The QAT (Queensland Assessment Task): Inventive authentic assessment designed to engage early adolescents of all ability levels.

Karen Harris and John Pitman

National Institute of Education, Singapore and University of Queensland, Australia

Abstract

The QAT is the first attempt at common statewide assessment (apart from basic skills testing) in the compulsory years of schooling in Queensland since the ‘70s.

The 2005 QAT for Year 9 students was made up of three standardised assessment tasks in different assessment modes, namely:

Task 1 – interactive, computer-based
Task 2 – constructed response, paper-based
Task 3 – performance-based.

It was designed to be intellectually challenging and have connections to the wide world. The QAT assessed generic skills and dispositions in the area of *Transforming ideas and/or information* with an emphasis on Processing. To a lesser extent it also assessed Knowledge (facts, concepts and procedures) from the curriculum areas of SOSE (Studies of Society & the Environment) and The Arts.

Authenticity was achieved by supplementing conventional testing with an interactive computer-based approach and a performance task — with the suite to be ‘administered’ over extended time and close to classroom conditions, to challenge a traditional negative concept and aura of ‘the test’. It was a foray into tasking.

The computer-based task required students to interact with the computer at all levels: comprehensive stimulus and support materials, access to questions, and provision of responses. The performance-based component of the QAT was not administered as its viability had been established from recent allied research in Queensland.

Student results showed a high level of reliability and provided intriguing data on sub-group differences.

The paper presentation will include illustrations of the student experience of the computer-based mode of assessment.

Context

The Queensland Assessment Task (QAT) was a recommendation of the Assessment and Reporting Framework Implementation Committee (ARFIC). This committee was established in 2002 to develop strategies for implementing the Recommendations of *The Report of the Assessment and Reporting Taskforce* which was endorsed by the Queensland Minister for Education in February 2002. (Education Queensland, 2003)

The aim of ARFIC was to establish a framework for curriculum, pedagogy and assessment for Queensland schools. The proposed framework was to comprise a set of interconnected and mutually supportive elements:

- a standards map, against which teachers assess and report student achievement;
- a range of assessment methods for collecting evidence of student achievement;
- procedures for collecting statewide data on this student achievement'
- processes for reporting achievement and monitoring progress over time;
- professional learning opportunities for teachers.

The standards map with its set of achievement standards was to provide a mechanism for teachers to capture the results of their assessments and to report student achievement in terms of the following three constructs:

- Knowledges (factual, procedural and conceptual)
- Processing
- Self & Others.

Evidence of student achievement in a nominated domain was to be collected through two distinct though related components:

- student responses to standardised tasks in the particular domain
- student responses to teacher-generated tasks in system-specified categories of the same domain.

The collection of standardised tasks is called a *Queensland Assessment Task* (QAT). The QATs are derived from the intents of existing Queensland Studies Authority syllabuses in the eight Key Learning Areas. QATs will be designed to capture rich information about student achievement in nominated domains and across them, in various mediums, using a variety of instruments, devices and strategies.

In this context, "standardised" means that the task is:

- the same for all students in the cohort;
- undertaken in all schools according to the same list of task parameters;
- marked according to a commonly applied marking scheme.

The pilot study was conducted in 2003. The 2003 QAT comprised a paper-based task and a performance task where the evidence was captured in a paper-based mode. The nominated domains were HPE – Physical Activity/Physical Education and Science – Natural and Processed Materials/Chemistry. The assessment data from the 2003 QAT were assigned mainly to the construct, Knowledges.

The 2003 QAT revealed that there was a lack of curriculum coherence across the pilot schools. This finding led to the decision to move the domain of transdiciplinary higher order thinking forward in the schedule to become the nominated domain of the 2005 QAT.

The intent was to capture rich information about student achievement on generic skills and dispositions in the area of *Transforming ideas and/or information* with an emphasis on Processing, using a variety of modes/mediums.

The design brief of the 2005 Year 9 QAT included:

1. The QAT will be made up of three standardised assessment tasks (SATs) in different assessment modes, namely:
   Task 1 — interactive, computer-based
   Task 2 — constructed response, paper-based
   Task 3 — performance-based (not to be administered)

2. The QAT will be intellectually challenging and have connections to the wide world.

3. The QAT will assess the student's achievements in the underlying generic skills in the domain *Transforming ideas and/or information*.

4. The QAT will draw on SOSE and The Arts.

5. The QAT must provide assessment data on:
   - Processing (transforming ideas and/or information).

6. Where the assigning of grades is not automated, the quality of student work will be judged by teacher-markers who will assign grades after referring to centrally-set, task-specific marking guides containing verbal descriptors of available grades.

Authenticity

In the Queensland context the question of authenticity is given by the following:

- *Is the context of the task appropriate for the cognitive demands? Is the context appropriate for the students?*

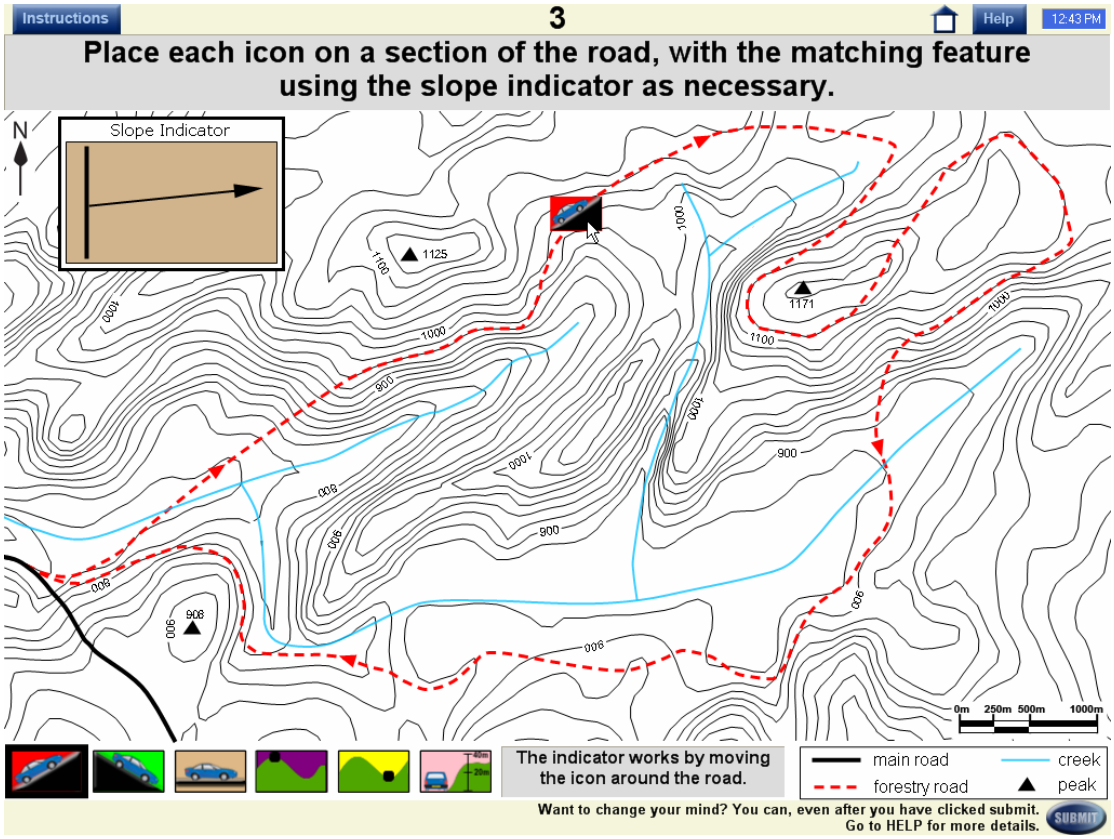
- *Does the context and mode of the task entice and engage students and target realistic audiences?*
- *Does the task have a genuine and valued purpose?*
- *Do students have opportunities to negotiate the assessment?*

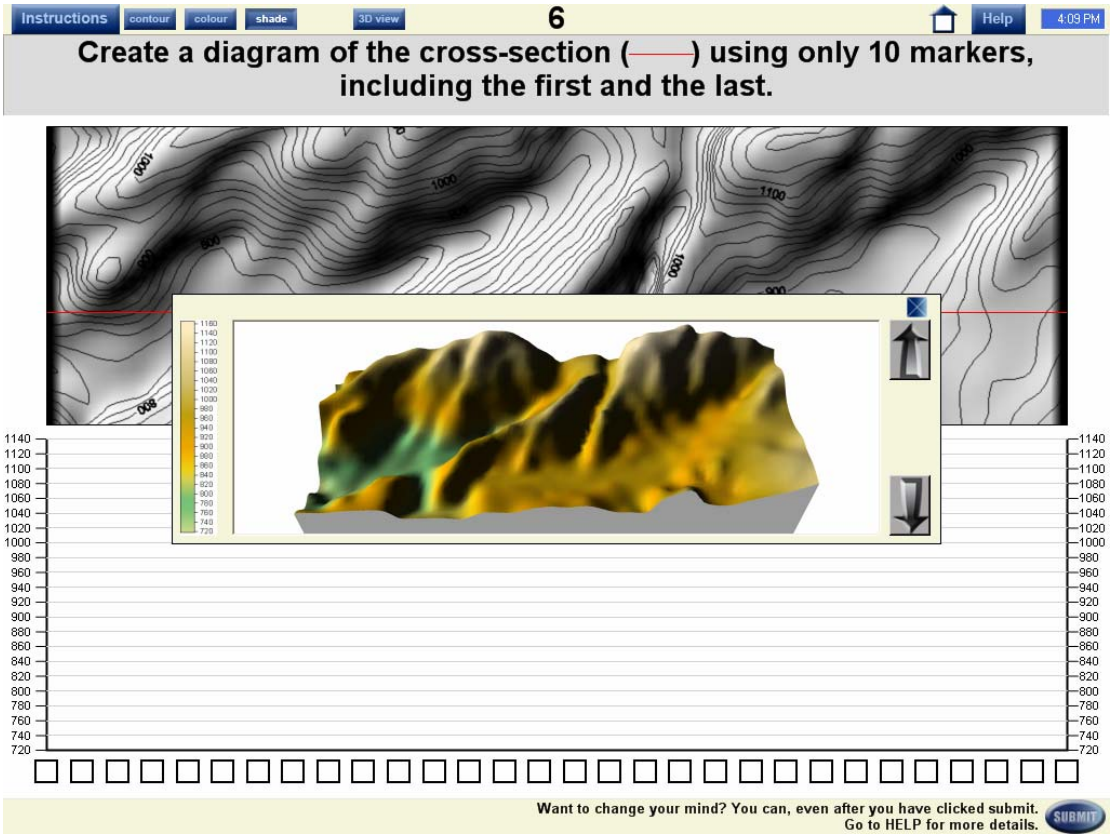
Students must recognise and be motivated by the purpose and relevance of the task. (Education Queensland, 2004)

The 2005 QAT broadly satisfies the above criteria. From the external evaluation, it can be claimed that the computer-based mode in particular engaged students through a task perceived by them as real and relevant. They were employing skills that are universally valued in this technological age. Students were required to problem-solve by accessing information in various forms, contour maps, 3D images, coloured versions of maps, written and graphical representations of the question and integrate the information to complete the question. The context of the computer-based task simulated real data about the terrain which students had to access to draw conclusions about it.

The contexts of the paper-based task and computer-based task appealed to somewhat different student groups. However they were not necessarily mutually exclusive. The pyschometric properties of the QAT indicated that it contained a range of cognitively demanding questions. The students would have had the opportunity to negotiate the assessment in the performance task.

Two questions from the computer-based task are included here for illustration purposes.





The computer-based task was self-contained and required the students to interact at all levels from drawing on comprehensive stimulus from a variety of sources and modes to accessing the questions and then providing responses.

A number of experiments with a computer environment gave the developers an insight into how the modern child interacted with this environment as an assessment mode. They were attracted to certain styles and they rejected certain attitudes and values that perhaps an older generation might hold dear. Students expected to **interact** with the computer environment and that the capacity of the software be the same as they access in everyday life. The environment had to provide quick responses, graphics, help screens, navigation capabilities and to have a professional appearance. The developers were cognisant of these aspects when designing the final computer-based task and tried to build these characteristics into the computer interface.

This project is contributing to an already established field. One aim was to tap into different ways that students learn and to provide opportunities to gather assessment data that could not be captured using a pen and paper mode.

The constructed response was designed with a conventional testing style. The view of test construction was a process of design which aimed to produce an authentic, integrated balanced and quality task covering a range of generic skills and dispositions in *Transforming ideas and/ or information.*

The performance task completed the suite of tasks for authenticity. This was not administered in the schools because the viability of this form had been established

from recent allied research in Queensland. The hub of the New Basics Project was Rich Tasks. The Queensland model of Rich Tasks draws on a range of ideas, including those of Ted Sizer which resulted in Rich Tasks being based on a demonstration model. A finding of the New Basics Research (Education Queensland June 2004) was

> *the assessment system proved to be powerful and effective, … it provides a research-based model for ensuring that students are actually exposed to the intended curriculum; student work is assessed against commonly applied standards; teachers have professional conversations about pedagogy and standards; and there is a reasonable degree of comparability in the grades awarded and reported on.*

Lead-in Activities

An aim was to avoid reliance on only a point-in-time measurement. The suite of tasks were administered over an extended period of time but mostly close to classrooms conditions. A Performance task obviously fits an extended time frame. A lead-in familiarisation activity was centrally designed and sent to schools a month before the administration of the computer-based task. This activity was designed to give freedom to schools to decide how best to use it. A school's decision was its own based on its content coverage and practised pedagogy . The activity was called *Contour Lines – a beginners guide to finding your way around them*. It was not intended to replace classroom teaching of the topic but to complement it.The activities could be used as a homework exercise, revision exercise or an introductory tool . Students could access the same material when they completed the computer-based task. It is important to note that the study of contour lines was included as core content in the syllabus but varying degrees of compliance to the syllabuses meant that the students might not have been equally familiar with this content.

The lead-in activities also orientated the teachers to the requirements of the QAT. A trial of the computer interface was conducted in each school.

The QAT design represented a striving for balance among the real-life experience component, the time-limited component and the point-in-time component while retaining the need for reliable data.

Reliabilty & Validity

The external evaluator stated "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. … Overall the standardised assessment tasks were evaluated as valid and yielded reliable results for reporting". (Klenowski, 2005)

To ensure reliability conventional methods of marking were implemented. Four of the seven questions on the computer-based task were marked by algorithms supported by human monitoring for anomalies. When the assigning of grades was not automated, the quality of student work was judged by trained teacher-assessors who assigned grades after referring to a centrally-set, question-specific marking guide containing verbal descriptors and/or graphical representations of available grades.

Quality control of marking was conducted and markers refocused if necessary. A third marking of a student response was obtained when the first two grades were not within an acceptable tolerance.

The method used to determine the final grades for each student incorporated traditional scoring techniques with appropriate aspects of criteria-based assessment as practiced in Queensland. The following standard referent was used in assigning overall grades.

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 judgments, operationalise these into accurate representations and products.*

The range of grades awarded on individual questions displayed characteristics of discrimination necessary to report results.

The two tasks, paper-based and computer-based, were administered on one occasion. Obtaining reliable results was a high priority. Using the same design criteria another version of the tasks should yield consistent results. The measure used to gauge internal consistency was Cronbach's Alpha which was calculated at 0.85.

Gender

Figure 1 shows a difference between the performance of boys and girls on the different modes of tasks. This result has to be considered in the light of the different contexts on each task.

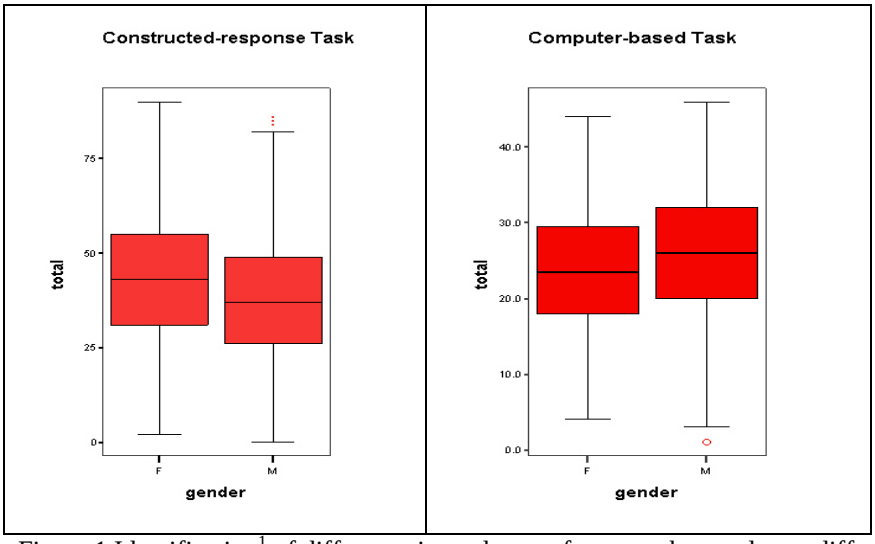


Figure 1 Identification[1] of difference in student performance by gender on different modes

[1] Using pairs of box and whisker plots. The box represents the upper quartile and the lower quartile, the heavy line in the middle of the box represents the median, and the whiskers extend to the nearest value not beyond 1.5x(Inter-Quartile Range). Points beyond this span are considered outliers.

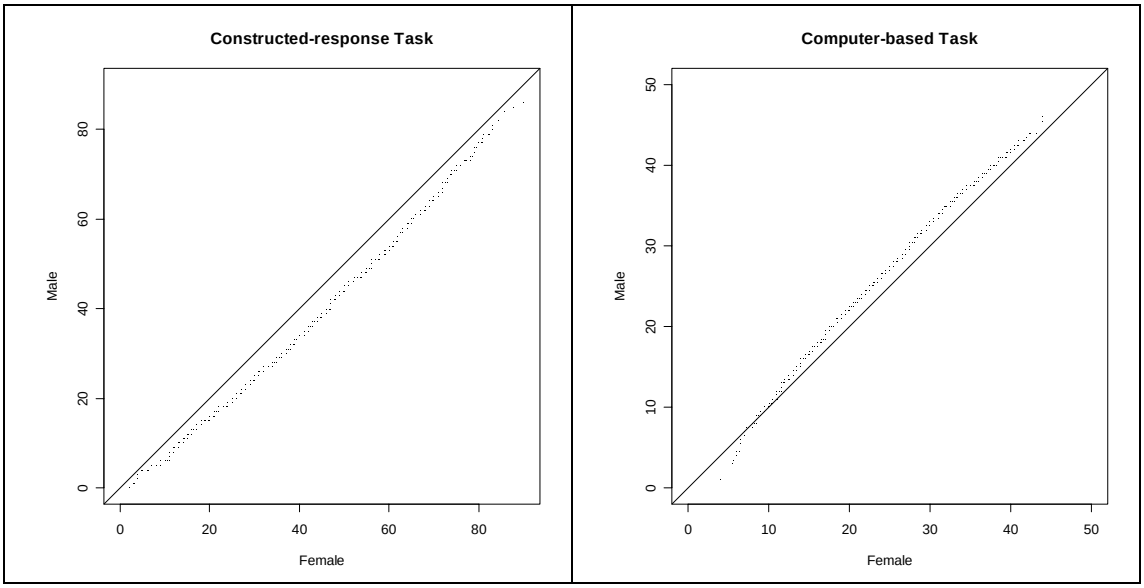


Figure 2 Comparison[2] of males and females total scores on mode of task.

The QQplot for results on the constructed-response task in Figure 2 shows a clear and consistent difference between the results of males and females and the females are always ahead. For the computer-based task the QQplot in Figure 2 shows a clear and consistent difference between the results of males and females and the males are nearly always ahead.

Conclusion

The 2005 QAT was multi-modal, rigorous, and authentic assessing underlying generic skills and dispositions in the domain of *Transforming ideas and/or information.* Both the teachers and students were orientated to the requirements of the QAT. It was a foray into "tasking" by employing the notion of nonspeeded tasks or ones without high reliance on time restrictions and supported by rigorous consistent and documented marking practices and processes to produce valid and reliable results.

[2] Using a Q–Q plot. This graphical data analysis technique compares the distributions of a variable in two groups by plotting the quantiles of one distribution against the quantiles of the other distribution. When two distributions are similar, the plotted points fall on a straight line through the origin at an angle of 45 degrees.

References

Education Queensland. (2003). *Assessment and Reporting Framework*, (draft version June 2003). Brisbane: Queensland Department of Education (Assessment & New Basics Branch).

Education Queensland (2004). *Growing an assessment culture* (issue no. 3). Brisbane: Queensland Department of Education (Assessment & New Basics Branch).

Education Queensland (June 2004). *The New Basics Research Report*, Brisbane: Queensland Department of Education (Assessment & New Basics Branch).

Klenowski V. (2005). Evaluation Report of the Pilot of the 2005 Queensland Assessment Task (QAT) (draft version 2005).