

An account of the inner workings of standards, judgement and moderation

*A previously untold
evidence-based narrative*

Professor Claire Wyatt-Smith

*Executive Dean, Faculty of Education and Arts
Learning Sciences Institute Australia*

&

Peta Colbert

Research Fellow, Faculty of Education and Arts

2014

Contents

Investigation of Queensland Senior Schooling: Context and focus.....	4
Standards, judgement and moderation in Queensland Senior Schooling: Findings	5
Part 1: Standards formulation and use.....	7
System: standards and judgement	7
Panellists: using standards to arrive at judgements.....	10
Part 2: Quality assuring student achievement.....	16
Work program approval.....	16
Monitoring	17
State and District Review Panel Chair's Report – Monitoring 2014	18
Verification	20
State and District Review Panel Chair's Report – Verification 2013	20
Comparability	23
State Review Panel Reports	23
Comparability LOA Review Summary	24
Unresolved reviews.....	24
Extraordinary reviews	25
Across the span of system quality assurance processes.....	25
Part 3: Assessment systems in Queensland secondary schools – Five Eras	30
Era 1: 1873–1969.....	30
Era 2: 1970–1978.....	31
Era 3: 1979–1985.....	32
Era 4: 1986 to 2010.....	34
Era 5: 2011 to future	36
Stengthening Queensland Senior Schooling: Recommendations.....	39
Methodology and design	43
Design principles	43
Participants	43
Data collected	43
Milestones.....	46
References.....	57

Appendices

Appendix 1	43
Appendix 2	47
Appendix 3	48
Appendix 4	49
Appendix 5	50
Appendix 6	51
Appendix 7	52
Appendix 8	53
Appendix 9	54

Tables

Table 1: Direct costs of externally moderated school-based assessment for QSA	27
Table 2: Expenditure on the moderation system 2000 and 2013	27
Table 3: Senior Education Officers, schools and subject groups 1991, 2000 and 2013	28
Table A1: Data collected to inform the Investigation	44
Table A2: Milestones and deliverables	46
Table A3: Frequency of District and SEO Support information provided in 2014 Monitoring reports	48
Table A4: Frequency of District and SEO Support information provided in 2013 Verification reports	49
Table A5: Comparability supported and unsupported in 2013 for nine subjects LOA SA	50
Table A6: Unresolved reviews by Authority-Subjects 2010-2013	51
Table A7: Extraordinary reviews 2009-2013	52
Table A8: Composite of data collected for the investigation	56

Figures

Figure 1: Syllabus extract on awarding exit levels of achievement	9
Figure 2: Quality assurance processes involving the work of teachers on panels	16
Figure 3: Monitoring reporting of outcomes flowchart	17
Figure 4: Verification reporting of outcomes flowchart	20
Figure 5: Comparability reporting of outcomes flowchart	23
Figure 6: QSA senior schooling quality assurance processes	26
Figure 7: Development of assessment practices on two dimensions, Queensland, 1912–2007	36
Figure 8: Key dimensions of assessment for student learning and reporting with confidence	37
Figure A1: Senior schooling quality assurance processes	43
Figure A2: Data collected across system quality assurance processes	45
Figure A3: Student profile examples	47
Figure A4: QSA Form R3 (2013)	53

Investigation of Queensland Senior Schooling

Context and focus

This paper was commissioned to inform the Review of Queensland Senior Assessment and School Reporting and Tertiary Entrance Processes undertaken by Australian Council for Educational Research (ACER). Specifically, the lead Investigator was tasked to address the hypothesis that the Queensland Studies Authority's (QSA, also referred to as system) approach to specifying assessment criteria and standards remains dominant and salient across disciplines. It was also to consider the holistic and analytic judgement models that operate in District and State review panels.

The original aims were therefore to consider the nature and function of standards in informing teacher judgements as reported through moderation panel processes. In support of this focus, the paper is both empirically informed and conceptual in nature, each in turn examining the nature and function of standards in judgements in senior schooling moderation contexts.

QSA's moderation panel processes as they involve standards-referenced judgements are of central concern as they are the means through which reliability, validity, comparability and transparency are realised in the senior schooling context. The District and State panels have responsibilities for:

- review and recommendation for approval of School Work Programs,
- monitoring teacher judgements using standards (mid-band decision-making) with reference to student achievement evidence from Year 11 (first year of senior schooling),
- the verification of student achievement assessed against exit achievement standards (mid-band and threshold), and certification of student achievement on course completion, and
- comparability with particular feed-forward for improvements in the operation of standards-referenced assessment and moderation in high-stakes assessment.

Consistent with the contract, the guiding questions for the paper are:

- What is the role of Expert Teacher panels in quality assuring assessment tasks for formative and summative purposes?
- How do standards function in panel judgements and decision-making?
- Do panels consider school characteristics and for comparability purposes, issues including like and unlike schools in considering portfolios?
- What are the official expectations about how standards are formulated and how they are to be used to arrive at judgements of quality in a range of disciplines?
- What assessment evidence is brought forth in a wide range of disciplines?
- What is the influence of the Standards Matrix in how standards and criteria are formulated for use in assessment tasks in Years 11 and 12?
- To what extent is the Matrix a controlling influence in the design and terminology of task-specific criteria and standards specifications?
- How do panels within disciplines and across disciplines operationalize the matrix approach, especially in regard to arriving at overall judgments of folios containing assessment evidence of different types?
- How do teachers in panels treat compensations or trade-offs evident in student folios and how do they relate this thinking to the Exit Achievement Standards?
- What are the characteristics of panel judgements using standards at threshold levels, with particular attention to Sound Achievement at the threshold?

For details of the methodological design guiding this investigation, the participants informing the data collection and the timelines for completion see Appendix 1.

Standards, judgement and moderation in Queensland Senior Schooling

Findings

Standards, judgement and moderation provide the lens for this discussion of the data collected to consider the expert panels charged with verifying judgement decisions for senior schooling and the processes in place to support this work. The discussion is organised into three main parts:

Part 1 considers how standards are formulated by panellists to inform judgements about student achievement.

Part 2 considers the quality assurance processes including work program approval, moderation including for monitoring and verification purposes, and comparability.

Part 3 presents a record of the history of externally-moderated standards-referenced assessment as practised in Queensland since its inception and looks back to the past for source information about the origins and foundation principles of the system, as well as potential for reinvigorating it. Like Parts 1 and 2, the discussion in Part 3 also represents an original contribution, though earlier versions have appeared in Smith's doctoral research (1995), and then in Wyatt-Smith and Matters (2007). As the latest historic record, Part 3 connects the main findings from the investigation to a suite of actions for consideration in strengthening the system going forward. It presents the view that the assessment model in Queensland has reached a further critical crossroads, heralded first in 1995 and again in 2007, though earlier action was not taken. The reasons for this are beyond the scope of the paper. It is worth emphasising here that this situation reflects how practice has continued to move well ahead of theory building as a continuing feature of the assessment model to date.

In this introduction readers are offered some background information to the present inquiry. At the commencement of the project it became clear that the work called for access to a considerable body of data. However, data sets or archival records of the type necessary for the work in many cases were not available, though the Investigators understand that some are in development. This situation meant that some data sets necessary for the work had to be purpose-built. The data had to be sourced, compiled, synthesised and examined to inform the discussion in Parts 1 and 2, and as such, this paper represents original work. It breaks new ground in how it constitutes a corpus of data not previously available in the form presented here. Further, this data-building approach was essential to constitute the empirical basis for the analysis. It has enabled a current and new, evidence-based account of the inner workings of the system. This situation reflects the limited nature of the digital assessment records relating to standards and moderation held by the QSA and the lack of sustained research and analysis into the system's quality assurance checks and balances and associated archival records.

The Investigators therefore wish to acknowledge the high level of cooperation of the Acting Director of QSA, the staff of QSA who gave generously of their time and also the teacher panellists who showed dedication and high levels of commitment to moderation for monitoring, verification and comparability purposes. The willingness of these participants to source and provide documents, some online and mostly hard copy, and to share their knowledge of system processes has been essential to the completion of this work.

Findings and recommendations are provided later in the paper. One recommendation is, however, worthy of pulling forward namely that the collection, collation and analysis begun in this work be built into QSA processes and practices as part of the necessary bolstering of the system. The continuation of QSA as a trusted accrediting and certifying agency hinges, in part, on this much needed work. Further, while it may have been appropriate in early implementation for practice to move ahead of theory building, mentioned earlier, the need for the necessary theory building and empirical data systems relating to standards and moderation are now high priority. The fact remains that a fully-theorised evidence-informed account of externally-moderated standards-referenced assessment in Queensland is not yet available, some 40 years after its introduction. The international interest in the model is well recognised, and there should be confidence that such research would be of high policy interest to a wide range of countries currently confronting issues of how to redesign assessment models to bring forth a wider range of assessment evidence (e.g., team work, problem-solving and creativity) than can be produced in time restricted pencil and paper examinations.

Reading the findings: Links to the guiding questions and recommendations

A key feature of the following discussion is the links created to the purposes of the informing paper and the recommendations reached in efforts to strengthen validity, reliability, comparability and transparency of Year 11 and 12 assessments and reported achievement. These links are signalled to the reader in two ways:



Boxes placed on the right edge of the page signal findings directly relevant to the questions guiding the investigation.

Key finding

The key findings heading on the left edge of the page signal direct links between findings and the recommendations.

In addition to these features of the discussion, in many instances, readers are directed to the appendices for substantive data analysis informing the discussion.

The purpose for creating these links is to assist readers as coverage of content occurs in numerous places and discussion of findings is informed by, but not restricted to, the guiding questions.

Part 1: Standards formulation and use

This part of the paper provides comments about how standards are formulated in syllabus documents. It also presents reported and observed practices about how teachers and panellists use standards in forming judgements of student achievement. The discussion extends to consideration of the nature and function of Profile Sheets as a managerial tool for recording judgements against criteria/dimensions and for combining grading decisions on these to formulate an overall or on-balance judgement.

What are the official expectations about how standards are formulated and how they are to be used to arrive at judgements of quality in a range of disciplines?

To what extent is the Matrix a controlling influence in the design and terminology of task-specific criteria and standards specifications?

System: standards and judgement

In looking at the syllabus advice about the formulation of standards and the information supporting teachers about forming judgements on student achievement, the following 18 Authority-Subject syllabuses were considered:

- | | |
|---|---|
| 1. Accounting (2010) | 10. Film, Television and New Media (2005) |
| 2. Biology (2004) | 11. Mathematics A (2008) |
| 3. Chemistry (2007/2014) | 12. Mathematics B (2008) |
| 4. Dance (2010) | 13. Mathematics C (2008) |
| 5. Drama (2013) | 14. Modern History (2004) |
| 6. Economics (2010) | 15. Music (2013) |
| 7. English (2010) | 16. Music Extension (2008) |
| 8. English Extension (2011) | 17. Physics (2007) |
| 9. English for ESL Learners (2007/2009) | 18. Visual Art (2007) |

In looking across the syllabuses, there was *high consistency in use of the matrix cell design* for presentation of the defined syllabus criteria/dimensions and standards. Terms used to refer to the matrix showed slight variation across syllabuses including Standards Matrix or Standards Associated with Exit Criteria. In all cases, the matrix showed a table layout with criteria/dimensions on the first column, and standards at intervals or levels across the page. The intention is to represent the intersection of the criteria/dimensions and standards written as verbal descriptors of quality. The consistent feature in design is that the standards are represented as mid-band descriptors. As discussed later in this paper, this feature—mid-band descriptors—necessarily falls short of guiding teacher's judgements at threshold levels.

There was a *predominance of a continuum representation* of each standard descriptor across A to E standards with common characteristics at various relative levels appearing in each cell. However, the design of the matrix in Mathematics A, B and C paid greater attention to distinguishing features of individual standards with the layout focused on aspects of performance expected for that standard with some cells remaining unfilled. This omission was deliberate so that, for example, B standard could be more readily distinguished from another level. The change in representation was not dependent on the year of syllabus development, that is, later years versus earlier years adopting either approach, but appeared to be subject specific in the subjects viewed (i.e. Mathematics). The carry forward of this to schools was not evidenced with school samples showing the predominance of the continuum representation of standards in criteria.

There are *multiple representations of standards* throughout the system's quality assurance processes. These have potential to compete for teacher attention, detracting from the proper influence of the

standards for awarding grades on course completion. Subject syllabus *Standards Associated with Exit Criteria* provide a five-level letter-grade A to E representation of standards, with a formulaic means for translating this to a five-category VHA to VLA representation. At Verification, the five-category VHA to VLA representation becomes a five-category VHA to VLA representation with ten differentiations of performance at each standard. Thus performance becomes represented against 50-rungs and recorded as VHA3 for example, showing the combination of categorical and relative rung representation.

In schools, teachers are guided by syllabus exit criteria when creating task specific representations of criteria and standards (i.e. the five-level letter-grade A to E). In practice, teachers move from the five-level representation through use of mathematical symbols to a fifteen-level representation with differentiation of performance (thresholds) indicated by use of plusses and minuses (i.e. +, -) with no official guidance provided. This fifteen-level representation of performance shows the combination of letter-grade and symbols. Student Profile examples provided in Appendix 2 show teacher's movement between these various representations of standards, with Verification prompting them to move to the five-category VHA to VLA representation with ten differentiations of performance at each standard for recording grades for summative assessment purposes. Teachers' movement between representation of standards recognised at the school level by students and parents and representations required at system level is an area for further investigation.

What is the influence of the Standards Matrix in how standards and criteria are formulated for use in assessment tasks in Years 11 and 12?

In school's assessment packages submitted for Monitoring, Verification, and Comparability *the matrix cell design was clearly dominant*. However, on several occasions, panellists were observed to be 'grappling' with the school's method for representing criteria and standards especially with their efforts to discern the demands of tasks or questions. The rigour and validity of assessment instruments were recurring subjects of panellists' discussion, especially in cases where tasks were considered to limit students' opportunities to demonstrate achievement against the full range of standards.

Overall, the *language of the system has changed* over time. A few specific examples include:

- use of the term *criteria/criterion* in syllabuses prior to 2010 with a movement toward using *dimensions/dimension* in syllabuses released during 2010 or later, and
- use of terminology of Standards Matrix and Dimensions and standards descriptors (i.e. English 2010) in some syllabuses, with others (i.e. Biology) referring to Standards Associated with Exit Criteria.

The variation in terminology reflects the period of development of the syllabus and approach in place at that time. Recently developed syllabuses use dimensions and Standards Matrix.

Consistency was evident in the category level of criteria/dimensions within some disciplines. For example, Mathematics (A, B and C) had common dimensions, as did Chemistry and Physics. Subject specific characteristics were however clear in the actual formulation of the matrices including subject content requirements at the various levels.

Syllabus advice provided to schools and teachers about the *method for determining exit levels of achievement* was highly consistent in 15 of the subjects examined. The preferred approach to guide student exit level of achievement decisions was the formulaic guide shown in Figure 1 (drawn from the 2007/2014 Chemistry syllabus, p. 30) noting that in syllabuses developed from 2010, the use of the term dimension/dimensions replaced criteria/criterion in this table.

Figure 1 points to the dominance of the analytic approach to arriving at on-balance judgement. The apparent assumption is that the judgement of overall quality of a student folio can be derived by adding achievement on component parts. In effect, the judgement of the whole is the same as the sum of the parts; judgement is a technician operation of combining parts according to the specified

formula.

Awarding exit levels of achievement

VHA	Standard A in any two criteria and no less than a B in the remaining criterion
HA	Standard B in any two criteria and no less than a C in the remaining criterion
SA	Standard C in any two criteria and no less than a D in the remaining criterion
LA	At least Standard D in any two criteria
VLA	Standard E in the three criteria

Figure 1: Syllabus extract on awarding exit levels of achievement

The two subjects forming the exception to this approach are English and English Extension. These subjects took what could be described as consistent with holistic approaches to judgement and decision-making. So in this approach, the judgement of the whole portfolio is based on considering the work as a whole, judging it against the stated requirements of the standards at the various levels. The standards themselves are required to carry the messages about how teachers are to arrive at discriminating judgements that the work satisfies the requirements of the higher or lower standard. In this judgement practice, the overall grade cannot be arrived at by adding or summing up achievement in distinct criteria as component parts. In short, the whole can potentially be more than the sum of the parts. Also of note is that English was the only syllabus that provided *Minimum requirements for sound achievement*. The empirical evidence which was used to inform this supposedly sharpened statement of standard at the minimum or threshold standard is not known.

Key findings

The matrix cell design for representing five standards is the dominant approach in syllabuses and in school assessment instruments. With few exceptions (identified in the preceding discussion), the standards are represented as existing on a continuum with each cell populated. The underpinning assumption is that quality exists on a continuum and that the focus of judgement is on discerning the degree to which prespecified features are present or absent.

Currently two terms are used to record indicators of quality, criteria and dimensions. As a consequence there is terminological variation across syllabuses in how standards are framed. Definitional clarity is needed.

There are two discernible approaches to teacher judgement implicit in syllabus documents namely analytic and holistic. There is limited information in syllabuses about how teachers and panellists are to arrive at overall judgements of quality beyond the type of information provided in Figure 1 above. For example, no information is provided about compensations or trade-offs in judgement whereby stronger features of performance can be used to offset weaker aspects.

While limited attention is given to judgement in syllabus documents, the assumption is made that it is possible to anticipate and prespecify all relevant features of quality. There is no official recognition of emergent criteria or those criteria that may not be prespecified but that may emerge legitimately in the course of evaluating actual student work, be it a single piece or portfolio. The syllabuses and other official materials used during panel activities are silent on such judgement issues, even though judgement research suggests that in arriving at judgements, stated criteria and standards may combine with previously unstated or latent criteria that can influence decision-making.

Panellists: using standards to arrive at judgements

*What assessment evidence is brought forth in a wide range of disciplines?
How do teachers in panels treat compensations or trade-offs evident in student folios and how do they relate this thinking to the Exit Achievement Standards?*

In moderating judgements of student work, panellists drew on a *range of documents* including: the relevant syllabus and standards matrix, student profile sheets and student work samples. During monitoring, panellists also drew on blank assessment tasks in efforts to familiarise themselves with the task for multiple reviewing purposes.

The Student Profile was a key device for mapping the assessment tasks scheduled in the Work Program and for recording judgements for formative and summative purposes through to awarding exit levels of achievement. Across subjects, there was *high consistency in the information contained on Student Profiles* reflecting QSA requirements for the profile to correspond to the assessment plan. Information on Student Profiles included:

- identifying information (school, subject, student),
- assessment instruments indicated by a number and/or topic/unit that aligns with and allows identification back to the Work Plan, sequentially placed according to plan and Year/Semester,
- spaces to record
 - standards achieved for each criterion/dimension for each instrument,
 - on-balance judgement according to each criterion/dimension,
 - Level of Achievement (LOA) for Monitoring, Verification and the recommended level of achievement at exit or course completion, and
- indication of the purposes of assessment, both formative or summative.

Using this information, panellists were able to track the standards (A-E) awarded for individual criteria/dimensions for each assessment at verification and monitoring. Examples of student profiles in four subjects are shown in Appendix 2.

Student Profiles acted as a managerial tool that served two main purposes for panellists, namely as a trigger or cue for closer scrutiny of assessment instruments, and as a formula for combining reported grades. These profiles were used in both monitoring and in verification panels. Examples drawn from observations in Information Technology Systems, Mathematics, and Physics follow:

- One panellist looked across all of the Profile Sheets as a record of teacher judgements with the purpose to “get a feel” for whether one assessment task was more discriminatory than others: “I get a feel across if enough time”. The panellist’s stated purpose was to identify if there was one assessment task common across folios that allowed him to discriminate the work of different standards. It was this task that became a concentrated focus for attention allowing him to manage or sort the considerable volume of material to be reviewed.
- One panellist reported referring to the Student Profile only after judging the student responses: “I only look at the summary and compare to my own judgement of the student. If it meets, then good. If not, I have to go back and look for the places of mismatch to examine the evidence against the standards”. In this instance, the Exit Achievement standards acted as the main reference point for judgement. In other cases, however, the profiles had a pervasive presence providing short-hand recordings of a series of judgements on components of tasks. This reduced judgement making to a formula whereby a string of letter grades with + and - (e.g., C, C+, B-) were used in a type of private compensation or trade-off scheme, with the rules remaining unstated.
- One panellist described the following: “I examine more fully student samples at A, C & E to see if my judgement is in line with the teacher and I can confirm their judgement by the evidence. I then glance at B & D samples. Then I look across the student's profile sheet: A A B = A with an eye on positively feeding back to student.”
- A panellist looking at a Profile Sheet commented on a further use: B B A A = A “I focus on the trend too. If they are lower grades earlier, then get better as the assessment instruments get harder, then I lean toward the higher level. This is important at verification where we have to consider placements – look at the trending of students across the year/s – if one is trending As

and the other is not, that helps with placement decisions". The latter refers to placement within the standard band and relative rungs.

The *dominance of the Profile Sheet* requires further investigation. Student achievement on all assessment tasks are represented on the Profile Sheet. Panellists reported that assessment tasks did not always offer opportunities for students to demonstrate achievement against the full-range of standards. This raises issues of the validity of the assessment instruments themselves. The Profile Sheet in effect sits between the panellists, the assessment instruments and the student work in the portfolios. This can dilute panellists' attention away from the student folios to the process of how to combine recorded teacher judgements across instruments, criteria and effectively 15 standards.

Panellists adopted an *individual approach to working across the assessment products* in efforts to form overall on balance judgements. For instance, examples below [recounted] are provided from Mathematics, Physics, Drama, and English:

- I would look at the task and see where that dimension is represented and then I would work out the emphasis I would give it. So I might say that a D was given in 1B but it was a question that did not show differences in responses – not developed enough of a question or it was very early on [in the year] and the same skills/knowledge are brought forth in a later instrument or in other questions in another assessment task and that was answered much more in depth by the student.
- I look at the Work Program first then all of the student work in each instrument. I average overall.
- I familiarise myself with the task, looking for questions relating to each dimension. Then I go to the student responses. I go through the A sample writing dimensions, then I tick when in questions, then I turn to the overview sheet [Student Profile].
- I look at clean assessments and standards of questions before looking at student responses.
- I get the syllabus out looking at the annotated standards. I look at the blank assessment task or A folio. I examine samples and determine where it sits: "you get an on-balance feel, you just know". I then look at the profile sheet and examine in tasks where the teacher has placed her ticks on the matrix to see whether I agree with the on-balance judgement for the task. I also view the teacher's comments to see whether the comments are at "loggerhead" with the ticks in the matrix. When looking through a sample, I get a good feel for whether the student is meeting the A standard, bumps stand out. Then focus is on interpreting the difference – I look across tasks to see what the student is having difficulty with. They may have been demonstrating a descriptor in another task, so they can do it.
- I familiarise myself with the assessment task. Then I go to the student work and read the teacher's comments, bearing in mind whether the comment is picky. I then look at the match between the teacher's comments and the teacher's judgement. When looking through, plot is not as important to me as implying, so there are key elements of the dimensions that are more important to me.
- I look at the Work Program overall, examining whether the tasks are challenging and appropriate. Then I look through the A student to match descriptors to the work. I examine the work task specifically – if there is a discrepancy, I take a step back and look at the overall folio, then I look at the Profile Sheet. I then repeat this process for a lower standard-level student. After this, I examine a threshold case. If all is ok after that, and I can confirm the teacher's judgements in these three cases, then no problems across the lot.

These examples show a range of approaches to working with the assessment evidence before panellists charged with the duties of monitoring and verifying judgements. Each example reveals an attempt by the panellist to manage the amount of information, applying individual discerning criteria, using compensations or trade-offs, and implicit weightings to tasks [type or sequentially based] or questions [dimension/criteria based].

Panellists' were observed repeatedly referring to the stated criteria and standards with an eye on *matching evidence in student work samples to features of standards*. In all observations undertaken, panellists were focused on finding evidence of criteria and features of standards in students' responses. An example from observations made during comparability shows two panellists scanning for evidence to support the school's judgement. They were observed to point physically to the identified features in the standard undertaking a running check together:

Evidence linked to feature	Panellist 1	Panellist 2
This has _____	Yes	Yes
This has _____	Yes	Yes
This has _____	Yes	Yes

Two further examples of panellists talking during verification showed similar focus:

- “the criteria [sp] specifically states ...”, “where did he do that?”
- “do we have other evidence that the student is as SA versus an LA”.

In instances when panellists viewed deficiencies with assessment task design, especially where it was considered to place limitations on the student’s opportunities to demonstrate a higher standard of work, *panellists sought evidence of that criterion in other assessments* of the student’s work. That is, though the criterion was indicated in the Work Plan to be assessed in the assessment task they were examining at the time, there were observed occasions when the criterion was not evident. As this was viewed as an assessment design issue, panellists sought evidence that the student had met the criterion in a different assessment task. In effect, the panellists worked to compensate for the flaws of assessment task design, where these occurred, to avoid disadvantaging students. This feature of panellists’ work is further discussed in *Across the span of system quality assurance processes* later in these findings.

Typically, *on-balance judgements were reached by combining* the suite of letter grades awarded to individual summative assessment items. In their efforts to calibrate their judgements, panellists drew heavily on the format of the Student Profile as an easy reference record of these grades.

In practice, they relied on a range of different *methods for combining the individual grades*. These included frequency based decision-making, cancelling out and averaging. It as also noted that the overall decision about level of achievement was more demanding in those cases where the quality was considered at *threshold levels*. Consider, for example, the following statements:

- “On balance judgement - all these Cs, then C. If a D and B amongst Cs, then cancel out the D and B, becomes a C.”
- “My maths thinking, five questions worth A, then 2/3 time show an A get an A”.
- “B ____ (_ = no response) = B [student 1]
B B B B B = B [student 2]
These are two different students. Syllabus says B in two dimensions required. Not met by first student. I would still do some sort of averaging (C) as the above two cases are miles apart.
B B E
B B B
Probably both a B but B- for first one and B for second.”
- One panellist had rewritten the teacher judgements as recorded on the Profile Sheet, essentially reproducing a table showing the assessment instrument and the judgement for each dimension without the on-balance judgement of the teacher. In doing this, the panellist ensured that the on-balance judgement of the school did not influence her own on-balance judgement. The panellist was observed to have written on a blank page:

	KCU	IP	EL
1B	D	C-	D
2B	C-	D+	C
3B	D+	C-	D+
4B	C	C-	C
5B	C-	B-	C

She reported that she would look across the teacher’s judgements to form her own overall judgement. In the example here, she said “they had more Cs than Ds, therefore a C student” ignoring relative placement judgements (+ & -).

These data point to how panellists rely on a range of judgement approaches to combine and trade-off reported judgements. More fundamentally, it shows the tension between the panellists working with the school's reported judgements and their own analysis of the evidence presented to them. It was clear that panellists understood their primary role as looking for evidence to support schools' reported judgements. This suite of insights is consistent with assessment research that shows that judgement of achievement can involve the interplay of explicit or stated assessment criteria and standards, latent or tacit criteria, and meta-criteria (readers are advised to see Wyatt-Smith & Klenowski, 2013).

*What are the characteristics of panel judgements using standards at threshold levels, with particular attention to Sound Achievement at the threshold?
Do panels consider school characteristics and for comparability purposes, issues including like and unlike schools in considering portfolios?*

Panellists sought to work in compliance with the official requirements of the standards in various syllabuses. It was also clear that some were challenged by the notion of the strength of adherence to a standard when reaching **overall judgments** as shown in the following English teacher's talk:

I suppose one of the biggest things that we had to overcome, or that some people had to overcome, was not wanting to penalise kids further for perhaps one particular skill that they obviously were weak in, like spelling, or punctuation and grammar, you know, having to look at it in a more holistic light, just honing in and saying 'But she can't possibly be VHA because she can't always spell correctly.' Whereas there's a lot more to it than making a few spelling errors.

Little research has been undertaken in the Queensland approach to standards-referenced assessment or in other assessment regimes into the largely private practice of arriving at judgements that involve trade-offs or compensations where stronger features of performance compensate for weaker aspects. In this investigation, the demands made of panellists were considerable as they worked to reconcile the teachers' comments and grades with their own appraisals of quality, as well as the talk of panel members, where a second opinion was sought. The demands were also increased by the range of quality that was considered acceptable within each of the standards. It is clear that there is a wide range of quality within each of the standards. That is to say, a threshold Sound represents work of discernibly different quality from work at high Sound. Further, threshold decision-making made particular demands on panellists well aware that their decision-making was consequential.

While syllabus and other materials produced by QSA provided detail on determining a standard, making achievement decisions and reaching on-balance judgements, they did not explicitly refer to placement higher or lower within the standard, that is, A+/A/A-. A decision about **relative achievement within a standard** was reported in the Form R3 in 2013, and then in Form R6 where achievement is transferred to a 10-point rung placement within a level. While no formal advice is provided in current syllabuses, a senior QSA officer indicated that production of two videos was being developed that addressed more directly placement within standard, that is, relative achievement decisions. Specifically the content of these materials were identified as pertaining to: *making 'relative achievement decisions' and decisions leading to placement on the Form R6, and making judgments where the student response matches more than one standard.* Videos addressing relative achievement decisions and completing the Form R6 are available through QSA's website (see www.qsa.qld.edu.au/29442.html).

During observations, much discussion occurred about relative placement on individual criteria or more specifically, on a feature or element of that criterion. One example occurred between two panellists in Physics with the focus of the discussion generally about placement decision-making before moving to consider a student's achievement to decide if it was higher or lower within C standard on one element:

And if there's consistency, if you can see it is, typically, you know, you open up a sample and you look at it and it's got typical of the B standard descriptors, then it's a B standard. If it's consistent with a little bit extra ... you might look at it as a top. And then, you know,

there is, I guess, around the thresholds there's some judgement to be made about is it consistently B standard or is it a little more A standard, so it becomes a low A rather than a high B, or VHA/HA. But that's a judgment to be made with an eye on the standards descriptors.

[The panellists then indicated, by physically pointing, movement within the C standard while talking about a feature of the student's work. Relative placement was discussed as needing to be moved lower on the feature of the standard.]

"That should be further down".

Criterion	A	B	C	D	E
<i>Knowledge and conceptual understanding</i>	<p>The student work has the following characteristics:</p> <ul style="list-style-type: none"> reproduction and interpretation of complex and challenging concepts, theories and principles comparison and explanation of complex concepts, processes and phenomena linking and application of algorithms, concepts, principles, theories and schema to find solutions in complex and challenging situations. 	<p>The student work has the following characteristics:</p> <ul style="list-style-type: none"> reproduction and interpretation of complex or challenging concepts, theories and principles comparison and explanation of concepts, processes and phenomena linking and application of algorithms, concepts, principles, theories and schema to find solutions in complex or challenging situations. 	<p>The student work has the following characteristics:</p> <ul style="list-style-type: none"> reproduction of concepts, theories and principles explanation of simple processes and phenomena application of algorithms, principles, theories and schema to find solutions in simple situations. 	<p>The student work has the following characteristics:</p> <ul style="list-style-type: none"> reproduction of simple ideas and concepts description of simple processes and phenomena application of algorithms, principles, theories and schema. 	<p>The student work has the following characteristics:</p> <ul style="list-style-type: none"> reproduction of isolated facts recognition of isolated simple phenomena application of simple given algorithms.

This example shows the close scrutiny that panellists gave to matching the work to the required specifications of the matrix.

Other panellists, in commenting specifically about factors they consider in threshold judgments, referred to the demands and modes of assessment instruments and the conditions under which the assessment was completed:

- If threshold – "I also look at whether the assessment is an assignment or an exam – I give more weight to the exam as they can get help with an assignment."
- Also consider assignment versus exam in threshold – "I give more weight to exams, but I also consider exams. If someone gets an A for the assignment and a C for tests, it is definitely a pass – but where is it placed, B or C?" [This remained an unanswered question.]

While relative placement decisions were made, during verification panellists were observed drawing on other means for *confirming placement*. One example, drawn from two panellists in Physics during verification shows an attempt to combine letter and numeric grade indicators in order to confirm placement.

A	B	C	D	E
14	11	8	5	2

Criterion			
(C-)	KCU	$7+8+5+11 = 31/4$	7.5
(C)	IP	$11+7+5+12 = 35/4$	8.1
(C-)	EC	$2+9+7+12 = 30/4$	7.3

"Just doing the #s which is an indication."

In this example, the panellists had reviewed the student work and arrived at a provisional judgement of placement as threshold C. They moved to then write the letter and numerical scores as they confirmed the judgement.

Overall, it was observed that *panellists use a variety of means to confirm their own judgements* including those based on formulaic calculations, their own evaluative experience, and their talk and interaction with other panellists. Different means for confirming placement were evident within panels. Some observations across each of the quality assurance processes include:

- During monitoring, panellists' talk tended to concentrate on supporting the school's judgements and confirming the placement of folios at mid-band standard. In some cases, direct comparisons were made between school placement decisions and the individual panellists' own evaluative history in their own schools in locating folios within the bands. One example provided by two

panellists making a comparison to another school provides a salient example: “The other school [folio] I reviewed was much higher than this, however, both are an A standard”.

- During verification, panellists adopted very explicit means of checking or confirming their own judgements using a range of methods, such as in the example above and in calibration checks with other panellists.
- During monitoring, verification and comparability, panellists were observed to draw on their own evaluative experience and they reflected on teacher judgements recorded for other schools. When placement became a paramount focus of discussions, these relative observations tended to increase, though there was no explicit talk of like and unlike schools.

*What assessment evidence is brought forth in a wide range of disciplines?
How do panels within disciplines and across disciplines operationalise the matrix approach, especially in regard to arriving at overall judgments of folios containing assessment evidence of different types?*

While many individual differences were identified in panellists approach to working with standards and assessment evidence in making judgements, a number of discipline differences were discernible. These observed differences related to the mode of assessment evidence and the varying emphases placed on certain aspects of the student portfolios.

- *Mode of assessment evidence* – A number of disciplines provided student assessment responses in electronic format. Panellists would work between hard copy materials in portfolios (Student Profile, assessment tasks and student responses) and electronic files provided on CD or DVDs (movie files, photos, website files). These files were observed to be used with copy materials in folios in conjunction with performance components in other media. This was observed in several subjects including dance, Drama, Music, Film, Television & Media, and Physical Education.
- *Relative emphasis on parts of the assessment package* – Panellists in Drama reported giving particular value to teachers’ comments to inform their valuation of the student’s assessment piece and for contextual information. When recorded student performance was the assessment evidence, teacher’s comments provided useful information to support the panellists’ appraisal of the performance captured in media files or other software. Panellists voiced the concern that when viewing recordings of student’s performances for assessment, they could miss subtleties in the live performance. The teachers’ comments addressed the difference between a live performance involving an audience and a record, digitally captured.

In cases where there was electronic evidence, the performance was watched routinely by one or more panellists who would give a running commentary of how the performance was matching against the expected features of quality. The critical need for ensuring panellists have suitable software is highlighted here. There were observed instances where panellists could not access the student response due to technological and software difficulties.

By far the dominant mode for panellists to work with assessment evidence was through print, hand-writing their judgements and advice to schools. This reliance on print or hard copy extended the time necessary for return of information to QSA. The much needed move from a paper-based system is discussed further in Part 2.

Key findings The Student Profile is a dominant assessment product that is used as a managerial tool by panellists. The profile records letter grades on which teachers rely, with no formal guidelines about how to combine what is in effect fifteen standards.

Panellists use a variety of means to confirm their own judgements, including those based on formulaic calculations, prior evaluative experience, and their attempts to calibrate the judgements across panellists. The latter would increase in cases where the judgements of panellists did not agree with each other or the school.

Judgements of student achievement at the threshold rely heavily on the panellists' tacit or in the head standards.

The dominant mode for panellists work is print-based, however, some subjects where live performance or demonstration is assessed, have moved to submitting digital records of student work. These include, for example, Dance, Music, Health and Physical Education, and Drama. In these subjects, teachers on panels worked between achievement data in different modes, with panel time including panellists' review of the performances themselves.

Part 2: Quality assuring student achievement

This part of the findings considers the four quality assurance processes undertaken by State and District panels. Consideration is given to the means through which reliability, validity, comparability and transparency is realised in the senior schooling context through these processes. Further, this part considers factors that span the four processes.

What is the role of Expert Teacher panels in quality assuring assessment tasks for formative and summative purposes? How do standards function in panel judgements and decision-making?

The work of teachers on panels occurs during four distinct quality assurance processes (see Figure 2). The discussion that follows focuses on each of these processes in turn, drawing on a range of data sources to inform the findings. The processes include: 1) Work Program Approval, 2) Monitoring, 3) Verification, and 4) Comparability.

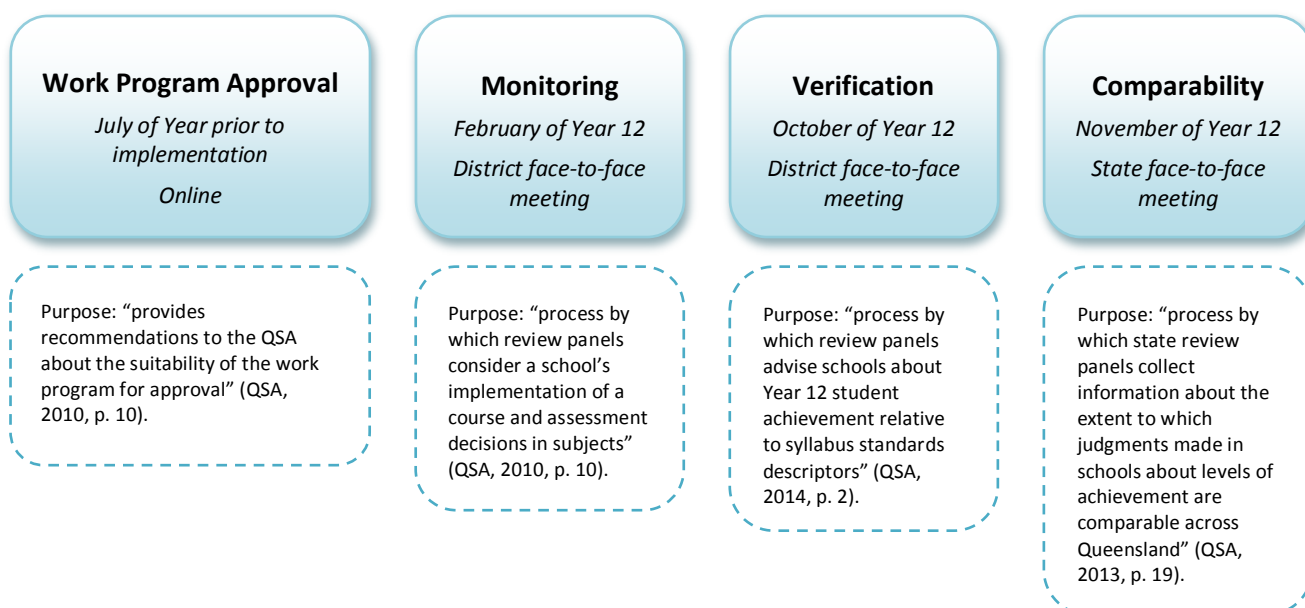


Figure 2: Quality assurance processes involving the work of teachers on panels

Work program approval

Work Programs are submitted the year prior to implementation in Year 11 (around July) by all schools in two circumstances: 1) when there is a new syllabus or 2) if the school is newly offering the subject to students. Work Programs include the following information: course organisation, outline of intended student learning, assessment plan, and sample student profile. At the time of Work Program submission, assessment instruments are not provided to the QSA for approval. This means that students are undertaking assessments for both formative and summative purposes without the instruments themselves having been checked or validated prior to use. This omission is of note, and

as indicated, has direct implications for how panels work to compensate for task design weaknesses, where these occur.

Outcomes of the Work Program approval process through the district and state panellists are fed-forward via the Form R2 to the school and QSA. The SEO uses this advice to schools to inform professional development content, or in some cases, instigate other actions as required, an example being school contact. It was reported that *contact with schools occurs throughout the year*, with 75% (approximately 366 schools) receiving some contact at least once per year. Following receipt of the Form R2 advice from the District Panel Chair, QSA makes contact with the school to “touch base” about actions they undertook as a result of the panel advice on the Work Program. The timing and requirements of Work Program Approval merit more detailed investigation. This work could focus on the assessment schedule in particular, how tasks for formative and summative assessment purposes build over the period of study, and the schools’ refinements or improvements to assessment instruments where these have occurred.

Key finding

Work Program approvals do not include formal systematic quality assurance processes applied to assessment instruments before they are implemented. While a school receives feedback about the quality of the curriculum, the assessments themselves are not subjected to review. This omission is of concern for several reasons. At the heart of the current model is teacher judgement and in particular, how judgement is standards-referenced. Currently, however, the review of Work Programs does not focus on teacher judgement and decision-making in assessment task design and the use of standards as applied to individual tasks and to folios. In effect, curriculum, teaching and assessment evidence are dislocated in the approval process.

Monitoring

District Panel Chair meetings for Monitoring purposes were held on 18-21 February 2014. Observations were made at the following six panels over this period: Chemistry; Drama; English; Information Technology Systems; Mathematics A; and Physics.

During these observations, discussions occurred with the six Panel Chairs and 29 panellists. These observations and interviews with QSA personnel are drawn on in the following discussion, along with additional materials prepared and provided to the Investigators. *Reporting of outcomes* from Monitoring (see Figure 3) occurs in the following way:

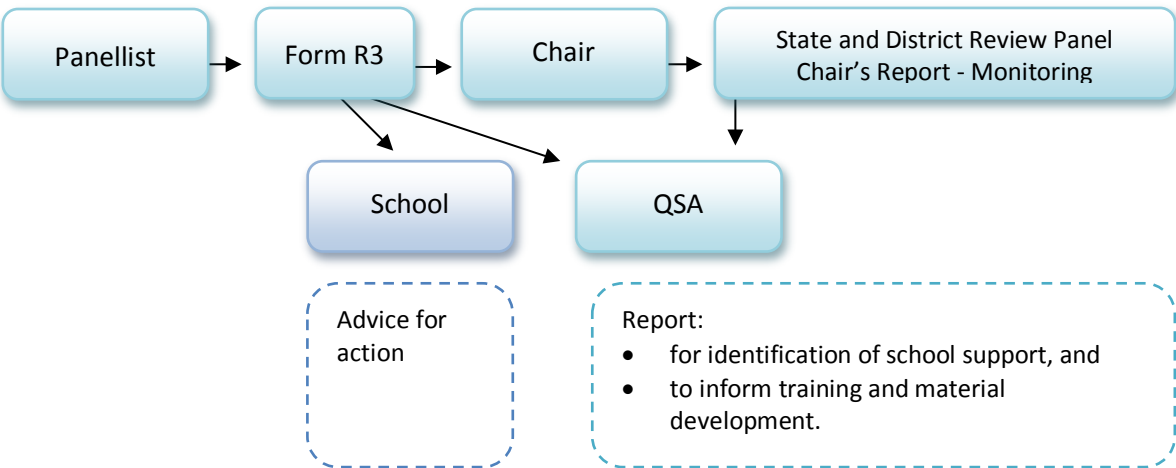


Figure 3: Monitoring reporting of outcomes flowchart

The flow of information as represented in this figure was confirmed by QSA staff in the Quality Assurance Unit as an accurate representation on 30 April 2014.

State and District Review Panel Chair's Report – Monitoring 2014

The *State and District Review Panel Chair's Report – Monitoring 2014* for 45 subjects were provided by QSA to the Investigators on 12 March 2014. The following analysis concerns only the sample provided.

The reports prepared by the District Panel Chairs contained information about the quality of school submissions and emerging issues of concern in schools' assessments and judgements. In turn, the information in reports was provided to the SEO for follow-up. The information requested by the form includes: school code, school, and tick select option for identifying issues with Assessment, Matching of syllabus standard descriptors, or interim Level of achievement (LOA) with no space allocated for further information. The following discussion considers the findings of a frequency analysis¹ of issues identified in the reports (see Appendix 3 for the frequency table). In summary, the reports considered were from 45 Authority-Subjects considered by 367 District Panels with 3,857 panellists during monitoring 2014.

First time analysis undertaken by the Investigators showed that from these 367 district reports, there were:

- 767 reported instances of issues related to the school's assessment,
- 664 reported instances of issues related to the matching of syllabus standard descriptors, and
- 521 reported instances of issues in school submissions related to interim LOA decisions.

Of the schools identified for follow-up by the SEO, 73% were relative to assessment design matters, 63% were relative to matching syllabus standard descriptors, and 77% were relative to interim LOA judgement, which indicates that schools were identified for follow-up on one or more issue. As there are only 489² Year 12 schools undertaking Authority-Subjects, it is also evident that schools were identified for follow up on one or all three matters in one or more subjects. Of note is that there were 95 districts (26% of districts) where no significant issues were reported.

On the completed reports, *provision of specific information relative to the issues* of concern was limited. As a result of the information requested on the form (described earlier), the majority of reports provided little or no information about the issue. Where there was information provided, it was apparent that Chairs adopted a similar manner of reporting to that requested on the Verification Chair Reports paperwork (discussed in *Verification* next).

When comments were made, several demonstrated *signalling to the system* through the SEO the importance of assistance required by the school in that subject. Examples included:

- "high concern"
- "big concerns"
- "(Priority 1) They need help."
- "Many issues. Student work is not following task sheets. Task sheets are not following the work program. Criteria sheets are not following the syllabus. Understanding of the syllabus is not present. I have sent a letter as an offer of help – they need a lot of it."

Other information provided on these forms related to:

- limitations of the assessment to provide opportunities for students to meet the related criteria/dimensions (or features of),
- alignment of assessments to the approved Work Program,
- alignment of the assessment to syllabus expectations, and
- explanations of circumstances obtained through contact with schools – for instance: "Lots of staffing issues last year. They should hopefully be back on track for 2014. I will be offering support throughout the year".

¹ It is noted that to date there has been no comprehensive analysis of District State Panel Chair Reports – Monitoring as a collective. This investigation represents the first consideration of the body of reports for identification of issues.

² Figure provided by QSA 10 March 2014. Schools identified have one or more students in Year 12 as at 2014 (QSA data).

The monitoring feedback also included instances of reporting non-alignment to the school's Work Program with the Year 11 assessments as implemented throughout the year.

A discussion with an SEO about the use of these forms for *informing school support* provides an *important contextualisation of reported issues*. It also brought to light the vital role played by various written reports, all currently existing in print form only. This importance was identified by one SEO who referred to the how these reports served to *filter* the flow of information across the QSA, panels and schools. The Chair's Report was characterised as acting as a 'first filter' reflecting advice provided to schools and identification of those schools to be followed up by QSA through the SEO. Using the identification of the school, the SEOs then referred to the panel's specific advice reported to the school on the Form R3. The Form R3 acted as the 'second filter'. The Form R3 was reported to be interrogated for the identified schools to ascertain the "severity and extent" of the issue of concern. In this way, monitoring acted as a mechanism for "feed[ing] into further support mechanisms". Effectively these filters worked to assist in managing the flow of information from in excess of 3,700 panellists across the 13 districts through to 20 SEOs in QSA who oversee 2-3 authority subjects and associated authority registered subjects (provided 29 January 2014 by P-12 Implementation Branch QSA).

What this meant in practice was that for a district in subject English where 23 schools were reported as having an issue requiring system support, upon interrogation of the Form R3 by the SEO, only 4 were confirmed as requiring support. Instances such as these are reflective of many factors including Chair experience or syllabus stage for instance and therefore, SEO expertise is relied on to interrogate and suitably action reported issues of concern.

While weaknesses have been highlighted in this discussion, strengths of the system are also noted as exemplified through the tracking of schools for support by the SEOs. An area for further work includes the work of SEOs, the action taken in schools in response to panel feedback, and closing the loop on necessary changes, especially as these relate to assessment task design.

Key findings

Monitoring provides an opportunity to provide feedback to schools on assessment undertaken for formative assessment purposes after these have been completed. Similarly, panellists have opportunity to identify issues with summative assessment after they have been implemented in classrooms and grading, as discussed further below. This timing for the validating or checking of assessment tasks to determine construct validity and cognitive demand is raised here as a matter for priority attention.

Currently monitoring focuses on expected quality at mid-band levels of the standards. This is a missed opportunity to strengthen the consistency of panel judgements at the necessary threshold levels. The rigour of monitoring would be increased with a shift in focus to the required features of quality for standards at threshold levels, and further at the tipping point into the higher standard. This would sharpen judgement capability across the full band width of each standard.

Verification

District Panel Chair meetings for Verification purposes were held on 21 October 2013. Observations were made at the following nine panels over this period: Accounting; Biology; Business Communication & Technologies; English; English Extension; Geography; Mathematics B; Music; and Physics.

During these observations, discussions occurred with the Panel Chairs and panellists. These and interviews with QSA personnel are drawn on in the following discussion, along with materials prepared and provided to the Investigators. *Reporting of outcomes* from Verification (see Figure 4) occurs in the following way:

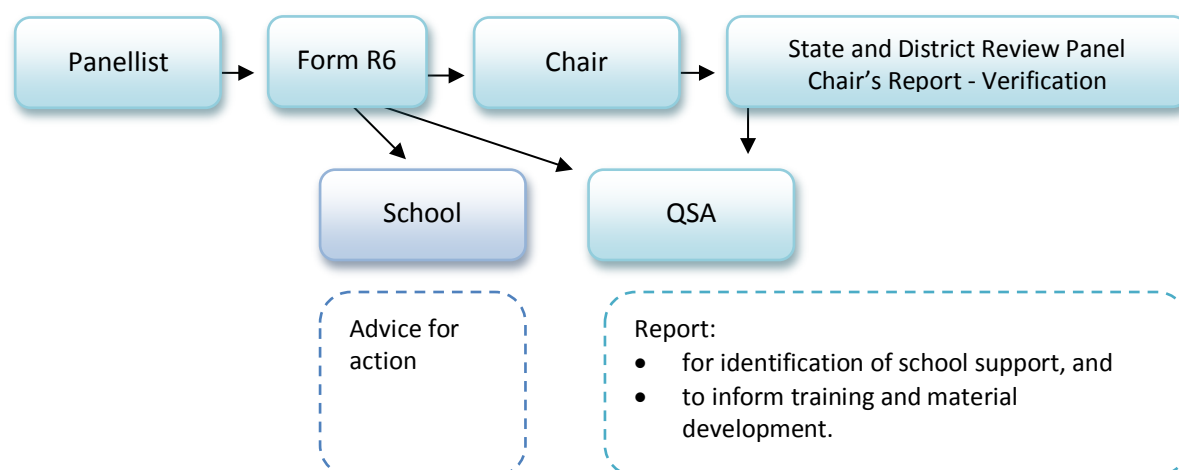


Figure 4: Verification reporting of outcomes flowchart

The flow of information as represented in this figure was confirmed by QSA as an accurate representation on 30 April 2014.

State and District Review Panel Chair's Report – Verification 2013

The *State and District Review Panel Chair's Report – Verification 2013* for 45 subjects were provided by QSA to the Investigators on 26 February 2014. The following discussion concerns only the sample provided.

The reports prepared by the District Panel Chairs contain information on significant issues in school subjects that require follow-up by the Chair and SEOs in QSA. The requested information includes: school name, tick select option to indicate issues relative to LOA or Assessment, with space allocated for Chairs to provide specific information on the issue. The following discussion considers the findings of a frequency analysis³ of issues identified in the reports (see Appendix 4 for the frequency table). In summary, the reports considered were from 45 Authority-Subjects considered by 344 District Panels with 3,717 panellists during verification 2013.

In these 344 district reports there were:

- 485 reported instances of issues in school submissions related to LOA decisions, and
- 374 reported instances of issues related to the school's assessment.

Of the schools identified for follow-up by the SEO, 77% of the reasons for contact related to LOA, and 60% were concerned assessment design matters. Follow-up for both LOA and Assessment were required for action post-verification. As there are only 489⁴ Year 12 schools undertaking Authority-Subjects, it is evident that schools were identified for follow up on both matters for one or more

³ At the time of writing there had been no comprehensive or sustained analysis across District State Panel Chair Reports – Verification. This investigation represents the first consideration of the body of reports for identification of issues.

⁴ Figure provided by QSA 10 March 2014. Schools identified have one or more students in Year 12 as at 2014 (QSI data).

subjects. Of note is that there were 100 districts (29% of districts) where no significant issues were reported.

LOA issues were related to discrepancies between the evidence in the student work and the proposed standards assigned by the school. That is, differences were identified between school judgement of the standard of student work and that of the panellists. Comments made about LOA include “school judgement in [criteria/dimensions]”, “inconsistent application of criteria”, “decisions regarding [criteria/dimension] were not substantiated with evidence in the folios”, “upward movement”, “downward movement”, and “the evidence in folio A does not match syllabus descriptors for an A standard” as examples.

Assessment issues covered a range of design and judgement matters including:

- *suitability of demands of assessment tasks* – “assessment didn’t provide students with opportunities to meet syllabus standards across all dimensions”, “ ‘A’ level opportunities limited”, and “tasks lack challenge”;
- *syllabus interpretation* – “not following sampling requirements”, “use of syllabus conditions on tasks”, “categorising elements of task to correct dimensions”, “assessment instruments do not meet syllabus requirements”;
- *matrices* – “criteria sheets not derived from exit standards”, “matching exit standards to task”, and “criteria sheets do not reflect syllabus standards descriptors”;
- *teacher judgement* – “poor alignment of standards to assessment task”, “matching evidence to descriptors”, “lacking evidence”, “incorrect application of standards”, and “inconsistent judgements against criteria”; and
- *teacher experience* – “lack of understanding of standards and the intent of the syllabus by the teacher”, “they need further support on their assessment”, and “new teacher”.

During verification, the observed practice was that panellists identified and reported issues to the Chair who, in turn, made a decision about reporting to QSA. The mechanism for doing this is the *State and District Review Panel Chair’s Report – Verification* form providing a record of the *Significant Issues*. This process of reporting of issues raises several considerations. First, the issues carried forward from panellists through to QSA were determined by Chairs. It was reported that Chairs consider the information provided to them by panellists to determine the nature of the issue and the appropriateness of the information prior to inclusion. This points to the role of the Chair as a critical mediator in the flow of information across District and State Panels, individual schools and QSA. Second, the information recorded about the issues varied in both quality and utility. The feedback ranged from Chair’s simply using the QSA provided ‘tick’ option (LOA, Assessment) with no or little elaboration of the nature of the issue to detailed description and references to related evidence in school submissions.

These observations provide an opening for reconsidering as among the highest priorities the selection and training of State and District Panel Chairs, and more broadly, the opportunity for them to become accredited for the role within QSA.

In Monitoring, a process for filtering information reported by panellists and Chairs was described. The *process for identification of schools for support* during verification similarly involves a process of filtering. Chair Reports act as the first filter, with the Form R6 acting as the second filter. However, it was reported that the purposes for verification were different from those of monitoring, and subsequently the Chair communicates more directly with the school in resolving issues of student placement within standards and on the related rungs. This once again highlights the important role of the Chair and recognises the expertise and time commitment of the Chair to this position.

As suggested earlier, *effective communication between the SEO and the Chair is essential* for effective feedforward from panels to the schools and in turn to QSA. This communication channel also highlights the depth of expertise required of the SEO in terms of discipline knowledge and assessment expertise. This extends to the use of standards in judgement and decision making and assessment

task design. Once the panel report is received, SEOs draw on the report and Form R6 within their subject portfolio to identify actions for supporting Chairs and schools. The SEOs map actions from the paper-based reports onto a recording system (word table, or Excel worksheet) to track contacts with schools. In this process the Investigators were advised that up to the time of this investigation, there was no holistic review and analysis of the performance issues across subjects for the purposes of identifying recurring and emerging system-wide assessment issues. Clearly the reports could be used for this analysis and represents a valuable opportunity for rigorous data mining as routine in QSA operations. Such analysis would inform both operational and strategic priorities.

The above observations regarding the role of the Chair and SEOs working with school staff points to the current heavy reliance of standards-referenced assessment on both assessment and discipline expertise, and the relational cultures within panels and between panels and schools. This was evident in cases where the finalising of agreement on school submissions at verification was undertaken by the Chair outside of panelling activities. The reliance of QSA on the Chair's critical work in resolving issues (e.g., judgements of level of achievement and matches between evidence and standards), merits further investigation.

Key findings

It is currently not routine practice to undertake comprehensive systematic analysis of District Panel Chair reports to inform State Panel Chair Reports – Verification. This is the case even though verification records provide critical information about key assessment knowledge, design issues, syllabus application and standard usage on student performance for the system to self-monitor.

The Chair has a key role during verification, working with schools to finalise agreement on reported achievement against defined standards for the award of exit achievement. This action is taken outside panel meetings, as occasion requires. In this regard, Chairs carry with them valuable corporate memory of the operation of panels.

SEO expertise is critical for ensuring appropriate and effective follow-up of issues identified through reporting. Their work serves to connect the moderation processes between Chairs and schools. As discussed later in this paper, it is timely to review how the vital operational role of Chairs is best undertaken, given that it is routinely undertaken by teachers and Heads of Department who have full-time teaching and leadership roles.

Comparability

State Panel Chair meetings for comparability purposes were attended by the Investigators on 4 and 6 November 2013. Observations were undertaken at the following 13 panels over this period: Biology; Dance & Drama; Economics; English; Film, Television & Media; Home Economics; Mathematics A; Mathematics B; Mathematics C; Modern History; Music; Physics; and Visual Arts. These observations along with interviews and documentation provided by QSA inform this discussion.

Reporting about the outcomes of comparability takes two forms: 1) State panellists complete Form C2 recording the outcomes of judgements for comparability purposes, and 2) State Chairs distil the information recorded by panellists on the Form C2, synthesising the strengths and areas for attention across the state. State Chairs and SEOs work together in preparing the published report of comparability namely the State Review Panel Reports. The *flow of reporting* during comparability (see Figure 5) occurs in the following way:

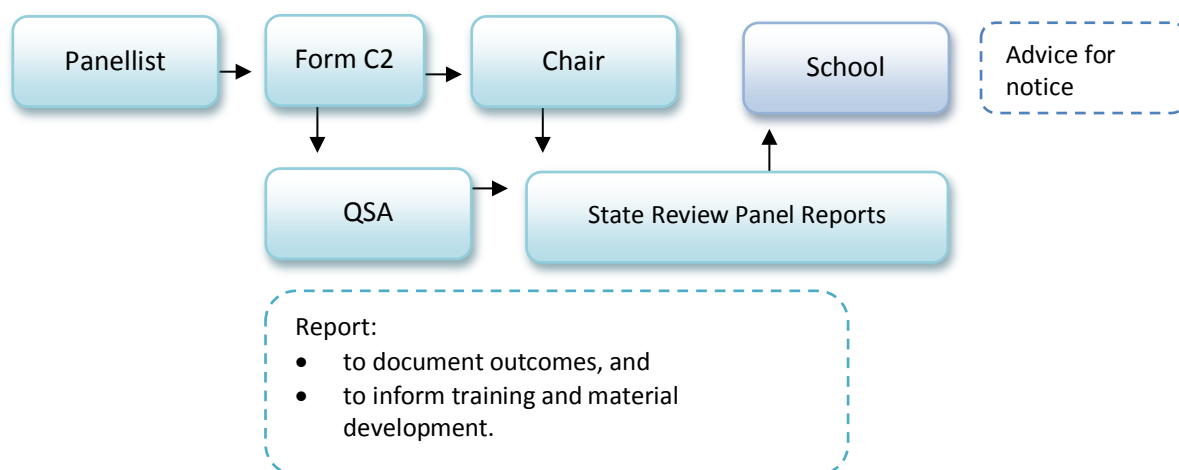


Figure 5: Comparability reporting of outcomes flowchart

The flow of information as represented in this figure was confirmed by QSA as an accurate representation on 30 April 2014.

In addition to undertaking comparability checks across threshold samples across the state for each subject, State Panels are charged with examining unresolved cases. Reviews of unresolved cases are required when the District Panel did not reach agreement on the LOA after three panellists' reviews.

Each of the processes of reporting outcomes of comparability is discussed further below.

State Review Panel Reports

This Investigation considered 2011, 2012 and 2013 *State Review Panel Reports* prepared for public dissemination of the outcomes of state comparability of judgement of student outcomes. The public reports present outcomes from the State Panel meetings for comparability purposes by subject.

From a review of the information provided across 34 subjects, advice in these reports was shown to adopt two descriptive styles:

- *advisory*, whereby information about the moderation process was provided through a number of examples prefaced with should, need to, must etc. Reports based on this style do not provide sufficient information on the performance of the State throughout the moderation processes.
- *reflection*, whereby comment was made in general terms about the performance of the State.

In the main the reports lacked specific information about the outcomes of comparability (i.e. number of supported/unsupported threshold judgements) instead opting to draw on general qualifiers; reports contained little or no evidence about frequencies or illustrative cases. Instead, comments of a general nature were used and included: evidence was found, significant variety, significant agreement,

significant alignment, sometimes incorrectly matched, in general, in most cases, and generally, for example.

The majority of the reports examined in 34 subjects across the three years of publication reviewed (2011-2013) incorporated both styles of reporting. As such, they could not be described as evidence-based accounts of validity, reliability and comparability and so their utility for informing panel operations is problematic. This observation suggests a need for evaluating the purposes and audiences of these reports.

What are the characteristics of panel judgements using standards at threshold levels, with particular attention to Sound Achievement at the threshold?

Comparability LOA Review Summary

Comparability LOA Review Summary (Form C2) records the outcomes of panellists' review at comparability. The Investigators were provided with Form C2s for the following nine subjects in the Sound (SA) LOA in 2013: Agricultural Science; Ancient History; Drama; English; Information Technology Studies; Mathematics C; Physics; Study of Religion; and Technology Studies. In these subjects across the 92 Districts providing assessment packages as samples to the State Panel, 159 (88.8%) were supported and 20 (11.2%) unsupported (see Appendix 5). The unsupported samples were restricted to four subjects. While the sample considered here is small, it included unsupported LOAs where the student samples were judged to be of higher quality, rather than at the threshold. When focus is placed on samples of work not supported in the judged LOA, it does not always follow that samples were judged at comparability as representing a lesser quality of work according to the standards. Instead, some samples were identified for higher placement in the standard.

Considering the high-stakes nature of senior schooling, judgements that have been agreed by the school and subsequently supported by two district panellists yet are unsupported at State Panel in comparability checks point to a number of concerns. These include understanding and application of standards by the District Panel, reliability of recorded student outcomes, and common or shared expectations of performance for standards at thresholds. At this point in the process of assuring student outcomes, unsupported LOAs do not affect the student's exit LOA. Comparability checking as currently undertaken does not constitute an element of quality assurance in the operation of standards-referenced assessment. Its primary purpose is to feed-forward to inform panel training and material development.

As these samples were submitted as meeting the requirements of standards at thresholds, comments elsewhere in this paper relating to the importance of how teachers' and panellists' understand and interpret thresholds when making judgements of student work according to standards (see *Standards formation and use*) apply here also. Overall it is clear that the focus on judgement of standards at the threshold requires priority attention. To complement this, there could be a focus on the features of quality that characterise the full band of the standard—at the lowest level (lower threshold), and the highest level (tipping point into the next standard). Suitable chosen illustrative exemplars of student folios could address this need.

Unresolved reviews

Unresolved reviews are required at State Panel during comparability in cases where the District Panel cannot reach agreement on the school's judged LOA after three panellists' reviews. The number of unresolved reviews for each Authority-Subject (38 subjects) for four years is 70 in 2010, 62 in 2011, 56 in 2012, and 52 in 2013 (see Appendix 6 for subject frequencies).

These figures show a reduction in unresolved cases by 26% in the last four years, with a reduction in cases reported each year over this period. This trend could be a positive indication of the traction of:

- valid assessment design,
- reliable application of standards for judging the quality of student work; and

- quality assurance processes.

The frequency of unresolved reviews is also an indicator for further investigation. Seven subjects (out of the 38) had more than 10 recorded unresolved cases in total across the four years. The three subjects with the highest number of cases were: Visual Art (29), Biology (15), and Legal Studies (15). Tracking of reviews longitudinally with comparison against other factors such as syllabus timeframe, Chair experience, panellist experience, and training could be considered when interpreting this data. Data systems including historic records of moderation outcomes are limited, restricted to comparability checks and records of extraordinary reviews. Public confidence would be better served through establishing data systems that permit the tracking of issues relating to the use of standards in moderation over time in individual subjects and curriculum domains.

Extraordinary reviews

Extraordinary reviews are conducted after comparability when agreement has not been reached between the State and the school. Information provided by QSA on 26 February 2014 documented the extraordinary reviews required in the last five years as: three in 2009, two in 2010, and nil required from 2011-2013 (see Appendix 7 for breakdown by subject).

The fact that no extraordinary reviews have been undertaken in the last three years is however a positive outcome of the work of QSA and schools in reaching agreement on judgements of student work and grades for award on exit or completion of a course of study. The work of the Chairs and SEOs in communicating panel advice and working with schools in reaching agreement is recognised in this outcome. Such agreement of itself should not however be taken as evidence of reliability, construct validity of assessment instruments or comparability.

Key findings

State Review Panel Reports provide little specific information on comparability outcomes in subjects. The reported information is of variable quality and utility in informing interventions at the school level.

Comparability checking as currently undertaken does not constitute an element of quality assurance in the operation of standards-referenced assessment. Its primary purpose is to feed-forward to inform panel training and material development.

A reduction by 26% in unresolved reviews is a positive trend, however, this finding should be considered in relation to outcomes of comparability reporting discussed earlier.

No extraordinary reviews for the last three years indicate the system secures agreement with schools on student outcomes at certification. The preceding discussion about unsupported judgements of placements within standards however shows a need for concentrated focus on the characteristics of quality at threshold levels, and as also suggested, the quality expectations at the top of each standard. Very High Achievement 4 to 10, for example, could be a particular focus, along with the top levels of High and Sound Achievement.

Across the span of system quality assurance processes

Organisation of panellists for review of school submissions varied based on the Panel Chair. In the main, panellists were organised around three criteria: experience, standards, and partnerships. There were strategies in place to ensure novice panellists were working with their more experienced colleagues, and that panellists experienced variety in terms of the standards they were examining and the colleagues with whom they worked in meetings.

The *experience of sitting on a panel* was reported as a means for improving teacher's knowledge of the criteria and standards, and for developing notions of quality at different standards. Working with

the standards and criteria, exposure to assessment instruments and professional conversations were reported to be invaluable for personal development. The recognised value of this was captured in a District Chair (DC) focus group:

- DC1: What it is, I say, is it's great professional development doing this process.
I think it makes us much more literate in terms of assessment, what makes good assessment.

And I have to say, I was, before I joined a panel I was sort of floating around not knowing what ...
- DC2: Hit and miss.
- DC1: Yeah, but once you get on a panel you can actually see, you get to see everyone else's work and get ideas. And you go, *oh*.

The *QSA system of eternally-moderated standards-referenced moderation is process-driven and paper-based*. The schedule of quality assurance processes spans the entire year as shown in Figure 6 (described by senior staff during interviews; Investigators’ representation). Central to this work are the key moderation processes: Work Program Approval, Monitoring, Verification and Comparability. The organisation required to support the 13 Districts and over 3,700 panellists meeting to discuss student submissions is extensive, as are the preparations required to support the State panel processes. In addition to the specific moderation processes described, the quality assurance schedule includes preparation of professional development and training, the annual Moderation Conference and a number of key areas of development.

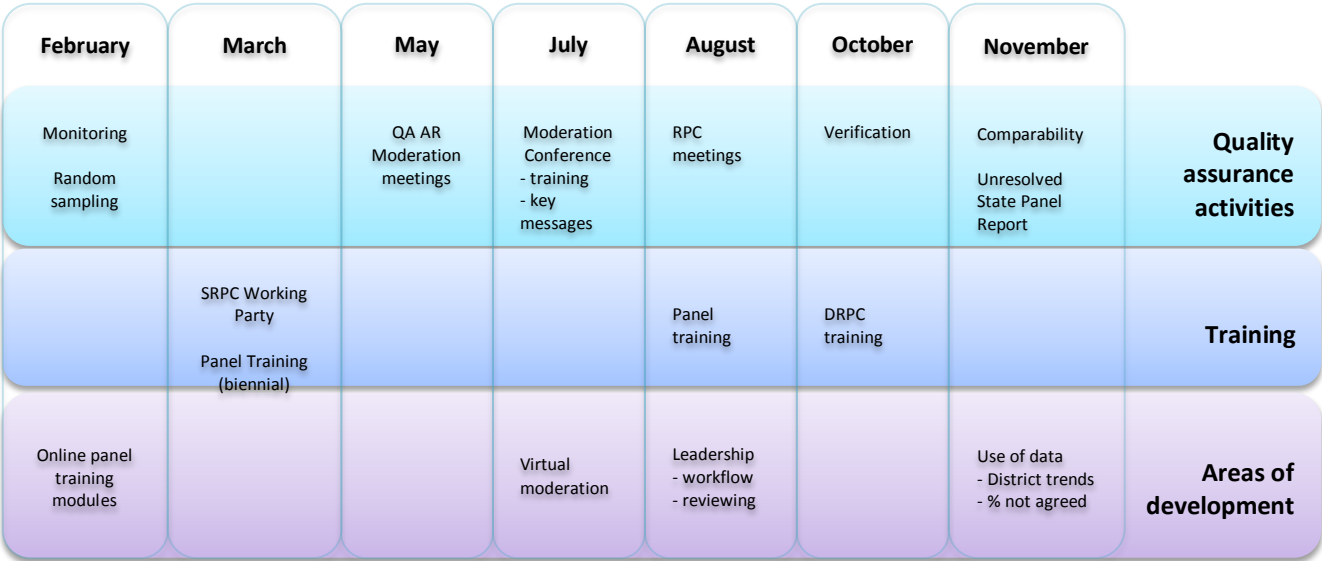


Figure 6: QSA senior schooling quality assurance processes

The paper-based nature of the system impacts efficiencies, with over 8,000 Form R3s (Monitoring) and Form R6s (Verification) reported to be submitted each year. The quality of completed documentation, as signalled elsewhere, was varied, impacting on the usefulness of the information in all reports provided to QSA. While intensive focus and energy are committed to processes, limited attention is given to data analysis at the key points of monitoring, verification and comparability. With the volume of information provided in paper-based form, an online system for immediate data capture and database storage is an advance that would make it possible to feed-forward monitoring and verification information, and permit longitudinal tracking of achievement data.

Information provided by QSA shows that direct costs have remained at a stable 22 to 24 per cent of the annual administration grant as shown in Table 1 prepared by the Policy Coordination Branch QSA. Of note in this table is that additional direct costs have been identified in some years and omitted in others, and amounts of expenditure have shifted markedly during this period. While detailed analysis of budget provisions for moderation in senior schooling is beyond the scope of this report, some

evidence-based commentary on budget was considered appropriate, in light of observations about current limitations in the system.

The flat lining of overall direct costs in Senior Schooling assessment system is of note when considered in relation to QSA responsibilities in curriculum and testing. These include developing curriculum resources for P-10 to support the move to the Australian Curriculum and Achievement Standards, and NAPLAN testing and reporting.

Table 1: Direct costs of externally moderated school-based assessment for QSA

Description	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	Total
QCIA	67,954	119,485	108,637	129,430	138,389	49,809	33,944	35,358	30,670	32,087	27,482	773,247
Services Co-ordination - Services	1,458,461	1,457,250	1,502,753	1,618,009	1,576,438	1,803,106	2,004,515	2,108,453	2,241,191	2,326,395	2,333,718	20,430,290
Panel Training	83,827	65,166	58,868	91,150	64,372	88,958	78,490	131,716	163,074	92,805	101,447	1,019,873
QCE Branch Directorate - Services	1,508,667	1,779,402	1,823,795	2,013,773	2,156,144	2,676,236	175,277	189,947	-	-	-	12,323,240
Panel Chairs Conferences & Seminar	214,235	220,688	205,106	223,789	237,045	266,215	222,241	227,536	207,523	216,955	194,861	2,436,195
Work Program Approvals	299,279	176,072	358,560	459,371	134,774	330,649	515,789	301,336	385,835	281,076	175,885	3,418,627
Monitoring	489,868	604,085	602,008	695,608	680,820	766,694	803,452	850,637	628,745	645,083	839,721	7,606,722
Verification	491,804	610,759	625,239	726,406	727,385	802,016	855,204	873,600	1,010,115	1,082,761	978,228	8,783,520
Random Sampling	20,853	24,001	16,410	33,155	35,884	45,683	69,701	70,211	111,275	87,291	97,695	612,160
Comparability Meetings	95,314	98,311	103,786	106,067	104,645	116,912	128,809	130,039	130,857	117,473	108,232	1,240,446
School Support - Services	27,811	17,326	17,497	16,544	9,163	5,642	1,965,665	2,253,898	2,497,221	2,608,892	2,582,993	12,002,652
QA Auth Registered Subjects	-	-	-	-	81,551	128,260	50,553	97,519	59,092	63,149	21,607	501,732
Learning Projects - WC&S	-	-	-	-	66,659	43,758	10,410	3,233	1,436	-	-	125,496
Total	4,758,072	5,172,547	5,422,660	6,113,303	6,013,269	7,123,939	6,914,049	7,273,486	7,467,035	7,553,968	7,461,871	71,274,198
Administered Grant	21,803,900	22,973,100	24,555,000	26,108,000	28,242,000	29,616,000	30,211,000	31,217,000	32,413,000	33,410,623	32,221,000	312,770,623
	22%	23%	22%	23%	21%	24%	23%	23%	23%	23%	23%	23%

Notes:

- * Administered Grant amounts verified by Department of Education, Training and Employment. Amounts exclude one-off supplementation for redundancy payments.
- * SES staff are not included in the figures above
- * No indirect (corporate) costs are included in the above figures
- * 2001-2002 were \$5,044,131 (23% of the Administered Grant funding for BSSSS, TEPA & QSCC)

While costs in most categories have increased, some significantly, notable exceptions are Panel Chairs, Conferences & Seminars, and Work Program Approvals. Both of these categories show a reduction in expenditure in the period 2012-2013, with funding being less than in 2002-2003. This could reflect the cycles of syllabus development, though further comprehensive analysis would be required to consider the impact of this cost reduction in light of the preceding comments on Work Program approvals and the lack of attention given to assessment tasks in these processes. It should be noted here that the Investigators were advised by QSA staff that historically, there had been no requirement to provide sample assessment tasks as part of approvals. However, the first author has direct experience of this requirement as part of accreditation processes applied to Work Programs.

Information on full-costs provides a more informed picture, with the focus on dollar per student, school and subject group as shown in Table 2 (prepared by QSA) and Senior Education Officers and subjects in Table 3 (prepared by QSA).

Table 2: Expenditure on the moderation system 2000 and 2013

Year	\$ per student	\$ per school	\$ per subject group
1999/2000	278	30,710	1,467
2012/2013	268	28,814	1,542

Notes:

- the number of students is the number who completed Year 12 (including visa students)
- a subject group is the group of students in an individual school undertaking the same subject
- expenditure has been corrected for inflation using Reserve Bank of Australia, inflation calculator and is in 2013 dollars

Table 3: Senior Education Officers, schools and subject groups 1991, 2000 and 2013

Year	Review/Senior Education Officers	All schools	Small schools (<50)*	Authority subject groups	Authority-registered subject groups
1991	26	293	15	6049	1949
2000	23	351	25	7350	3580
2013	20	452	52	8444	3219

Notes:

- the number of small schools is included as they typically require a disproportionate level of support and assistance, mainly due to the small number of experienced staff and the greater impact of staff turnover
- a subject group is the group of students in an individual school undertaking the same subject.

Table 2 shows that expenditure per student and per school has decreased while there has been an increase in expenditure per subject group. While direct expenditure has remained stable, Table 3 clearly shows that there are decreasing numbers of SEOs in place to support the growing number of schools and subjects in the quality assurance processes.

Additional information provided by QSA about costs associated with research positions clearly shows that since phasing out the Research and Policy Unit over the period of 2005-2007, there has only been one year of expenditure that has an allocation for such staff. The lack of a coherent approach to data-driven interrogation of moderation has been noted throughout these findings. It is therefore timely to revisit the foundation principles of the assessment approach, discussed in the next section, and examine the sufficiency of the investment in and the staffing profile for effective implementation.

QSA's moderation processes are currently understood as based on a *partnership* between itself and the school (QSA, undated ppt). In this partnership, schools are expected to develop high quality Work Programs consistent with syllabus requirements, implement valid assessment instruments and formulate and report judgements about student performance using stated standards. QSA has the role of ensuring reliability and comparability through syllabus development and enacting the four key quality assurance processes. This report indicates it is timely to revisit this partnership and consider in particular the accountabilities of schools and QSA in quality assurance processes and more specifically, the requirements of the assessments that students undertake throughout their period of senior schooling. This observation extends to assessment for both formative purposes and summative purposes. An initiative to strengthen the model would be the inclusion of summative assessment instruments in School Work Program Approvals.

This paper raises questions about the absence of attention currently given to assessment task design and to assessment tasks for summative purposes, in particular. This omission in Work Program Approvals has the effect of weakening quality assurance overall. The paper has also brought to light instances where panels have noted issues with task design, including cases where limitations of assessment instruments have been raised in panel reports in successive years, with little or no action reported to be taken in the school. As discussed later in this paper, establishing a bank of high quality assessment instruments as exemplars of expectations of instruments, together with a commentary on how they match syllabus requirements, is one way to address this current gap in quality assurance.

Supporting schools was a premise underlying the quality assurance work of panels, as reflected in predominance in the talk of panellists that they were "looking for evidence to support the school", and "trying to agree with what the school said; we try not to rock the boat". This recurring message had a number of influences.

- Panellists supporting students in cases of poor assessment task design* – Panellists were observed to search through student responses in other assessments to find evidence of the capability represented in a criterion when it was not evident in the assessment. This was the case where the assessment task created to capture that performance failed to do so due to limitations in assessment design.

- *Panellists refrained from reporting details of issues with assessment instruments* – during monitoring panellists were observed to refrain from providing fine tuning advice or comments on the design of a particular assessment instrument, especially when the schools' on-balance judgement did not change for the student as a result. One panellist commented: "end of day, is school on track. If at verification, I would question it a bit more".
- *Panellists viewed changes to Reports as 'channelling' their decision-making* – panellists reported that changes to monitoring paperwork from 2013 to 2014 narrowed their options for recording judgements and providing feedback to schools. The 2013 form represented five standards, VHA to VLA with threshold indicators, essentially allowing placements of students in 15 LOA (see Appendix 8 for 2013 Form R3). In 2014, this changed to five standard representation leading one panellist to comment, "they want us to agree with the school".

Overall, it appears that the quality assurance feed-forward aspect of the system is not leading to improvements in practice. That is, the loop is not always closing. Questions are therefore raised about whether: the system has the capability with its current mechanisms for information management or the regulatory capacity to close the loop; or schools have the willingness or capability to attend to the feedback provided. Some evidence informing this discussion includes the following extracts:

- District Chair Monitoring Reports:
 - "Schools ignoring panel advice?"
 - "attributes ticked [for school action] but were not done, it's chronic, it's the second successive year."
 - "school still has made no amendment for verification folios to meet syllabus requirements ... I am just 'flagging' this early, as feedback was delivered at Verification 2013, but no adjustment has been forthcoming".
 - "... evidence has consistently not met A standard for a number of years."
 - "... has had these issues raised now for several years..."
- District Chair Verification Reports:
 - "They need further support on their assessment."
 - "Assessment tasks still lack rigour and depth..."
 - "Ongoing issues with the [school], being handled at SEO level..."
 - "The school has been given previous advice about this."
 - "There are ongoing issues..."
 - "The school has continued to have difficulty with how to put together a submission using QSA guidelines. This was raised in Monitoring earlier this year and was not addressed in the verification package."
- Investigator observations also confirmed that Chairs and panellists reported that some schools did not take up the advice offered by the panel, with reoccurrence of the same issue/s occurring in some cases for several years.

A number of possible contributing factors are identified:

- The paper-based nature of the system prevents timely data capture and follow-up of issues. Some examples follow.
 - Issues for follow-up, identified by panellists and reported to QSA by Chairs, are reliant on quality and clarity of information, which is varied. Chair synthesis of panellist advice and timely provision to QSA are all factors affecting import of messages about improvements.
 - Issues are transcribed by SEOs onto a tracking excel database with synthesis of issues recorded.

The intensive process-driven nature of the system, with QSA staffing committed to this work while less attention is given to analytic and evaluative work.

Noting QSA and individual Chair's committed efforts, further investigation should be undertaken to examine the factors that impact on how schools attend to panel feedback and the workforce needs of QSA staff to inform school action through to changes in assessment practices, if required.

Key findings

Currently there are no archival databases established for the purpose of longitudinal tracking, analysis and evaluation of moderation practices and protocols involving standards in disciplines and knowledge domains.

The ability of Chairs to communicate effectively with schools and SEOs about issues identified in packages submitted for quality assurance is critical.

Panellists report that participation in moderation panels substantially contributes to teachers' knowledge and skills in assessment design, syllabus understanding and application of standards in judgement of student work.

An underpinning notion of the work of moderation through each of the quality assurance processes undertaken by panels was that it was locating evidence to support school judgements.

While the percentage of the annual administration budget allocated to direct costs associated with externally-moderated school-based assessment has remained stable over the last 10 years, the numbers of Senior Education Officers to support the growing number of schools and subjects has decreased.

There are examples where school submissions come to panels at verification showing that required assessment changes as noted in earlier panel reports have not been undertaken.

In Part 2 of the findings, many features of the current model of externally moderated standards-referenced assessment have been considered. A Panel Chair identified the system's strength as allowing the "flexibility to design and implement teaching and assessments to suit [their] students". This flexibility is critical in ensuring a broad range of assessment opportunities for summative assessment, including traditional pencil and paper examinations done under wholly supervised conditions, performances and demonstrations, assignments undertaken over time, and multimodal assessments. This broader range of assessment evidence is achieved through explicit connection of summative assessment requirements with defined standards in syllabuses. It is the standards that hold as the common yardstick across schools. It is also true that teacher judgement in classes and in moderation panels lies at the heart of the system. It is long overdue for QSA to address the uneven quality of assessment evidence coming from schools, including variability in the quality of school submissions of student folios to panels. The fact is that the assessment tasks for formative and summative purposes are of variable quality, with some having potential to limit student opportunities, mentioned above. Assessment task design is a critical issue for priority attention. It is also time for formally recognising the demands made of District and State Panel Chairs and panellists, and the high level of assessment literacies they need to bring to moderation. Attention now turns to consider assessment in Queensland secondary schools across five eras, with recommendations presented in the concluding section.

Part 3: Assessment systems in Queensland secondary schools – Five Eras

Introduced in the early 1980s, the current system can be considered as developing across five eras or phases. An overview of these phases is provided in the following discussion, originally published by Smith (1995), with an update in Smith and Matters (2007).

Era 1: 1873–1969

Public examinations were first held in Queensland secondary schools in 1873 and persisted for most of the following century. The examinations had a primary function as gate-keeping, the practice being that the examination worked to sort students for the purposes of determining (non-) entry into

education pathways and the workforce. The examinations were traditional in that they relied on students working within fully supervised, time restricted conditions, with no access to resources other than pencil and paper. The setting of the examinations was done centrally. The University of Sydney was responsible for setting the Senior Public Examination until 1912 when The University of Queensland took over this role after coming into existence as Queensland's first university.

Over time, the examinations had an inevitable and strong backwash effect on the curriculum and classroom teaching, learning and assessment. Routinely the teaching year was staged to build student knowledge of the type required for display in the examination, with rehearsal for the type of questions and for managing time restrictions in examination conditions. In large part, the examination items focused on student control of content knowledge, the time restricted examination genre not permitting opportunities for problem-solving or evaluative thinking that require more extended time and access to material resources. In retrospect, it is fair to state that irrespective of the quality of the examination in any given year, the scope of the knowledge, skills and capabilities assessed was very narrow, relative to that routinely taught and assessed in accordance with current Queensland syllabus materials. Also of interest here is how the examinations worked to define the roles of the teacher and student as both pitting themselves against the demands of the examinations, with past papers providing rehearsal opportunities. Further, the grading of student work relied on numeric scoring tied to a reporting framework using letter-grades, in the absence of any sense of quality represented in standards stated as verbal descriptors. In the latter phase of the public examination system in Queensland, student results in the form of letter grades were published in newspapers, the grade appearing with the student name. In part, as a legacy of this era, there remains in the community and to some extent in the media residual understandings that numeric scores captured as percentages have an absolute or at least intrinsic meaning.

Era 2: 1970–1978

By the late 1960s however, 'teachers, students, the press and the public at large [had become] increasingly disenchanted with the public examination system' (Sadler, 1991:3). In July 1969, the State Government established an expert committee, chaired by Dr William C. Radford, Director of the Australian Council for Educational Research, to review the system. In May the following year, the Committee submitted its report, which contained 47 recommendations for change. One of these recommendations called for the abolition of public examinations at both Year 10 and Year 12 levels, and the introduction of a system of school-based norm-referenced assessment. In 1970, the Government of the day accepted the recommendations, and made appropriate legislative provision for their implementation. The fully school-based assessment system was introduced in 1971 and became known as the Radford Scheme.

The Radford Scheme represented a radical change, which was without precedent in Australia, and pioneered norm-referenced school-based assessment using teacher-made tests. In essence, the scheme involved a significant devolution of authority for assessment to the classroom teacher, the school and review panels, and a shift in emphasis from terminal (final) to continuous (ongoing) assessment. No longer was it the teachers' responsibility to prepare students as candidates for external, centrally controlled examinations. Rather, for the first time in the history of secondary education in Australia, Queensland teachers were required to document the main aspects of a course of study; to develop and implement a range of test instruments including assignments and examinations; and to report on student achievement using a norm-based method.

The determination of grades under norm-based procedures appeared simple in principle. The distribution of grades in each subject for the State as a whole was fixed, and more or less followed a normal (bell-shaped) curve. Teachers ranked students and allocated grades from 7 (highest) to 1 (lowest). The Radford Scheme also involved selected teachers in a review or moderation process, the express purpose of which was for teachers to check that each school's proposed grades were roughly comparable with those proposed by other schools. The process was managed by the Board of Secondary School Studies (BSSS) and required that each school appoint a teacher representative (usually the Subject Master) to attend a moderation meeting. At the meeting, each representative

presented sample work from students in Years 10 to 12, and it was ‘moderated’ or compared with work from other schools. Responsibility for assessment was therefore vested in the teaching profession as a whole, not within a central bureau or agency, even though the BSSS played a significant organisational role.

Sadler (1991:3) made the point that ‘the change from external examinations to school-based assessment has been described in retrospect as the greatest influence on the professional development of secondary teachers in Queensland’s history’. The Radford System was not without its problems, however, as was evident in two research studies (Campbell et al., 1975; Fairbairn, McBryde, & Rigby, 1976) undertaken to inquire into the implementation of the system. The reports of both studies (Some Consequences of the Radford Scheme for School, Teachers and Students in Queensland, Campbell et al., 1975; Schools Under Radford, Fairbairn et al., 1976) concluded that norm-referenced school-based assessment had not realised many of the expectations of the Radford Committee. Furthermore, they indicated that tests and examinations had, contrary to expectations, increased in frequency, while the norm-based awarding of grades contributed to unhealthy competition and even animosity among students. One of the primary concerns was the erosion in teacher–student relationships caused by school-based assessment practices.

In February 1976, the BSSS commissioned an expert panel chaired by Professor Edward Scott to review the two research reports named above, together with Board members’ comments thereon, ‘with a view to advising the Board on implications of the reports for Board Policy and practices’ (Scott et al., 1978). In April 1978, the panel tabled its final report entitled *A Review of School-based Assessment in Queensland Secondary Schools* (acronym ROSBA). In 1979, the Queensland Cabinet accepted in principle the ROSBA report. Implementation did not begin until 1981, after which it occurred in three successive phases, each phase involving an increasing number of schools. By 1986, all Queensland secondary schools had come ‘on line’.

What follows does not attempt to provide a detailed analysis of the differences and similarities between the Radford Scheme and the current system, ROSBA, as this is outside the scope of this paper. It focuses, however, on the conceptual transition teachers were required to make in this second change-over.

Era 3: 1979–1985

Under the Directorship of John Pitman, this era was marked by the shift from a norm-based to a criteria-based approach to assessment. Some of the principles of the Radford Scheme are retained in ROSBA. For example, assessment is to be continuous, and teachers’ responsibilities are to include the planning and implementation of an approved course of study and the reporting of student achievement. However, the similarities between the two systems were not sufficient for school personnel to make an easy and trouble-free transition from one system to the other. Scott et al. (1978:3) pointed to the substantial nature of the transition:

We believe that, while maintaining the spirit of the Radford Report, some major changes in the implementation of that spirit are essential.

Essentially, the transition from Radford to ROSBA required that teachers make a significant conceptual break from a norm-based approach to assessment, which relied heavily on direct inter-student comparisons for determining quality, to an objectives-driven curriculum and a criteria-based system of assessment. For the latter, direct comparisons among students were replaced by criteria and standards as the yardstick for awarding grades and reporting student achievement.

Whereas the Radford Scheme pioneered school-based assessment, ROSBA did the pioneering for non-norm-referenced assessment. In particular, ROSBA explicitly set out to focus teacher attention on curriculum objectives, and the performance criteria for those objectives against which students should be assessed (Scott et al., 1978). In part, this was (and is) achieved by requiring teachers to write comprehensive and detailed school Work Programs that specify various aspects of a course of study.

These aspects include: the variety of learning experiences to be undertaken; the available resources necessary for satisfactory completion of the course; the range of assessment items and the conditions under which they will be undertaken; and the means by which the teachers will determine grades on course completion. In its role as the centralised certifying authority, the then Board of Secondary School Studies (BSSS) accredited Work Programs if they fulfil syllabus requirements and cater for the interests, needs and abilities of the schools' students.

As previously mentioned, a major distinction between the Radford and ROSBA schemes is that teacher judgments about student work no longer rely on direct inter-student comparisons, ranking of student performances, or the aggregation or weighting of schools. The comparison is between the work to be assessed (either a single piece or a representative sample) and defined criteria and standards. A basic premise of the system is that student performance can be improved if teachers make available the criteria to be used in judging the quality of student performance. In practice, ROSBA requires that teachers prescribe and publish detailed criteria prior to students commencing an assessable task. Whether teachers use letters, grades or other symbols to communicate summary judgments of performance on a task or a collection of tasks is a matter determined by individual schools.

A small study of the early period of ROSBA implementation was undertaken by a 12-member research team headed by Professor W. Jack Campbell (Campbell et al., 1983) concluded that many school personnel were not prepared for the conceptual break from Radford to ROSBA. The study reported that school staff considered that 'they were engaged in a major innovation without a sufficient understanding of the philosophical and psychological rations' for such change (Campbell et al, 1983:25).

Teachers' ill-preparedness for the change-over can be accounted for, in part at least, by the fact that the implementation of ROSBA did not take place within an established theoretical framework. The assumption was that practice would, and indeed in some respects, should proceed ahead of theory. To illustrate this point, it is useful to contrast the finding of Campbell et al. (1983) concerning the lack of preparedness of those responsible for the implementation, with the BSSS's perception of the demands made on teachers in the implementation of the ROSBA system. Speaking as the Director of the BSSS, Pitman exhorted teachers to 'see themselves as embarking upon a trail-blazing exercise' during which important insights related to the implementation of ROSBA proposals would be generated (Pitman, cited in Campbell et al., 1983: 3). Referring to the BSSS, he also claimed that 'we are quite prepared to admit we are learning at the same time as the Phase I schools are learning', and 'the Board openly admits that there are areas in which decisions cannot be made at this stage for lack of information' (Pitman, cited in Campbell et al., 1983:3). Taken together, these comments indicate that the expectation, at least on the part of the BSSS, was that teachers were the pioneers of a new approach to assessment. As such, they were 'licensed' to work through the curriculum and assessment implications of so-called criteria-based assessment, outside any existing theoretical framework for the system.

In hindsight the wisdom of this aspect of implementation can be called into question because of the assumptions it made about the relevance of teachers' experience in a norm-referenced system to one requiring a criteria-based approach to assessment. Consider, for example, the assumption concerning teachers' understanding of ROSBA's five Levels of Achievement, which replaced the Radford 7-point scale. The labels for these categories are: Very High Achievement; High Achievement; Sound Achievement; Limited Achievement; and Very Limited Achievement. Although many teachers had considerable experience in, and therefore felt comfortable with, the Radford procedure of rank-ordering students for grading purposes, they were inexperienced in determining individual achievement by matching a sample body of work with standards that are 'non-numerical, and made according to multiple criteria using the human brain as the primary evaluative instrument' (Sadler, 1987: 191). Campbell et al. (1983: 29) made the point that 'the belief that teachers know, either intuitively or from experience, what constitutes a Very High Achievement or any other qualitative assessment is not well founded'. He further claimed (Campbell et al., 1983: 37) that 'the attention of the Board concentrated too narrowly on the details of implementation and action'. Accordingly, the Campbell report advised that 'high level and continuous reconceptualisation of what standards-based

assessment means in practice' was essential. Sadler (1986: 4) similarly pointed out that the Queensland system of school-based assessment was 'sufficiently distinct from the most fully developed existing varieties of criterion-referenced assessment in the USA for it to require independent developmental work'.

Work on the conceptualisation of what criteria- and standards-based assessment means in practice was formally begun in 1985. Pitman argued for funds to establish a four-person 'think tank' known as the Assessment Unit. The Unit was established in 1986 with a brief to:

- establish a sound theoretical foundation for a school-based assessment system using defined criteria and standards; and
- clarify and make suggestions about the practical aspects of such a system in secondary schools. (Board of Secondary School Studies, Brisbane, 1986:1)

Era 4: 1986 to 2010

Since 1986, a school-based approach to assessment, known as criteria-based assessment, has been implemented in all Queensland secondary schools. A key feature of the Queensland model of criteria-based assessment is that, in judging the quality of student work (either a single piece or a representative sample), teachers no longer rely on direct inter-student comparisons, ranking of student performances, or the aggregation or weighting of scores. The comparison is rather between the work to be assessed and explicit criteria and standards.

A basic premise of this approach is that student performance can be improved if the teachers define and make available to students the criteria against which assessable work is to be judged. In principle, this means that students no longer need to guess at teacher expectations for a successful performance. Another related premise is that, in criteria-based assessment, students will feel as if their performance has been more judged against the specified criteria than against the teacher's implicit criteria (and standards).

This fourth era was characterised by developments in the conceptualisation of school-based assessment that took as its centre stated criteria, and in turn, defined standards, written as verbal descriptors of quality. This conceptualisation and consideration of the policy and practice implications were undertaken initially in the Assessment Unit. Between 1986 and 1989, the Unit produced a number of Discussion Papers that addressed some of the theoretical issues confronting school personnel in their implementation of the system. These included such matters as defining achievement levels, the autonomy of the school in school-based assessment, the nature of standards; and the value of teachers' qualitative judgments. Indeed, it was in one of these papers that the meaning of the terms criteria and standards as used in ROSBA and as defined below, came to be understood by Queensland secondary teachers.

criterion: A distinguished property or characteristic of any thing, by which its quality can be judged or estimated, or by which a decision or classification may be made. (From Greek *kriterion*, a means for judging).

standard: A definite level of excellence or attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognised measure of what is adequate for some purpose, so established by authority, custom, or consensus. (From Roman *estendere*, to extend). (Sadler, 1987:194)

The Unit's discussion papers were written primarily for an audience of teachers, and in 1986, multiple sets of the papers were distributed to each Queensland secondary school. If the ideas and, more importantly, answers contained in these discussion papers had been disseminated to schools earlier, the implementation of ROSBA could have been considerably more effective and efficient, and the gross ill-preparedness of teachers to use criteria and standards may have been avoided. However, although the Assessment Unit Discussion Papers have gone some way to providing a theoretical framework, a comprehensive and fully articulated version of the underlying theory of criteria- and standards-based assessment in Queensland is not available some fourteen years after the

implementation of ROSBA. This situation can be accounted for, not only because the Assessment Unit was disbanded in the late 1980s as a result of funding cuts but also because that powerful model of partnership between academe and the bureaucracy has not been able to be repeated. Since then, there have been no significant developments in the underlying theory of the system, from either a curriculum perspective or its assessment dimension. The set of 21 discussion papers were again made available to schools and the general education community in the mid-1990s but the extent to which these valuable documents were read and digested would not appear to be great.

Although many of the ideas of the discussion papers have influenced the organisation and administration of system and, indeed in the formulation of policy, as yet they have not been endorsed as Queensland's official assessment policy. However, they are 'recommended to the teaching profession for consideration in the formulation of curriculum and assessment policy within secondary schools' (Board of Secondary School Studies, 1986:2). Whether teachers have read the papers, and indeed, how their interpretations have an impact on classroom practices remain unclear. For example, one of the principles underpinning criteria-based assessment, as presented in those papers is that by making criteria and standards explicit, students would feel more judged by the standard than by the teacher (Sadler, 1986). There are two assumptions related to this: that defined criteria and standards could make teachers' grading practices more explicit and hence more accountable, and that available, agreed-upon criteria and standards could enhance the reliability and credibility of teachers' judgments. As mentioned previously, it is not yet known whether the principles and assumptions underpinning criteria-based assessment match those underpinning classroom practices as there has been no sustained research on the issue of the organisation of the Queensland assessment system as a whole and its potential to impact on classroom. Relevant discussion papers, including McMeniman (1986a, 1986b), make clear that in principle, ROSBA enlists criteria in the service of instruction and the improvement of learning (formative assessment), as well as the more traditional use of evaluative criteria to judge learning outcomes (summative assessment). ... It should be mentioned here, however, that formative assessment cannot be directly equated with diagnostic assessment, although it shares with the latter an interest in checking on student progress. Specifically, formative assessment refers to those tasks completed by a student that will not form the basis of reporting achievement on course completion. Hence, formative⁵ assessment has a teaching focus, whereas summative assessment is exclusively concerned with the reporting and certifying functions of assessment.

A key premise underlying this organisational feature of the system is the proposition that formative and summative assessments are not mutually exclusive but complementary approaches to providing a reliable indication of student achievement (McMeniman, 1986b). A related premise is that classroom teachers are in the ideal situation to monitor their students' learning, and also to provide informed judgements and reports on student achievement. To date, the distinction Queensland secondary teachers make between formative and summative assessments in particular subject areas and how they stitch the two together have not been the subject of research. Similarly, the role of stated standards in how grades are awarded has not been researched, although the move to link assessment criteria and standards did become firmer in the final era, as discussed below.

Towards the end of this era there was a discernible move in research, policy and to a lesser extent practice, to move stated standards and the issue of quality to the centre. One catalyst for this move was the Viviani Report (Viviani, 1990) that called for establishing an evidentiary base reflective of the education system's efforts to subject itself to scrutiny and to provide data useful for evaluative and improvement purposes, which resulted in the formation of an Evaluation, Research and Development function within the (then) Board of Senior Secondary School Studies. There were two other noteworthy catalysts for the increasing emphasis on standards-based assessment (for commonly applied tasks as well as for teacher-devised tasks); namely, the New Basics research project (Queensland Department of Education and the Arts (DETA), 2004) and the work done under the banner of Education Queensland's Assessment and Reporting Framework Implementation Committee (2002–05). Although these two initiatives were radically different in nature, purpose and scope,

⁵ It is the case in practice in Queensland that formative assessment can have a summative function.

common to them was the commitment to install a system that aligned curriculum, assessment and reporting, with the strong focus on teacher knowledge of task demands and stated standards. Indeed, it is worth noting that the current policy priority relating to alignment across these three facets—curriculum, assessment and reporting—existed in Queensland well in advance of the federal government decision about a common standards reporting framework, and well in advance of the Queensland Curriculum Assessment and Reporting Framework developments.

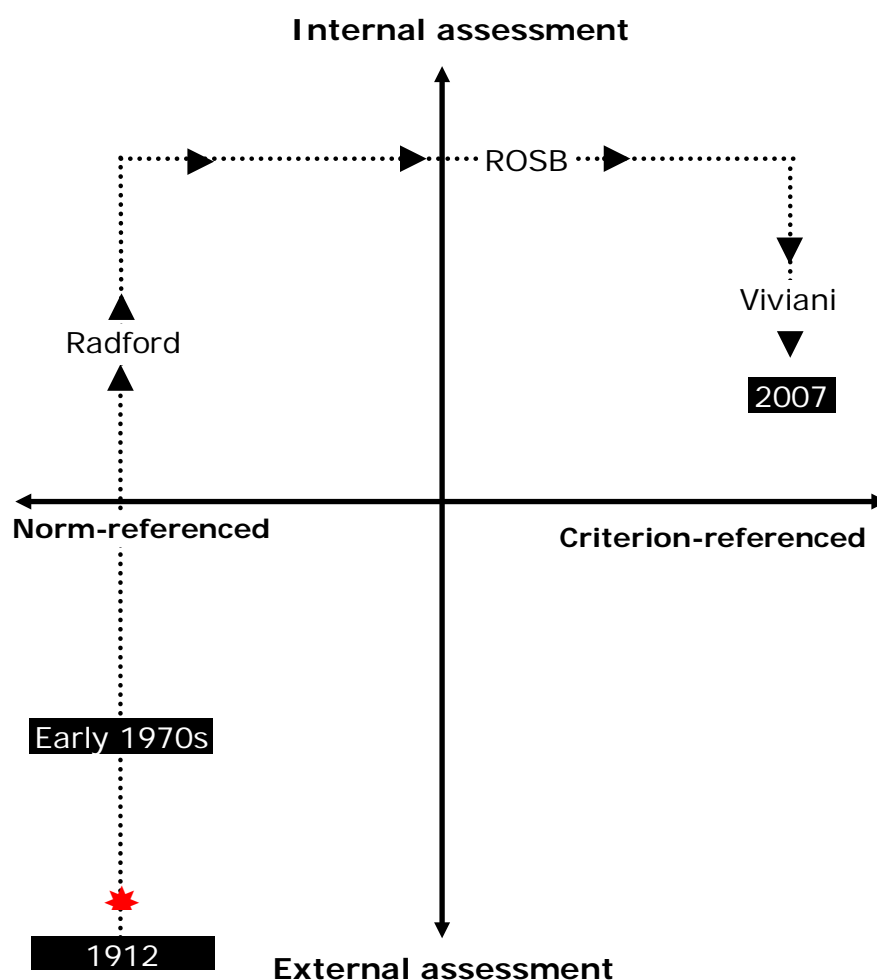


Figure 7: Development of assessment practices on two dimensions, Queensland, 1912–2007

(Source: Adapted from John A Pitman, QBSSSS, 2002)

Era 5: 2011 to future

The preceding historical overview has profiled the historic shifts in Queensland senior schooling away from public examinations to school-based assessment, known in its first iteration as Radford, and then on to an internationally distinctive form of standards-referenced assessment. Recurring observations across the eras include:

- practice has moved ahead of theoretical development of the system;
- little sustained research has been undertaken into the operation of the system, including approaches taken to standards-referenced assessment in classrooms and in moderation; and
- increasing emphasis on process management with considerably less attention given to developing system infrastructure and self-evaluation.

The absence of a well-developed theoretical rationale and a sustaining infrastructure could be expected to cause difficulties, including for teachers. It has led to a current situation where practical application has moved well in advance of model building and system theorising.

Opportunities exist to reinvigorate the assessment system as identified in findings already discussed. Focus now turns to consider the key elements of this paper to bridge the gap between theoretical framing and accountability measures currently in place. Figure 8 provides a means to capture the four key elements of practice applied in this system: assessment, standards, judgement and moderation. These elements, taken together in a coherent approach, hold promise for the new era to realise the enabling power of assessment.

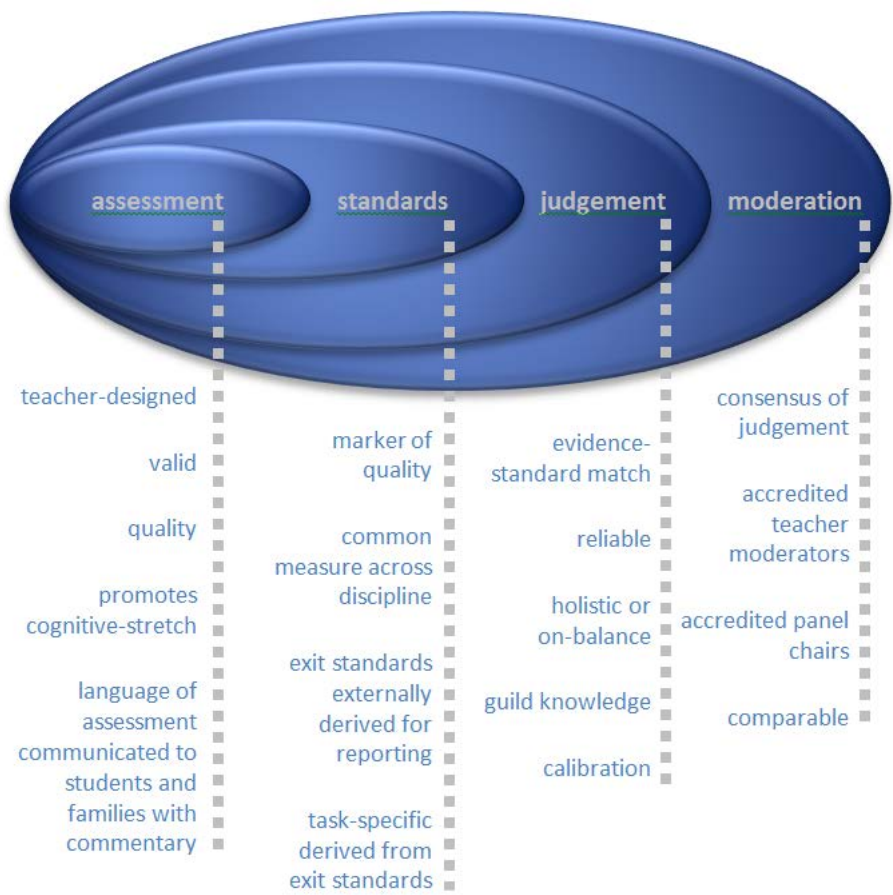


Figure 8: Key elements of assessment for student learning and reporting with confidence

A focus on quality assessment is central to reform efforts. Quality teacher-designed assessment that is valid and allows students to stretch cognitively and have opportunities to demonstrate their learning across all standards is not only critical, but the right of students. Inducting students into the language of assessment through the provision of assessments with stated standards as represented in questions, accompanied by exemplars with teacher commentary about features of performance is a related priority. Exemplars of this type could then be used to communicate to parents and the wider community, thereby building a local assessment community that is distinguishable from others.

Exit achievement standards representing markers of quality should continue to inform task design and guide task-specific standards as applied in teachers’ assessments. This is the integral linking of quality expectations from individual task to assessment portfolio. However, it is time to challenge the dominant matrix as the one-size fits all approach. How standards are represented should be discipline specific, ensuring that key features of quality are clearly identifiable, providing a clear sense of level—what is a B and what is a D, for instance. It is a reasonable expectation that the same standards that work to facilitate judgements of students’ achievement, should work to inform the student about how they could self-monitor and improve their own learning.

With a focus on consensus of judgements at a system level, focus turns to ensuring appropriately trained and experienced moderators. Accrediting teacher moderators and panel chairs recognises the expertise that is built and continually developed to assure comparability of judgements of student achievement across the state.

The recommendations discussed next call forward a strengthening of quality assurance processes to ensure that the state can deliver meaningful information marked by high-quality assessment instruments and a strong focus on standards that meet the needs of high levels of construct validity, reliability and comparability. Essentially, the proposition on offer is that the Queensland model of externally-moderated standards-referenced assessment has moved through the identified eras to a point of readiness for clarified, considerably strengthened messages about assessment literacy in the context of standards-referenced assessment and moderation. These include but are not restricted to:

- teacher judgement as being at the heart of the approach taken to externally-moderated standards-referenced assessment
- deliberate alignment of curriculum, learning and teaching, and assessment
- standards as external referents of quality—common yardsticks for determining quality
- assessment criteria and standards and the role they play in making available information about desirable features of performance at the task level and within folios
- features of high quality assessment task design and construct validity, ensuring that all students have opportunities to demonstrate achievement across the full range of the standards
- the notion of senior schooling as a period during which students’ meta-cognitive abilities are developed as they are given explicit opportunities to use standards and related exemplars for improvement purposes.

Finally, in regard to standards-referenced moderation, the preceding paper indicates the clear need for clarifying authority relations between the QSA and schools. The discourse of panels ‘supporting’ school and teacher judgements has become potent. While recognising that the partnership between the Authority and schools is central, the discourse of accountability through rigorous valid assessments, reliable judgements and quality assurance systems needs to be reasserted and evidenced to sustain and build the confidence of parents, students, the wider public and the teaching workforce in Queensland senior schooling.

Strengthening Queensland Senior Schooling

Recommendations

The distillation of key findings, identified throughout the *Findings* section of this paper, point to a number of recommendations and related actions for strengthening moderation processes and practices.

It is recommended that years of duration of panel service be reviewed to ensure a sustainable approach to maintaining a cadre of well qualified and trained Chairs and teacher panellists.

Actions to consider relate to:

- *Timeframe of service as a panellist* – restriction of panellist service to three to five years could ensure that a panel would maintain a mix of expert and novice panellists throughout the syllabus cycle.
- *Timeframe of service as a Chair* – service as a Chair could be restricted to the syllabus cycle for the subject. On implementation of a new syllabus, an existing panellist with five years' experience could be one requirement for the Chair position.

It is recommended that teacher's service on panels be formally recognised as a measure of professional esteem confirmed through membership to an *Institute of Accredited Assessors*.

Actions to consider relate to:

- *Service recognition* – formal accreditation of teacher panellists.
- *Mentoring* – expert panellists would adopt a mentor role for novice panellists to assist in inducting them into the processes of quality assurance with a focus on ways of working across the source documents including the syllabus, standards matrix from the syllabus, student profiles, assessments, student responses, and the relevant quality assurance forms.

The work of panellists is valued systemically, however, it is accompanied by limited professional recognition. While panellists' expertise in this high-stakes moderation context is critical, there is also high benefit to schools. Both the knowledge obtained through the act of training and exposure to varied assessments and targeted professional conversations, and the skills developed to identify syllabus expectations in assessments and judge student work against standards, places these panellists in an expert group.

It is recommended that professional development of Chairs and panellists occur each year, rather than the current two-year cumulative cycle.

Actions to consider relate to:

- *Cycles of training* – two layers of training are required for panellists in order to support previous recommendations. Each year, one layer would be for novice panellists with the second layer of training for more experienced panellists.
- *Content of training* – Chair training should include a specific focus on calibrating judgement against standards, managing communications and ensuring quality expectations in reporting. Similarly, panellists require calibration training about making relative achievement decisions within standards and according to rungs, where relative placement is required. Specific

calibration training on how to apply the standards at threshold levels and in moderation discussions should also occur.

- *Resource site development* – a bank of high-quality assessments tasks and related statements of standards and exemplars of student work should be established. A further option is for the exemplars to be annotated and be accompanied by ‘a cognitive commentary’ (Smith, 1995; Wyatt-Smith & Bridges, 2008) that describes how judgement was arrived and the influence of compensations or trade-offs in applying the standards.

As procedural forms are the communication linchpin between the aspects of quality assurance being undertaken and follow up of issues identified, panellist and chair training could extend to expectations for accurate and effective communication of assessment design matters and those related to level of achievement. This would attend to the gap identified in the findings about the high variability of actionable information in Chair reports.

It is recommended that judgement at thresholds be a concerted focus of action. An aim would be to ensure common understanding of how aspects of performance can be combined, including the process of matching work with the requirements of standards. Related areas for attention include the typical features of work considered to be at the highest or aspirational levels (e.g., Very High Achievement levels 6– 10).

Actions to consider relate to:

- *Create a bank of assessment examples of student work at threshold level* – panellist training should be supported through provision of a bank of assessment work samples recognised as illustrative of quality at threshold levels. These could be accompanied by a brief cognitive commentary about on balance judgement with a particular focus on compensations or trade-offs, as they influence decisions about grading.
- *Undertake research* into alternative approaches to formulating and promulgating standards as suited to disciplines. This will involve critical investigation into the continuing utility of the continuum representation of A to E standards and the dominant matrix approach.

It is recommended that information management systems for moderation be implemented to ensure time efficient capture of data for research, analysis and reporting purposes.

Actions to consider relate to:

- *Development of a central database* – information captured should be stored in a central database.
- *Development of online forms* – all stages of the quality assurance process should be linked to an online capture of information. Online forms fit for purpose are required to be developed, ensuring that all information feeds into the central database.
- *Development of database reporting* – automated reporting should be implemented to allow information to be drawn from the database relative to reporting purposes.

An online system of data reporting in moderation processes is essential and would allow timely access to data at key junctures. The findings indicate that forms are not always accurately completed, they sometimes lack sufficient information to ensure useful feed forward functions, and are at times not submitted to QSA as required. An online process for capturing data, with required fields of information, would assist in addressing this gap. Online capture of information would also allow for internal tracking of issues for follow up by the SEOs as identified during panelling and reported by Chairs. Additionally, this recommendation and associated actions allow for development of an archival record to permit system and self-analysis as more fully considered next.

It is recommended that a Research and Development Data Analysis Unit be established to undertake sustained and ongoing quantitative and qualitative research and analysis into standards-referenced moderation.

A data analysis unit would facilitate deeper understanding, tracking and reporting of issues for intervening in practice. When considered alongside the recommendation concerning information management system development, the opportunities for information capture and timely use point to the critical need for development of such a unit. An established R and D Unit would enable the QSA to undertake systemic and continuous review of the system in operation. Key data capture moments could include Work Program Approvals, Monitoring (DPC Reports), Verification (DPC Reports), Comparability (SPC Reports), and Random Sampling. This unit would also act in support of key positions, including Chairs and the SEOs who work as the linchpin between the panels and the schools.

The findings identified key stages in the operation of senior schooling assessment where there is no systematic analysis of the advice provided by Chairs. Opportunities exist at these stages to collect rich empirical evidence into the operation itself, enabling self-monitoring and the identification of needed improvements in ways currently not available.

Beyond the process and data-driven analysis opportunities already described, the Unit could also undertake wider investigations concerning areas of need to ensure appropriate, accurate and transparent communication of information. Some opportunities include:

- Research and description of relative achievement decisions about placement within standards and according to the 10-point rung.
- Determining how classroom teachers understand, engage with and interpret syllabus documents and publicly available State Review Panel Reports. The purpose here is to consider effective strategies for disseminating findings from moderation processes to teachers to further inform the use of standards and how assessment aligns to curriculum and learning and teaching at the school level.

It is recommended that priority should be placed on developing and implementing common terminology across syllabuses.

It is recognised that the language of the system has changed over time with different terminology adopted as shown by the use of Standards Matrix and Dimensions and standards descriptors versus Standards Associated with Exit Criteria in syllabuses. While this is a consequence of timing of syllabus development, common terminology would assist in ensuring consistent understanding and reference across schools and the wider community.

It is recommended that Work Program Approval processes make explicit provision for reviewing assessment instruments for construct validity and fitness for purpose.

Actions to consider relate to:

- *Submission of the suite of assessment instruments* – at Work Program Approval, schools could be asked to submit a sample of assessment instruments illustrative of those used for formative assessment purposes, as well as the suite of assessment instruments used for summative assessment purposes.
- *Digital folios as exemplars* – digital exemplars could be developed in schools, with these available to students and parents as concrete examples of the standards and expectations of quality.

This recommendation recognises that currently, no quality assurance checks are applied to the full range of Year 11 and Year 12 assessment instruments during Work Program Approval processes and prior to the assessments being implemented in classrooms. Findings indicate that assessment instruments have a direct impact on opportunities for students to demonstrate achievement against the full range of achievement standards. Beyond the curriculum, it is clear that there are many

contextual factors involved in schools' own review of assessment tasks. These should not diminish the common expectation of demonstrated construct validity and the requirement for fitness for purpose.

A final recommendation relates to clarifying assessment purposes especially as they relate to formative and summative assessments over the course of the senior program of study. This would, in turn, ensure that standards could be reinstated as being as much concerned with student learning improvement as with reporting achievement on course completion. In short, the system could realise the potential envisaged for it more than four decades ago in centring on standards to inform learning and student self-monitoring, as well as being the stated, common yardstick for measuring student achievement.

Methodology and design

Design principles

A number of design principles informed the data collection for the investigation.

1. Staging of data collection was consistent with the schedule in place for quality assuring school-based assessment in Years 11 and 12. Ensuring a diverse range of views were represented in the corpus of data was a priority to ensure a balanced view.
2. The participants invited to contribute to the data considered included panellists at State and District levels, and other staff in relevant sections of QSA who were key to the ongoing development, delivery, and maintenance of processes to support quality assurance processes.
3. A main focus was the extent to which policy was carried forward through to practice. This focus required attention to the official policy messages and the enacted messages.
4. While there are 10 discipline areas, with 49 subjects falling within, specific emphasis was placed on English, Mathematics and Science subjects with lighter sampling around other subjects in parts of the data collection and analysis supporting this investigation. The focus in the targeted analysis was to reveal discipline differences in materials and practices.

Participants

Participants included panellists working on State and District panels for monitoring, verification and comparability purposes, and key staff in QSA. Participants in formal interviews or focus groups were provided with an information sheet and consent form (see Appendix 9).

Data collected

As QSA quality assurance processes were influential in the data collection, an interview was held with senior QSA staff who described the range of processes supporting senior schooling processes (see Figure A1 for the Investigator's representation). Opportunities for data collection were identified with a focus on ensuring that the investigation considered as many components of the quality assurance cycle as possible within the timeframe available.

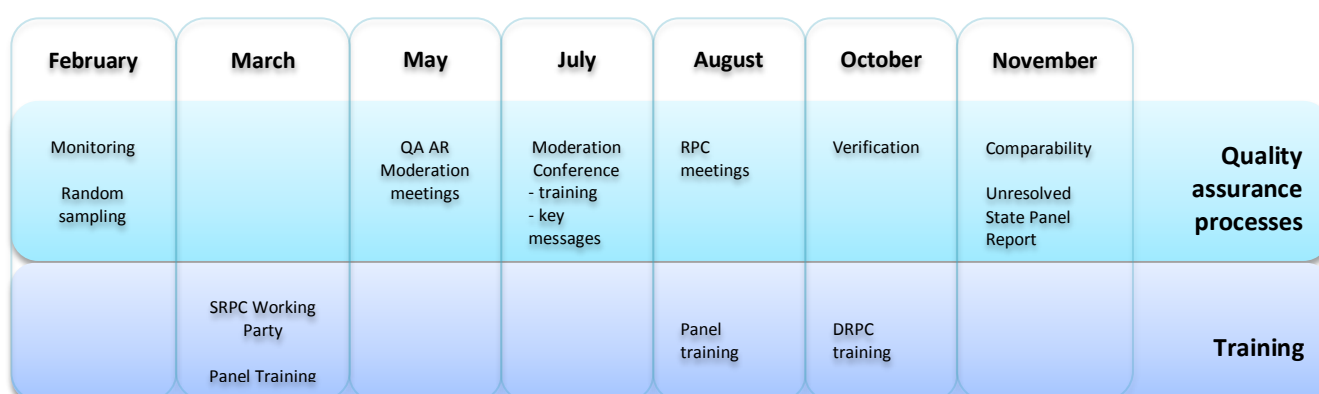


Figure A1: Senior schooling quality assurance processes

To ensure sufficient depth and breadth of coverage of these processes, data included observation, interviews, focus groups and document collection. Figure A2 maps the data informing this investigation as aligned to the quality assurance processes. In terms of observations, the following three quality assurance processes were attended:

- District Panels for verification purposes were attended at Ipswich Girls Grammar School and Ipswich Grammar School on 21 October with nine disciplines areas observed, a focus group undertaken with five Panel Chairs, and two individual interviews.

- State Panels for comparability purposes were attended at QSA and Nudgee on 4 and 6 November with 13 disciplines areas observed as well as training for panellists relating to comparability and unresolved cases.
- State Panels for monitoring purposes were attended on 18-21 February 2014 with six subjects observed.

Table A1 provides more information on the corpus of data collected; Appendix 10 provides a tracking of all data informing this investigation.

Table A1: Data collected to inform the Investigation

DATA CATEGORY	DETAILS	BREAKDOWN	
Observation	District Panel Chair meeting for Verification purposes 21 Oct 2013	Accounting Biology Business Communication & Technologies English English Extension	Geography Mathematics B Music Physics
	State Panel Chair meeting for Comparability purposes 4 & 6 Nov 2013	Biology Dance & Drama Economics English Film, TV & Media Home Economics Mathematics A Mathematics B Mathematics C	Modern History Music Physics Visual Arts Chair Chats Unresolved Training Comparability Training
	District Panel Chair meetings for Monitoring purposes 18-21 Feb 2014	Chemistry Drama English	Information Technology Systems Mathematics A Physics
Interviews	Verification 21 Oct 2013	Biology Music	Mathematics A, B, C Science
	QSA personnel 4 Nov 2013	Assistant Director, P-12 Implementation Branch and Manager, Quality Assurance Unit	
Focus groups	Verification 21 Oct 2013	Mathematics A, English, Chemistry, Physics and Biology	
Document collection	Senior Syllabuses	18 syllabuses examined	
	Package 1 prepared by QSA 4 Nov 2013	Assorted materials	
	Package 2 prepared by QSA on request	Forms R3 & R6 plus Forms C2; school support materials for Ancient History	
		Published assessment instruments, original school documents and a completed Tool for Schools	
		Report on Building Professional Capacity in Educational Assessment by Paul Kilvert	
		PD materials from Review Panel Chair meetings in Districts 2013	
	Package 3 prepared by QSA delivered 7/2/14	State and district review panel chair's report – Verification 2013 (344 reports)	
	Package 4 prepared by QSA delivered 12/3/14	State and district review panel chair's report – monitoring 2014 (367 reports)	
	QSA email 1/2/2014	Position descriptions and work profiles for the Review Officer (RO), Standards and Assessment Officer (SAO), and Senior Education Officer (SEO)	

DATA CATEGORY	DETAILS	BREAKDOWN
	QSA email 26/2/14	Information on extra ordinary reviews for the last 5 years
	QSA email 10/3/14	Information on: unresolved reviews for last 4 years, and schools undertaking Year 12 Authority-subjects in total and by subject
	QSA email 3/3/14	Financial information related to the cost of quality assurance of Year 11 and 12

An important qualification is required at this point as it is pertinent to reading of the findings. Due to the timing of the review and this subsequent investigation, data collection began with verification observations in October 2013. Data considered could not be tracked in a linear fashion—for instance, from monitoring through verification through comparability—thus preventing specific identification of matters from origin to finalisation. Mapping of the data collected according to the quality assurance processes of QSA is shown in Figure A2.

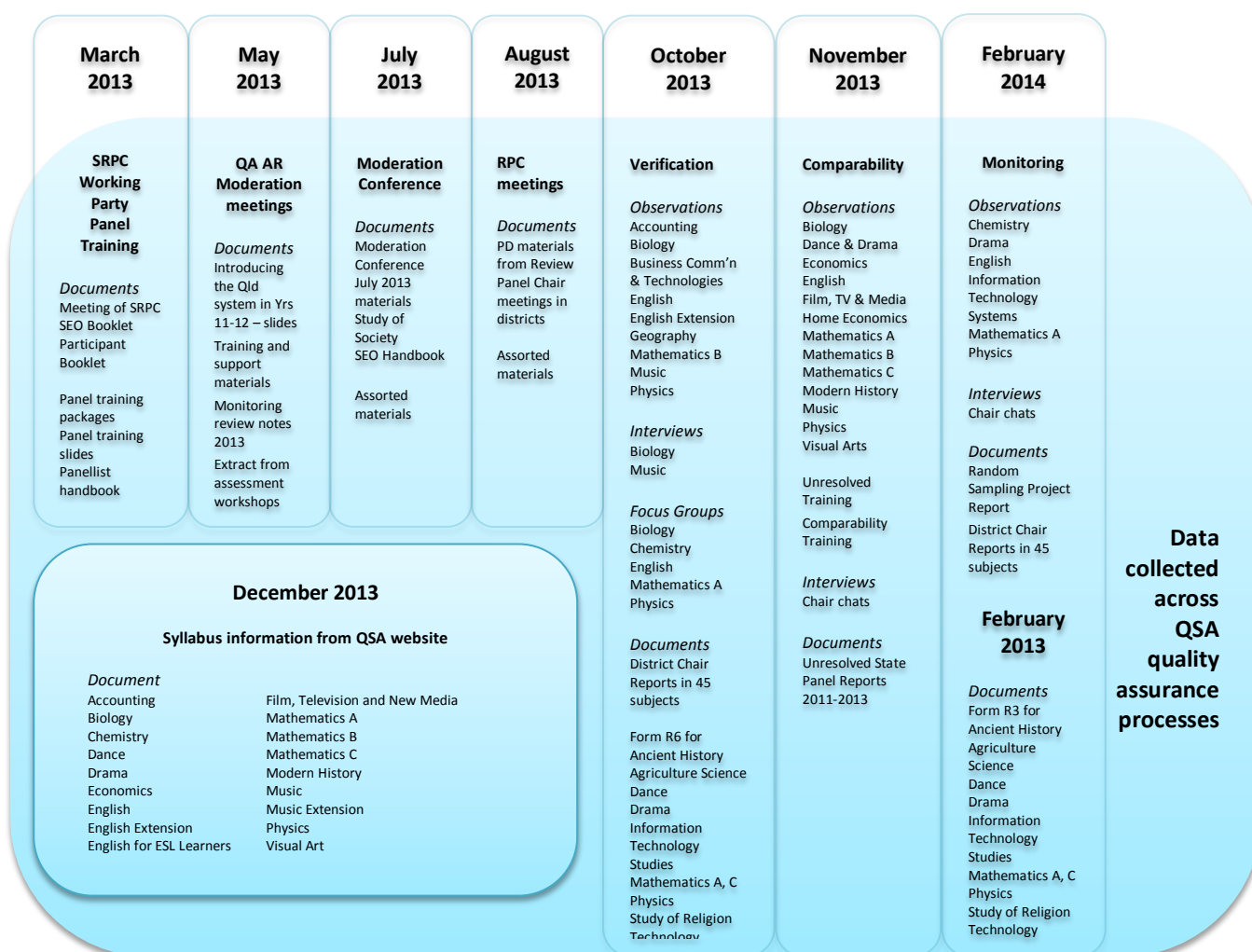


Figure A2: Data collected across system quality assurance processes

Milestones

Information on the investigation's milestones are specified Schedule 1 of the subcontract. The work met all of the deliverables as specified in the subcontract as shown in the following table (Table A2).

Table A2: Milestones and deliverables

TIMELINE	MILESTONE	DELIVERABLES	COMPLETION
	Appoint Sub-Contractor Sign contract Revise and refine methodology/project plan	Appointment Signing of contract Methodology discussion	Completed
31 Dec 2013	First oral briefing	Emerging findings	Completed 12 Dec 2013
31 Mar 2014	Second oral briefing	Contents of informing paper	Completed 15 Apr 2014
30 Apr 2014	Final written report	Informing Paper	Completed 24 April 2014

Appendix 2

Information Technology Systems (ITS)

Student Name: _____						
Teacher: _____						
Teacher: _____						
Semester	Instrument	Submission	Familiar Application	Problem Solving	Communication	Level of achievement
1	Item 1.1 Exam Computer Processes and Hardware	F,M	C+	C		
	Item 1.2 Minor Project ECommerce and Websites	F,M	D	D-	D+	
	Item 1.3 Written Folio User Documentation		D+	D+	C-	
	Semester 1 LOA		C-	D+	D+	
2	Item 2.1 Graphics Folio Multimedia / Digital Design	F,M	C	C	C	
	Item 2.2 Team Oral Presentation Network Components and Design	F,M	B	B-	B-	
	Item 2.3 Written Report / Practical Task Computer Systems	F,M	C	C	C+	
	Semester 2 LOA		C+	C+	C+	
3	Semester 2 Monitoring					
	Item 3.1 Exam Network Architecture and Systems	S,V	C-	C-	C-	
	Item 3.2 Written Task Network Architecture and Systems	S,V	C	C-	C-	
	Item 3.3 Practical Group Project Network Install, Configure and Manage	S,V	B-	C	C	
4	Semester 3 LOA					
	Item 4.1 Major Project Multimedia Authoring and Production A		S,V	C	C-	
Exit level (verification)						

Mathematics A Profile

Teacher's Name: _____

Mathematics A Profile

Student's Name:				Teacher's Name:			
Year of Entry:				Year of Exit:			
Semester	Instrument Number	Formative / Summative	Assessment Instrument	Knowledge & Procedures	Modelling and Problem Solving	Communication & Justification	Level of Achievement
One	1.1	F	Mid-Semester Exam				
	1.2	F	Extended Modelling Problem and Solving Response				
	1.3	F	End-Semester Exam				
	Semester 1			Rating			
Two	2.1	S	Mid-Semester Exam				
	2.2	S	Report				
	2.3	S	End-Semester Exam				
	Semester 2			Rating			
Moderation				Summary			
Three	3.1	S	Mid-Semester Exam		C-	D	C-
	3.2	S	Response		E	E	E
	3.3	S	End-Semester Exam		C	C	C
	Semester 3			Rating			
Four	4.1	S	Mid-Semester Exam		C	D	C
	4.2	S	Report		C-	D+	C-
	4.3	S	End-Semester Exam				
	Semester 4			Rating			
Verification				Summary @ Exit			
Rating							

Figure A3: Student profile examples

Senior Dance Profile for

COURSE COMPLETION 2013

COURSE COMPLETION 2015			CRITERIA		
YEAR	UNIT	ASSESSMENT INSTRUMENT	Choreography	Performance	Appreciation
Yr 11 Formative (F)	1. <u>Functions of Dance and Musical Theatre</u> "The Evolution of Dance"	1. Performance of a repertoire			
		2. Written Appreciation task			
		3. Group choreography task communicating a storyline			
	2. <u>Duncan, Graham and Weir</u> "Emotional Contemporary"	4. Contemporary Performance task			
		5. Appreciation assignment			
		6. Contemporary Choreography task (family connection focus)			
Interim standards					
Interim level of achievement					
Yr 12 Formative (F)	5. <u>Popular Dance</u> "Popular Dance of the Youth Culture"	1. Individual performance of hip-hop dance		C	
		2. Choreography task on popular culture item/ product	C-		
		3. Extended written essay			D
	6. <u>Ailey & Bruce</u> "Cultural Differences"	4. Performance of a contemporary repertoire		C	
		5. Extended written response in response to a quote			D+
		6. Individual choreography in partners focussing on contact work	C		
C	Standard verification Standards		C	C	D+
	Standard level of achievement		SA-		
SA3	Individual style standards	7. Individual choreography task for a public audience			
		of av			

STUDY OF RELIGION STUDENT ASSESSMENT PROFILE

ASSESSMENT PROFILE

I

Yr 11 Teacher:

Yr 12 Teacher:

UNIT STUDIED

ASSESSMENT INSTRUMENTS

Knowledge Processes

CRITERIA

Evaluative Processes

Research and communication skills

Signed

Nature and significance of religion drawing examples from a range of world religions and religious traditions.
Religion in Australia and the local community

SEMESTER 1 RESULT

write relations: case study

relations: Modern

tions

T

Appendix 3

Table A3: Frequency of District and SEO Support information provided in 2014 Monitoring reports

SUBJECT	DISTRICT		SEO SUPPORT REQUESTED					SCHOOLS	
	# Reports	# Panellists	Assessment	Matching syllabus standards	Interim LOA decisions	# Schools	Districts no issue	# per subject	% schools with issue
Aboriginal & Torres Strait Islander Studies	1	4	1	1	2	2	0	12	16.7
Accounting	13	106	11	10	4	12	8	243	4.9
Aerospace Studies ²	1	8	0	0	0	0	0	17	0
Agricultural Science	4	29	2	3	1	4	3	52	7.7
Ancient History ¹	13	117	29	27	23	37	3	263	14.1
Biology	13	199	28	19	13	32	5	405	7.9
Business Communication & Technologies	13	101	45	30	14	49	0	218	22.5
Business Organisation & Management	6	45	19	10	13	24	1	123	19.5
Chemistry	13	181	57	30	29	61	1	384	15.9
Chinese	3	26	8	5	4	10	0	54	18.5
Dance	7	75	22	21	7	32	2	141	22.7
Drama	13	170	33	40	26	55	2	350	15.7
Economics	9	61	15	6	9	20	1	113	17.7
Engineering Technology	1	11	0	0	0	0	1	42	0
English	13	215	33	44	32	64	4	419	15.3
English for ESL Learners	1	14	5	1	1	5	0	33	15.2
Film, Television & New Media	10	81	18	12	13	19	5	165	11.5
French	4	31	5	13	7	15	1	63	23.8
Geography ⁴	13	126	28	17	11	34	1	264	12.9
German	2	14	3	2	2	3	0	45	6.7
Graphics ³	13	144	27	28	28	38	2	283	13.4
Health Education	6	41	15	8	6	16	1	93	17.2
Home Economics	13	97	25	22	20	31	4	195	15.9
Hospitality Studies	3	22	8	7	2	9	0	47	19.1
Indonesian	1	4	2	2	2	2	0	8	25
Information Processing & Technology	12	76	16	7	6	18	5	146	12.3
Information Technology Systems	7	75	14	14	13	23	1	149	15.4
Italian	1	8	3	1	1	3	0	24	12.5
Japanese	10	84	10	9	6	16	4	170	9.4
Legal Studies	13	134	19	18	14	31	3	313	9.9
Marine Studies	6	35	6	5	3	7	2	68	10.3
Mathematics A ¹	13	187	49	24	25	60	2	420	14.3
Mathematics B ¹	13	185	25	19	20	26	6	400	6.5
Mathematics C	13	152	22	20	7	25	7	313	8
Modern History	13	150	28	25	23	43	3	317	13.6
Music	13	143	32	55	50	61	1	311	19.6
Other Languages	1	9	0	0	0	0	1	18	0
Philosophy and Reason	1	7	5	5	4	7	0	17	41.2
Physical Education	13	178	33	27	14	44	2	388	11.3
Physics ¹	13	157	13	19	13	22	3	373	5.9
Science ²¹	7	43	12	5	5	13	2	83	15.7
Studies of Religion	6	50	17	13	5	20	1	96	20.8
Study of Society	1	8	1	1	1	1	0	29	3.4
Technology Studies ⁴	9	55	5	4	7	8	4	115	7
Visual Art	13	199	18	35	35	43	3	378	11.4
Totals	367	3857	767	664	521	1045	95		

¹ district/s did not indicate number of panellists; ² Chair did not identify support reason-asked SEO to contact to discuss;

³ school identified for follow-up on 'other' matter; ⁴ No information about issue provided for one or more district.

Appendix 4

Table A4: Frequency of District and SEO Support information provided in 2013 Verification reports

SUBJECT	DISTRICT		SEO SUPORT REQUESTED					SCHOOLS	
	# Reports	# Panellists	LOA Decisions	Assessment	# schools identified	# Districts no issues	% Districts no issue	# per subject ⁵	% schools with issue
Accounting ²	12	117	5	13	14	4	33.3	243	5.8
Agricultural Science	4	28	0	0	0	4	100	52	0
Ancient History	13	138	23	17	27	3	23.1	263	10.3
Biology	12	171	26	11	28	3	25	405	6.9
Business Communication & Technologies	12	101	10	9	14	4	33.3	218	6.4
Business Organisation Management	6	44	7	2	8	2	33.3	123	6.5
Chemistry	13	195	15	30	31	5	38.5	384	8.1
Chinese ⁴	3	27	0	1	1	1	33.3	54	1.9
Dance ⁴	7	65	11	6	15	2	28.6	141	10.6
Drama	13	166	20	16	24	4	30.8	350	6.9
Economics ⁴	8	57	11	7	12	3	37.5	113	10.6
Engineering Technology	1	10	0	0	0	1	100	42	0
English ³	13	219	20	13	24	1	7.7	419	5.7
English Extension ⁴	3	26	8	1	8	0	0	63	12.7
English for ESL Learners	1	15	0	2	2	0	0	33	6.1
Film, Television & New Media ⁴	4	31	6	5	6	0	0	165	3.6
French	3	23	5	7	7	0	0	63	11.1
Geography	11	121	25	19	27	2	18.2	264	10.2
German ⁴	3	23	3	3	4	1	33.3	45	8.9
Graphics ³	13	150	50	26	53	1	7.7	283	18.7
Health Education	5	29	10	7	13	0	0	93	14
Home Economics	11	85	19	18	23	1	9.1	195	11.8
Hospitality Studies	3	20	2	3	5	1	33.3	47	10.6
Indonesian	1	5	1	1	1	0	0	8	12.5
Information Processing & Technology ⁴	12	79	6	6	9	4	33.3	146	6.2
Information Technology Systems	5	45	4	6	7	2	40.0	149	4.7
Italian ³	1	10	2	1	4	0	0	24	16.7
Japanese ⁴	9	73	4	2	4	4	44.4	170	2.4
Legal Studies ¹	12	113	17	13	18	3	25	313	5.8
Marine Studies	6	34	2	0	2	4	66.7	68	2.9
Mathematics A	13	211	17	26	32	5	38.5	420	7.6
Mathematics B	13	202	8	10	12	6	46.2	400	3
Mathematics C	13	155	9	8	14	8	61.5	313	4.5
Modern History ¹	12	132	22	5	23	3	25	317	7.3
Music	13	150	30	18	39	2	15.4	311	12.5
Music Extension	7	57	5	6	10	3	42.9	129	7.8
Other Languages	1	7	0	0	0	1	100	18	0
Philosophy & Reason	1	7	6	3	7	0	0	17	41.2
Physical Education ³	10	153	19	14	23	2	20	388	5.9
Physics	11	153	18	20	28	1	9.1	373	7.5
Science 21	7	37	4	6	6	2	28.6	83	7.2
Study of Religion ^{1 4}	5	39	4	6	6	2	40	96	6.3
Study of Society	1	10	1	1	2	0	0	29	6.9
Technology Studies	10	70	13	3	13	4	40	115	11.3
Visual Art ³	7	114	17	3	19	1	14.3	378	5
Totals	344	3717	485	374		100	29.1		

¹district/s did not indicate number of reviewers; ²school/s identified for follow up on matters not related to LOA or Assessment; ³district/s did not identify issue; ⁴no information-back page blank; ⁵2014 Year 12

Table A5: Comparability supported and unsupported in 2013 for nine subjects LOA SA

SUBJECT	DISTRICTS ¹	SUPPORTED	UNSUPPORTED	NO SAMPLE
Agricultural Science	4	8	0	0
Ancient History	13	26	0	0
Drama	13	20	4	2
English	13	26	0	0
Information Technology Studies	7	14	0	0
Mathematics C	13	22	1	3
Physics	13	17	9	0
Study of Religion	6	12	0	0
Technology Studies	10	14	6	0
Totals	92	159	20	5

¹ In some subjects District panels were combined.

Table A6: Unresolved reviews by Authority-Subjects 2010-2013

SUBJECT	2010	2011	2012	2013	TOTAL
Accounting		1			1
Ancient History	4	4		1	9
Biology	4	5	4	2	15
Business Communication & Technologies			1		1
Business Organisation & Management		1	1	2	4
Chemistry	2	1	4	2	9
Chinese				1	1
Dance	1		1		2
Drama	2		1	1	4
Economics	2	1	3	2	8
English	4	2	3	4	13
English Extension	2			3	5
Film, TV and New Media		2	2	1	5
French				2	2
Geography	1	3	1	2	7
Graphics	2	2	4	2	10
Health Education	1	2		3	6
Home Economics	4	1		2	7
Hospitality Studies	1		1		2
Indonesian		1			1
Information Processing & Technology	1		2	3	6
Information Technology Systems	4	1	1		6
Japanese	1	2	1		4
Legal Studies	4	4	5	2	15
Marine Studies			1		1
Mathematics A	1	1		1	3
Mathematics B	2	1	2		5
Mathematics C		2	2	3	7
Modern History	1	1			2
Multi strand science	1	2			3
Music	2	1	2	1	6
Music Ext	5	2	2	3	12
Physical Education		1	3		4
Physics	4	1	2	1	8
Science 21	2	2			4
Study of Religion	2	5	2	2	11
Technology Studies	2				2
Visual Art	8	10	5	6	29
TOTAL	70	62	56	52	

Table A7: Extraordinary reviews 2009-2013

YEAR	#, SUBJECT
2009	1 Indonesian submission
	1 Graphics submission
	1 Marine Studies submission
2010	1 English Extension submission
	1 Information Technology Systems submission
2011-2013	Nil


		<h2>Form R3 (Year 11)</h2>			
<h3>School proposal for interim levels of achievement</h3>					
Interim levels of achievement of sample students	School:	School code:			
	Subject:	Subject code:	0		
	District:	Panel code:			
Very high achievement	<p>The focus of the monitoring review is to provide advice to the school about implementation of the course of study and judgments concerning student achievement at this time. Five students should be identified by code letter A, B, C, D, E and placed appropriately to indicate the interim Level of Achievement. Further information regarding the preparation of the submission is available in <i>The Moderation Handbook</i>.</p> <p>Number of students completing subject at Year 11: <input type="text"/></p> <p>School comments:</p> <p>Signed by the Principal Date: / /</p>				
	High achievement	<p>Panel advice may include consideration of the following:</p> <ul style="list-style-type: none"> • implementation of the course of study • conditions of assessment • standards of sampled student work • concordance with ACACA guidelines for quality and equity in assessing student achievement • quality and standard of assessment instruments • appropriateness of assessment criteria • proposed interim levels of achievement <p>Review panel advice:</p>			
		Sound achievement	<p>Recommended action(s):</p>		
Limited achievement			<p>Signed by Chair Date: / /</p> <p>for <input type="checkbox"/> state, or <input type="checkbox"/> district review panel.</p>		
	Very limited achievement		<p>Copies:</p> <ul style="list-style-type: none"> • Before meeting: Original, duplicate and triplicate forwarded with samples to be reviewed. Quadruplicate retained by school when making initial submission. • After meeting: Original returned to school. Duplicate retained by Chair and returned to District Coordinator. Triplicate retained by Chair. 		

Figure A4: QSA Form R3 (2013)

Disciplinarity of Judgement

INFORMATION SHEET

Project team and contact details

Professor Claire Wyatt-Smith
Executive Dean of Education and Arts
Australian Catholic University
Claire.Wyatt-Smith@acu.edu.au

Peta Colbert
Research Fellow & Doctoral Candidate
Faculty of Education and Arts
Australian Catholic University
peta.colbert@acu.edu.au

Please contact members of the team if you have any questions.

Project focus

This research is being undertaken as part of the approved review of school-based assessment in senior schooling. The research focuses on how stated criteria and standards are used by experienced teachers to arrive at judgements of quality in different discipline areas and in the context of standards-referenced moderation. The main focus is on standards, judgement and disciplinarity. The research will provide essential information about discipline responsive ways in which experienced teachers apply stated standards in the work they undertake on moderation panels.

They will be gathering information in the course of moderation meetings next week. They will also be seeking your agreement to participate in individual interviews or focus group meetings. Their attendance at the moderation meetings will be for observing standards-based judgement in operation. They are also interested to hear from teachers who would agree to be interviewed, either individually or in focus group meetings.

What this means for teachers involved in the QSA verification processes

The researchers will be undertaking observations of panel meetings and are seeking teachers willing to participate in interviews and focus groups held during the day.

The interviews are planned to take approximately 15 minutes each and focus groups 30 minutes each with each recorded with permission to allow the researchers to refer to responses to ensure accuracy of representation. Copies of the recordings can be provided to individual participants upon request.

All data are de-identified to ensure no details are contained in the materials that would assist in identification of participants. *Please note:*

- *that participation in this research is voluntary and you are free to withdraw at any time, without comment or penalty;*
- *that all data from schools and staff will be confidential and de-identified;*
- *pseudonyms will be applied in publications to ensure the privacy of schools and teachers.*

Communication of the findings

Data from observations, interviews and focus groups will be used in the informing paper written by Professor Claire Wyatt-Smith. Please note that names of individual teachers will not be used in reporting and thus no teacher, school or student will be individually identified. Findings will be focused on disciplinarity differences in applications of standards to student work to reach judgements about quality.

Disciplinarity of Judgement

PARTICIPANT CONSENT FORM

Research team and contact details

Professor Claire Wyatt-Smith
Executive Dean of Education and Arts
Australian Catholic University
Claire.Wyatt-Smith@acu.edu.au

Peta Colbert
Research Fellow & Doctoral Candidate
Faculty of Education and Arts
Australian Catholic University
peta.colbert@acu.edu.au

Please contact members of the team if you have any queries about the project.

Participant consent

By signing below, I confirm that I have read and understood the *Information Sheet* and in particular have noted that:

- I understand that all information collected will be treated confidentially and the anonymity of myself will be maintained in any data, reports or publications resulting from this research;
- I have had any questions answered to my satisfaction;
- I understand that no foreseeable risks are involved for myself;
- I understand that my participation in this research is voluntary;
- I understand that I am free to withdraw at any time, without comment or penalty;
- I understand the risks involved, having read the information provided to me; and
- I understand that if I have any additional questions I can contact the research team.

Name: _____

Signature: _____ Date: _____

Table A8: Composite of data collected for the investigation

DATA TYPE	SOURCE	DETAILS
Observation	District Panel meeting for Verification purposes	Areas observed: <ul style="list-style-type: none"> Accounting Biology Business Communication & Technologies English English Extension Geography Mathematics B Music Physics
	State Panel meeting for Comparability purposes	Areas observed: <ul style="list-style-type: none"> Biology Dance & Drama Economics English Film, TV & Media Home Economics Mathematics A Mathematics B Mathematics C Modern History Music Physics Visual Arts Chair Chats Unresolved Training Comparability Training
	District Panel meeting for Monitoring purposes	<ul style="list-style-type: none"> Chemistry Drama English Information Technology Systems Mathematics A Physics
Interviews	Verification	<ul style="list-style-type: none"> Biology Music
	QSA personnel	<ul style="list-style-type: none"> Assistant Director, P-12 Implementation Branch Manager, Quality Assurance Unit
Focus groups	Verification	Chairs present from Mathematics A, English, Chemistry, Physics and Biology
Document collection	Senior Syllabuses – downloaded from QSA website	<ul style="list-style-type: none"> Accounting (2010) Biology (2004) Chemistry (2007) Dance (2010) Film, Television and New Media (2005) Mathematics A (2008) Mathematics B (2008) Mathematics C (2008)

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> • Drama (2013) • Economics (2010) • English (2010) • English Extension (2011) • English for ESL Learners (2007 trial, amended March 2009) • Modern History (2004) • Music (2013) • Music Extension (2008) • Physics (2007) • Visual Art (2007)
	QSA package 1	<ul style="list-style-type: none"> • QSA Organisation Structure • Verification Timeline • Diagram of moderation process • Authority Themes and Strategies 2013 – theme communication and consistency • Previous Themes and Strategies 2008-2012 • Meeting of state review panel chairs 10 May 2013 <ul style="list-style-type: none"> ○ SEO Booklet ○ Participant Booklet • Moderation Conference July 2013 <ul style="list-style-type: none"> ○ Study of Society ○ SEO Handbook • Review panel chair meetings – Consistent and quality communication Aug 2013 • Sample of completed R6 at exit • Review panel chair workflow • Verification workflow • Verification – Preference for RTS (form) • State and district review panel chair's report – Verification 2013 (form) • Panel training packages – Quality Assurance Unit March 2013 • Panel training – Subject package 3 (slides) Oct 2013 • Panellist handbook – Extract of A-Z of Senior moderation March 2013 • Panel chair letter – verification update • Panellists letter • Verification review notes 2013 – Part B-Pre-review notes • Moderation protocol – Form R6 advice

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> • Form R6 process – Verification to exit – A step-by-step guide • Preparing a verification submission – A step-by-step guide • Verification submission checklist • Verification – Frequently asked questions • Unresolved submission review notes 2012 – overview and form • Comparability meetings – QAU Officer Handbook November 2012 • Welcome to the QSA – Suzhou International Foreign Language School • Quality Assurance Unit meeting – Making judgments about student achievement October 2013 • QSA list of Memos 2013 • Memos <ul style="list-style-type: none"> ○ Year 12 verification for Authority subjects (11 Oct 2013) ○ From verification to exit including Forms R6 and R7 procedures (11 Oct 2013) ○ Final submission of Forms R6 and associated materials. • School moderator meetings – 2012 – slides • Enhancing assessment capacity – for teachers – slides • Screenshots from QSA website <ul style="list-style-type: none"> ○ moderation handbooks ○ Senior moderation hub ○ Moderation forms ○ State review panel & random sampling reports ○ District and state review panels ○ Moderation resources ○ English 2010 assessment ○ Physics 2007 authority subject • Introducing the Qld system of externally moderated school-based assessment in Yrs 11-12 – slides • Training and support materials for 2013 monitoring including letters to chairs and panellists, Form R3 monitoring advice strategy, and Form R3 monitoring advice protocol • Monitoring review notes 2013 – Part B Pre-review notes • Sample R3 advice and action spreadsheet • Extract from assessment workshops based on feedback from monitoring

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> • Calculating Overall Positions (OPs) – The basic principles • P-12 assessment overview • A-Z of Senior Moderation October 2013 • School-based assessment – The Queensland System • Marine Science 2013 – Work program requirements – revised April 2013 • Comparability Meetings – QAU Officer Handbook November 2013
	QSA package 2	<ul style="list-style-type: none"> • 10 Packages containing Form R3-Year11, Form R6-Year12, Student Profile, Comparability LOA Review Summary, Database list of Issues and actions for – <ul style="list-style-type: none"> ○ Agriculture Science (Brisbane South, Toowoomba, Rockhampton (2), Wide Bay) ○ Ancient History (Rockhampton, Brisbane South, Mackay, Brisbane East, Wide Bay, Gold Coast, Cairns) ○ Dance (Brisbane Central, Brisbane South (2), Brisbane North (2)) ○ Drama (Brisbane South, Brisbane Central, Brisbane-Ipswich, Toowoomba, Gold Coast) ○ Information Technology Studies (Brisbane South, Brisbane North, Mackay, Townsville/Cairns, Townsville) ○ Mathematics A (Toowoomba, Brisbane South, Wide Bay, Mackay, Rockhampton) ○ Mathematics C (Brisbane South, Rockhampton, Brisbane North, Brisbane Central, Cairns) ○ Physics (South Coast/Gold Coast, Sunshine Coast (2), Brisbane South, Brisbane East) ○ Study of Religion (Rockhampton, Brisbane Central, Ipswich-Brisbane, Brisbane East, Brisbane North) ○ Technology Studies (Wide Bay, Ipswich-Brisbane, Sunshine Coast (2), Toowoomba) • Published assessment instruments, original school documents and completed Tool for Schools including: <ul style="list-style-type: none"> ○ Physics 2007 sample assessment instrument ○ Senior Assessment Hub – Quality assuring senior assessment instruments: A tool for schools (t4s) <ul style="list-style-type: none"> ▪ Physics ▪ English ○ 2013 Verification # overview ○ Breakdown of Panel membership by district, state and sector as at 22 January 2013 ○ Application for membership of a review panel – form ○ Application for review panel chair position ○ Economics 2010 – sample assessment instrument and student response ○ English Yr 12 Sem 1, 2012 sample assessment and annotated student response and judgments across dimensions ○ Information Processing and Technology 2012 – sample assessment instrument and A tool for schools

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> ○ Meeting paper – QSA – 19 June 2013 – Review panel system and extended experimental investigations • Building Professional Capacity in Educational Assessment – Chief Executive Study Leave Report – Paul Kilvert – April 2012 • Professional development materials from Review Panel Chair meetings in districts 2013 – presentation, Key Messages and participant booklet (Consistent and quality communication August 2013).
QSA materials in response to email request 29 January 2014		<ul style="list-style-type: none"> • Random sampling project - 2013 Report on random sampling of assessment in Authority subjects (rec 30 January 2014) • Memo - District and state Year 11 monitoring meetings 2014 (except extension district review panels) (rec 30 January 2014) • State Review Panel Report 2013 (February 2014)
QSA email 1 February in response to request		<ul style="list-style-type: none"> • Review Officer (RO) position description and work profile • Standards and Assessment Officer (SAO) position description and work profile • Senior Education Officer (SEO) position description
QSA package 3 delivered 7 February 2014		<div>State and District review panel chair's report – Verification</div> <div>2013</div> <div> <ul style="list-style-type: none"> • Accounting • Agricultural Science • Ancient History • Biology • Business Communication & Technologies • Business Organisation Management • Chemistry • Chinese • Dance • Drama • Economics • Engineering Technology • English Extension4 • English for ESL Learners • English • Film, Television & New Media4 • French • Geography • Hospitality Studies • Indonesian • Information Processing & Technology • Information Technology Systems • Italian • Japanese • Legal Studies • Marine Studies • Mathematics A • Mathematics B • Mathematics C • Modern History • Music • Music Extension • Other Languages • Philosophy & Reason • Physical Education • Physics </div>

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> • German • Graphics • Health • Home Economics
		<ul style="list-style-type: none"> • Science 21 • Study of Religion • Study of Society • Technology Studies • Visual Art
	QSA email 26 February 2013	Information on extra ordinary reviews for the last 5 years.
	QSA package 4 delivered 12 March 2014	State and District review panel chair's report – Monitoring 2014 <ul style="list-style-type: none"> • Aboriginal & Torres Strait Islander Studies • Accounting • Aerospace Studies • Agricultural Science • Ancient History • Biology • Business Communication & Technologies • Business Organisation & Management • Chemistry • Chinese • Dance • Drama • Economics • Engineering Technology • English • English for ESL Learners • Film, Television & New Media • French • Geography • German • Graphics3 • Health Education • Hospitality Studies • Indonesian • Information Processing & Technology • Information Technology Systems • Italian • Japanese • Legal Studies • Marine Studies • Mathematics A • Mathematics B • Mathematics C • Modern History • Music • Other Languages • Philosophy and Reason • Physical Education • Physics • Science21 • Studies of Religion • Study of Society • Technology Studies • Visual Art

DATA TYPE	SOURCE	DETAILS
		<ul style="list-style-type: none"> Home Economics
	QSA email 3 March 2014	Financial information about the costs associated with quality assuring Year 11 and 12 Senior Schooling

References

- Board of Secondary School Studies (BSSS). (1986). *Introduction to discussion paper*. Brisbane: Author.
- Campbell, W.J., Archer, J., Beasley, W., Butler, J., Cowie, R., Cox, B., Galbraith, P., Grice, D., Joy, B., McMeniman, M., Sadler, R., & Wilkes, R. (1983). *Implementation of ROSBA in schools*. Unpublished report to the Minister for Education, Brisbane.
- Campbell, W. J., Basset, G. W., Campbell, E. M., Cotterell, J. L., Evans, G. T., & Grassie, M. C. (1975). *Some consequences of the Radford Scheme for schools, teachers, and students in Queensland*. Final Report of Project. Brisbane: Australian Advisory Committee for Research and Development in Education.
- Fairbairn, K., McBryde, B., & Rigby, D. (1976). *Schools under Radford: A report on aspects of education in secondary schools in Queensland since the introduction in 1971 of school-based assessment*. Brisbane: Department of Education.
- McMeniman, M. (1986a). *A standards schema* (Discussion paper 3). Brisbane: Assessment Unit, Board of Secondary Schools Studies.
- McMeniman, M. (1986b). *Formative and summative assessment: A complementary approach*. Assessment Unit Discussion Paper 6. Brisbane: Board of Secondary School Studies.
- Pitman, J. (2002). *The Senior Certificate: A new deal*. Brisbane: Education Queensland and Board of Senior Secondary School Studies.
- Queensland Studies Authority (QSA). (2010). *School-based assessment: The Queensland system*. Brisbane: Author.
- QSA. (2013). *A-Z of senior moderation*. Brisbane: Author.
- QSA. (2014). *What is verification* (ppt). Brisbane: Author. <<http://www.qsa.qld.edu.au/2138.html>>
- QSA. (Undated). *Quality-assurance processes* (ppt). Brisbane: Author.
- Sadler, D.R. (1986). *Subjectivity, objectivity, and teachers' qualitative judgments*. Assessment Unit Discussion Paper 5. Brisbane: Board of Secondary School Studies.
- Sadler, D.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D.R. (1991). *Rethinking certification policies for school leavers*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Gold Coast, 29 November.
- Scott, D.R., Berkeley, G.F., Howell, M.A., Schuntner, L.T., Walker, R.F., & Winkle, L. (1978). A review of school-based assessment in Queensland secondary schools. Brisbane: Board of Secondary School Studies.
- Smith, C.M. (1995). *Teachers' reading practices in the secondary school writing classroom: A reappraisal of the nature and function of pre-specified criteria*. Unpublished PhD Thesis. Brisbane: University of Queensland.
- Viviani, N. (1990). *The review of tertiary entrance in Queensland 1990*. Submitted to the Minister for Education, Queensland Board of Senior Secondary School Studies, Brisbane.
- Wyatt-Smith, C. M., & Bridges, S. (2008). *Meeting in the middle – assessment, pedagogy, learning and students at educational disadvantage*. Final Evaluation Report for the Department of Education, Science and Training on Literacy and Numeracy in the Middle years of Schooling.
- Wyatt-Smith, C., & Klenowski, V. (2013). Explicit, latent and meta-criteria: types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy and Practice*, 20(1), 35-52.
- Wyatt-Smith, C., & Matters, G. (2007). *Proposal for a new model of senior assessment: Realising potentials*. Report prepared for the Queensland Studies Authority. Brisbane: Griffith University and Australian Council for Educational Research.