

Evaluation Report of the Pilot of the 2005 Queensland Assessment Task (QAT)

Val Klenowski



2006

This work was funded by the Department of Education and the Arts.

**Queensland government
Department of Education and the Arts (2006)**

CONTENTS

Tables	iv
Figures	iv
Executive Summary	v
1. Introduction	1
Background	1
The Assessment and Reporting Taskforce	1
Assessment and Reporting Framework	1
ARFIC 2003	2
Queensland Standards Map	4
Queensland Assessment Task	4
The 2003 QAT Pilot (June – December 2003)	8
(ARFIC) Working Parties	17
2. The 2005 QAT Pilot (2004 – 2005)	23
A Futuristic QAT	23
The 2005 QAT	25
Queensland Curriculum and Assessment Reporting (QCAR) Framework	30
3. Evaluation Methodology	36
Context	36
Evaluation Standards	36
Framework	37
Change Process	38
Aims	40
Methodology	41
Code of Practice	45
4. Design Brief and Specifications	46
The Construct of Transforming	46
Did the QAT meet the design brief and design specifications?	46
The Standardised Assessment Tasks and the TGTs	48

5. Technical Considerations	54
Key Determinants of Validity	54
How valid was the 2005 QAT?	57
Quality of Student Data	60
Ensuring Reliability	61
Reliability	62
Marking Guides	63
Standards	65
Marker Advisors and Markers	66
Marker Training Programme	66
Management of the Marking Operation	68
Reporting	69
Equity Issues	69
Conclusion	70
6. The Students' Perspective	
Student Sample	72
Data Sets	72
Analysis	73
Computer-based Task	74
Differences from School Based Assessment	75
Level of Challenge	76
Level of Support	77
Perceived Skills Required	78
Level of Difficulty	79
Learning	80
Recommended Changes	81
Paper-based Task	82
Level of Difficulty	83
Level of Interest	83
Learning	83
Required Skills	84
Suggestions for Improvement	84
Timing for the Collection of Student Responses	84
Overall Opinions	84
Conclusion	85

7. Teachers' Perspectives

Interviewee's Role	87
Relevance	88
Intellectual Challenge and Motivation	90
Equity Issues	93
Alignment of Performance and Targeted Intentions	95
Preparedness of the School	96
Usefulness	97
Implications for a Common Assessment Task	98
Conclusion	98

8. Implications for Policy

Key Issues	100
Significance of the Pilot	107
Opportunities	114
Challenges	124
Implications for Action	130
Appendices	133
References	154

TABLES

Table 1	Types of transformation	47
Table 2	Overall 2005 QAT results	61

FIGURES

Figure 1	Framework for the Evaluation of QATs	39
Figure 2	Interactive factors affecting implementation	40
Figure 3	Percentage of student responses to Questions 1-18, Computer-based task	75
Figure 4	Extent of skills used in completing computer-based task	78
Figure 5	Percentage of student responses to Questions 25-30, Paper-based task	82
Figure 6	Threats to validity: the Eight Stage Model (reversed)	126

EXECUTIVE SUMMARY

Introduction

The 2005 Queensland Assessment Task (QAT) emerged from a background of earlier developments such as the trial of the New Basics (2000-03), the establishment of the Assessment and Reporting Framework Implementation Committee (ARFIC) (2002) and the pilot study of the 2003 Assessment and Reporting Framework (ARF) via the 2003 Queensland Assessment Task. These initiatives aimed to establish a framework for curriculum, pedagogy and assessment for Queensland schools. Another aim was to provide a common frame of reference against which teachers could assess and report student achievement in certain constructs, at certain junctures in a common format to parents and the system. These developments continued in the pilot of the 2005 QAT which explored how the overall design, development, implementation and continuation of the use of common assessment tasks could be improved.

This evaluation of the 2005 QAT is contextualized in a large-scale reform agenda of 'national consistency in education' that involves assessing students against national benchmarks and reporting these to parents. Consistent with this is the Queensland Government's planned introduction of the Queensland Curriculum and Assessment Reporting (QCAR) framework for the purposes of aligning the syllabus related component of school curriculum (the intended learning outcomes) with the assessment and reporting of student achievement.

As a pilot study, the 2005 QAT, aimed originally to explore whether statewide assessment data related to student achievement in the Key Learning Area (KLA) syllabuses could be provided and reported. The intent was to capture rich information about student achievement in nominated domains and across them, in various mediums using a variety of instruments, devices and strategies. Given the dynamic policy arena, a reduction of resources for the project and the finding of the 2003 pilot study that curriculum coherence could be lacking in Queensland schools, the specifications for the 2005 QAT were adjusted.

In 2005 the focus was on Year 9 students and the QAT, which was pitched at Level 5, was made up of two, rather than three, standardised assessment tasks in the different assessment modes of computer-based and constructed response (paper-based). It was intended that there would also be a performance-based task and that these three standardised tasks would be complemented by a corresponding teacher generated task, however, these latter tasks were not incorporated and do not form part of the evaluation.

A futuristic QAT, consisting of a computer-based task and a constructed response, was developed to assess the student's achievements in transforming ideas and/or information. This construct of processing, which includes the underlying generic skills and dispositions, was chosen as the focus for assessment over the construct of knowledge (facts, concepts and procedures) because the developers were uncertain about the level of curriculum coherence. They were aware that teachers use a wide-range of curriculum materials and approaches to provide students with a variety of learning experiences.

The construct of processing was particularly suited to the use of technology. Students were required to work in multiple modes to transform ideas and information and their achievement was measured in the underpinning repertoires and generic skills. The QAT drew on the Study of Society and the Environment (SOSE) and the Arts which coincided with the roll out of these syllabuses from the Queensland Studies Authority (QSA). The 2005 QAT was administered in August in 56 schools. In October trained teachers marked the students' responses and the results of student achievement were reported in December.

The External Evaluation

The framework for the evaluation is based on large-scale curriculum and assessment reform efforts that appear to impact on changing school and classroom practice both at the central policy level and at school and district levels. It takes into consideration the context for change, policy levers (curriculum, standards, tasks) and local challenges (relevance, readiness and resources). The design and implementation of the QAT was evaluated across four dimensions: the design brief and specifications; technical considerations including the validity and reliability of student achievement data; the alignment of curriculum and assessment and the policy implications.

Qualitative and quantitative methodologies were adopted. The data collection methods encompassed interviews with students, teachers, principals, QAT coordinators, QAT developers, marker advisors, policy officers and academics, surveys of students, site visits to two schools, QAT team meetings and the central marking operation involving teacher assessors. Documents were another rich source of data that informed this evaluation. These included policies, QAT documents including proposition papers, project reports executive documents, ministerials and project plans. Other important data sources were the computer-based task, the constructed response task, the QAT administration guidelines, records and reports of ARFIC meetings and its working parties, marking manuals and guides, students' results and final QAT reports (on the computer-based task and the marking and grading).

The pilot of the 2005 QAT set out to explore the potential of combining the more traditional paper-based mode of assessment with electronic medium to assess students' generic skills. In many ways the QAT team has been successful. A summary of what the analysis of the data revealed as successes and challenges

follows. Much has been achieved and should be celebrated but in the context of the QCAR framework there are many challenges ahead.

Successes

Meeting the Design Brief and Specifications

The evidence suggests that the 2005 QAT provided intellectual challenge and made connections to the wide world for Year 9 students. The two standardized tasks, pitched at Level 5, assessed generic skills including the use of ICTs and multi-literacies. It was possible to assess students' achievements in transforming ideas and/or information and in the underlying generic skills and dispositions. This was evident from the students' overall results and the way in which these were reported.

Students received a certificate which included a grade for both tasks and an overall grade. Grades A-E were used, however, only Grade A was defined on the statement of results. The percentage of grades awarded based on all task items (overall), computer-based items and paper-based items were also graphically represented on certificates.

Collectively A-grade students
Exhibit knowledge of key aspects of history, geography and media
Extract information from prose, diagrams, maps and symbolic text; clarify it and transform it to display meaning in multiple media.
Discern patterns and relationships in verbal, pictorial and symbolic text (alone or in combination); make significant decisions and judgements, operationalise these into accurate representations and products.

The most significant development has been the computer-based task. The relevance of the electronic medium, multi-literacies, online marking and the cost effectiveness of computer-based assessment were identified as successes.

Addressing Technical Considerations

To be meaningful measurement must be replicable, comparable and consistent. The tasks of the 2005 QAT provided students with considerable diversity by using a variety of modes (such as diagrams, graphs, pictures and symbols). All questions of the constructed response and computer-based tasks were designed to assess transformation skills. Validity was achieved by increasing the range of contexts for student performance and by providing more extensive sampling of the skills enabling the selected constructs to be assessed. Validity includes the evidence available for assessment interpretation and the potential consequences of assessment use. As data collection stopped when the certificates were sent to schools it was not possible to establish how the results were interpreted or used at the school and classroom levels.

The method used to determine the student's grades incorporated traditional scoring techniques together with appropriate aspects of criteria-based assessment as practised in Queensland. The range of grades achieved both on the computer-based task and the constructed response suggest that the tasks were of varying difficulty so that the students were given an opportunity to demonstrate the extent of their achievements and markers were able to discriminate between achievements of different quality.

The QAT was administered on one occasion, so the importance of reliability was concerned with how reliable the results were so that they could be generalised to the point that the students' results would be consistent if the same design criteria were used and other items were on the test. The form of reliability was internal consistency. The consistency of performance on each test item was considered. The measure used is called "Cronbach's Alpha" and for the 2005 QAT is 0.85.

To ensure reliability all of the following methods were observed in use during the central marking operation of the 2005 QAT.

- Documented, field tested marking guides;
- Specified criteria;
- Annotated examples of all score points;
- Ample practice and feedback for markers;
- Multiple markers with agreement prior to marking;
- Periodic reliability checks throughout;
- Retraining if necessary and
- Arrangements for the collection of suitable reliability data. (Forster & Masters, 1996:43)

Overall the standardized assessment tasks were evaluated as valid and yielded reliable results for reporting.

Aligning Curriculum and Assessment

The 2005 QAT focused on a range of generic skills, attitudes and dispositions considered essential for the learner of the 21st century. Teachers and students were asked about its utility, relevance, intellectual challenge, the motivation provided and whether the intentions of the QAT aligned with the required performance of students.

The students' quantitative and qualitative responses give a clear sense of acceptance of the computer-based task, with the paper-based task being generally acknowledged as the harder and less interesting of the tasks. The qualitative data gives a more mixed sense of the students' responses. There is a strong sense that the QAT is different from other forms of assessment, both in terms of its use of computers and its challenging nature.

The teachers' opinions provide a useful insight into the educational climate in which the QAT was piloted and in which any future common assessment task will

be implemented. The QAT encountered a climate of conflicting views about the very thing it was intended to be - an external form of assessment in addition to the established school-based assessment of the Queensland system. Where some teachers saw this as potentially providing a firm point of reference, or at least useful preparation for high-stakes tests in later years, others saw mainly a lack of congruity between the external assessment and internal curriculum.

The positive response to the computer-based task seems to indicate a climate that is conducive to innovations that enhance the students' motivation, while the response to the paper-based task was more mixed. There is obviously no agreement among teachers from a wide range of schools on just how challenging a Level 5 task should be, however, all teachers felt that the QAT was in fact intellectually challenging (in some cases, too challenging). In addition, the teachers believed that in general the students were strongly motivated by the task. The most powerful message is that despite a common assumption that motivating the range of students in the middle years necessitates a lowering of standards these students can in fact be strongly motivated by work of genuine intellectual challenge.

Policy Implications.

Changing Policy Context

The rapidity of change in the policy-making arena and the consequent impact on the design and development of intended assessment strategies was highlighted. A dynamic policy context requires the team responsible for the design and development of the task to be supported and directed with responsible policy leadership.

Co-ordination of Roles, Responsibilities and Relationships

A key message was the uncertainty associated with the policy context and with the roles and role relationships among policy-making groups. Some hindrances to communication and inefficiencies occurred due to the lack of articulation of professional boundaries. It is important that the roles and role relationships for all parties are made explicit and that channels of communication are kept open throughout the trial and implementation phases of the QCAR framework.

Standards

Teachers involved in the marking of the 2005 QAT engaged in standards-referenced assessment. They found their experience of moderation richly rewarding. Identified benefits included: teachers' improved understanding of the standards expected at Level 5 and possible gaps in their own teaching and learning programmes and curriculum priorities. Teachers appreciated the value of common assessment tasks and their use in the context of standards-referenced assessment.

The standards will provide a common frame of reference for making judgments about the quality **and** progress of student learning, and a common language for

reporting. Teachers will need to meet regularly to discuss work for moderation purposes. Communication of the centrality of the use of standards to parents, carers, students, teachers, principals and the system will be fundamental.

Maximising the Prospect for Innovation

The QAT demonstrated the possibility of assessing generic skills, applied in year 9, with the results reported relative to an A-standard performance. These processing skills are integral to one's learning capacity. A graphical component was developed and incorporated into the marking schemes for the constructed-response to reward these higher order skills used in the short answer responses. With greater recognition of the importance of the nature of learning and the need to teach learning skills, innovative tasks that are designed to assess such skills, represent an important step. The electronic medium proved a favourable environment for demonstrating these skills.

This affirms the need for the statements of essential learnings, at key junctures, to identify the generic skills and attributes **as well as** the domain-specific knowledge, skills, understandings and dispositions. Coherence across the key junctures is required and generic skills will need to be incorporated into the assessment tasks so that students and teachers understand their value and importance in learning.

Teacher Professional Development

Teachers and schools gained professionally from participating in the 2005 QAT. Some key areas for support that have been identified for continued development are: teachers' assessment skills and literacy; 'growing an assessment culture'; training in marking, moderation practices; computer based assessment; use of ICTs; curriculum development; building student's learning capacity and developing pedagogy. Those who had extensive experience in moderation exercises, and the training of teachers for the central marking operation, stressed the teacher development benefits in terms of establishing validity and reliability in assessment practice. An ongoing and sustained professional development program should parallel the development and trialing phase of the QCAR framework.

Bank of Quality Assessment Tasks

Quality resources developed for the QAT are plentiful and comprise: common assessment tasks, marking guides for both tasks, commentaries on each item, grade distributions and commentaries, model solutions, exemplars based on student work samples, analyses of results and marker training guides. Teachers respond to principles and practices that they can relate to and that are grounded in their own contexts. They do not change their practices based on research or evaluation evidence alone, they need examples of implementation by teachers with whom they can relate. A strategic approach in the establishment of the bank of assessment tasks should be adopted.

Challenges

The Need for a Trial

Regardless of the expertise on the development team there is always a need to trial the assessment tasks. The importance of trialing the assessment tasks prior to administration was acknowledged by many teachers and developers. The shortened timeline for the development of 2005 QAT meant there was no trial of the constructed response task. This impacted on the validity of some questions. If a common task is to be developed then it must be trialed with students prior to administration. This includes students with special needs.

Teacher Capacity

Building teacher capacity is a major challenge that will require the provision of systematic training, resources and consultant support for teachers and principals. Professional learning communities should be established at the local level to meet the particular, identified needs such as for those teachers in remote and rural settings. Too often these rich cultural contexts are neglected by policy developers who give the impression that policies can be implemented in a homogeneous manner without attending to the disparities that exist. The pilot indicated that gaps or weaknesses in subject knowledge, assessment practice or pedagogical understanding will limit the extent to which teachers can use the QCAR framework.

Teachers and principals will need to understand the new policies and the QCAR framework, key terminology and concepts associated with these policies, the theoretical underpinnings of the proposed changes, the implications for practice and the fundamentals of assessment theory and practice in relation to those proposed changes.

Validity

The nature and format of the 2005 QAT was considered valid and data suggests that worthwhile skills and processes were assessed. Threats to validity need to be addressed in developing common assessment tasks. The tasks need to:

- address important aspects of the target domain of the essential learnings;
- motivate students and
- act as models for assessment tasks to be developed by classroom teachers.

If a narrow focus is adopted in the common assessment tasks then teachers, parents and the community will focus on the areas assessed rather than the full spectrum of essential learnings that aim to include: cross-curricular aims, generic skills and attributes and domain-specific knowledge, skills, understandings, dispositions and values.

Reporting Results

Given the innovative character of the tasks and the nature of a pilot, reporting should be approached with caution. With the move towards the inclusion of some external assessment in the form of common assessment tasks it may well be that some students are more prepared than others for this form of assessment. Other factors that impact on the approach taken with reporting relate to:

- The purpose of the assessment;
- The connection between the assessment task and the curriculum;
- Teachers', parents',/carers', and students' expectations.

These factors highlight the importance of explaining what the results relate to when reporting.

Understanding the Change Process

Purpose and Concept

Students and teachers did not appear to understand fully the **purpose** of the QAT. The tradition of school-based assessment in Queensland will need to be harnessed and teachers will need to understand how valued learning can also be assessed both by internal and external means. The fact that the QAT was external, standardised and, for some teachers seen as 'just a trial', diminished the relevance of the experience.

A poorly conceptualised change or one that cannot be demonstrated will be difficult to implement. From the outset teachers, students, parents, carers and administrators need to know the intended purposes of the QCAR framework and why it is being introduced. They also deserve to know who will benefit from it and how. What will be achieved for students needs to be made explicit.

Resources

A shortened timeline for completion and a reduction of resources made the intended outcomes for the 2005 QAT more difficult to achieve. If a change is poorly resourced or resources are withdrawn then this will impact negatively on the implementation of the change.

Communication and Commitment

Communication and collaboration among policy-making groups is required and strategic leadership is needed to move forward and to prevent feelings of frustration from students, teachers, principals, curriculum developers and policy officers. In 2005 there appeared to be no long-term commitment for the QAT to carry people through the anxiety and frustration of early experimentation or unavoidable setbacks. Commitment is an important factor to consider.

Co-ordination of the Change

Parallel changes in curriculum, assessment and pedagogy will need to be carefully co-ordinated and supported in the implementation of the QCAR

framework. What was highlighted was the necessity for co-ordination of the changes and understanding of how parallel changes intersect.

Pace and Scope of the Change

If the change is introduced too quickly teachers, students and parents may not be able to cope. Conversely, if the change is implemented too slowly impatience or boredom can set in with teachers or students moving on to something else.

If the intended change is too broad and ambitious and teachers have to work on too many initiatives this can be problematic. Alternatively, the change can be too limited and specific so that not much changes in classroom practice.

Student and Parent Involvement in the Change

Students and parents need to be involved in the change or have it explained to them. This did not happen to a satisfactory extent in the 2005 QAT pilot as was clear from students' and teachers' varied explanations of what they thought was being assessed and why.

Leadership of the Change

The leadership of the change needs to be consistent, organised and managed effectively. To a considerable extent the management and policy leadership of the 2005 QAT suffered from a lack of consistency and support.

Implications for Action

Alignment of the curriculum and assessment

Inform teachers of the essential learnings and the standards. A clearer understanding of what students need to know and be able to do, and what they should be given the opportunity to learn, will help teachers focus their assessment practices. The standards will provide a common frame of reference for making judgments about the quality of student work **and** the progress of student learning, while providing a common language for reporting.

Validity and reliability

Address the threats to validity and reliability. Assessments based on a thoughtful process, grounded in a multifaceted body of evidence are more valid and can lead to continuing improvement in teaching and learning. Such a comprehensive approach to assessment prevents high-stakes decisions being made on the basis of a single test. The validity of the passing scores and the achievement levels of common assessment tasks need to be addressed.

Establish processes to develop consistency and comparability of teacher judgments at the local and district levels. With the emphasis on standards as statements that indicate different levels of quality of performance, teachers will need to meet regularly to discuss work for moderation and reliability purposes. This will involve teacher discussions focused on the exemplification of standards in student work. Sufficient reliability will be required for each intended use of the

results. Teachers, principals, students, parents and carers need to know the intended uses of the common assessment tasks; an important consideration for the improvement of teaching and learning.

Teacher support

Provide adequate resources and opportunities for teacher development and valuable learning so innovative research and development can progress. Encourage the growth of teacher professional learning communities both within and across schools. Provide a bank of assessment tasks that includes evidence of student work illustrative of the standard achieved.

Equity and fairness

Give appropriate attention to students with special needs and language differences. Where the quality of student work or progress is lacking provide opportunities for meaningful remediation. Provide explicit rules for determining which students are to be assessed to ensure careful adherence.

Strategic leadership and implementation plan

Provide a coherent implementation plan for the QCAR framework and disseminate widely to the education community. Encourage the development of a dedicated team approach in districts and schools to support and monitor the implementation at the local professional level. At the central level nurture new talent to ensure there is sufficient capacity and to provide the necessary infrastructure for implementation during the trial phase.

To help establish coherence across the policy making arena reinstate the work of the ARFIC or a similar reference committee to provide guidance and support throughout the implementation and to inform policy leadership. This will also help to maintain 'organisational memory' and capitalize on the wealth of conceptual and intellectual development achieved to date.

Ongoing evaluation, research and development

Engage in ongoing evaluation of intended and unintended effects of the QCAR framework, in particular the common assessment tasks. Collaborate with schools, districts and academics for research and development purposes.

Possible areas for collaborative research in the context of policy-making

Equity and Fairness Issues

With the release of the 'Disability Standards for Education 2005' it will be important to ensure that the QCAR framework incorporates these standards. Research and development will help to identify how best to adapt assessment to support the student with special needs.

Moderation and Teacher Judgment

It will be important to explore the ways in which teachers make judgments using standards-referenced assessment in order to inform policy about how these judgments ought to be made.

Ongoing Monitoring and Evaluation

As was apparent in the evaluation of the 2005 QAT important insights and understandings of the intended and unintended consequences of policy change need to be understood to provide strategic policy leadership.

Teacher Support Strategies.

The development of professional learning communities for the implementation and evaluation of the changes introduced by the QCAR framework should be researched to ascertain the level of resources and the policy support required.

Computer-based Assessment

Important developments have been achieved in the development of computer-based assessment for the 2005 QAT and should be continued. Important lessons were learnt for future computer based assessments and emerging technologies have the capability to reduce and simplify the deployment to an insignificant cost. This potential should be harnessed and explored for possible policy implications.

Conclusions

The 2005 QAT has the potential to alert teachers, principals, parents and the community to the important skills of learning. It also constitutes an application of multi-modal assessment with computer-based tasks, constructed response tasks, performance and teacher-generated-assessment. It will be of interest and use to those who were not involved in the pilot as it has provided important insights into other exciting dimensions of the assessment arena. That is, the assessment of key generic skills and the use of computer-based assessment. These findings are useful to inform the implementation of the QCAR framework and to illustrate stimulating innovative steps for possible directions in the development of assessment in the state of Queensland.