Developing the Enabling Context for School-Based Assessment in Queensland, Australia

Reg Allen
Developing the Enabling Context for School-Based Assessment in Queensland, Australia

Reg Allen
# Contents

Abbreviations ................................................................................................................... v  
About the Series .............................................................................................................. vii  
About the Author ............................................................................................................ ix  
Acknowledgments ......................................................................................................... xi  
Executive Summary ....................................................................................................... xiii  
Introduction .................................................................................................................... 1  
Key Features of School-Based Assessment in Queensland ........................................... 3  
Enabling Context ............................................................................................................. 10  
Drivers for Change ......................................................................................................... 12  
Lessons Learned ............................................................................................................. 13  
Bibliography .................................................................................................................. 17
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>ISCED</td>
<td>International Standard Classification of Education</td>
</tr>
<tr>
<td>NAPLAN</td>
<td>National Assessment Program—Literacy and Numeracy</td>
</tr>
<tr>
<td>OP</td>
<td>Overall Position</td>
</tr>
<tr>
<td>PIRLS</td>
<td>Progress in International Reading Literacy Study</td>
</tr>
<tr>
<td>PISA</td>
<td>Program for International Student Assessment</td>
</tr>
<tr>
<td>QCE</td>
<td>Queensland Certificate of Education</td>
</tr>
<tr>
<td>QCS</td>
<td>Queensland Core Skills</td>
</tr>
<tr>
<td>QSA</td>
<td>Queensland Studies Authority</td>
</tr>
<tr>
<td>READ TF</td>
<td>Russia Education Aid for Development Trust Fund</td>
</tr>
<tr>
<td>SAI</td>
<td>Subject Achievement Indicators</td>
</tr>
<tr>
<td>SABER</td>
<td>Systems Approach for Better Education Results</td>
</tr>
<tr>
<td>TIMSS</td>
<td>Trends in International Mathematics and Science Study</td>
</tr>
</tbody>
</table>
About the Series

Building strong education systems that promote learning is fundamental to development and economic growth. Over the past few years, as developing countries have succeeded in building more classrooms, and getting millions more children into school, the education community has begun to actively embrace the vision of measurable learning for all children in school. However, learning depends not only on resources invested in the school system, but also on the quality of the policies and institutions that enable their use and on how well the policies are implemented.

In 2011, the World Bank Group launched Education Sector Strategy 2020: Learning for All, which outlines an agenda for achieving “Learning for All” in the developing world over the next decade. To support implementation of the strategy, the World Bank commenced a multi-year program to support countries in systematically examining and strengthening the performance of their education systems. This evidence-based initiative, called SABER (Systems Approach for Better Education Results), is building a toolkit of diagnostics for examining education systems and their component policy domains against global standards, best practices, and in comparison with the policies and practices of countries around the world. By leveraging this global knowledge, SABER fills a gap in the availability of data and evidence on what matters most to improve the quality of education and achievement of better results.

SABER-Student Assessment, one of the systems examined within the SABER program, has developed tools to analyze and benchmark student assessment policies and systems around the world, with the goal of promoting stronger assessment systems that contribute to improved education quality and learning for all. To help explore the state of knowledge in the area, the SABER-Student Assessment team invited leading academics, assessment experts, and practitioners from developing and industrialized countries to come together to discuss assessment issues relevant for improving education quality and learning outcomes. The papers and case studies on student assessment in this series are the result of those conversations and the underlying research. Prior to publication, all of the papers benefited from a rigorous review process, which included comments from World Bank staff, academics, development practitioners, and country assessment experts.

All SABER-Student Assessment papers in this series were made possible by support from the Russia Education Aid for Development Trust Fund (READ TF). READ TF is a collaboration between the Russian Federation and the World Bank that supports the improvement of student learning outcomes in low-income countries through the development of robust student assessment systems.

The SABER working paper series was produced under the general guidance of Elizabeth King, Education Director, and Harry Anthony Patrinos, Education Manager in the Human Development Network of the World Bank. The Student
Assessment papers in the series were produced under the technical leadership of Marguerite Clarke, Senior Education Specialist and SABER-Student Assessment Team Coordinator in the Human Development Network of the World Bank. Papers in this series represent the independent views of the authors.
About the Author

Dr. Reg Allen is the Chief Executive Officer of the Tasmanian Qualifications Authority in Australia. He has 30 years of experience as a teacher, subject master, board member, teacher representative, consultant, deputy director, and director in the diverse education contexts of Australia, England, and the United States. He has acted as an expert education advisor in Australia and the United States, including as a leading member of the Framework Research Advisory Group for Queensland’s New Basics in 2001–04. In 2012, the National Center for the Improvement of Assessment (New Hampshire, United States) invited him to provide an international perspective on key technical issues in the use of performance assessments for university entrance at a major colloquium in Boulder, Colorado. His main previous role was as Deputy Director of the Queensland Board of Senior Secondary School Studies, where he played a key role in the design and implementation of Queensland’s Student Education Profile for senior school students, including the integration of vocational education into post-compulsory education. He has produced over 40 publications across diverse education specialties, including national studies and the development of policies for the Australasian Curriculum Assessment and Certification Authorities.
Acknowledgments

Peter Luxton, Acting Director, Queensland Studies Authority
Dr. Gabrielle Matters, former Deputy Director, Queensland Board of Senior Secondary School Studies
John A. Pitman, former Director, Queensland Board of Senior Secondary School Studies
Executive Summary

The State of Queensland in Australia offers a unique example of how an assessment program can embody the principles of assessment of, as, and for learning. Queensland’s assessment program can serve as a model for other countries wanting to explore more effective ways to measure student learning at the secondary level. This is especially relevant given the global discussion on how to create assessments that are more valid, demanding, and not limited by the constraints of traditional, multiple-choice, paper-and-pencil formats. It also is relevant given the discussion on how to improve the teaching force and foster professional development amongst teachers.

In the early 1970s, Queensland replaced its external examinations program with a system of externally moderated, school-based assessment. The school-based assessment in each subject is informed by student achievement during upper secondary—years 11 and 12. To ensure that all teachers judge and grade their students’ achievement using the same standards, samples of student work are reviewed by external moderation panels.

The school-based assessment program is a shared responsibility between a state authority and the schools. The former is an independent government body responsible for the integrity and credibility of the results that it certifies. It ensures that moderation works appropriately, and it ranks students across the state as an input to the tertiary education selection process. Schools are responsible for collecting evidence of student work, judging that work based on the standards, and submitting sample work to external moderation panels. Teachers carry out these assessment-related tasks as part of their normal duties.

What drivers allowed for this unique assessment program to take root? In the 1960s in Queensland, as elsewhere, there was increasing concern amongst educators about the educational value of centrally-run examinations, especially in the context of more students continuing on to year 11 and year 12. Political and social circumstances of the times allowed this concern amongst educators to lead to the establishment of a radically different approach to high-stakes assessment. The idea that the person best qualified to judge a student’s achievement level was his or her teacher became well established early on among key stakeholders. Low direct costs due to teachers carrying out the assessment activities as part of their regular job further contributed to institutionalizing the program.

Other countries aiming to reform their assessment systems by incorporating school-based assessment features can benefit from the Queensland experience. The Queensland approach requires consensus among key stakeholders on the following beliefs or guiding principles: that teachers are best qualified to judge the achievement of their students; that assessment activities should never be separated from curriculum and instruction; and that the construct and consequential validity of assessment results (and the impact on learning) should take priority over a narrow focus on psychometric concerns about reliability and equating and the value of standardized testing.
Developing the Enabling Context for School-Based Assessment in Queensland, Australia

Reg Allen

Introduction

In the global debate about how to improve education quality, traditional assessment systems are often criticized because they typically do not contribute to improved teaching and learning, or because they do not measure the most relevant skills that students need to be successful in life, such as problem solving, critical thinking, creativity, communication, and teamwork. The State of Queensland in Australia has put in place a student assessment program that greatly overcomes these criticisms. This program represents very progressive ideas about education, and about assessment of, as, and for learning. Countries aiming to reform their assessment systems may greatly benefit from the experience of this state, which already has transited the road to a very innovative assessment system.

This case study has three main purposes: (1) to describe the main features of the Queensland program of externally-moderated, school-based assessment at the upper-secondary school level;\(^1\) (2) to analyze the enabling context and drivers for change that allowed for such a program to be introduced and maintained;\(^2\) and (3) to draw lessons for other countries aiming to reform their assessment systems.

To understand the main features of the Queensland program, it is important to understand its context. Queensland is one of six states comprising the Australian federation\(^3\). It is a rich state in a rich country: the third largest economy in a nation that has a GDP per capita of US$60,000. International assessments of student achievement levels show that Australia is among the higher-performing countries in the world, with Queensland having similar performance to the other five Australian states (Australian Curriculum,

---

\(^1\) Queensland has other assessment programs that are not discussed in this paper. For instance, Queensland regularly participates in national (NAPLAN) and international (PIRLS, PISA, TIMSS) large-scale assessments at the primary and lower secondary levels to monitor the quality of education relative to other states and countries. For an overview of assessment programs, see Queensland Study Authority webpage at http://www.qsa.qld.edu.au/3111.html.

\(^2\) For a general discussion of the enabling context for student assessment, see Clarke (2012).

\(^3\) Under the Australian constitution, power over education is a matter for the states, although the Commonwealth government has been increasingly active in this area over the last 50 years.

---

Developing the Enabling Context for School-Based Assessment in Queensland, Australia 1
Assessment and Reporting Authority, 2011). As of 2011, Queensland public schools served nearly half a million students (70 percent of the student population) in 1,237 schools (primary and secondary) staffed by more than 36,000 teachers. The other 30 percent of the student population attended private (independent or Catholic) schools. There are about 500 secondary schools, with around 400 schools with students in years 11 and 12 participating in Queensland’s program of externally-moderated, school-based assessment.

The initiation, development, and maintenance of Queensland’s assessment system have to be seen within the context of its geography and demography. Queensland is a geographically large (about 1.8 million square kilometers) state in the northeast of Australia. While the southeast corner of the state is where the bulk of the population lives, the state as a whole is relatively decentralized. The capital, where the state education authorities are located, is close to the southern border. This means that there are many small, and several large, schools 1,000 to 2,000 kilometers from the “head office.” In the context of the Queensland system, it is essential, for organizational and political reasons, to maintain effective relationships with these distant schools and with clusters of schools in provincial centers a long way from the capital.

In 1972, Queensland abolished its external examinations for year 12 students in schools, replacing it with a program of externally-moderated, school-based assessment (the “Queensland system”) for all schools (public and private), with students seeking certification of their year 12 results. The assessment is used to inform teaching and learning, certify\(^4\) achievement, and make decisions about entry to tertiary education. There are no external examinations for students in schools in Queensland; instead, high-stakes decisions about individual students completing learning programs at ISCED level 3A\(^5\) are made based on their achievements during the last two years of secondary school (upper secondary—years 11 and 12). Teachers judge and grade their students’ work using centrally-set curricular standards and assessment guidelines. To ensure that all teachers grade their students using the same standards, samples of student work are reviewed by external moderation\(^6\) panels. Student results are then officially certified, contribute to meeting high-school graduation requirements, and are used as the basis for tertiary education entrance decisions.

\(^4\) The term “certify” is used here to refer to the idea of recording results on an officially-issued document (a certificate)—a legal document of record—that is accepted as proof of its contents.

\(^5\) ISCED is the international standard classification for education. Programs at ISCED 3A are oriented to direct access to ISCED 5A (university).

\(^6\) Moderation means the procedures needed to bring one school’s assessment decisions in line with another’s before the decisions are finalized. In some systems, this takes the form of “statistical moderation”—using a statistical process. In some, it takes the form of “consensus meetings”—meetings of teachers to agree on the application of standards. In others, it takes the form of “external review”—usually by panel. The term “social moderation” is sometimes used to distinguish the latter two forms from statistical moderation.
In years 11 and 12, Queensland students follow courses of study (subjects with names like “Physics,” “Modern History,” “Japanese”) at a school. Their achievement in these subjects is assessed by the school as they go through the course. At the end of year 12 they receive a formal certificate of their results in these subjects from a central state authority—a statutory board set up by legislation—that testifies to their achievement in each subject against a set of standards that apply to all students doing this subject at all schools across the state.

While the school is responsible for assessing student achievement, the central authority is responsible for making sure that it can certify the results with confidence in their comparability across the state—it is the authority’s task to make sure that, for example, two students with the same result in Physics from schools thousands of kilometers apart have, in fact, met the same standards.

In recent years, Queensland also uses results from its program of school-based assessment as counting towards meeting the requirements of a certificate confirming completion of senior secondary school studies. Since 2009, the Queensland Certificate of Education (QCE) is awarded based on completion of a sufficient amount of study to a satisfactory standard. Results in subjects assessed through externally-moderated, school-based assessment are a cornerstone of this.

The rest of this case study describes in detail the main features of this assessment program at the secondary level; identifies the key factors that allowed for such a program to be introduced and maintained; and draws lessons for other countries aiming to reform their assessment systems.

Key Features of School-Based Assessment in Queensland

In Queensland, externally-moderated, school-based assessment is used for multiple purposes: to support teaching and learning, to officially certify the achievement of upper-secondary school students in the subjects they study, and to provide results that are used for selection into tertiary education.

High-stakes assessment that informs ongoing teaching and learning is a major feature of the school-based assessment program in Queensland. Teachers regularly assess their students in subject-based courses that are aligned with the state curriculum. As in many other countries, teachers use their classroom assessment to monitor their teaching, provide feedback to students on their

---

7 This central authority has had several names over the last 40 years. It is currently called the Queensland Studies Authority—QSA. For this case study, it is most relevant to emphasize that it is a central authority, independent of the state education department. The use of such authorities was and is common practice in Australian states. There were, therefore, no difficulties in establishing such an authority to run the new year-12 assessment system: all that was needed were changes in the functions of a previously-existing authority.
learning and grade student work. Unlike the situation in most other countries, in Queensland these teacher decisions about the standards their students have shown in their work directly drive the final, officially-certified, results, unmediated by combining the teacher’s decisions with results on an external examination or by scaling.

To ensure the credibility, integrity, and quality of the official, formally-certified, statements of students’ achievements, there is a rigorous quality assurance procedure (external moderation) to ensure that all students doing the same subject and being awarded the same grade have met comparable standards.

Responsibilities regarding school-based assessment are clearly delineated. Schools are responsible for ensuring that student achievement is assessed according to official and agreed-upon standards and requirements for content and assessment. The teacher of a particular subject at a school must make decisions about the details of assessment as part of the courses she or he teaches. This means being responsible for the assessment instruments, the timing of assessment events, the criteria for marking, the marking, and the contribution of the results to the school’s final decision about a student’s achievement in the subject.

In Queensland, the teacher is part of the formal, official, summative assessment system and has responsibility to the school for ensuring that decisions about students’ achievements are soundly based on evidence. This presents a challenge: the teacher must at once be on the side of the student and on the side of the interests of the system. At times, this tension appears in expressions of anger about the external panel’s “rejecting” the school’s (in effect the teacher’s) judgments, or in statements of concern that teachers at other schools are manipulating the system.

The formal, official certification process is a shared responsibility between the schools and the Queensland Studies Authority (QSA), which is the state-level authority. While the schools are responsible for assessing student achievement in subject-based courses, QSA is responsible for certifying that assessment results in each subject are comparable for all schools across the state. It is QSA’s task to make sure, for example, that two students with the same result in a physics

---

8 In the first 10 years of the system, students received a result on a scale of 1 (lowest) to 7 (highest). Since the 1980s, students receive one of five results from “very limited achievement” to “very high achievement.” Each of these is defined by explicit verbal standards for each subject.

9 Sadler (1995) remarks that “comparability has to do with whether the performances of all students who are awarded a particular grade in a subject are, within the range of performances associated with a designated grade level, of equivalent quality regardless of which agency undertook the assessment or in which year the assessment took place.”

10 References to “the teacher” should not be misunderstood as meaning that each teacher operates independently of other teachers teaching the same subject at the same school. It is the school that is responsible for the quality of the decisions it makes about students’ assessments. The teacher or teachers of a subject at the school are a means by which it meets these responsibilities.
course from schools thousands of kilometers apart have, in fact, met the same standards. It does this through a rigorous system of external moderation.

This system of external moderation requires a partnership between the central authority—the QSA—and the school. The central authority:

- is set up by legislation
- is independent from government in the details of its operations
- is funded by government
- provides students with certification\(^{11}\) of their achievements
- sets the curriculum framework ("syllabus") for each subject within which schools develop their courses of study
- sets and operates procedures required to ensure sufficient comparability of subject results across the state
- provides tertiary institutions (primarily universities) with a rank order of students in terms of overall academic achievement\(^{12}\) derived from individual students’ subject results
- designs, develops, and administers a test of generic skills (the Queensland Core Skills test—QCS test) with the primary purpose of generating information about groups of students (not individuals).

For each of the high-stakes subjects (generally those of relevance to entry to university education):

- the central authority sets the curriculum framework for each subject (the "syllabus")
- the school determines the details of the program of study in this subject, including the intended program of assessment (the "work program")
- the central authority approves the work program as meeting the requirements of the syllabus, including the assessment that will be used to determine the final result ("the exit level of achievement") against standards defined in the syllabus
- the school delivers the work program
- the school provides to the central authority samples of its decision making about the levels of achievements for each of a small number of students on two occasions during the course (once in year 11 and once in year 12), with additional information, if required, at the end of year 12

---

\(^{11}\) The term “certification” refers to the idea of issuing a formal document that is widely accepted as proof of achievement. For example, a university that requires a student seeking entry to have previously demonstrated a certain minimum achievement in mathematics (a "pre-requisite") can see that this requirement has been met when the student produces (acceptable) formal certification of having achieved at this level in year 12.

\(^{12}\) The term “overall academic achievement” is used here in the sense that a combination (an aggregate or an average) of results across a student’s different subjects represents a measure of achievement overall. Grade-point average is an example of a measure of overall achievement, one based on the assumption that grades are comparable across subjects without any scaling.
through its district and state panels, the central authority reviews the adequacy of the school’s decision making about student levels of achievement on three occasions (once in year 11 and twice in year 12). Such reviews may lead to recommendations to the school for changes in its decisions.

- the central authority certifies students’ achievement in a subject where it is satisfied that the standards required by the syllabus for that subject have been applied by the school to the work of students in that subject.

In years 11 and 12, students may also follow programs in lower-stakes subjects (ones that are not used in compiling the tertiary entrance rank order). These have a less elaborate system of quality assurance.

For the first ten years of the school-based assessment program, for each high-stakes subject, the central authority used a system of consensus moderation meetings, supervised by district and state moderators. At these meetings, teachers discussed work samples, assessment instruments, and standards. Overall, the central authority in those days expected that the distribution of grades state-wide would align with a “normal” distribution and would therefore be roughly the same in each subject. In practice, there were subjects where the state-wide distribution of the highest results was persistently less than required. That is, regardless of the official expectation that there would be the same percentage of top grades in each subject, in some subjects teachers were applying a shared notion of standards and deciding that they did not have enough students who met their idea of the appropriate standard for the highest result—the opposite of the grade inflation that might have been expected.

A key policy that has continued to the present day was present from the start. Each of the work samples taken to moderation meetings was a representative selection of the work of an individual student, rather than a sample of students’ responses to particular assessment instruments. This allows reviewers to focus on the overall standard shown in the work of an individual student rather than a teacher’s marking of a particular test or project.

13 The geography of Queensland, and the number of schools, together with the policy context that the school is the decision maker, pushes the use of a local district model, with state supervision of the comparability between districts. There are costs associated with this, of course, but the use of a central model would remove the need for teachers to take responsibility of their judgments of standards.

14 A standard can be (and is today) defined in the words of a syllabus. A standard is also a set of expectations internalized by a teacher that the teacher applies to the work of students. At these moderation meetings, teachers discussed student work in terms of their notions of standards—syllabuses did not then have any explicit definitions of standards. It is important to recognize that definitions alone do not establish standards—assessment practices and decisions define and reinforce standards.

15 At that time, results were given on a scale of 1 (lowest) to 7 (highest).

16 People unfamiliar with the idea of looking at a body of work against a set of standards often go straight to re-marking individual assessment items or criticizing individual tests rather than looking at the body of work as a whole against the set of standards. Equally, those inexperienced in the criteria and standards approach of the Queensland system feel unable to reach a judgment about standards unless students have done the same set of tests.
Participating in consensus moderation meetings or, today, district review panel meetings, is a core activity of “teachers as professionals,” where they examine evidence about student performance, judge that evidence based on curricular standards, and give advice to schools about grades accordingly. Moreover, teachers secure significant professional recognition (and professional development) through participation in moderation panels.

There were significant tensions in the first years of the Queensland secondary school certification program. Teachers were concerned about the validity of the moderation processes while others were concerned that the teachers were still wedded to traditional notions of what an assessment should look like.

After a series of inquiries and reports, changes were made to the program’s policies and procedures. These included:

- explicit recognition of the purposes of assessment as, of, and for learning
- syllabuses that include a description of higher-order processes as well as content and skill relevant to the subjects
- district and state review panels to accredit work programs and approve standards shown in schools’ decisions about students’ results.

In the early 1990s, the state education authority investigated how effective its procedures were in achieving comparability within subjects. It did this (and still does) by having district panels review a random sample of student work and judge to what extent that work was rated according to the state standards. Recent reports show a level of agreement between the school’s and the district panel’s judgment of around 84 percent. The extent of really serious disagreement is around 7 percent of the sampled folios.

Queensland also has a unique approach to selecting students for tertiary education. While most countries rely on standardized examinations for selecting students, Queensland relies on the same school-based assessment program already described. That is, the selection is fully based on student course work during the last two years of secondary school.

Since 1974, students apply to tertiary institutions with a statewide ranking based on their secondary school achievement. The ranking, called since 1992 an

---

17 Assessment of learning: summative, formal certification; assessment as learning: students reflecting on their performance against criteria for good practice in what is being learned; assessment for learning: formative, gathering information about what is being learned as a foundation for determining what to do next. A quality assessment program combines these functions rather than regarding summative assessment as the one that “really counts.”

18 In this context, “comparability” refers to the idea that students with a particular result in a particular subject have completed work of essentially equivalent quality, even though they have experienced different assessment tasks and have been marked by different teachers at different schools.

19 There are 5 levels of achievement. Each level is divided into 10, giving a 50-point scale. A serious disagreement is one where the difference is one or more levels of achievement and at least 8 points on the 50-point scale.
Overall Position (OP), reflects the overall academic achievement of students based on their achievement in different subjects.

This provision of a ranking in terms of overall achievement reflects a policy principle, not commonly publicly espoused as such,\(^{20}\) that was put in place within two years of the start of the program. The principle is that of designing the system to avoid putting too much pressure on any one component. Too much pressure happens when the results from a single component—an examination, a moderation system—are relied on alone for high-stakes decisions. The consequences of such excess pressure can show up in different ways: for external tests it can lead to a preference for reliability over validity; for moderation systems it can lead to malpractice of one kind or another.

Since 1974, the central authority has administered a test (since 1992 called the Queensland Core Skills test—QCS test) of general academic achievement to all (eligible) year 12 students in late August or early September. The principal purpose of the test is to gather group (school and subject class) information (measures of central tendency and spread) to allow the calculation of this statewide rank order of students for tertiary entrance purposes (the OP).

The QCS test is not an external examination in the sense that it is used for individual high-stakes certification: students receive individual results but, unlike external examinations, these are not used for any high-stakes decisions for the individual—two individuals with identical subject results and the same tertiary entrance rank can have very different QCS results. Students’ rankings or scores in the QCS test are not combined with their subject results.

Until 1989, this test was a 100-item, multiple-choice test kept secure and designed around specifications requiring the testing of verbal and quantitative reasoning. The restricted nature of this test—multiple-choice only—and the fact of its being kept secure—which leads to suspicion about its nature and validity—produced pressures that led to the following significant changes:

- the inclusion of a writing task from 1989
- redesign of the test from 1992 to be a test of 49 common curriculum elements of the Queensland senior curriculum
- the inclusion of a short-response item component from 1992
- release of all test papers each year from 1992.

These changes significantly increased the test’s face, construct, and consequential validity (at the cost of reduced reliability—Cronbach’s alpha is typically around 0.86 to 0.89) and, of course, the costs: a new test every year plus human marking of the writing task and short-response items. At the same time, the test has a positive impact on teachers’ assessment practices and on schools’ focus on the skills (the common curriculum elements) tested.

\(^{20}\)Recognition of this appears in the report of a 1990 inquiry into the system.
Results on the test are an individual grade (A to E, with an A'), reported on students’ certificates and a set of scores.\textsuperscript{21} Schools provide for each student in each subject a number on an interval scale\textsuperscript{22} comparing the achievement of students in that school in that subject. This is now known as a Subject Achievement Indicator (SAI). These subject interval scales have no absolute value—they represent relative achievement within the group of students. Differences between schools within subjects have no impact, provided that each produces fair comparisons of its own students.

Information about the groups in each school in each subject (from those students’ scores on the test) is used in scaling to place of each of those interval scales on a common scale (of overall achievement, not subject-specific achievement). An aggregation of these scaled subject scores produces a school-wide interval scale, and rescaling of these school-wide interval scales puts them on a common basis state-wide. This state-wide interval scale is then used to determine a state-wide rank order. The purpose of the scaling is to create fairness by \textit{removing} any influence of the different levels of academic\textsuperscript{23} strengths across schools and subjects.\textsuperscript{24}

The tertiary entrance rank procedure began after two years of experience with the use of selection based on average grades, in response to university needs for a fair and reasonably robust approach to selection based on “academic merit.” The use of average grades in subjects has the obvious problem that it treats the grades in all subjects as essentially equivalent\textsuperscript{25} and it puts too much actual and perceived pressure on the comparability of individual subject grades.

On a state-wide basis, the scores on the core skills test and separate estimates of overall achievement from the assessments provided by teachers correlate sufficiently well (~0.7–0.75) for the use of the test as a scaling test in principle: in practice there are issues about the validity of its use for small groups and for all

\textsuperscript{21} While these scores have been seen, unofficially, as providing information about school performance, they are not an official indicator of education quality.

\textsuperscript{22} Interval scales show order and gaps. There is no zero in these scales. From 1992, the same range of values is used in all subjects in all schools; the least successful student in a subject in a school is assigned 200 and the most successful in that subject in that school 400. Up to 1991, schools used a scale of 1 to 99—creating the sense that this was a percentage.

\textsuperscript{23} The underlying construct of the state-wide rank order is “overall academic achievement,” not subject-specific achievement. The underlying construct of the QCS test is, correspondingly, overall academic achievement. The test construct is given in more detail in terms of a set of 49 \textit{common} curriculum elements. When QCS test group results are used to compare the academic strength of, for example, a school’s group of students doing Physics with its group doing French, the underlying construct is neither French nor Physics, but general or overall academic achievement.

\textsuperscript{24} Note that an individual’s place in the state-wide rank order is determined by subject achievements, not by that individual’s result on the QCS test: two individuals with the same subject results in the same school, but different QCS test scores, will have the same place in the state-wide rank order.

\textsuperscript{25} However these grades are determined, there is no reason to suppose that this is a good assumption. If the grades are determined “normatively,” they are only comparable across subjects if the cohorts doing the subjects are similar—and this is known not to be so. If the grades are determined by criteria and standards, these are subject specific—a grade in French reflects achievement in French, a grade in Physics does not reflect achievement in French.
groups of students—a generally satisfactory overall relationship may conceal anomalies.

Key assumptions about having schools provide an interval scale within each subject for the students in that school (SAI) are:

- subject teachers can fairly compare the achievement of the students they have taught and assessed in a fairly fine-grained way\textsuperscript{26}
- any errors in this process are not correlated across subjects—each subject teacher assigns results independently of teachers in other subjects.

The detailed procedures of the program over the years have been made more elaborate, driven by the need to ensure that adjustments to students’ results are made where there is any significant departure from these assumptions.

Because the primary purpose\textsuperscript{27} of the test is its use in scaling, there is little pressure on the year-to-year comparability of grades. This is now done through equating based on information from item trials.\textsuperscript{28}

Individual grades on the test are, however, used to some extent in alternative admission schemes for students seeking entry to some university courses. This appears to be growing. The proportion of students being eligible for the standard state-wide rank order is declining (for example 84 percent of year 12 students in 1987, 72 percent in 2001)—which puts pressures on the validity of the procedures used for scaling the subject assessments.\textsuperscript{29} Both of these factors are producing pressures for change.

**Enabling Context**

What ingredients are required for maintaining such a comprehensive and sophisticated student assessment program? This section describes the policies, institutions, human and fiscal resources, and broader social context that have enabled Queensland to put in place its externally-moderated, school-based assessment program.

\textsuperscript{26} The nature of state-wide external moderation is that it can deliver reasonable comparability for grades within a subject if there are around five to seven distinct grades. It cannot deliver reasonable comparability for a 50- or 100-point scale and, more significantly, it is obvious to participants that it cannot. Examinations have a related problem—roughly captured by ideas of the standard error of measurement and standard error of estimate—but this is not obvious to teachers, students, and parents.

\textsuperscript{27} Students can also use their results in the QCS test to show that they meet the requirements of the new certificate—the QCE—for standards in literacy and numeracy.

\textsuperscript{28} Items are trialed, usually in multiple forms, with students outside Queensland. Analysis of the results of these trials (which are secure) is used to estimate the likely property of these items in the actual test. Each year, the test itself is publically released.

\textsuperscript{29} For example, smaller proportions of students being eligible means more small groups (subject classes with fewer than 15 students); more cases where the group of students in a school covers only a limited range of achievement.
The Queensland Studies Authority (QSA) oversees the quality of the program at the state level. This authority was established by law in 2002, but before that year it existed under different names and with different functions. It is a highly stable and institutionalized organization funded by the state, but independent from the government in the details of its operations.

QSA has a stable and qualified staff. In 2011, there were 248 full-time-equivalent employees, mostly based at the central office, and 13 based at district offices across the state. Positions are filled through a mix of permanent and temporary appointments, and secondments.

It is not easy to estimate the total costs of the school-based assessment program in Queensland. From one perspective, a full costing of the program would include the costs of teachers’ time in preparing work programs, assessing student achievement, and contributing to the processes that ensure adequate comparability of statewide results. From another perspective, virtually all of these items are part of the role teachers do or should carry out in the normal course of their professional work. These are not part of the expenditure of the central authority, and are not in the cost estimate that follows.

The following cost estimates are intended to cover all direct costs related to certification of senior secondary school subject results and the determination of a rank for use in selection for tertiary education. The current approximate annual expenditure of QSA on externally moderated, school-based assessment is US$19 million, of which the external assessment for scaling (Queensland Core Skills test) accounts for around US$5.3 million and the external moderation program for around US$10.7 million.30

Over the years, around half of the total expenditure of the state authority has been on the employment of staff, mostly based in the head office, with a few in regional centers. It is estimated that the state authority’s costs have more than doubled while the total number of students has increased by 50 percent since 1987 (Board of Senior Secondary School Studies, 1987; and Queensland Studies Authority, 2009).

This approach to the certification of secondary school students appears to be cheaper than traditional external examination programs used for the same purpose. In fact, the current direct cost per student assessment (one result in one subject for one student) appears to be about US$70. The school-based assessment program in Queensland is less than 60 percent of the estimated cost per student assessment of a comparable external examination program in Australia.

---

30 This is less than half of the current total expenditure of the central authority, which now has a wide range of other responsibilities, including the years 3, 5, 7, and 9 statewide testing; design and organization of the Queensland Comparable Assessment Tasks (QCAT) in years 4, 6, and 9; registration of schools as Vocational Education & Training providers; and syllabuses for all levels of Queensland schooling.
Drivers for Change

The Queensland school-based assessment program first issued results to year 12 students in 1972, following the adoption by the state government of the recommendations of a report from a committee of educators. This committee had been set up in response to public concerns about the validity of external examinations as a measure of student achievement,\(^{31}\) the strictness with which the exams were graded,\(^{32}\) the growing diversity of the upper-secondary cohort, and the need to prepare youth for the labor market.

The public debate that followed the release of the report had its predictable elements, but the reform was supported by the teachers union (which noted that its younger members were more likely to be in favor) in terms of the enhanced professionalism it involved, and was not opposed by the leading university in Queensland. A series of meetings across the state—both public, and with organizations including business organizations and parent groups—appear to have helped assuage potential concerns. The capacity, skills, personal commitment, and seniority of education department personnel attending these meetings were instrumental in bringing people “on board.” The travel distances involved made consensus more difficult, but Queensland’s small population and closely connected networks of influential people made it more feasible.

Politics played an important role in introducing the new school-based assessment program. From 1968 to 1988, the Queensland government was dominated by a political party with a non-metropolitan and rural support base. Education much beyond grade 8 was not common amongst adults in this support base, and unusual for most ministers. The politics of the introduction of the Queensland assessment program must be seen in the context of:

- concern in the government’s non-metropolitan and rural support base about meeting the education needs of their children, and their suspicion of capital city authority
- limited personal awareness amongst government ministers about the later stages of education, with a tendency therefore to accept expert advice where there was no conflict with other political issues
- an education minister willing and able to act on advice from the most senior bureaucrat, who took an educational leadership role, seeking out and acting on advice from prominent education experts.

---

\(^{31}\) It is clear that educators of that period would have recognized their philosophies in ideas such as assessment of, as, and for learning; the importance of assessing the big and important aspects of learning, not restricting assessment to those aspects most easily measured through standardized tests; and the need to improve student learning through connecting curriculum, instruction, and assessment.

\(^{32}\) In particular, the physics examination paper of 1967. Examinations were set by university academics. This particular physics paper was set and marked in such a way that most students failed; an outcome that made no sense to teachers and the community, who saw this group of students as having high ability. After public outcry, the pass rate was changed and results reissued.
The introduction of the school-based assessment program relied on the belief that teachers were the best qualified persons to judge student performance. Stakeholders did not trust external assessments. There was consensus among stakeholders about the importance of empowering teachers in the new assessment program.

Putting a school-based assessment program in place requires time. Teachers need to learn to integrate them into their professional practice. It is necessary to keep on renewing teachers’ understanding, acceptance, and practice of these policies.

Producing a state-level ranking of students based on school-based assessment presents important technical challenges. Although the idea seems quite simple and intuitive, ensuring the validity and fairness of the ranking for all students requires a lot of sophistication. A key assumption of this approach is that teachers are the best-qualified people to compare the achievements of the students they have taught.

Some people doubt that Queensland’s program is, or could ever be, effective or correct. Some think that a standardized test anonymously marked is the only way to ensure sufficient reliability, consistency, and fairness. Support for anonymous, standardized tests persists in spite of the well-known effect of narrowing of the taught and learned curriculum (and hence reduced validity, where the intended curriculum retains the learning outcomes that are not part of the standardized test). To supporters of these tests, the Queensland program must a priori be unworkable, lacking sufficient reliability and rigor. The data about the effectiveness of the moderation process—the annual random sampling study—from this viewpoint cannot be sufficient evidence to the contrary.

But Queensland’s program does work, and has for a long time. Its effectiveness depends not only on the processes and procedures of the program, but on sufficient community confidence—a feature of a culture and a context—that teachers can act fairly and reasonably when reaching judgments about students’ work. Building such confidence takes time, carefully designed procedures, and leadership.

**Lessons Learned**

In one form or another Queensland has had a functioning program of externally-moderated, school-based assessment for 40 years. This section reviews the main lessons that other countries can extract from this experience.

**Introducing and Sustaining School-Based Assessment Takes Time and Effort**

Education leaders in Queensland took advantage of the cultural, social, and political context; a climate for change; and particular events to implement what must seem radical proposals today. The reforms went to the heart of assessment
as cultural practice, practices that have a profound impact on what is learned. Comparably radical changes could be difficult to implement today, for various reasons. For example, there is the short timeframe in which education reforms today are expected to demonstrate their impact, the nature of public discourse about education in 24-hour media, and the direct involvement of political leaders in the details of education decision making.

To introduce such a reform represents a particular challenge. If people do not see that the change is really necessary, then there will not be sufficient pressure to make it happen. On the other hand, if people are convinced that the change is really necessary, then there will be a corresponding pressure for the change to be fully in place immediately, without delay and with results.

The continuing existence of the Queensland program has been enabled by several factors:

- low direct costs to the state government
- the state education authority’s treating schools as partners
- long-term continuity of the senior leadership in the central authority
- the small number of schools in Queensland, which allows for close connections between leaders at the center (the central authority) and the periphery (the schools)
- the state education authority’s being independent and hence able to act in partnership with all schools, public and private.

**Consider Starting Off Small**

The externally moderated, school-based assessment is a high-stakes certification that was introduced for all schools (public and private) at the same time. Assessment’s impact on teacher practices only becomes widespread and significant when the assessments have real impact—that is, when the results are a high-stakes turning point for students. Can such a program be implemented gradually?

A preliminary step could be to start building experience and expertise among teachers with a school-based assessment program which does not have any high-stakes purposes and therefore will not raise public concerns. It then might be possible to implement such a program progressively, drawing on the lessons learned in Queensland in the 1980s about the importance of phasing in significant changes in the assessment. The challenge would come in reassuring those left behind in the previous program that they were not disadvantaged and those in the new program that they were not being subject to experiment.

It makes sense to design a program to be as simple as possible at the outset, allowing room for an increase in complexity as needed. A program that is very complex at the outset may be difficult to introduce and it has no room to develop further.
**Maintain Integrity**

The assumption that teachers and their schools can and do act professionally (when supported by the checks and balances of effective external moderation procedures) includes a vote of confidence for integrity. That is, teachers and schools are expected to resist attempts by persons outside the school to exert undue influence on their decision making. The Queensland program assumes that attempts to exert undue influence will be constrained by intrinsic notions of legitimacy, as well as by the checks and balances that ensure transparent decisions.

The Queensland program’s acceptance of school decisions on the basis of their grounding in the application of explicit standards to the evidence of student work, evidence that must be produced when required, is the process by which decisions are shown to be transparent. Such a mechanism is effective when set within a context where attempts to exert undue influence are not common and are commonly seen as inappropriate, even corrupt. In the late 1980s, the considerable elements of corruption then found in other areas of the state government did not appear to spread to the operation of the upper-secondary assessment program. This may have been helped by the statutory independence of the central authority and its commitment to integrity. Perhaps too, those involved in corruption may not have seen year 12 results as a source of income or influence.

**Emphasize Teacher Professionalism as a Key Ingredient**

From the outset, the Queensland program located not only responsibility for high-stakes assessment decisions at the school level, but also responsibility for the details of the courses and assessment program that students at individual schools would follow. Local responsibility assumed and depended on the professionalism of teachers. It also embraced the idea that learning is best achieved when the teacher is designing and implementing the learning program, determining how information about student achievement will be gathered, and gathering and using that information.

Positioning the teacher as a fair and reasonable arbiter, who applies standards to students’ work, places additional responsibility on the teacher. It also prevents the teacher from taking the traditional role in an external examination program of being in partnership with the student, seeking the best advantage in a contest with the examiner.

There is some concern about an “excessive work load” for teachers. After all, their responsibilities include framing goals for learning, determining how they will gather information about their students’ achievements in relation to these goals, and playing a part in the external moderation required for these assessments to be part of a statewide certification program. Workload represents
a continuing challenge for teacher unions. They must give suitable expression to members’ concerns while encouraging the view of teachers as professionals.33

It might be assumed that the theory and practice of these professional responsibilities would have been a routine part of teacher pre-service courses for a long time. Until the last decade or so, this seems to have been more the exception than the rule.

In terms of in-service education, one of the key lessons of the Queensland assessment program is that the best professional learning occurs as a side effect of teachers’ participation in high-stakes student assessment, whether at the school level or in the external review panels.

The value of having an assessment program based around the ideal of teacher professionalism should not be underestimated. Professional teachers ensure that classroom practices foster the development of the deep learning considered essential for students’ futures. Programs that espouse one view of teachers but imply another in the way they act, or systems that behave as if most teachers cannot be professional, will find that many teachers will live down to this expectation—though there will be honorable exceptions. Programs designed around the expectation of professional behavior will find, over time, that many, though not all, teachers will live up to this expectation.

**Prioritize the Impact of the Assessment on Teaching and Learning**

The initiation and maintenance of the Queensland year 12 assessment is based on an assumption that the effectiveness of an assessment and certification program lies primarily in its capacity to drive excellence in student learning, rather than in its accuracy of measurement. Of course, certification programs must strive to ensure that the results they certify have the level of validity and reliability expected by those who use these results to make decisions.

Of the factors that an education system might hope to influence directly, teacher practices are perhaps the most important. What happens in classrooms—the practices and enacted standards found there—directly affect learning. Managing the practices and enacted standards in the classroom is the teacher’s role. The design of an effective assessment and certification program should therefore take account of the critical importance of what teachers do and how the program can contribute to excellence in these practices.

What teachers do in their classrooms is shaped by their participation in “communities of practice,” that is, the networks through which ideas, understanding, knowledge, techniques, and approaches to learning are introduced, spread, shaped, and changed through experience. The Queensland program creates and maintains highly effective communities of practice, through

---

33 Developments at the national level in 2005 implying a move to an external examination system for Queensland saw strong statements opposing such changes from Queensland political leaders, sector leaders, teacher unions, and parent groups.
which teachers develop, refine, and enhance what they do and how they know what, how, and how well their students are learning.

Judgments of the value of the Queensland program depend on whether it is seen primarily through a psychometric or standardized testing lens, or primarily through a “theory of action” lens. The latter has a focus not on accuracy and consistency of measurement as such, but on the impact of an assessment program on what is taught and learned.

Perhaps the most important lesson to be drawn from the Queensland school-based assessment program is that the integration of curriculum, instruction, teacher/classroom practices, and statewide high-stakes assessment can be done. Making it work requires giving priority to the impact of assessment programs on learning and teaching practices. It means giving priority to validity (including both construct and consequential) rather than to issues of technical reliability and equating. By adopting this focus, countries will be in a better position to have a strong assessment system that effectively contributes to education quality.

Bibliography


———. 2010a. 2009 Data Summary: Year 12 Enrolment and Certification. Brisbane, Queensland: QSA.


List of papers published in the current Working Paper Series


Forthcoming papers

The State of Queensland in Australia offers a unique example of how an assessment program can embody the principles of assessment of, as, and for learning. Queensland’s assessment program can serve as a model for other countries wanting to explore more effective ways to measure student learning at the secondary level. This is especially relevant given the global discussion on how to create assessments that are more valid, demanding, and not limited by the constraints of traditional, multiple-choice, paper-and-pencil formats. It also is relevant given the discussion on how to improve the teaching force and foster professional development amongst teachers.

**Reg Allen, Chief Executive Officer,**
*Tasmanian Qualifications Authority in Australia.*