A design process for constructing the Queensland Core Skills Test

G N Matters
A design process for constructing the Queensland Core Skills Test

G N Matters
Computer Page Formats and Typesetting
Chris Cumming, Diane Delaney, Tina Poole

This document was produced "inhouse" on Unisys and Hewlett Packard computer equipment.
Printing and Binding: Watson Ferguson & Co.
FOREWORD
In August 1991, following the Queensland Government's acceptance of the major recommendations of *The Review of Tertiary Entrance in Queensland 1990* (The Viviani Report), this Board was charged with the development and administration of a core skills test.

The Queensland Core Skills (QCS) Test will replace the Common Scaling Test (Australian Scholastic Aptitude Test plus the Writing Task) as a device for scaling school-based assessments. The last Tertiary Entrance Score will be issued in December 1991. As from 1992 Overall Position and up to five Field Positions will be used for university selection.

Unlike the Australian Scholastic Aptitude Test (ASAT) component of the Common Scaling Test (CST) before it, the QCS Test will become publicly available after its administration.

The QCS Test also has the function of certifying individual achievement. Student results on the QCS Test will be reported on a five-point scale.

The test will be administered to Queensland Year 12 students for the first time in September 1992.

This paper presents a view of test construction as a design process, one with evaluation of both the process and the product as an integral part from the beginning. It is not enough to attempt to produce a good test - quality as well as public accountability requires that decisions be recorded, reported and assessed.

An important feature of the design process is the involvement of teachers both before and after the test, in evaluating items and reacting to the testpaper.

The design process, and indeed the actual preparation of the QCS Test, is managed by Gabrielle Matters, Associate Director (Tests and Examinations). The substantial progress made to date is due largely to her care, industry and enthusiasm for the project.

Gabrielle presented this paper to the Board's Core Skills Steering Committee at its July 1991 meeting.

There are five supplementary papers. These appear on buff paper. It is suggested that, for continuity and coherence, they be read at the time they are referred to in the main paper.

John A. Pitman
Director
ACKNOWLEDGMENTS
For ideas, information and/or inspiration, the author gratefully acknowledges:

Dr J (Reg) Allen  
Board of Senior Secondary School Studies  
Brisbane, Queensland

Dr HPJ Kreeft & Henk Moelands  
Examinations Department, Cito  
(National Institute for Educational Measurement)  
Arnhem, The Netherlands

Doug McCurry  
Australian Council for Educational Research Ltd  
Melbourne, Victoria

John O'Brien  
Board of Senior Secondary School Studies  
Brisbane, Queensland

Dr Günter Trost  
Institute for Test Development & Talent Research  
Bonn, Germany

Professor Ingemar Wedman & SweSAT program team  
Umeå University  
Umeå, Sweden

Alison Wolf  
Institute of Education, University of London  
London, UK

Test Development Unit  
Board of Senior Secondary School Studies  
Brisbane, Queensland
Table of Contents

FOREWORD i
ACKNOWLEDGMENTS ii
INTRODUCTION 1

THE DESIGN MODEL 1

1. DESIGN CRITERIA 2
   1.1 Integration 2
   1.2 Balance and range 3
   1.3 Difficulty 4
   1.4 Discrimination 5
   1.5 Bias 5
   1.6 Demands 5
   1.7 Face and content validity 6
   1.8 Sequence 7

2. APPLYING THE CRITERIA 8
   2.1 Item characteristics 8
   2.2 Quality control 9
   2.3 Decisions 10

3. EVALUATION 10

CONCLUSION 11

SUPPLEMENTARY PAPERS
One - Established decisions about the test 13
Two - Acquiring an item bank 17
Three - Proposed list of attributes 27
Four - Timeline for construction and evaluation of sample test 31
Five - Layout of units and items 33

REFERENCES 39
INTRODUCTION
Constructing a test involves more than merely selecting some attractive test items. It is necessary to make decisions about what is to be tested, about what material can be used in the items, about the criteria for selecting and sequencing the items, and about the procedures to be used in assembling the test.

One view of test construction sees it as production by working through a set of largely mechanical rules. Another view, and the view urged in this paper, sees the task of constructing a test as a process of design. Furthermore, because any result of a design process is a somewhat imperfect realisation of the intentions of the designers, it is appropriate to include evaluation both of the process and of its outcomes as a key component of the process itself.

This paper, therefore, sets out a design process intended to produce an authentic, integrated, balanced, and high-quality test covering the range of curriculum elements while at the same time providing appropriate differentiation. The design process includes an evaluation component.

A number of decisions relating to features of the Queensland Core Skills Test (QCS Test) have already been established. These appear in a consolidated form in Supplementary Paper One (page 13).

This discussion paper assumes that a bank of items already exists. Supplementary Paper Two (page 17) documents the process that has resulted in an initial bank of items.

As a result of previous work done at the Board, 49 common curriculum elements have been identified. While the test will be audited against the list of these 49 elements, it has been agreed that it is useful to group them under headings entitled 'attributes'. There are three purposes for doing so. First, the attributes provide a shorthand description of the domain of the test. Second, they provide a set of criteria each of which can be elaborated into standards to yield a practicable set of criteria and standards for defining the final grades. Third, they provide an additional means of assessing the range and balance of the set of items that are included in any particular test. Supplementary Paper Three (page 27) gives a proposed list of 13 attributes, in addition to the list of the 49 curriculum elements.

THE DESIGN MODEL
The proposed construction model involves five steps:

- listing the design criteria
- constructing the test using these criteria, keeping records of decisions (and their rationale) involved in the application of the criteria
- evaluating the adequacy of the criteria and the effects of decisions made along the way
- publishing this evaluation
- adjusting the application of the design criteria.
It is suggested that the production and publication of a sample test, due in February 1992, can and should be used as a trial of much of the complete design model. Furthermore, it seems that it may be possible to arrange for a small number of students to take the test and to be, along with their teachers, part of the evaluation of the design model. Supplementary Paper Four (page 31) gives a proposed timeline for the construction and evaluation of a sample test.

1. DESIGN CRITERIA
The items should be selected such that the test

- is an integrated whole
- is balanced and includes a suitable range of items
- is suitably difficult
- discriminates
- is without bias
- makes appropriate demands on the test takers
- has face and content validity.

The items should be arranged within each subtest so that their sequence assists students to demonstrate their optimum performance.

1.1 Integration
An integrated test is one which 'holds together'. It is one where the items are chosen by the test constructors so that the test, as a whole, reflects a suitable range and balance of the 49 curriculum elements. Suitability is, in turn, defined in terms of student experience (of each subtest) and in terms of the properties of the (total) test as measured by psychometric analysis.

Questions that could arise in regard to student experience are:
- Does the whole test capture a large proportion of things I do at school?
- Is the subtest that I am now doing a coherent whole?
- Is the item of the subtest that I am doing now recognisably part of the subtest?

Student experience is discussed in detail later in 1.6.

Statements that could be made in regard to properties of the total test as measured by psychometrics are:
- Items are not merely replications of one item.
- The test is internally consistent.
  (Measure coefficient alpha.)
- It is sensible to extract a single result from the test.
  (Do principal components analysis of measures of association of items.)

Psychometric measures for any particular test can only be 'estimated' before its administration. It is anticipated that the analyses of trials will be a guide to such measures.
1.2 Balance and range

*Balance* refers to the relative proportions of the various dimensions listed below. *Range* refers to the span of the sample space.

While range implies coverage of the various dimensions, balance implies the relative importance within this coverage. The relative importance or balance can be gauged from the frequency of occurrence.

Balance and range can be assessed in terms of the following nine *dimensions*:

i. curriculum elements
ii. attributes
iii. epistemic content
iv. text of stimulus material
v. length of stimulus material
vi. context
vii. type of response
viii. difficulty
ix. number of grades for marking student responses (essentially applies only to grading of Short Response Items).

Each item can be classified in terms of each of these nine dimensions as follows:

i. up to three curriculum elements - showing primary/secondary/tertiary emphasis
ii. up to three attributes
iii. one epistemic area - using four of Phenix's (1964) six-fold division (symbolics/empirics/aesthetics/synoptics)
iv. one of four types for the text of the stimulus material - numerical/verbal/tabular-graphical/visual
v. one of three lengths - short/medium/long, together with number of items per unit of stimulus material
vi. one difference of context - everyday/esoteric
vii. one of three types of response activity -
   - free-active
   - constrained-active
   - constrained-passive
   plus at least one of three types of response mode -
   - verbal
   - numerical
   - tabular-graphical
   - visual
or
   - open (i.e. not stipulated)

viii. one of three difficulties - easy/medium/hard
ix. the number of grades for marking student responses - 5/4/3/2.

NB: vii and ix are relevant for SRIs only.
For each of the nine dimensions it is necessary to determine, on a policy basis, the desired balance.

It is anticipated that as the test is constructed a simple frequency tabulation (or matrix) using the item characteristics in terms of these nine dimensions will be made. This matrix will allow progressive monitoring of the balance and the range.

1.3 Difficulty
While using this model, difficulty can be assessed before the use of the test in terms of

- the judgment of intrinsic difficulty for students at the senior level [a subjective (easy/medium/hard) rating by a panel of teachers]
- the distribution of students' response grades
- IRT analysis of difficulty level. [estimated from trialling]

An alternative to setting items which have a particular difficulty is to set the so-called 'questions of neutral difficulty' (Good 1989, p.73). Such questions (i.e. items) which are necessarily of the short answer and extended writing type, set tasks which can be completed by most students but at different levels and are marked by assigning responses of individual students to distinct levels. The word 'level' in this context does not refer to the grade of response as assigned by the marker, but rather to the category (such as advanced or ordinary) a student 'enters' for being tested.

Even so, there are serious difficulties in constructing and applying an appropriate marking scheme. Experience in the United Kingdom of measured reliability when using this model suggests some loss of marking accuracy.

One significant problem associated with test items of neutral difficulty seems to lie in ensuring that students from both ends of the ability range respond at the highest level of the marking scheme of which they are capable.

Good's (1989, p.81) belief that 'provided the mark allocations reflect, at least approximately, the difficulty of the questions and enough time is allowed, appropriate grades can be awarded, even if the most and least able are not given sufficient opportunities to demonstrate the full extent of their abilities and achievements' does not seem to promise enough to commend the inclusion of this sort of item in the 1992 QCS Test. Profound difficulties would arise in appropriately entering students for that category of the test designed to test their particular achievement level.

Supplementary Paper Five (page 33) presents a sample of the layout for short response items on the QCS Test. Instruction blocks are attached as marginalia to each item to cue students as to what is required of them. Spaces are provided to indicate the length of an appropriate response. Hence the student is aware of the length and the nature of the task, if not its complexity.
1.4 Discrimination
The discrimination of the test can be assessed in terms of:

- item discrimination: point bi-serial, bi-serial, IRT analysis
- distribution of final scaling scores.

The former information about the items is obtained from analysis of trial data. The latter information about the test is what we require. Obviously, the distribution of final scaling scores cannot be obtained until after the test is marked. The discrimination of the test can only be estimated during the construction process. This estimate, or prediction, will come from the item discriminations and difficulty levels of the items chosen for inclusion in the test.

Also, should an item exhibit low discrimination or low level of difficulty when trialled, it will not necessarily be excluded from the item bank or from the test. The item might still be included in the test in order to satisfy other criteria such as face validity. Viviani (1990, p.44) dictates that the QCS Test aim 'to test basic English expression and numeracy'. Items designed to do this may be neither difficult nor discriminating, but nevertheless do have a place in the test.

1.5 Bias
Bias can be assessed in terms of:

- expert insight into the individual items
- the analysis of student performances over the whole test.

Analysis of test performances cannot be made until after the test. In any case, the test is a test of achievement, not of aptitude, and the assessment of bias is fraught with problematic assumptions.

It is proposed that items be scrutinised by expert panels in an attempt to remove biases which could be possibly inherent in the stimulus material or in the language used within the items.

The sensitivity review procedures are described in the paper Towards Equity in the Queensland Core Skills Test - The Queensland Sensitivity Review Process (O'Connor and Robotham, 1991).

1.6 Demands
The test seeks to provide students with an opportunity to demonstrate their optimal performance in what the test declares itself to be testing. A student's 'test-wiseness' is not, therefore, something which should be overtly tested.

It is useful to consider the demands placed on students by the test by posing the questions students might reasonably ask during the test:

- Are there adequate opportunities for me to show what I understand and what I can do? (for both the most and the least able)
- Am I able to choose the mode of expression for my response on some
occasions?

• Am I able to vary the mode of expression for my response and still be adequately rewarded?

• Do I have the opportunity to use the skills I have learned in my schooling? (both cognitive, and by employing my 'tools of the trade')

• Are the cues appropriate?
  . Thirty thousand students should have a clear idea of what they are supposed to do, not only for the test as a whole, but also and for each and every item (this is not to say that their idea of what to do is always correct, it is to say that it should be clear).
  . Students must understand what are reasonable responses. For example, students should be told to 'write exhaustively' if that is the intention; not have to think in 'secret' testing language as to what their response should be.

• Is the meaning of the item clear?

• Have I experienced this type of item before?

• Am I aware of the value/worth of an item?

• Is the layout helpful?

• Am I able to organise my time (even though the test is unspeeded)?

• Are the things referred to in the stimulus material accessible in terms of my background: gender, culture, geography etc.?

• Are the questions accessible (to the full achievement range of me and my peers) thus:
  - was the expression clear?
  - were many words too difficult?
  - were sentence structures sometimes over-complicated?
  - were unfamiliar technical words left unglossed?
  - was the stimulus material wordier than is desirable?

• Are some apparently simple problems really simple?

• Are some problems merely simple arithmetic in disguise?

• Do the items seem authentic?

• Are there indications of the relative importance of items in terms of the proportion of my effort I should invest in them?

The last consideration is of considerable importance in a speeded test, where there is insufficient time for most students to complete attempts at all items. The QCS Test is not intended to be a speeded test, although the SRI subtest may need special attention in this area.

1.7 Face validity and content validity

The term 'face validity' implies that a test should, in addition to having pragmatic or statistical validity, appear pertinent and related to the purpose of the test as well. That is, it should not only be valid, it should also appear to be valid (Mosier, 1947).

Anastasi (1966) and others emphasise two things about face validity: (i) face validity should be separated from, and not confused with, criterion-related validity, content validity or construct validity; and (ii) although not measurable by the analysis of test data, face validity is a very important feature of any test.
It is further argued that a test with high face validity has a better chance of inducing cooperation and positive motivation among students, of reducing dissatisfaction and feelings of injustice among low scorers, of convincing policymakers, employers and administrators to accept the implementation of the test, and of improving public relations with the mass media and the courts. This list certainly translates to the Queensland testing scene.

Face validity can be gauged by interviewing a sample of test-takers. A more direct approach involves using some willing students, teachers and the other users of the test to assess face validity of the test as a whole as well as the validity of particular items. Quantitative analysis of these ratings provides an indication of the agreement among the raters and their relationships with other aspects of the test and item analysis.

Often, what appears valid is what people have been conditioned to expect. It seems essential, therefore, that those who are asked about face validity (students, teachers, users of test results) are properly informed before the event. After all, face validity is no more than a belief or an opinion.

The dictates of security limit what can be done with any item in advance of its use. However, the publication of a sample test will permit such an analysis of its face validity and a consequent guide for the construction of the main test.

It is proposed that:

▪ there be published a comprehensive Student Information Bulletin
▪ there be systematic post-test studies of face validity
▪ these studies be published
▪ they be used as a guide in test construction.

Content validity at the test construction stage can be seen as part of the quality control of items. This is discussed further in 2.2.

1.8 Sequence
While the sequence in which students attempt a paper cannot be controlled, it is possible to make judgments about a suitable sequence in terms of

▪ the smoothness of the transition from section to section of the paper
▪ the order of items within a section.

When items are arranged on an incline of difficulty, it is assumed that the less able students will succeed with the earlier items and the more able will answer all the paper, taking opportunities to demonstrate positive achievement by attempting the more difficult questions or parts. This is in-paper test-wiseness.

Otherwise, it is possible that the most able students might include more information in their responses than is necessary to achieve the highest grade, or might misinterpret items to make them more complex than intended by the test constructor. The problems could conceivably be more acute at the other (lower)
end of the ability range with students gaining their marks in a piecemeal fashion.

The method of sequencing, by arranging items on an incline of difficulty, assumes that the same incline of difficulty suits all students, and that it is possible to write and arrange items to conform accurately with such an hierarchical order. (It is of course assumed that item statistics obtained from trialling and/or teacher judgments from panelling are available when questions are arranged on an incline of difficulty.)

Items on the MCQ and SRI subtests could be differentiated in terms of their difficulty (easy, medium, hard), their epistemic content (four categories which go beyond the traditional verbal/quantitative split) and the format or type of response (as in 1.2. vii).

It is suggested therefore that, within a subtest, units be presented in the sequence easy verbal, easy quantitative; medium verbal, medium quantitative; hard verbal, hard quantitative, using as the intrinsic difficulty levels those obtained from teacher judgment. The verbal/quantitative split is a rough *a priori* classification. Students should be informed of this presentation in the *Student Information Bulletin*.

2. APPLYING THE CRITERIA
Applying the design criteria discussed in Section 1 requires that:

- there be a systematic body of information about each item: item characteristics
- there be extensive quality control
- in assembling the test the decisions made along the way in terms of the criteria be both deliberative and recorded.

2.1 Item characteristics
The characteristics as known or estimated for each item should be entered into a computer database so that systematic searches and selections may be made. To parallel the computer records, a card system should also be designed to capture the item characteristics and the history of the item.

Item characteristics relate to the nine dimensions listed in 1.2 plus a 'short name' for easy identification at the test construction stage and an estimate of the time required for a student’s response.

The latter estimate is a necessity at the test construction stage, the aim being to produce an unspeeded test.

The history of the item captures the following: panelling information (date and type of panel - inhouse, expert, equity, language etc.); trialling information (population, date, item analysis); author of item; and details of public use (date, test, reactions etc.).
2.2 Quality control

Quality control is related to successfully meeting the documented specifications for the QCS Test. When the issue of quality control is explored, three themes recur:
- items
- production
- security.

Quality control procedures should ensure that no item on the test is irrelevant, 'silly', ambiguous, obviously biased, or at an inappropriate level of assumed knowledge (particularly with respect to vocabulary or mathematical operations; vide Test Specifications attached to Supplementary Paper Two on page 19.)

Quality control through expert panelling provides a form of content validation. Content validation seeks to determine the extent to which items test the features that they appear to test and are intended to test.

Guidelines issued to panellists relate to: technical aspects (content validity, completeness or clarity, marking schemes, time and length); subject matter aspects (technical terms, degree of difficulty, ambiguity, educational soundness); bias (gender, ethnic, socioeconomic, geographical etc.); editorial and linguistic aspects.

The multiple-choice distracters should be appropriate, the 'best' answer being the key; the Writing Task stimulus material should ensure that all students are able to respond in the genre of their choice; the Short Response Item paper should contain appropriate cues and should allow the use of the skills that students have learned throughout their (senior) schooling.

Item analysis (both statistical and qualitative) from trialling should result in the discarding of items which do not 'work' or in the revising of items until they do. And ultimately the suitability of an item should be rationalised in terms of the 'wholeness of knowledge', a characteristic defined in conversation with Graeme Withers at ACER in 1986.

Once the test items are, after feedback from the many people involved in the panelling and trialling 'loops', deemed acceptable, then the total test is constructed with the desired balance along the nine dimensions as described in 1.2. Mechanisms are set in place to control the production line so that an accurate rendition of the test as constructed is printed.

During and after printing of the test, and in packing and despatching the papers to schools, strict procedures are enforced in order to maintain security on the test.

The essence of quality control procedures is three-fold:

- repetitive checks at every stage of the process (loops)
- not having people check their own work (i.e. as the sole check); but having independent checks at all stages.
- sharing information about the test on a 'need-to-know' basis; and having all persons involved sign declarations related to confidentiality and security.
2.3 Decisions
The process of test construction involves a step-wise refinement. The model proposes that at each stage, decisions - to select, to reject, to rearrange, to compromise - should be made explicitly in terms of the criteria as listed above and should be recorded as such. While this may seem to slow down the process it is suggested that it provides for accountability and for the need to improve, since the quality of the decision-making can be evaluated and reported.

3. EVALUATION
Often test evaluation is in terms of the psychometric properties of the test and its results - a product emphasis. That is, the evaluation asks the following questions:
(a) does this test meet its specifications (does it require modification)? (b) do the test takers know the required material (should they take the course again because they 'failed')? For the QCS Test, the test as taken is the one and only for the student so these questions are essentially irrelevant. The most useful evaluation is that which informs the matters which can be changed: the processes which generate the next test. After all, the specifications are a given (a), and the students are not in a Queensland-in-the-sixties mode of 'pass/fail' (b).

It is proposed that resources be allocated so that a report on the construction of the test can be available within four to six weeks of the release of the sample test.
It is proposed that this immediate evaluation use:
. interviews with students about their experience of the test
. interviews with teachers about aspects of the test
. face validity studies
. content validity reports
. an expert independent, external evaluator
to evaluate
. the adequacy of the criteria used in the construction of the test
. the accuracy of the decisions made in terms of these criteria.

It is also proposed that if possible there be a later report as the result of an evaluation using
. marking of a sample of student scripts
. statistical item and test analysis.

Construct validity studies are necessarily a longer-term consideration.

It should be noted that the primary reason for composing a sample test for schools is to present students with a familiarisation exercise, not a trial test.

However, public concern may be triggered by the viewing of the sample test. This concern could increase if not allayed. The reports should be reasonably comprehensive so that they address both surface concerns - what was the right answer - and substantive issues - was this a good test.

It is envisaged that the reports be presented to the Board. Should their contents not amount to a 'seal of approval', the public should be informed of any proposed
changes in item characteristics, selection and sequencing - any changes that impact on students.

CONCLUSION
The design model proposed in this paper presents a method of developing the new QCS Test at the highest possible professional level whilst at the same time acknowledging that test development is a process full of ambiguities. Test development is arguably both an art and a science: it requires both imagination and discipline.

This paper refers to the design process for creating one, the first such, test. The process as described is therefore in itself experimental.

It is argued that test construction should be seen, not as a set of mechanical procedures, but as a process of design. The important features of this design process are teacher involvement (both before and after the test), built-in quality control loops, documentation of decisions made in implementing the design criteria, and evaluation of both the process and the product as an integral part from the beginning.

Within this model it should be possible to produce a high-quality test for Queensland students, accepted by them and the education community as having value and integrity.
Supplementary Paper One
Established decisions about the test

The Queensland Core Skills (QCS) Test, required by the Viviani Report, is a common statewide (Queensland) test for Year 12 students to be administered for the first time in September 1992. It is a test of achievement in the common elements of the Queensland senior curriculum. It is not a test of specific subject-related achievement, nor is it an intelligence test or an aptitude test. It is, however, fully syllabus-based: it does not test core skills as a 'bolt-on extra', but is accessible to all Year 12 students regardless of subjects taken at this level.

The 49 common curriculum elements represent the common testable elements in the experienced Queensland senior curriculum. They were identified from the syllabuses in Board subjects in terms of elements which are
- common to two or more syllabuses
- required by the syllabus to be included in work programs
- likely to be accessible to most students through the set of subjects they actually take
- testable in the formats available.

The process of identification involved extensive consultation with teachers about the identification, description and occurrence of the elements. The process also involved consultation with teachers and test developers about testability.

The seven-phase curriculum scan is described in detail in a report in preparation at the time of writing this paper.

The QCS Test will comprise seven hours of testing over two consecutive days. These two days are the Tuesday and Wednesday of the third last week of the third term of the state school year.

There will be four papers in three modes of assessment:

<table>
<thead>
<tr>
<th>Paper</th>
<th>Day and Time</th>
<th>Type of Assessment</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Day 1 morning</td>
<td>Writing Task (WT)</td>
<td>2.0 hr</td>
</tr>
<tr>
<td>2</td>
<td>Day 1 afternoon</td>
<td>Multiple-choice questions (MCQI)</td>
<td>1.5 hr</td>
</tr>
<tr>
<td>3</td>
<td>Day 2 morning</td>
<td>Short response items (SRI)</td>
<td>2.0 hr</td>
</tr>
<tr>
<td>4</td>
<td>Day 2 afternoon</td>
<td>Multiple-choice questions (MCQII)</td>
<td>1.5 hr</td>
</tr>
</tbody>
</table>

Papers 2 and 4 together constitute the MCQ subtest; paper 1 is the extended writing subtest; paper 3 is the SRI subtest.

Certain learnings are not accessible via this testing model. Given the three modes of assessment together with the pen-and-paper nature of the test, one criterion for determining the properties of the test as a whole, and for what can be on the test, is testability.
Actually the equipment allowances for the SRI paper are such that the QCS Test no longer fits the mould of traditional multiple-choice and extended writing papers.

Draft instructions read thus:

A. Students must bring a black pen, a pencil, an eraser and a ruler.

B. Students are permitted to use a programmable or non-programmable calculator. (Calculators which make enough noise to disrupt other candidates are NOT permitted.)

C. Students may choose to bring along any other academic 'tools of the trade' which they routinely use in their OWN course(s) of study (e.g. coloured pencils, set square etc.) [The intention of this is to allow students to use the equipment they have mastered to demonstrate the skills acquired during their instruction. The selection may differ form student to student. This is acceptable.]

D. Students should not use pencil cases or opaque containers for their equipment. Encourage students to use plastic bags or transparent containers.

E. Students may NOT borrow equipment from each other during the test. Ensure that students are reminded to bring all of the equipment they are liable to need.

[Draft only. Pilot Study testpaper, August 1991.]

The test has three purposes. It will be used to provide

- group parameters (subject-group and school-group) for scaling required in the calculation of Overall Positions (OPs) which are independent of subjects studied or school attended
- group parameters on components of the test (yet to be identified) for scaling required in the calculation of Field Positions (FPs), up to five per student
- individual results on a five-point scale, probably reported on the Senior Certificate (vide Viviani Report p.43).

The three subtests, identified according to mode of assessment, will be marked in the following manner.

1. The MCQ (I,II) answer sheets are optically scanned at the University of New South Wales. Each correct answer is given one mark. Each incorrect answer is given zero. Unattempted questions also score zero. There is no penalty for wrong answers. For each question there are four alternative answers (options). Students are asked to select the 'best' answer. Students blacken the circle representing their answer (A, B, C or D) on mark-sensor sheets.

2. The SRI subtest is to be marked in Queensland according to an assessment model called acceptable minimum standards, at least twice, with an item-specific marking scheme, exemplars and detailed instructions being given to
markers. A referee marking occurs in the case of discrepant markings.

3. The WT is marked at a central location in Brisbane by a pool of approximately 300 markers who are trialled, trained and accredited by the Board of Senior Secondary School Studies. Each student script is marked at least three times, according to holistic criteria-based assessment, by different markers working independently. A process of monitoring marker consistency identifies markers who are out-of-step.

Results on subtests or components will not be made available as they have no validity.

Students who are otherwise eligible for an OP (i.e. have 20 semester units of Board subjects studied over four semesters, including three subjects for four semesters) MUST sit the QCS Test. Their results will contribute to the scaling parameters as well as being reported individually on the five-point scale. Other ineligible students MAY sit the QCS Test. Their results will not contribute to any scaling parameters, but individual results on the five-point scale will be reported. Thus students from across the ability range will sit the QCS Test.
Supplementary Paper Two
Acquiring an item bank

The QCS Test is developed and owned by the Queensland Board of Senior Secondary School Studies (BSSSS) which, for 1992, has contracted the Australian Council for Educational Research (ACER) to compose a pool of multiple-choice questions from which 300 will be selected: 100 for a sample test, 100 for the ‘real’ test and 100 for a backup test.

Items for the Short Response Item subtest are written at BSSSS by the Board’s test development unit, by ACER and by the University of Canberra. Items written within BSSSS, together with those purchased elsewhere, are panelled and trialled before use.

The Writing Task is chosen from a number provided by teams of people who work to specifications set by the Board, including the criteria and standards established for its marking. An extract from *The Writing Task: Specifications* together with *Criteria and Standards for Marking* are attached on pages 23 and 25 respectively.

Test developers involved in writing items in the new SRI format are given the following general directions.

Note the *Test Specifications* for the QCS Test.

The 49 common elements symbolise the unifying theme of the Queensland senior curriculum. The whole test will eventually be audited against this list of 49.

The QCS Test is a cross-curriculum test. It is not an examination with a syllabus.

Take account of the ability profile of Queensland 17-year-olds, both eligible and ineligible for an Overall Position (OP).

Take particular note of the characteristics of the new SRIs.

Ensure that stimulus material is self-contained.

Go through the literature seeking stimulus material, noting that some topics are depleted.

Do not use popular textbooks or the subject content of Queensland Board syllabuses as a source of stimulus material.

Note that quality control loops result in about 20 per cent of the items being discarded and a larger proportion of the items being revised.

Obey copyright rules.

Document bibliographical details of stimulus material.
Shred working documents.

Do not discuss any aspects of the items at home or with work colleagues outside the section.

Follow security provisions for panelling of items.

Follow rules for layout.

Enter all information about item characteristics and item history on a card system in the item bank file of the Associate Director (Tests and Examinations).

Items written within the Office of the Board and items purchased elsewhere become the property of the Board of Senior Secondary School Studies.

The SRI paper will increase the validity of the test by expanding the range of testing formats, traditionally only multiple-choice format but recently (1989) incorporating extended writing. Hence, test developers should exploit the advantages of this new format in comparison with the closed nature of multiple-choice questions and the open-ended nature of the Writing Task.

The Viviani Report gives no blueprint for this new testing format therein referred to as 'short written answers' (Viviani 1990, p.45). The following list provides a summary of the characteristics of short response items as defined post-Viviani. The change of name became a necessary part of the conceptualisation of SRIs.

SRIs...

- complement the multiple-choice questions and the Writing Task to give an integrated test
- test as wide a range of curriculum elements as possible
- deal with the whole student ability range
- reflect some school-based assessment practices
- can be constructed along the dimension closed/open
- can be constructed along the dimension convergent/divergent
- involve students in generative thinking and/or production of answers
- involve markers in assessing the quality of a response/conclusion and/or the accuracy of an answer
- are marked according to criteria and standards
- may affect gender-linked discrepancies in student performance.

(The list is not considered to be exhaustive.)
1. NAME

The Queensland Core Skills Test

2. GENERAL DESCRIPTION

The Queensland Core Skills (QCS) Test consists of two multiple-choice question papers (MCQ) of one and a half hours each; one short response item (SRI) paper of two hours, and a Writing Task (WT) of two hours. It is intended for administration to students in Year 12 in Queensland for use as a scaling test in the process required to obtain Overall Positions (OPs) and up to five Field Positions (FPs) from students' school assessments. Students' individual results on the QCS Test are also reported.

2.1 All students eligible for an OP must sit for the QCS Test. Students not otherwise eligible for an OP may sit for the QCS Test if they wish.

2.3 The QCS Test is grounded in the Queensland Senior curriculum through testing the 49 common elements identified from the Queensland Senior curriculum.

2.4 The QCS Test assumes students possess both an elementary level of 'general knowledge' and a knowledge of vocabulary and mathematical operations at a level of sophistication consistent with that of a student with a sound general Year 10 education.

2.5 Unless the nature of the question requires otherwise, any substantive vocabulary of a higher level of sophistication whose meaning cannot be determined from the context is glossed.

2.6 Mathematical operations, as well as including the basic arithmetic operations involved in calculation, also include fundamental mathematical concepts such as percentage, ratio, area and angle.

3. CONTENT

3.1 The MCQ and SRI papers cover a wide range of the common elements.

3.2 Students are expected to respond to a wide variety of stimulus material from a broad range of subject areas. Stimulus material could include: prose passages, direct quotations, film/television transcripts, poetry, dialogues; tabular, mathematical and graphical material; visual material such as maps, diagrams, photographs, cartoons, reproductions of works of art.

3.3 The SRI paper requires a wide variety of responses from students: choosing a key; writing one word, a sentence, a paragraph; producing a
3.4 The WT paper presents students with written and visual material centred around a common theme of interest and relevance to Year 12 students. Cues to focus students' thinking are given where appropriate. The WT tests the students' expressive and productive skills, requiring them to produce an extended piece of continuous prose.

4 CONSTRUCT

4.1 The QCS Test measures a student's ability to

GATHER AND PROCESS

- comprehend facts and literal meanings
- comprehend inferences
- comprehend causal relationships
- extract and process information from styles such as poetry, prose and instructional text
- extract and process information from styles such as cartoons, diagrams, tables, graphs and symbolic text

ANALYSE AND ASSESS

- order and select information
- discern patterns and relationships
- apply techniques, rules and models
- organise and evaluate arguments
- draw conclusions using skills such as evaluate explicit & implicit assumptions, distinguish factors, evince & assess principles, predict conclusions
- make judgments

CREATE AND PRESENT

- write at length
- produce pictorial, diagrammatic, verbal, tabular and graphical information.

4.2 It is assumed that these developed abilities are measured using the MCQ, SRI and WT formats.

4.3 It is assumed that students who achieve well overall in senior studies generally perform well in the QCS Test.

4.4 The validity of all items in the SRI and MCQ papers of the QCS Test is checked at each stage of their development. All items are examined by expert panels.

4.5 Every item in the MCQ and SRI papers is trialled on a population at the appropriate year of schooling in a school system or population closely
4.6 Construct validity of the test is in terms of overall achievement in the 49 curriculum elements. The major item of information to be derived from the test is a single result. Valid inferences about other constructs cannot be drawn from single items, single units, single subtests, or single components.

5. MARKING

5.1 The MCQ answer sheets I and II are optically scanned and marked according to a key. Each correct answer scores one mark. Each incorrect answer scores zero. Unattempted questions also score zero. There is no penalty for wrong answers.

5.2 Markers of the SRI paper are trialled, trained and accredited by the Board of Senior Secondary School Studies. Each script is marked at least twice using an acceptable minimum standards marking scheme. There is a process for monitoring marking discrepancies during the marking operation.

5.3 The WT scripts are marked by markers accredited and appointed by the Board of Senior Secondary School Studies. Each script is marked at least three times using holistic criteria-based assessment whereby a single level is assigned to each script according to the standards developed around a set of criteria. The marking scale has six levels ranging from 1 (highest) to 6 (lowest) with +, 0, - qualifiers for each level. A marker makes a decision in two stages with the first stage involving the allocation of a level from 1 to 6 and the second stage involving the assigning of a qualifier. There is a process for monitoring marker quality during the marking operation.

5.4 Students' individual results, each expressed as one of five grades, are determined using a process which involves the use of a set of criteria and standards.

5.5 'Scaling scores' are recovered from the rank order of students based on their results in the 131 items on the test, viz. 100 multiple-choice questions, 30 short response items and one Writing Task.

6 DISSEMINATION

6.1 The Board of Senior Secondary School Studies holds the copyright for the QCS Test. A new edition is prepared each year comprising items submitted by item writers appointed by the Board.

6.2 In association with the development of the QCS Test, an information bulletin is circulated to Year 12 students.
6.3 Once an edition of the test is administered it is no longer secure and students may retain their test papers. Also, copies may be purchased from the Board. A report on that edition of the test, including the answers, is published as soon as possible after its administration.
1. NAME The Writing Task

2. GENERAL DESCRIPTION

The Writing Task is a two-hour test involving the preparation of one piece of writing, of approximately 600 words.

It is intended for administration to students in Year 12 in Queensland...

2.1 The Board of Senior Secondary School Studies holds the copyright for the Writing Task. A new edition is prepared each year by setters appointed by the Board.

2.2 In each edition students are presented with stimulus materials in a variety of genres. Visual materials are also used. Cues to focus students' thinking are given, where appropriate. Students are required to base their writing on one or several of the items presented in the stimulus materials and to give the writing a title.

2.3 Only one set of materials, grouped around a common theme, is presented in each paper although the richness of the stimulus materials should allow considerable variety of response.

2.4 The common theme for the task is chosen in such a way that it is of interest and relevance to most Year 12 students.

The way in which the task is presented will reflect what is generally regarded as good teaching - learning practice in Queensland. The theme and presentation will allow the whole range of Year 12 students the opportunity to respond.

3. CONTENT

The Writing Task will show the students' ability to express themselves in written English language. The stimulus materials will be drawn from a wide variety of sources appropriate to Year 12 students in Queensland.

Students may use any genre (except poetry) for their writing. This is so that students may respond in the genre in which they feel most comfortable. (Research has tended to indicate that poetry cannot be reliably and speedily marked.)
4. **CONSTRUCT**

   The Writing Task provides a measure of students' capabilities to respond in writing given written and visual stimulus materials..., it tests the students' expressive and productive powers, requiring them to produce an extended piece of continuous English prose. It tests the students' achievements as measured against the criteria and standards for the task.

5. **MARKING**

   Scripts are marked by markers accredited and appointed by the Board of Senior Secondary School Studies. Each piece is marked four times using holistic marking. A single level is assigned to each script according to standards developed around a set of criteria. The marking scale has six levels ranging from 1 (highest) to 6 (lowest), with +, 0, - qualifiers for each level. A marker makes a decision in two stages with the first stage involving the allocation of a level from 1 to 6 and the second stage involving the assigning of the qualifier.
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>ACHIEVEMENT OF PURPOSE</th>
<th>CONTROL OF STRUCTURE</th>
<th>PROFICIENCY WITH A WRITTEN LANGUAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The student &lt;br&gt; 1. writes so that the central idea is clear and well developed throughout &lt;br&gt; 2. writes with sensitivity and direct relevance to the stimulus material</td>
<td>The student 1. organizes ideas and/or images effectively &lt;br&gt; 2. uses the conventions of the chosen genre for a desired effect &lt;br&gt; 3. writes at an appropriate length</td>
<td>The student 1. maintains cohesion and shows flexibility in sentence construction to gain a desired effect &lt;br&gt; 2. uses an extensive vocabulary with discrimination and precision &lt;br&gt; 3. consistently maintains the conventions of the written language</td>
</tr>
<tr>
<td>2</td>
<td>The student &lt;br&gt; 1. writes so that the central idea is clear to the reader and well developed for the most part &lt;br&gt; 2. writes with direct relevance to the stimulus material</td>
<td>The student 1. organizes ideas and/or images effectively &lt;br&gt; 2. follows the conventions of the chosen genre with some success in creating a desired effect &lt;br&gt; 3. writes at an appropriate length</td>
<td>The student 1. maintains cohesion and shows command of sentence construction &lt;br&gt; 2. uses vocabulary with discrimination &lt;br&gt; 3. consistently maintains the conventions of the written language</td>
</tr>
<tr>
<td>3</td>
<td>The student &lt;br&gt; 1. writes so that the central idea is clear to the reader and shows some development &lt;br&gt; 2. writes with direct relevance to the stimulus material</td>
<td>The student 1. succeeds, for the most part, in organising ideas and/or images &lt;br&gt; 2. follows the basic conventions of the genre chosen &lt;br&gt; 3. writes at an appropriate length</td>
<td>The student 1. maintains cohesion &lt;br&gt; 2. uses vocabulary appropriate to the subject-matter &lt;br&gt; 3. maintains the conventions of the written language for the most part</td>
</tr>
</tbody>
</table>

continued over
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>ACHIEVEMENT OF PURPOSE</th>
<th>CONTROL OF STRUCTURE</th>
<th>PROFICIENCY WITH A WRITTEN LANGUAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>The student</td>
<td>The student</td>
<td>The student</td>
</tr>
<tr>
<td></td>
<td>* writes so that the central idea is clear to the reader*</td>
<td>* uses a structure which is discovered only with effort from the reader*</td>
<td>* shows occasional lapses in cohesion*</td>
</tr>
<tr>
<td></td>
<td>* writes with relevance to the stimulus material for the most part*</td>
<td>* uses some of the conventions of the genre chosen*</td>
<td>* uses appropriate vocabulary*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>* writes at an appropriate length*</td>
<td>* maintains the basic conventions of the written language with occasional lapses*</td>
</tr>
<tr>
<td>5</td>
<td>The student</td>
<td>The student</td>
<td>The student</td>
</tr>
<tr>
<td></td>
<td>* writes so that the central idea can be discovered with effort from the reader*</td>
<td>* shows frequent lapses in organising ideas or images (e.g. through faulty paragraphing)*</td>
<td>* shows uneven cohesion*</td>
</tr>
<tr>
<td></td>
<td>* writes with occasional relevance to the stimulus material*</td>
<td>* writes at an inappropriate length*</td>
<td>* uses a restricted vocabulary*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>* shows serious lapses in maintenance of conventions of the written language*</td>
</tr>
<tr>
<td>6</td>
<td>The student</td>
<td>The student</td>
<td>The student</td>
</tr>
<tr>
<td></td>
<td>* writes without any sense of a central idea*</td>
<td>* fails to organise ideas or images*</td>
<td>* writes so that cohesion is lacking as shown by improper pronoun choice, tense confusion and fragmented sentence structure*</td>
</tr>
<tr>
<td></td>
<td>* writes with little relevance to the stimulus material*</td>
<td>* writes so little that no structure can be perceived*</td>
<td>* uses immature vocabulary*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>* fails to maintain the conventions of the written language*</td>
</tr>
</tbody>
</table>

This standard describes the top of the first level.

1+ The student
* writes so that the central idea is clear and well developed throughout*
* writes with sensitivity, authority and direct relevance to the stimulus material*

The student
* organises ideas and/or images effectively*
* uses the conventions of the chosen genre for a desired effect*
* writes at an appropriate length*

The student
* maintains cohesion, and shows flexibility and originality in sentence construction to gain a desired effect*
* uses an extensive vocabulary with discrimination and imagination*
* consistently maintains the conventions of the written language*

Genre: a form (of writing).

For the Writing Task, the student may have attempted to: write a traditional composition around a theme, convey information, argue a point, describe personal experience, write a short story, make diary entries, write a television script, write a play, script an interview, write a newspaper or magazine article, write a long letter or a series of letters, prepare a report, write fantasy or science fiction, write a literary criticism and so on. Other possibilities exist.

Students were told not to write a poem (verse), even if a poem is included in the stimulus material.
The 13 attributes represent a shorthand version of the domain of the test - the skills or behaviours which are being sampled in any instance of the test. They are presented as a convenience, an heuristic, rather than as being ontologically or epistemologically necessary.

GATHER AND PROCESS

- comprehend facts and literal meanings
- comprehend inferences
- comprehend causal relationships
- extract and process information from styles such as poetry, prose and instructional text
- extract and process information from styles such as cartoons, diagrams, tables, graphs and symbolic text

ANALYSE AND ASSESS

- order and select information
- discern patterns and relationships
- apply techniques, rules and models
- organise and evaluate arguments
- draw conclusions using skills such as evaluate explicit & implicit assumptions, distinguish factors, evince & assess principles, predict conclusions
- make judgments

CREATE AND PRESENT

- write at length
- produce pictorial, diagrammatic, verbal, tabular and graphical information.

The list of attributes serves as a short descriptor for each criterion in the second attached schedule of criteria and standards (pages 29 and 30). The 49 common elements can be grouped under these headings.

The first attached schedule (page 28) lists the common curriculum elements.
COMMON CURRICULUM ELEMENTS (49)

Recognising letters, words and other symbols
Finding material in an indexed collection
Recalling/remembering
Interpreting the meaning of words or other symbols
Interpreting the meaning of pictures/illustrations
Interpreting the meaning of tables or diagrams or maps or graphs
Translating from one form to another
Using correct spelling, punctuation, grammar
Using vocabulary appropriate to a context
Summarising/condensing written text
Compiling lists/statistics
Recording/noting data
Compiling results in a tabular form
Graphing
Calculating with or without calculators
Estimating numerical magnitude
Approximating a numerical value
Substituting in formulae
Setting out/presenting/arranging/displaying
Structuring/organising extended written text
Structuring/organising a mathematical argument
Explaining to others
Expounding a viewpoint
Empathising
Comparing, contrasting
Classifying
Interrelating ideas/themes/issues
Reaching a conclusion which is necessarily true provided a given set of assumptions is true
Reaching a conclusion which is consistent with a given set of assumptions
Inserting an intermediate between members of a series
Extrapolating
Applying strategies to trial and test ideas and procedures
Applying a progression of steps to achieve the required answer
Generalising from information
Hypothesising
Criticising
Analysing
Synthesising
Judging/evaluating
Creating/composing/devising
Justifying
Perceiving patterns
Visualising
Identifying shapes in two and three dimensions
Searching and locating items/information
Observing systematically
Gesturing
Manipulating/operating/using equipment
Sketching/drawing
### Suggested Criteria & Standards for 1992 CS Test (Draft Only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gather and Process</td>
<td>Comprehend Facts &amp; Literal Meanings</td>
<td>comprehends some facts or some literal meanings.</td>
<td>comprehends most facts or most literal meanings.</td>
<td>comprehends most facts and most literal meanings over a restricted range of material.</td>
<td>comprehends (almost) all facts and (almost) all literal meanings over a wide range of material.</td>
<td>comprehends (almost) all facts and (almost) all literal meanings over a wide range of material.</td>
</tr>
<tr>
<td></td>
<td>Comprehend Inferences</td>
<td>shows an awareness of probable inference.</td>
<td>comprehends inferences requiring either deduction or induction.</td>
<td>comprehends inferences requiring both deduction and induction over a restricted range of material.</td>
<td>comprehends inferences requiring both deduction and induction over a wide range of material.</td>
<td>comprehends inferences requiring both deduction and induction over a wide range of material.</td>
</tr>
<tr>
<td></td>
<td>Comprehend Causal Relationships</td>
<td>perceives explicit relationships between two factors.</td>
<td>deduces a causal relationship between two explicit factors.</td>
<td>deduces causal relationships between a number of factors.</td>
<td>deduces and induces causal and other relationships between many and varied factors.</td>
<td>deduces and induces subtle causal and other relationships between factors from inter-related material.</td>
</tr>
<tr>
<td>Extract and Process</td>
<td>Extract and Process Verbal Information</td>
<td>extracts some explicit information.</td>
<td>extracts and defines some explicit information.</td>
<td>extracts, defines and clarifies information.</td>
<td>extracts information, clarifies it and transforms it to display meaning.</td>
<td>extracts information, clarifies it and transforms it to display meaning.</td>
</tr>
<tr>
<td></td>
<td>General Information</td>
<td>extracts some explicit information.</td>
<td>extracts and defines some explicit information.</td>
<td>extracts, defines and clarifies information.</td>
<td>extracts information, clarifies it and transforms it to display meaning.</td>
<td>extracts information, clarifies it and transforms it to display meaning.</td>
</tr>
<tr>
<td>Create and Present</td>
<td>Write at Length</td>
<td>demonstrates little cohesion, a restricted vocabulary and limited proficiency with language in writing with little relevance.</td>
<td>demonstrates moderate proficiency with written language using some generic conventions in writing with a central idea which has some relevance.</td>
<td>demonstrates, for the most part, the ability to maintain cohesion in the conventions of written language, to organise ideas and to follow generic conventions in developing a clear, relevant central idea.</td>
<td>demonstrates proficiency with the conventions of genre and written language, an effective control of structure, and an ability to develop a clear, relevant central idea.</td>
<td>demonstrates a confident and flexible proficiency with written language, a skilled and effective control of structure, and a consistent ability to develop, clearly and sensitively, a relevant central idea.</td>
</tr>
<tr>
<td></td>
<td>Produce Pictorial, Diagrammatic, Verbal, Tabular and Graphical Information</td>
<td>produces information with little clarity or visual appeal including items with errors in spelling, punctuation and grammar.</td>
<td>produces information that has some clear parts and visual appeal but with items that contain errors in spelling or punctuation or grammar.</td>
<td>produces visually appealing, clear and coherent information that may contain some errors in spelling or punctuation or grammar.</td>
<td>produces visually appealing, clear, coherent and accurate information containing few errors.</td>
<td>produces clear, coherent and accurate information of the highest visual appeal.</td>
</tr>
</tbody>
</table>

continued over
<table>
<thead>
<tr>
<th>ORDER AND SELECT INFORMATION</th>
<th>DISCERN PATTERNS &amp; RELATIONSHIPS</th>
<th>APPLY TECHNIQUES, RULES AND MODELS</th>
<th>STRUCTURE &amp; EVALUATE ARGUMENTS</th>
<th>DRAW CONCLUSIONS</th>
<th>MAKE JUDGMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;A&quot; STANDARD</td>
<td>Identifies explicit and relevant information from a restricted range of material.</td>
<td>Recognizes and selects patterns and relationships.</td>
<td>Uses some techniques for making calculations.</td>
<td>Identifies the essential aspect of a verbal argument.</td>
<td>May be aware of possibilities.</td>
</tr>
<tr>
<td>&quot;B&quot; STANDARD</td>
<td>Identifies explicit and relevant information from a wide range of material.</td>
<td>Recognizes patterns and relationships.</td>
<td>Uses some techniques for making some calculations.</td>
<td>Identifies the fundamental aspect of a verbal argument and appropriately evaluates its worth.</td>
<td>Exercises comparisons when making judgments.</td>
</tr>
<tr>
<td>&quot;C&quot; STANDARD</td>
<td>Identifies explicit and relevant information from a large range of material.</td>
<td>Recognizes simple patterns and relationships.</td>
<td>Uses some techniques for making approximate calculations as well as solving some problems.</td>
<td>Identifies the essential aspect of a mathematical argument.</td>
<td>Exercises comparisons when making judgments.</td>
</tr>
<tr>
<td>&quot;D&quot; STANDARD</td>
<td>Identifies explicit and relevant information from a wide range of material.</td>
<td>Recognizes patterns and relationships.</td>
<td>Uses some techniques for making approximate calculations as well as solving some problems.</td>
<td>Identifies the essential aspect of a mathematical argument.</td>
<td>Exercises comparisons when making judgments.</td>
</tr>
<tr>
<td>&quot;E&quot; STANDARD</td>
<td>Identifies explicit and relevant information from a wide range of material.</td>
<td>Recognizes patterns and relationships.</td>
<td>Uses some techniques for making calculations.</td>
<td>Identifies the essential aspect of a mathematical argument.</td>
<td>Exercises comparisons when making judgments.</td>
</tr>
</tbody>
</table>

**Highest Proficiency** → **C Criterion**
Supplementary Paper Four
Timeline for construction and evaluation of sample test

SELECT DESIGN CRITERIA (tentative)    July 1991

RESEARCH PILOT STUDY (SRI format)    August 1991 - September 1991

SELECT DESIGN CRITERIA (final)      October 1991

APPLY CRITERIA TO
CONSTRUCT SAMPLE TEST      December 1991 - January 1992

PUBLISH SAMPLE TEST      Beginning of first term 1992

EVALUATE PROCESS
teacher interviews    February 1992
student interviews    March 1992
external evaluator

PUBLISH EVALUATION      Late March 1992

ADJUST APPLICATION OF
DESIGN CRITERIA      April 1992
Supplementary Paper Five
Layout of units and items

The attached sample illustrates the layout of a unit on the SRI paper.

The sample has been annotated to explain the function of each part of the unit. Each unit in the test is constructed in a similar fashion. A unit consists of a common piece of stimulus material and a set of items. These may be preceded, where necessary, by an introduction.

Arrows point to the following parts of the unit:

- **item**: is a combination of stem and cue.

- **introduction**: gives the referent for the unit to follow and may indicate the number of items in that unit. The introduction, when necessary, occurs before the presentation of the stimulus material.

- **stimulus material**: is comprised of verbal, numerical, visual (pictures, photographs, works of art, diagrams) or tabular/graphical material selected by the test developers with the aim of promoting students' responses.

- **stem**: a command or a question which indicates the task or the answer required from the student.

- **cue**: an instruction block attached to an item as marginalia (beside the space provided for the student response). The cue gives students a clear idea of what is required of them.

- **response**: the student's reaction to an item as communicated to the marker in writing/drawing/calculating etc. The student response is graded by the marker.

Source: Pilot Study testpaper (August 1991)
UNIT FOUR

INTRODUCTION
The next two items refer to this passage.

STIMULUS MATERIAL (VERBAL) FOR Items 5 and 6
In certain rice growing areas, the disease schistosomiasis has been common for centuries. Schistosomiasis in humans results from infection by tiny parasites which, after eating through the skin, move through the body in the bloodstream, feeding and growing to mature worms.

Because of the methods employed in planting and cultivating rice paddies, there is a high incidence of schistosomiasis among rice farmers. While working in flooded fields and irrigation ditches, their legs, hands and arms are exposed to water infested with the parasites which cause the disease.

Most of the parasites eventually lodge in or near the liver where they mature and lay thousands of eggs. The infection impairs the liver's function, which results in fevers, severe abdominal pain, nausea and lethargy. Victims may live for many years but death can follow in five to ten years after infection.

Some of the schistosoma eggs pass out of the bloodstream of an infected person, reach the digestive tract, and are eventually excreted in human wastes. Once these eggs reach rivers and streams they hatch into small larvae which are the schistosoma parasites. In order to survive, the larvae must find an aquatic snail in which to develop. When development is complete a second larval stage passes from the snail into the water and seeks another human host. If the newly hatched larvae fail to find a host snail within 24 hours of hatching, they die.

UNIT FOUR CONTINUES ON THE NEXT PAGE...
ITEM 5  

STEM FOR ITEM No. 5

In the form of a diagram summarise the life cycle of the schistosoma.

INSTRUCTION BLOCK OR CUE

Draw a clear, accurate and well-presented diagram.

SPACE FOR RESPONSE
Suggest one practical way in which the life cycle of the schistosoma could be broken.

*I hypothesise that the life cycle of the schistosoma could be broken by*

*My reasoning for this suggestion is*

Write in sentences.
REFERENCES


Information regarding this publication may be obtained from the

Associate Director (Tests & Examinations)
Gabrielle Matters

Telephone: (07) 864 0258

© This material is copyright. It may be copied freely for the use of schools in Queensland. It may not be reproduced for sale without express permission.

Board of Senior Secondary School Studies
Floor 9, 135 Wickham Terrace, Spring Hill, Brisbane
PO Box 307, SPRING HILL Q 4004

Telephone: (07) 864 0299
Facsimile: (07) 832 1329

ISBN 0-7242-4688-6
Library Digitised Collections

Author/s:
Matters, Gabrielle N.

Title:
A design process for constructing the Queensland Core Skills Test

Date:
1991

Persistent Link:
http://hdl.handle.net/11343/115740