confident that a computer-based version of the QCS Test could be delivered without problems in schools. Furthermore they argued that, at this point, not all students have had the opportunity to develop their skills in the use of computers to a high level and this would mean that the process of responding online would disadvantage some students.

Other groups made the point strongly that the ability to produce the type of response required by the WT is an important skill and should not be changed in spite of the increasing use of computers. They expressed the view that creating a document with a computer involves a different approach to that involved in creating a document with pen and paper. They felt that the current task's requirements for planning, drafting, editing and then producing a final version involve skills, processes and procedures that should be retained.

A concern expressed by some teachers was that it would be difficult to continue to include some of the SR items currently included in the test if they were required to be delivered and answered online. In their opinion this would mean that the nature of the test would change. On the other hand, the use of computers would be likely to increase the range of items that could be included in the test.

In summary, the overwhelming opinion of those consulted was that for now the test should remain a pen-and-paper test. It must be said, however, that there was a strong body of opinion

that at an appropriate point in the future a change should be made to the QCS Test so that at least part of it can be undertaken online.

One advantage of moving part, or all, of the QCS Test to an online delivery is that it would address the third reason given by Trost (1992) for his recommendation to not release the test paper each year. Each year a proportion of the multiple-choice items could be items that were being trialled during the actual test. These items would not count towards the students' result in the test and so could remain secure. The other items could be published after the test. In this way, item statistics could be obtained on the trial items using Queensland Year 12 students under actual examination conditions. Those trial items that functioned well could then be included in the QCS Test in later years.

The Review Team believes that the QSA should start looking at this issue by establishing a project team to investigate it and develop some options for consideration. The project team should have some members in common with the project team discussed in section 5.2.

The Review Team makes the following recommendation:

Recommendation 24

that the QSA establishes a project team to investigate the issue of having students complete part of, or the whole, QCS Test by computer and develop feasible options and timelines for further consideration

6 Concluding comments

The QCS Test has been running for 20 years. It was previously externally reviewed in 1992 and 1996. Both of these reviews were positive concerning the test and its associated processes and practices.

In addition to reviewing operational manuals and reports, and analysing data provided by the QSA, the Review Team met with a wide range of external and internal stakeholders to ascertain their views on the test. The Review Team is grateful for the cooperation and assistance of everyone involved, and for their willingness to share their thoughts and views.

The Review Team considers the QCS Test program to be sound and well conducted. The test itself is of high quality due to its design criteria, the care and expertise of those involved and the disciplined procedures used in its development. The administrative procedures are well established and run smoothly and effectively. The marking processes are thorough, carefully applied and well monitored.

The Review Team is of the view that overall the QCS Test continues to perform well the functions for which it was designed and introduced. The recommendations in this report are designed to enhance and further improve the program and the test instrument, and extend the test's viability and effectiveness as a scaling instrument well into the future.

A number of recommendations relate to the test itself, including investigating potential enhancements of the trialling program and conducting further research into how different groups perform in the test. Other recommendations focus on the importance of ensuring that the CCEs remain an important and prominent feature of the secondary curriculum so that their connection between the curriculum and the test is not weakened. Some other recommendations are aimed at improving the understanding of both teachers and the public about the purposes of the test and its processes. A further group of recommendations is aimed at providing schools with richer information on the performance of their students and showing teachers how they can use this to improve teaching and learning. In addition, there are recommendations relating to the introduction of the Australian curriculum and the potential of greater use of ICT in the testing and marking programs.

APPENDICES

APPENDIX A: Documents considered by the Review Team

1. General information

- *About the QCS Test*, pamphlet for Year 11 students and their parents
- *Student Information Bulletin – All you need to know about the Queensland Core Skills Test*, booklet for Year 12 students
- *Common Curriculum Element – occurrence matrix*
- *CCEs within baskets*
- *Schedule for grading the QCS test*
- Matters, G. 1991, *A design process for constructing the Queensland Core Skills Test*, Board of Senior Secondary School Studies, Queensland
- Trost, G. 1992, *The Queensland Core Skills Test: Evaluation of Design Criteria and Process*
- Trost, G. 1996, *The Queensland Core Skills Test; A Follow-up Evaluation of the Design Criteria and Process*
- QSA, 2007, *How schools prepare students for the Queensland Core Skills Test*, report

2. Copies of the test and retrospectives

- 2008 Queensland Core Skills Test paper
- *Queensland Core Skills Test – Retrospective 2008*
- 2009 Queensland Core Skills Test paper
- *Retrospective – 2009 Queensland Core Skills Test*
- 2010 Queensland Core Skills Test paper
- 2011 Queensland Core Skills Test paper

3. Administrative information and instructions

- 'Directions for Administration – 2010 Queensland Core Skills Test'
- *Marker Manual – Queensland Core Skills Test, 2010*
- *Marker Manual – Queensland Core Skills Test, September 2011*
- *QCS Test 2011 – Welcome to the 2011 Queensland Core Skills (QCS) Test marking operation*, general information for markers
- 'WT marking procedure' instructions
- 'Writing Task marking guide: Criteria and standards'
- Short Response Marking Scheme – Unit Two Item 2, Unit Eight Items 15, 16 and 17

4. Data reports

- *Queensland Core Skills Test Yearbook – 2009*
- QSA 2009, *QCS Test – OP-eligible QCS candidates – Data for the State*
- QSA 2009, *Data summary – QCS attendance and grades*
- *Queensland Core Skills Test Yearbook – 2010*
- QSA 2010, *QCS Test – OP-eligible QCS candidates – Data for the State*
- *QCS performance of OP-eligible students at a Sample School 1992–2010*, report showing standardised difference between school and state means

APPENDIX B: QSA officers interviewed by the Review Team

1. Officers involved in item writing and test development

Shiron Dixon	Manager, QCS Test
Greg Wightman	Principal Testing Officer (MC)
Pam Brennan	Principal Testing Officer (WT)
Maryanne Gurieff	Principal Testing Officer (SR)
Paul Christensen	Principal Testing Officer (Itembank)
Andy Edwards	Senior Testing Officer
Kirsti Ellerton	Senior Testing Officer
Peter Panteleit	Senior Testing Officer

2. Marker monitoring, marking operations (logistics), data collection

Brian Nott	Assistant Director, Analysis and Reporting
Marg Veerman	Marker Monitoring Team
Kevin Offer	Marker Monitoring Team
Karen Wilson	Marker Monitoring Team
James Cousins	Marker Monitoring Team
Wayne Poole	Marker Monitoring Team
Danielle McPherson	Marker Monitoring Team
Katy Wong	Marker Monitoring Team
Kaye Draper	Marking Operations Coordinator
Allan Hill	Marking Operations
Bev Palmer	Marking Operations
Lenny Ingle	Data Collection
Christine Moore	Data Collection
Arty Bryant	Data Collection
Ann Dorey	Data Collection
Ian Redpath	Data Collection

APPENDIX C: Meetings held by the Review Team with supervisors of marking, immersers/supervisors and markers

1. Writing task coordinator and supervisors (trainers)

Pam Brennan	Coordinator of Writing Task
Marg Veerman	WT Supervisor/Marker Trainer
Richard Dobrenov	WT Supervisor/Marker Trainer
Edna Galvin	WT Supervisor/Marker Trainer
Luke Higgins	WT Supervisor/Marker Trainer
Christine Hill	WT Supervisor/Marker Trainer
Carole Houston	WT Supervisor/Marker Trainer
Tracey Jarrett	WT Supervisor/Marker Trainer
Linda Kempe	WT Supervisor/Marker Trainer
Scott McDonald	WT Supervisor/Marker Trainer
Lauren Newman	WT Supervisor/Marker Trainer
Russell Smerdon	WT Supervisor/Marker Trainer
Ann Vitale	WT Supervisor/Marker Trainer
Sam Watts	WT Supervisor/Marker Trainer
Peter Panteleit	WT research project
Michael Madden	WT research project

2. Ten markers of the Writing Task

3. Short response coordinator and immersers (trainers)

Maryanne Gurieff	Coordinator of Short Response
Nev Glynn	Unit Manager
Shiron Dixon	Unit Manager
Paul Christensen	Unit Manager
Wendy Amy	Immerser/Marker Trainer
John Oxley	Immerser/Marker Trainer
Helen Best	Immerser/Marker Trainer
Pam Lanham	Immerser/Marker Trainer
Sharon Crone	Immerser/Marker Trainer
Kevin Gray	Immerser/Marker Trainer
Andy Edwards	Immerser/Marker Trainer
Lydia Commins	Immerser/Marker Trainer
Paul Hathaway	Immerser/Marker Trainer
Jack White	Immerser/Marker Trainer
Wendy Mackay	Immerser/Marker Trainer
Kirsti Ellerton	Immerser/Marker Trainer
Jan Findlay	Immerser/Marker Trainer
Bob Chrisopherson	Immerser/Marker Trainer
Rob Forsyth	Immerser/Marker Trainer

4. Ten markers of the short response items

APPENDIX D: Details of stakeholders involved in the consultations

1. School visits

The Review Team visited five schools between Monday 12 September and Thursday 15 September and held discussions with:

- the principals, the QCS Coordinators and other senior teaching staff
- teachers with experience in teaching Year 12 students
- Year 12 students
- parents.

The schools visited were:

- Rochedale State High School (Principal: Karen Tanks)
- Marist College Emerald (Principal: Marie Martin)
- Emerald State High School (Principal: Angela Collins)
- Clayfield College (Principal: Brian Sabins)
- Kenmore State High School (Principal: John Fitzgerald)

2. Other meetings

a. Principals from Brisbane North District:

- Jeff Major (Wavell Heights State High School)
- Marker Stower (Mt Maria College)
- Vince Webb (St Columban's College)

b. Representatives of the Teacher Unions

From the Queensland Teachers Union:

- Karyn Hart
- John Fitzgerald
- Paul Penny

From the Independent Education Union:

- Jennifer Winn
- Dr Paul Giles
- Miriam Dunn

c. Members of the Tertiary Entrance Committee

- Professor Claire Wyatt-Smith (Griffith University)
- Margaret Price (Griffith University)
- Rod Varnham (TAFE Operations)
- Dale Nicholas (Brisbane Grammar School)
- Hayman Lui (Loreto College)
- Mick Cutler (Queensland Council of Parents and Citizens' Association)
- Greg Thurlow (Department of Education and Training)
- Pat Smith (Queensland Tertiary Admissions Centre)
- Peter Jordan (Queensland Studies Authority)
- Brian Nott (Queensland Studies Authority)

d. Representative of the Joint Council of the Queensland Teachers Association

- Neville O'Brien from INTAD

e. Representative of the Queensland Tertiary Admissions Centre (QTAC)

- Phillip Anthony

3. Phone interviews with QCS community representatives

- William Simpson
- Ross MacDonald
- Kevin Murgatroyd
- Stuart Napier
- Tracey Tappenden
- Jane Wallace
- William Sperring

APPENDIX E: Psychometric analysis of the QCS Test

1. Psychometric analysis of the 2009 and 2010 QCS Tests

In 1991 Matters wrote a paper entitled *A design process for constructing the Queensland Core Skills Test* in which she articulated the design criteria for the test. In the paper she identifies a number of components that need to be taken into account when constructing a psychometrically sound test that will be used for relatively high stakes assessment. The components include the following:

The QCS Test must:

1. be an integrated whole
2. be balanced and include a suitable range of items
3. be suitably difficult
4. discriminate among students
5. be without bias
6. make appropriate demands on the test-takers
7. have face and content validity
8. present items within each paper so that their sequence assists students to demonstrate their optimum performance.

These criteria reflect good test development methodology and are just as pertinent today as they were in 1991. It was important, therefore, for the Review Team to seek information on these criteria to ensure that the current QCS Test is still fulfilling the requirements of good test development practice. The first two components are considered separately below. However, components 3 to 8 relate to good test design and test functioning. As such, they are considered as a group in Part 4, 'Test and item functioning'.

2. An integrated whole

In the Trost Report (1992), the test being an integrated whole was interpreted as the test being:

[a] renunciation of homogeneous subtests and thus differs from principles that are guiding similar programs in other countries where different subtests are each designed to measure circumscribed aptitudes or areas of achievement and where in some cases not even total test scores are computed. Yet it is in line with the major recommendations of the Viviani Report and the subsequent political decisions to use a cross-curriculum core skills test, not a specific subject-related achievement test nor a scholastic aptitude test yielding an aptitude profile.

(Trost, 1992, page 3)

The Queensland Studies Authority (QSA) provides feedback to all schools on the whole test and on five separate criteria (Criterion α , Criterion β , Criterion θ , Criterion π and Criterion Φ). The Review Team strongly believes that the criterion-related data are useful at an individual school level as they provide feedback to schools and students that can be integrated into the teaching and learning program of schools to improve learning.

The potential advent of technology into the delivery of the QCS Test raises the prospect of providing schools with feedback at the item level within the criteria. This could add a powerful new dimension to the usefulness of these data in relation to the teaching and learning activity.

The fact that the QCS Test is linked directly to the curriculum means that the feedback is potentially more pertinent and relevant to the teaching and learning program than it would be if the test were a more generic aptitude or ability test.

When reviewing the effectiveness of the QCS Test in achieving its aim of being an integrated whole, the Review Team examined the internal consistency (reliability) of the test and each of its

A Review of Aspects of the Queensland Core Skills Test - February 2012

nominated criteria. The team also examined the correlation matrix between the overall test results and the criteria (α , β , θ , π and Φ) – which are referred to as ‘baskets’.

TABLE E1

Item types and reliability coefficients for the 2009 and 2010 QCS Tests

Year	QCS		Multiple choice		Short response	
	No of Items	Reliability	No of items	Reliability	No of items	Reliability
2009	118	0.86	100	0.89	17	0.85
2010	119	0.87	100	0.89	18	0.81

Table E1 shows that the reliabilities for the 2009 and 2010 QCS Tests are both very close to 0.90. This relatively high reliability suggests that the QCS Tests have a high degree of internal consistency and is evidence that the **test is operating as an integrated whole**.

Tables E2 and E3 examine the correlations between the various ‘baskets’ and the overall QCS Tests.

TABLE E2

Correlation coefficients between the baskets (criteria) and the 2009 QCS Test

	Baskets (criteria)					Total test
	α^1	β^2	θ^3	π^4	Φ^5	QCS
α	1.0000	0.7256	0.7648	0.5801	0.6483	0.8588
β	0.7256	1.0000	0.7390	0.5712	0.6151	0.8571
θ	0.7648	0.7390	1.0000	0.5792	0.6983	0.8938
π	0.5801	0.5712	0.5792	1.0000	0.4424	0.6201
Φ	0.6483	0.6151	0.6983	0.4424	1.0000	0.7638
QCS Test	0.8588	0.8571	0.8938	0.6201	0.7638	1.0000

It can be seen in Table E2 that all the criteria correlated quite highly (above 0.75) with the QCS Test except for the π criterion (Create and Present). This is not surprising as the Writing Task is included in this basket. This criterion was still relatively highly correlated with 0.6201.

¹ Criterion α comprises Comprehension and Collect

² Criterion β comprises Structure and Sequence

³ Criterion θ comprises Analyse, Assess and Conclude

⁴ Criterion π comprises Create and Present

⁵ Criterion Φ comprises Apply Techniques and Procedures

TABLE E3
Correlation coefficients between the criteria and the 2010 QCS Test

	Criteria					Total test
	α	β	θ	π	Φ	QCS
α	1.0000	0.7041	0.7561	0.6509	0.6661	0.9030
β	0.7041	1.0000	0.6758	0.5640	0.6372	0.8366
θ	0.7561	0.6758	1.0000	0.6000	0.6181	0.8782
π	0.6509	0.5640	0.6000	1.0000	0.5361	0.7548
Φ	0.6661	0.6372	0.6181	0.5361	1.0000	0.8250
QCS Test	0.9030	0.8366	0.8782	0.7548	0.8250	1.0000

The correlations between the baskets for the 2010 test, shown in Table E3, are relatively high and uniform (minimum = 0.536, maximum = 0.756).

Table E4 shows the correlations between the item types of the various papers for the 2010 QCS Test. As would be expected, the MC and SR are highly correlated, and the correlations with the Writing Task relatively lower.

TABLE E4
Correlation between item types of the various papers for the 2010 QCS Test

	MC	SR	WT	QCS
MC	1.000	0.807	0.511	0.936
SR	0.807	1.000	0.531	0.918
WT	0.511	0.531	1.000	0.718
QCS	0.936	0.918	0.718	1.000

The Review Team believes that these data support the conclusion that the test is operating as an integrated whole, yet the component parts (criteria) are still contributing uniquely to the underlying construct.

3. Balanced and includes a suitable range of items

Matters (1991) defined *balance* as the relative proportions of the dimensions listed below and *range* as the span of the sample space. She stated that the balance and range could be best assessed in terms of the following nine *dimensions*:

- a. Curriculum elements
- b. Attributes
- c. Epistemic content
- d. Text of stimulus material
- e. Length of stimulus material
- f. Context
- g. Type of response
- h. Difficulty
- i. Number of grades for marking student responses (for short response items)

Matters then specified how each item in the QCS Test could be classified according to the following criteria.

A Review of Aspects of the Queensland Core Skills Test - February 2012

- a. Up to three curriculum elements – primary, secondary or tertiary emphasis
- b. Up to three attributes
- c. One epistemic area – symbolics, empirics, aesthetics or synoptics
- d. One of four types for the text of the stimulus material – numerical, verbal, tabular/graphical or visual
- e. One of three lengths – short, medium or long, together with the number of items per unit of stimulus material
- f. One difference of context – everyday or esoteric
- g. One of three types of response activity – free active, constrained active or constrained passive PLUS at least one of three types of response modes – verbal, numerical, tabular/graphical, visual or open
- h. One of three difficulties – easy, medium or hard
- i. Number of categories for marking responses – 2, 3, 4 or 5

The Review Team has not tried to reconstruct a full table that captures all the information above. Rather the intention is to show how the items and marks have been distributed across the criteria with the intention of identifying any appreciable difference in the distribution as this could suggest a change in the nature of the construct being measured.

Table E5 shows the summary of the composition of the 2010 QSC Test by item type, basket, number of items and number of marks. It is effectively a Table of Specifications for the 2010 QCS Test.

TABLE E5
Composition of the 2010 QCS Test

Basket	MC marks	MC items	SR marks	SR items	WT marks	WT items	Total marks	Total items
α	56	28	25	4			81	32
β	38	19	27	3			65	22
θ	64	32	30	4			94	36
π	14	7	25	3	136 ⁶	1	175	11
Φ	28	14	30	4			58	18
Total	200	100	137	18	136	1	473	119

Table E5 shows that most of the marks (37%) are assessing the CCEs in the π (Create and Present) criterion. This is primarily due to the contribution of the WT.

Table E6 provides a cross-temporal summary of the composition of the QCS Test (balance and range of items and which criterion they primarily assess) for the years 1992 to 2010. A review of this table shows whether the blueprint (Table of Specifications) has changed. A significant change could indicate a change in the construct over time, as it would suggest that the items are measuring something different.

⁶ The Writing Task is always allocated to the π basket

TABLE E6
Composition of the QCS Test from 1992 to 2010

Year	α	β	θ	π	Φ
1992	19	13	20	36	13
1993	16	13	25	35	11
1994	15	15	22	37	10
1995	17	19	16	37	11
1996	15	16	22	33	14
1997	21	15	20	34	10
1998	14	15	22	33	10
1999	18	21	16	32	13
2000	16	17	23	33	11
2001	17	11	21	39	12
2002	12	18	23	33	14
2003	19	18	19	40	5
2004	17	17	23	36	7
2005	17	11	25	36	12
2006	17	13	19	39	12
2007	16	17	21	33	13
2008	18	15	21	34	13
2009	17	15	23	34	11
2010	17	14	20	37	12
Average	17	15	21	35	11
Standard deviation	2	3	2	2	2

It can be seen from Table E6 that the composition of the test has remained relatively stable over time. The criterion with the highest content weighting is π (Create and Present) every year. The second most frequently assessed criterion most years, particularly the last 12 years is θ (Analyse, Assess and Conclude). The criterion that has the least weighting in most years is Φ (Apply Techniques and Procedures).

These data strongly suggest that there has been no significant shift in the mix of criteria that constitute the variable and this adds support to the argument that the test is basically assessing the same construct as it was when it was first developed nearly 20 years ago. Even when there has been a curriculum change, because the criteria that comprise the construct are generic, they likely to remain embedded in the new curricula.

It also suggests that the items that have been included in the test cover a suitable range of the Common Core Elements and they represent a comprehensive range of item types.

The Review Team is confident that the test was well balanced when it was introduced, that it has retained this balance over time, and that the construct being measured by the 'balanced' test is well supported by the various jurisdictions as being appropriate for purpose.

4. Test and item functioning

This section considers the QCS Test requirements to 'be suitably difficult', 'discriminate among students', 'be without bias', 'make appropriate demands on the test-takers', 'have face and content validity' and 'present items within each paper so that their sequence assists students to demonstrate their optimum performance'.

Difficulty and discrimination (components 3 and 4 of Matters (1991)) have been included in this section because they interrelate at the level of an item.

Overall test difficulty

The first part of this section looks at the overall difficulty of the QCS Test.

Table E7 shows the QCS performance data for various groups of those OP-eligible students who were in Year 12 and completed all four papers of the QCS Tests in 2009 and 2010.

It can be seen from Table E7 that the mean for the total group was 117.8 out of a possible 235 (50.1%) in 2009 and 109.3 out of a possible 237 (46.1%) in 2010. These scores suggest that the tests are of average difficulty (which is the requirement for a scaling test) for the population of students. The spread of scores is relatively large which means that it is discriminating across the range of scores.

A second point worth noting from Table E7 is that males outperform females in all categories in both years, except for the written task and the related π (Create and Present) criterion. These differences are not statistically significant but they are persistent and consistent and should warrant further investigation. The main cause of these test statistics is likely to be the fact that the population of males and females is quite different. There are significantly fewer males meeting the OP-eligibility criteria than females. This would suggest that there is some self-selection occurring in the total population whereby the less able males are opting out of the OP program.

The next section of the review examines the item-level information for the QCS Tests.

Item statistics

Table E8 provides the item number, criterion, item facility (difficulty), and item discrimination (item-total correlation) for each of the multiple choice items that comprise the 2009 and 2010 QCS Test.

A Review of Aspects of the Queensland Core Skills Test - February 2012

TABLE E7

Means and standard deviations of the raw scores for OP-eligible students who were in Year 12 and completed all four papers of the 2007–2010 QCS Tests

Paper	2010			2009			2008			2007		
	All	Females	Males									
QCS Test	109.3 (26.5)	106.4 (25.7)	113.0 (27.1)	117.8 (27.3)	115.3 (27.3)	121.1 (27.1)	123.8 (23.9)	121.9 (23.6)	126.2 (23.9)	121.2 (27.6)	119.7 (27.3)	123.1 (27.9)
Multiple choice	52.0 (13.5)	49.9 (13.0)	54.8 (13.7)	53.7 (13.1)	51.9 (12.9)	56.1 (13.0)	54.6 (13.3)	52.8 (13.1)	56.9 (13.3)	55.0 (13.7)	53.3 (13.6)	57.2 (13.6)
Short response	27.6 (10.1)	26.7 (9.7)	28.8 (10.4)	34.6 (10.8)	33.3 (10.9)	36.2 (10.5)	31.2 (8.2)	30.7 (8.2)	31.9 (8.2)	35.0 (9.1)	34.4 (9.0)	35.8 (9.3)
Written	29.7 (7.0)	29.9 (6.9)	29.5 (7.1)	29.5 (8.1)	30.1 (7.9)	28.8 (8.4)	38.0 (6.2)	38.4 (6.0)	37.4 (6.4)	31.2 (9.2)	32.1 (8.9)	30.1 (9.4)
Criterion α	23.0 (5.9)	22.6 (5.9)	23.6 (5.9)	23.1 (6.0)	22.6 (6.1)	23.7 (5.9)	25.6 (5.5)	24.8 (5.5)	26.6 (5.3)	21.4 (5.3)	20.9 (5.2)	22.1 (5.3)
Criterion β	12.9 (4.4)	12.5 (4.1)	13.4 (4.6)	18.9 (4.9)	18.5 (4.9)	19.5 (4.8)	21.0 (5.0)	20.8 (4.9)	21.3 (5.1)	21.2 (5.5)	21.0 (5.4)	21.5 (5.5)
Criterion Φ	12.4 (6.5)	10.7 (5.9)	14.7 (6.5)	15.6 (5.5)	14.5 (5.5)	17.1 (5.1)	13.7 (5.9)	12.5 (5.6)	15.3 (5.8)	17.6 (5.9)	16.2 (5.5)	19.2 (5.9)
Criterion π	38.9 (9.0)	38.9 (8.9)	38.9 (9.1)	34.3 (9.1)	34.6 (8.8)	33.9 (9.5)	42.5 (6.7)	43.1 (6.5)	41.8 (7.0)	36.4 (9.7)	37.2 (9.4)	35.4 (9.9)
Criterion θ	22.0 (6.1)	21.7 (6.0)	22.4 (6.2)	25.9 (7.6)	25.2 (7.6)	26.9 (7.5)	20.9 (6.7)	20.7 (6.7)	21.2 (6.7)	24.6 (7.6)	24.4 (7.6)	24.8 (7.6)

TABLE E8
Summary of item statistics for the multiple choice items in 2009 and 2010

Item no.	2009			2010		
	Criterion	Item facility	Item discrimination	Criterion	Item facility	Item discrimination
1	α	0.81	0.23	α	0.65	0.30
2	θ	0.67	0.17	Φ	0.26	0.39
3	α	0.36	0.21	θ	0.54	0.33
4	β	0.46	0.32	α	0.57	0.31
5	β	0.50	0.23	Φ	0.43	0.35
6	β	0.30	0.14	Φ	0.53	0.24
7	β	0.81	0.27	α	0.53	0.19
8	β	0.88	0.20	θ	0.52	0.34
9	β	0.73	0.24	α	0.60	0.26
10	Φ	0.50	0.31	α	0.42	0.39
11	Φ	0.66	0.31	α	0.68	0.32
12	Φ	0.46	0.39	β	0.42	0.28
13	α	0.80	0.27	α	0.58	0.28
14	β	0.67	0.43	α	0.62	0.31
15	θ	0.26	0.24	θ	0.37	0.17
16	θ	0.42	0.12	Φ	0.59	0.30
17	β	0.55	0.32	β	0.47	0.28
18	θ	0.31	0.14	θ	0.29	0.26
19	α	0.63	0.16	β	0.55	0.30
20	Φ	0.72	0.32	θ	0.44	0.15
21	α	0.64	0.36	α	0.75	0.36
22	Φ	0.50	0.32	θ	0.66	0.27
23	θ	0.60	0.27	π	0.40	0.33
24	θ	0.26	0.07	α	0.52	0.26
25	φ	0.29	0.08	θ	0.33	0.20
26	α	0.79	0.31	β	0.74	0.27
27	α	0.50	0.27	θ	0.58	0.16
28	θ	0.35	0.20	θ	0.44	0.04
29	α	0.66	0.25	α	0.73	0.20
30	α	0.37	0.33	Φ	0.41	0.29
31	θ	0.40	0.30	Φ	0.42	0.38
32	α	0.86	0.26	θ	0.53	0.26
33	α	0.60	0.40	π	0.75	0.36
34	α	0.67	0.36	θ	0.40	0.28
35	β	0.51	0.34	α	0.70	0.17

continues

TABLE E8 (continued)
Summary of item statistics for the multiple choice items in 2009 and 2010

Item no.	2009			2010		
	Criterion	Item facility	Item discrimination	Criterion	Item facility	Item discrimination
36	θ	0.43	0.23	θ	0.24	0.14
37	α	0.52	0.34	β	0.30	0.22
38	α	0.59	0.33	β	0.32	0.17
39	θ	0.47	0.33	Φ	0.55	0.36
40	β	0.37	0.19	Φ	0.31	0.24
41	β	0.41	0.24	θ	0.32	0.39
42	α	0.29	0.30	β	0.23	0.10
43	π	0.25	0.19	θ	0.49	0.24
44	θ	0.56	0.20	θ	0.48	0.17
45	θ	0.33	0.20	θ	0.36	0.24
46	π	0.48	0.22	θ	0.44	0.34
47	θ	0.31	0.22	θ	0.29	0.14
48	π	0.70	0.29	θ	0.31	0.10
49	θ	0.35	0.26	β	0.34	-0.01
50	θ	0.44	0.24	β	0.40	0.17
51	θ	0.56	0.18	α	0.71	0.18
52	α	0.62	0.17	Φ	0.73	0.40
53	θ	0.44	0.13	Φ	0.69	0.46
54	π	0.61	0.16	Φ	0.53	0.45
55	β	0.63	0.27	θ	0.40	0.18
56	θ	0.84	0.34	α	0.76	0.21
57	α	0.74	0.25	β	0.69	0.29
58	α	0.71	0.22	π	0.54	0.23
59	α	0.67	0.34	β	0.52	0.31
60	α	0.59	0.21	θ	0.59	0.15
61	β	0.51	0.32	α	0.52	0.32
62	α	0.73	0.27	α	0.76	0.36
63	θ	0.40	0.17	α	0.63	0.39
64	θ	0.28	0.03	Φ	0.35	0.26
65	Φ	0.51	0.38	β	0.29	0.17
66	Φ	0.37	0.30	θ	0.73	0.30
67	Φ	0.49	0.45	α	0.56	0.19
68	θ	0.50	0.34	α	0.62	0.27
69	θ	0.49	0.35	θ	0.40	0.15
70	θ	0.37	0.19	θ	0.54	0.18
71	α	0.56	0.32	π	0.63	0.13
72	β	0.67	0.25	β	0.35	0.23
73	θ	0.62	0.21	θ	0.40	0.09
74	β	0.62	0.26	θ	0.50	0.42
75	β	0.36	0.38	α	0.78	0.39
76	Φ	0.74	0.30	α	0.59	0.32
77	θ	0.45	0.31	β	0.46	0.37
78	β	0.34	0.25	α	0.51	0.43

continues

TABLE E8 (continued)
Summary of item statistics for the multiple choice items in 2009 and 2010

Item no.	2009			2010		
	Criterion	Item facility	Item discrimination	Criterion	Item facility	Item discrimination
79	Φ	0.50	0.22	π	0.51	0.36
80	θ	0.34	0.25	α	0.65	0.39
81	α	0.45	0.36	α	0.45	0.32
82	θ	0.37	0.25	α	0.53	0.12
83	θ	0.37	0.18	β	0.62	0.39
84	π	0.60	0.28	β	0.46	0.37
85	β	0.35	0.32	θ	0.37	0.28
86	β	0.61	0.36	β	0.34	0.20
87	β	0.61	0.30	θ	0.72	0.39
88	β	0.38	0.30	Φ	0.51	0.51
89	θ	0.46	0.22	β	0.41	0.30
90	θ	0.49	0.31	θ	0.53	0.40
91	θ	0.54	0.24	π	0.45	0.25
92	π	0.71	0.31	β	0.36	0.30
93	θ	0.29	0.12	α	0.54	0.21
94	α	0.56	0.33	θ	0.42	0.19
95	β	0.47	0.25	θ	0.33	0.16
96	θ	0.48	0.29	α	0.48	0.41
97	β	0.43	0.29	α	0.57	0.36
98	Φ	0.47	0.24	Φ	0.55	0.42
99	α	0.47	0.25	α	0.45	0.24
100	θ	0.65	0.33	π	0.44	0.30

The purpose of the test is to place all students' scores onto one scale so that the scores can be added as the basis for the calculation of an OP. The scaling test should effectively spread the scores along the measurement scale. In order to maximise spread, the intention is to have multiple choice items that have a facility of 0.5 as this maximises the spread of the cohort across the scale. It is also important that the items have a high discriminating power. The point biserial correlation is used to give an indication of the item discrimination. Items with a point biserial value less than 0.2 would be considered to be of dubious value in a test that requires the items to discriminate so that the resulting distribution is appropriate for scaling. While the point biserial values in Table E8 cannot be determined until after the test, the items that are used in the test are trialled and those items that are problematic in terms of their discriminating power would generally be reviewed with a view to improving the discrimination power of the item before it goes into the final QCS Test.

Table E8 shows the item facilities (difficulties), point biserial correlations, and criterion measure by each of the multiple choice items in 2009 and 2010. Items that have point biserial values less than 0.2 have been highlighted. It can be seen that in 2009 there are 18 items (18 percent of the multiple choice items) with point biserial values less than 0.2, and in 2010 there are 25 items (25 percent of the multiple choice items) with relatively low point biserial correlations. In 2010 there is one item with a negative point biserial correlation (item 49). While the point biserial correlation is not independent of the difficulty of the items, the effect is most noticeable with items that are very hard or very easy for the cohort of students and there are very few items of this type in the tests.

Ideally the test developers should be working towards minimising the number of items with point biserial values less than 0.2. It would be worthwhile for the QSA to

A Review of Aspects of the Queensland Core Skills Test - 31 January

investigate whether the point biserial values that are obtained in trialling are providing the test developers with information that is pertinent to identifying those items that fail to discriminate effectively in the final tests. If the trialling is not identifying such items because of differences between the trialling population and the test population, the value of trialling is limited.

A second point that can be noted from Table E8 is that the majority of the items have facilities centred around 0.5. This is encouraging. In fact the mean of the facilities for 2009 is 0.52 and the spread of the facilities around this mean is 0.15. If the assumption were made that the distribution of item facilities should be normally distributed around a mean facility of 0.5 then there would be an expectation that 68% of the items would lie between 0.35 and 0.65 (0.5 ± 0.15) with 17% of the items having values below 0.35 and the same percentage above 0.65. As it stands, 65% of items lie within the band defined by the 0.35 to 0.65 item facilities, 22% of the items have facilities higher than 0.65 and 13% have facility values less than 0.35. In 2010, the mean of the facilities is 0.50 and the spread of the facilities around this mean is 0.14. This distribution is very close to being normal with 68% of the item facilities within 0.5 ± 0.14 . This is an excellent result and shows that the multiple choice items are extremely well targeted for the cohort of students and consequently this component of the test is more than meeting the test development requirements for the scaling purpose of the QCS Test.

A minor issue regarding test development that might be considered by the test developers is to construct the test in a way that minimises, as far as possible, dependencies among items. Dependencies can sometimes result from the ordering of items in the tests. If, for example, there are difficult items or, in the case of the SR paper more difficult units, at the start of the test, the less able students might struggle with them and lose confidence immediately. As a consequence, later they might be unable to correctly answer items that are well within their ability level. Sound test practice generally suggests that when composing tests it is better to have the easier items first in the test.

Unless there are compelling reasons to do otherwise, when arranging the units within the SR paper, the easier units should be placed first. Then the easier items within the unit would be placed first where it makes sense to do so.

Table E9 shows the unit number, item number, maximum mark, criterion, item mean, and facility for each of the short response items and units that comprised the 2009 and 2010 QCS Tests.

An analysis of the facility values for the 2009 QCS Test shows that the order of units from easiest to hardest is 4, 1, 5, 2, 3, 6, 7 and 8. Of course these are only ranks and the value of the differences between the means may be relatively small. This observation is not meant to be a criticism but rather an aspect of test construction that should be explored with a view to seeing if the test can be improved by more closely ordering the units according to their relative difficulty.

It is acknowledged that the location of the items (units) in the test has to be done with the 'trial data'. If these data are not producing estimates that are stable enough to order the units appropriately in the test then perhaps the usefulness and quality of the results from the trialling data items need to be evaluated.

It also can be seen from Table E9 that the units in 2010 were appreciably harder than in 2009 (assuming there hadn't been any significant change in the distribution of abilities in the cohorts). In 2010 only 3 of the 18 short response items had item means above 50%; whereas in 2009, 9 of the 17 items had means above 50%. Comparisons to the earlier years also suggest that the short response items of 2010 were particularly difficult. It is encouraging that the items in 2010 are fairly well ordered within a unit in terms of their relative difficulties. The order of units from easiest to hardest for 2010 is 2, 1, 3, 6, 8, 4, 7 and 5.

TABLE E9
Summary of mean scores for short response items for 2009 and 2010

2009						2010					
Unit	Item no.	Max. mark	Crit.	Mean mark	Facility ⁷	Unit	Item no.	Max. mark	Crit.	Mean mark	Facility ⁷
Unit 1	1	8	β	5.11	0.64	Unit 1	1	5	α	2.23	0.45
							2	4	θ	1.75	0.44
Unit 2	2	5	Φ	3.51	0.70	Unit 2	3	8	β	4.14	0.52
	3	8	Φ	5.29	0.66		4	7	α	2.61	0.37
Unit 3	4	9	θ	3.92	0.43	Unit 3	5	8	θ	4.13	0.52
	5	8	θ	5.25	0.66		6	5	α	2.11	0.42
Unit 4	6	8	θ	4.02	0.50	Unit 4	7	8	θ	2.71	0.34
	7	8	α	4.83	0.69		8	8	Φ	3.26	0.41
Unit 5	8	5	Φ	3.90	0.78	Unit 5	9	10	Φ	1.45	0.14
	9	6	α	2.92	0.49		10	7	π	3.23	0.46
Unit 6	10	8	Φ	4.74	0.59	Unit 6	11	8	β	3.14	0.39
	11	9	β	4.28	0.48		12	10	θ	4.52	0.45
Unit 7	12	6	α	3.60	0.60	Unit 7	13	4	Φ	2.34	0.58
	13	7	θ	3.27	0.47		14	8	Φ	2.84	0.36
Unit 8	14	12	π	2.39	0.20	Unit 7	15	7	π	3.09	0.44
	15	8	β	2.60	0.32		16	11	β	0.99	0.10
Unit 8	16	9	α	2.63	0.29	Unit 8	17	8	α	3.85	0.48
	17	10	θ	3.27	0.33		18	11	π	3.88	0.35
Overall		134		69.7	0.52			137		55.2	0.40

⁷ Facility = mean/ maximum mark

APPENDIX F: The scaling procedure

1. Purpose of scaling

The purpose of scaling is to place all students' scores onto one scale so that they can be added as the basis for the calculation of an Overall Position (OP) or Field Position (FP). In Queensland, assessment is school-based so a common test is required to provide a basis for producing scores on a single scale.

To be viable as an instrument for placing all subject assessments across all schools on a common scale the QCS Test must provide valid estimates of:

- the relative performance of, and spread of, subject groups within a school
- the relative performance of students across schools.

In addition, the impact of scaling against the QCS Test should be independent of:

- subject choice and therefore have minimal effect on the curriculum choice of students
- the school attended and student characteristics including gender and ethnicity.

2. Overall Position (OP) and Field Positions (FPs)

OPs and FPs are statewide rank orders. OPs are a measure of overall achievement in senior secondary school studies and give equal weight to all subjects. FPs are statewide rankings of OP-eligible students in specified fields of study.

There are five fields as follows.

- A extended written expression involving complex analysis and synthesis
- B short written communication involving reading comprehension, pictorial interpretation and English expression
- C basic numeracy
- D solving complex mathematical problems involving symbols and abstractions
- E substantial practical performance involving physical and creative arts or expressive skills

Scaling for OPs uses scores derived from total test scores. Scaling for FPs uses scores derived from subsets of items in the QCS Test.

3. Calculating OPs and FPs

There are three broad stages in the calculation of OPs and FPs:

1. Schools provide for each student a set of Subject Achievement Indicators (SAIs), which are numbers lying between 200 and 400 that indicate students' relative achievements in a subject group. For large schools at least one student must get 200 and one must get 400; other students' SAIs are distributed between these two numbers as determined by teacher judgement.

In each school SAIs are scaled using scores that are aggregates of QSC Test item scores, suitably weighted using item weights. There are six scaling scores: one for the derivation of an Overall Achievement Indicator (OAI) and the remaining five for the derivation of Field Achievement Indicators (FAIs).

A Review of Aspects of the Queensland Core Skills Test - 31 January

The first scaling score is the total score on the QCS Test but in the calculation of the remaining scaling scores the items are weighted according to their importance in relation to the five fields. The calculation of these weights (item weights) will be described in a following section.

The SAIs are scaled using subject-group scaling parameters, the weighted mean and mean difference, derived from the distribution of scaling scores of the subject candidature.

2. For each student the best SAIs are averaged to yield measures of the student's OAI and their five Field Achievement Indicators (FAIs). Each of these indicators is a weighted average: the OAI is the average of the best scaled SAIs to give a weight of 100, and the FAIs are weighted averages of the best SAIs to give a total weight of 60. For the OAI the weights are the number of semester units for which the subjects have been studied (unit of credit), whereas for the FAIs the weights also take into account the importance of the subjects in relation to the five fields. The calculation of these weights (Field weights) will be discussed in a following section.
3. The OAIs and FAIs are then scaled using school-group scaling parameters to give measures that are comparable across schools. The scaling parameters are the weighted mean and mean difference, with QCS Test weights (isoweights) determined by the relationship of each student's overall within-school measure (WSM) and their QCS Test scores.
4. These measures are then banded into OPs and FPs: 25 bands for OPs and 10 bands for FPs.

Of necessity, this description of the scaling procedure is simplistic and neglects the variations due to size of school and subject candidatures. These are well documented in the QSA's documents and reports.

4. Weights

There are three sets of weights: item field weights, field weights and isoweights.

Item field weights

As noted above, there are six scaling scores, each of which is a weighted sum of the scores on the QSC Test items. The weights are termed item weights.

The purpose of the FPs is to assist the processing of applications in different faculties by having measures based on quantitative and verbal skills in addition to the measure of overall achievement. The FPs should then be as independent as possible to discriminate between applicants on different dimensions. Five sets of item weights are determined so that the scaling scores used to derive the corresponding FAIs should be as independent as possible.

QCS Test items are first allocated to fields according to the cognitive demand of the items and the skill sets required for the subjects with the highest field weights for specific fields.

Item weights are then determined from the results of principal components analysis of students' results on the items. The first principal component represents performance in the overall test, and the second component represents the difference between the verbal and quantitative attributes.

Items allocated to fields A to D are determined using the plot of the loadings on the second principal component versus the loadings on the third principal component.

A Review of Aspects of the Queensland Core Skills Test - 31 January

The weight of an item is the length of the projection of the position of the item on the line $y = x$, or $y = -x$, whichever is appropriate.

The weights for the item for Field E are determined by the nature of the item.

The following table shows that the correlations between the scaling scores for the five fields are as small as would be expected from a test of the nature of the QCS Test. Fields A and B are focused on verbal skills, while fields C and D are focused on quantitative skills. The observed correlations between A and B, and between C and D are higher than the others (excluding Field E).

TABLE F1
Correlations between the fields of the 2010 QCS test.

	B	C	D	E	QCS
A	0.6356	0.5043	0.3713	0.7566	0.8043
B	1.0000	0.5410	0.4738	0.7594	0.7809
C	0.5410	1.0000	0.6723	0.7847	0.8296
D	0.4738	0.6723	1.0000	0.6726	0.6966
E	0.7594	0.7847	0.6726	1.0000	0.9483

While the Review Team believes that there is obvious merit in this approach, it is unnecessarily complicated. The weights for fields A, B, C and D could be found more directly by a confirmatory factor analysis based on the design of the test.

Field weights

In the calculation of Field Achievement Indicators each subject is weighted for every field. Subjects that have a high weight for a field are those that have more emphasis on assessment in the specified skill area defined by that field. Table 4 provides an example.

TABLE F2
Field weight example

	OP	A	B	C	D	E
English	5	5	3	1	0	4
Mathematics B	5	1	2	5	5	1
Biology	5	5	4	3	3	3
Modern History	5	5	4	2	0	2

Isoweights

Once the six scaling scores have been determined, scaling parameters are calculated for the OAI and FAIs. SAs for each subject group are scaled to the mean and mean-difference (a measure of spread) of the QCS Test scaling scores for that subject group.

In the calculation of the subject group means and mean differences, the QCS Test scores of individual students are weighted according to the difference between their QCS Test score and their overall within-school ranking. Students whose ranking on the basis of the QCS Test is significantly different from their ranking on the within school measure (WSM) will have their scores down-weighted. The weights applied to student scores are termed isoweights.

Calculation of the WSM is based on *The ranking from unbalanced paired-comparison data*, a paper by H.A. David (1987). The calculation of WSMs for OP-eligible students consists of comparing the subject achievements of students at the

A Review of Aspects of the Queensland Core Skills Test - 31 January

school, keeping a tally of the number of times a student 'wins' or 'loses' these comparisons.

$$WSM = (W_1 + W_2) - (L_1 + L_2),$$

where

- W_1 is the number of wins of the student
- W_2 is the total wins of persons beaten by the student
- L_1 is the number of losses of the student
- L_2 is the total losses by persons to whom the student lost.

The distributions of WSMs are transformed to the same mean and mean difference as the school's distribution of QCS Test scores for OP-eligible students. A linear transformation is used for this purpose.

The weight given to a student's QCS Test score in the calculation of school group and subject group scaling parameters is based on the magnitude of the difference between their QCS Test score and scaled WSM.

Polyscores

Student achievement is also reported against a set of grades: Very High Achievement, High Achievement, Sound Achievement, Limited Achievement and Very Limited Achievement. As a quality control measure the QSA derives a score from the grades awarded within a school and compares these scores (polyscores) with the other within school measures of achievement.

Derivation of the polyscores involves two steps. The first is to derive a set of average scaled percentiles for students in the school and the second is to transform these percentiles so that they have the same mean and mean difference as the scaled OAI for OP-eligible students.

Deriving an average percentile is based on a method known as polyweighting whereby, for each subject, expected percentiles are calculated for each of the five grades (Sympson, J.B & Haladyna, T.M. 1988, *An evaluation of 'polyweighting' in domain-referenced testing*). These percentiles are then scaled, taking into account the difficulties of obtaining high grades across subjects. For each student the average scaled percentile is calculated.

The distributions of polyscores, WSMs and OAIs are compared to identify anomalous results, and adjustments made if necessary.