

Submission to the Queensland Parliament Inquiry into Assessment Methods for Senior Mathematics, Chemistry and Physics

Gary Turner B.Sc.(Hons) (Qld) Ph.D. (Auck) B.Teach (Hons) (Tas) MZIP
Physics Teacher

Terms of reference

1. That the Education and Innovation Committee inquire into and report on the assessment methods used in Senior Mathematics, Chemistry and Physics in Queensland schools.
2. That, in undertaking this inquiry, the committee should consider the following issues:
 - Ensuring assessment processes are supported by teachers
 - Student participation levels
 - The ability of assessment processes to support valid and reliable judgments of student outcomes.

This submission is from one who is generally happy with the **content** of the syllabus, and I do not care whether we use marks, grades or ticks to decide whether a student's work is correct or incorrect.

I care only that a system is clear to both students and teachers, and easy to use.

The intent of the Physics and Chemistry Syllabi as presented in the syllabi has been undermined by the uncertainty around the Criteria used for Assessment, and the consequent investment of time, by students and teachers, in the new assessment pieces.

From the nature of the uncertainty, and attempts at rectification since the introduction of the Syllabi, the inference is that the Queensland Studies Authority has been, and remains, under-funded and through being pushed into burgeoning number of areas (Junior High School and Primary), over-stretched, such that workers in any one year have been processing material without familiarity with what may have been issued in a previous year.

The uncertainty in the Criteria that are used for assessment means that the processes that are intended to create a verifiable and reliable judgement of students cannot do so. Review panels do try to create that certainty, yet it is a continuing (and to the frustration of teachers, retrospective) bi-annual attempt. Despite the Criteria being labeled "standards", they have not been standard as several interpretations and glossaries have had to be issued.

I support the six underlying principles of exit assessment, that is, continuous assessment, balance, mandatory aspects of the syllabus, with significant aspects of the course of study, selective updating for fullest and latest information.

The expectations around the development of a school assessment program would be reasonable and workable in principle, except that in the current syllabi (Physics and Chemistry) the new assessment-tasks (i.e. Extended Experimental Investigations EEI and Extended Research Tasks ERT) require from teachers additional marking-time which has never been provided but instead has had to come from teacher's own time.

Worse, a draft (a concept which was merely offered in the syllabi as one of several options through which to ensure authenticity of a student's-work) has come to be seen as a formative teaching-tool of its own and thus requires a full marking-session (with many comments written in detail in the Grade 11 version) which is then followed a week later by a full-marking of the final version. Marking and grading in total thus takes much longer than for a traditional exam. In addition, the time-constraint on commenting on drafts is much more rigid than for marking the final assignment, since the draft has to be returned to students within a few days so that students can finish the assignment (the dead-line is set usually within a week of submitting the draft) so teachers are more-greatly pressured for time.

Verification and post-verification requirements are reasonable, but the timing, and allowed modes of communication cause a problem if a Panel identifies a problem with a school's folio as presented at Verification (at the end of Term 3) for then it is difficult for a school to remedy any deficiency(ies) by using the opportunity of the remaining assessment-item in Term 4.

The recommendations for change that could ensure more teachers support the assessment processes include the provision of certainty of assessment-criteria upon release of syllabi – and ensuring contact with teachers of the subject, especially if new interpretations are released as teachers new to the subject cannot be expected to know to ask.

To ensure the changes are brought to teachers' attention, each and every change (to date made by memos and release of glossaries) should be made by way of release of an amended syllabus. Only in that way would schools be sure to give to those teachers who are new to the subject all the current-changes. The current practice of releasing changes as memos and glossaries (rather than as full reprints of the syllabus) could be interpreted as a cost-saving measure.

In addition, the certainty of assessment-conditions (i.e. length of assignments) with a standard example of what an "A" for each criterion is in a sample task could have been provided in 2007 if the release of the syllabus had not been rushed (for whatever was the reason).

Student participation levels

Assessment practices influence participation levels to some extent – but the students who know the career-path just "grin and bear" the assessment-load. Their participation is determined by that career-path.

For those students who are influenced by the assessment, it is not the tasks within a single subject that is the determining factor but the sum of work in all 5 (or 6) subjects.

The stated intent of the syllabi was to encourage more students to undertake senior science and to do so with the expectation that because of increasing number of options in life, teachers could do so without the expectation that they were teaching to a homogenous university-track cohort. Yet perversely, the introduction of the Queensland Certificate of Education, has caused the opposite to become the case.

(Because schools need to show that they are producing a large percentage of school-leavers who have achieved the QCE, then students are re-directed at the end of Term 1 in Grade 11 to subjects

that they are most likely to pass – thus leaving in Physics (and Chemistry) only those students who ARE on the University-track).

The ability of assessment processes to support valid and reliable judgments of student outcomes

The processes around assessment-implementation do not entirely support valid and reliable judgments of student outcomes, because of the nature of some of the assessment-types.

This uncertainty is because it IS difficult to know how much outside support has been given to students, although to a large extent that can be overcome by the class-room teacher taking to students –

but more to the point it is difficult to know, between schools, how much assistance or direction has been given to students in forming the projects, in analysis and in writing reports.

For example, how many drafts (or parts of drafts) have a student been allowed to submit for comment;

and how detailed were the comments, are they Socratic questioning (and if so, whether too pointed and leading), or direction instruction of what to do.

Certainly one does not see any comment about the level of assistance given to an individual student annotated onto any student-report of an Extended Experimental Investigation (EEI).

How can an “A” given for responses in an exam for which a formulae-sheet has been provided, be compared to the grading for an exam for which a formulae-sheet has not been supplied? And the QSA has given no guidance on how to compare the grading for exams that have, and do not have, a formulae-sheet?

How does it look that teachers who have set the exam, or assignment, can offer after-school tutoring to those of their own students? [Consider the reaction if the setters of an external exam were holding small tutorial-sessions?]

The assessment criteria could allow teachers to make valid and reliable judgments if they had been properly defined, and the words had a clear, precise meaning, and even that is after years (7) of attempts at refinement – which period itself is a censure of the inadequate process of the implementation and release of the syllabus-document, a short-coming that could be symptomatic of a lack of resources (which has consequently caused a great problem in a very precise and powerful subject that cannot operate efficiently when treated in such a superficial, under-resourced manner).

Common-conditions could be enforced, for example, as to length of report. How can an “A” in a 3,000 word report be compared to a “A” in a 8,000 word report?

How can teachers with a class of 20 students reliably and consistently read through a set of reports which can be from 20 to 60 pages long? Egregious mistakes in concepts are being buried in a mass of words, and teachers and reviewers do not notice.

For EEIs I talk to students as they work, ask to see drafts, and read a log-book, which are the same claims that I read for other schools.

However it is not knowable how much assistance is given verbally in-class by teachers (or after-school tutors), or whether a second, longer, more informal set of instructions are given to students but not included with the official set of task-sheets sent to panel-meetings.

Teachers have difficulty applying the standards.

A teacher who has some experience is still set-back by receiving comments that one's judgement is not matching the syllabus. In one subject that I taught, and sent for a Verification a Folio that I had graded overall at just above border-line "VHA" [that is, low "A"], was returned as high "SA" [i.e. C+]. I telephoned the panel-chair and was told that the two panelists had graded it separately as mid "B" and mid-"C" and that the assigned grade was simply an average.

I was told that the "B" was given by the pre-reviewer and that on that basis that the person would have spent more time on it (i.e. more than the two hours allowed in the panel-meeting) than the second panelist, that grade could be accepted.

First – that means that three experienced teachers (the two panelists and myself) had separately graded that Folio overall as "A", "B" and "C".

Second – the time allowed for reviewing is inadequate. The time-allowance was not changed when the Syllabus was changed – despite the introduction of two new assessment-types, i.e. ERT & EEI which each might require half-an-hour to read and grade, of which the reviewer has to read four (plus the exams) in 2 hours. Again – evidence that the implementation of the new syllabus was under-resourced, under-funded, and no proper assessment of its requirements made.

When a set of Folios from Random Sampling (of another District) is received, the reviewer has two hours in which to grade 8 Folios (i.e. a year of work from 8 students) which includes grading 8 EEI-reports (and possibly as many ERT-reports) and several exams. Doing half of the Folios in about 4 hours is barely possible. Two issues arise: it is difficult to ensure then complete comparability across the districts; and when egregious examples of conceptual errors are seen in an assignment one realises that the error has passed unremarked by the student's teacher and the two reviewing-panelists, again emphasizing the time-shortage which militates against the process working as intended.

Despite the obvious inconsistencies, even if a Folio is re-judged as VHA1 instead of VHA10, the official verdict is still that the process has produced a comparable result (since the grading is still "A") and so is said to be working successfully.

Across panelists, and panels of diverse districts, the discernment of the difference between "analysis" (= B Grade) and "systematic analysis with relationship between patterns" (= A Grade) in the Criterion for Investigative Processes is fraught with tension and uncertainty.

And this has been demonstrated as recently as within the past week, when teachers realized that a sample of a Chemistry EEI that has been posted by QSA used data that was qualitative and yet was annotated to suggest that some form of systematic analysis had been displayed and some patterns related, thus being graded as "A" standard – which was news to many teachers commenting on the same.

And similarly in the difference between Knowledge & Understanding v Evaluation & Concluding. I have seen an EEI graded at A+ (in a Folio of VHA 10) with an egregious conceptual error (of size of electrical current depends on speed of electrons) not noticed by the original teacher nor by two panelists –

And similar errors in all topics covered in Reports from EEI – over-sights driven by both the size of the EEI with inadequate time given to teachers to mark properly (ie. comprehensively) and panelists having insufficient time to review (and errors in understanding which would have been obvious in even the most straight-forward exam-question).

And there are many questions in exams categorized as “A” by the teacher setting the exam, and panelists saying the setting-teacher has not understood (or usually, in practice, not seen) the QSA re-definition (subsequent to the release of the syllabus) of the meaning of the criteria. The teacher will not have seen the memos, and glossaries-memos, upon being told to teach the subject, because the teacher will have simply read the syllabus assuming that it is complete unto itself, and having no reason to guess that there are other necessary material.

Scaffolding, i.e. guidance and feed-back, is given during the learning and drafting process.

Scaffolding for Grade 11 can be quite detailed,

going as far as specifying what should be the contents of each part of a Scientific Report; and how to go about working out from some quantity of theory, how a researchable question can be worked out;

and how some amount of data is needed to have enough to analyse.

Students have to be talked through the jargon of each criterion, because the criteria are not written in language that is either easily understood by students, or is capable of only a single meaning – yet the QSA insists that the criteria-sheets attached to the students’ task-sheet should not stray from the official wording in the syllabus.

The difficulty for teachers who are given the syllabus at the start of a school-year and told to teach it, is that the lack of a single, certain meaning is that the teacher is most likely to choose the “wrong” one and find out only at the end of the year (if Grade 12) or next year (if Grade 11) – and yet being an educated person, and the wording of the criteria often having a plain-meaning, the teacher has no reason to suspect that an “interpretation” is required.

The necessity for an interpretation, for students or new teachers, is itself an alarm that the criteria have not been well-written and were not well-tested before release.

Students have to be told that if one wants to comment on a relationship between patterns in data, then one needs two patterns (at least) in the data, which means two trends or trends in a pair of related variables – subtle but important points that students do not otherwise deduce from that criterion. [An alternative is to have a pattern in one set of data and to then discern and explain a relationship between that pattern and an anomaly in the pattern].

On reading a draft, for Grade 11, the comments can be quite specific, such as “there is no theory presented which provides justification for the hypothesis” or “how does this hypothesis follow on from the previous theory”;

Or “what is the trend in the graph” or “do you have enough data to construct a graph” or “have you thought about what the intercept on the y-axis means?”

In Grade 12, students are supposed to remember what they did (by way of working out what to do and how to read criteria) from the one Grade 11 experience which may be more than 12 months prior, and so is very unlikely – so in practice the students might ask enough questions that the teacher is supplying as much detailed information as in the previous year, even though the explicit direction in the syllabus is that less scaffolding must be given in Grade 12 (and for such reason teachers do not WRITE as much scaffolding in the task-sheet).

The verification requirements could allow for valid and reliable judgments of student outcomes, but in practice, not.

Teachers say that the best Professional Development is to be on a Moderation Panel. That necessity is an indictment of the implementation of the syllabus, for it means that the non-panel teachers have been given criteria that they cannot understand at face-value. So the panels are reading exam-questions written by well-intentioned teachers who take at face-value the KCU criterion that “A” questions have complex and challenging concepts – and choose concepts that they think are challenging - but which in the time of the use of the syllabus the QSA decided should be read as “in complex & challenging circumstances” (on the basis that concepts cannot be categorized as challenging or non-challenging) – and as for many similar on-the-run definitions, the changes are “announced” by memo within the QSA system, instead of re-writing the syllabus (despite there having been one re-write simply for purporting to reduce the length of the Discussions in the EEI and ERT, and the definitions in the memos could have been included as changes at that time of re-writing). The importance of that dereliction is that while schools would ensure that new syllabi were brought to the attention of a new teacher, there exists no mechanism for ensuring new (nor indeed continuing) teachers receive or read memos on any topic.

For exams, it is difficult to know how much practice has been given in a question which, when it appears on an exam, is claimed to be “unseen” such that the Question is not then routine (and thus categorized as “A” standard), but which many students (even those who have not done well on any other question) do well – but more significantly, all answer in a particular way when there may be several ways to attack the question. Or in some cases, the standard-answer supplied has a specific mechanical error, and the students have been graded as “correct” when including that error, so that then the only way that the students could know that was to have seen that answer to the question (or the answer to a nearly identical question) – so the question has been practised, despite the claim of being “A” grade for not being seen.

Panelists are told not to re-mark submissions – but if a 68-page Report of an EEI is submitted with 74 ticks scattered about the pages, and the teacher awards a B, A, B+ in each of the respective criteria, what is the panelist to do except read it carefully to decide what has caught the teacher’s eye, and to decide to which criterion the sentence or paragraph in the student’s response might correspond, and what grade that might match (i.e. in official jargon, which part of the criterion-standard does that part of the response match). So the panelists are forced to “mark”. The panelists are told to “look for evidence to support the school’s judgement” – but if the school has not indicated clearly the evidence that supports its judgement, then “looking for the evidence” does amount to marking the report (for reviewers need to make notes of pieces that align with a criterion and at what standard). So all the while, the panelists are conscious that they are doing the contrary to their instructions, yet knowing that is the only way that makes the system work. For persons concerned with being precise, and following official instructions, that conflict of orders versus necessity creates a lot of angst, and a concern that the order-giver is ignorant of (or denying) the reality of what the panelists are being expected to do. That angst, added to the uncertainty of the criteria, and that each year they see coming from new teachers work that demonstrates the resulting confusion (for teachers, and thus for the students of those teachers), has generated a lot of tension in panelists who have to try to make the system work.

Recommendations or suggestions for change to ensure assessment processes support valid and reliable judgments of student outcomes.

Precise, single-meaning, criteria (determined, with examples supplied) before the syllabus is released.

Removal of barely distinguishable criteria – e.g. EC from KCU – when even the syllabus-committee itself, pre-release of the syllabus, wanted to have only 2 groups of criteria, not three – but which the QSA for ideological reasons was fixed upon having three (for no educational or Phys-Chem reason other than to match the humanities subjects).

More time for reviewers in panel (if the same assessment-techniques are retained).

QSA to have a budget that allows communication with the subject-teacher(s) in each school – rather than having teachers say that the best PD is to be on panel which is an indictment of the syllabus, and the implementation thereof, that the non-panel teachers have been given criteria that they cannot understand at face-value.

That those DRPCs who might be willing to do so are allowed to circulate physics-teachers within the District to communicate a summary of general issues seen in each stage of Monitoring – and that doing that in writing is allowed – at present QSA forbids those actions, saying it is not part of the job of the DRPC.

Physics, and commenting in detail on critical technical questions of a large Report or difficult question in an exam, is very difficult to do orally – physics-teachers prefer to have something on paper to see and read and mull over, so as to come to a considered opinion – not be compelled to debate orally by phone. So now we have the position that many DRPCs (and as a school-moderator for 3 years I saw notes from a lot of them) at whatever cost, ignore the annual written instruction from QSA to not so communicate – just to make the system work - as so many teachers acknowledge in public or semi-public fora (i.e. discussion-lists).

Details and examples follow.

My experience is as a Teacher (of Physics for 5 years of this syllabus, and Chemistry of 3 with this syllabus), and 3 years as a school-moderator), in:
first, basic mechanics of the system
and second, the new syllabi

It would not be so bad if the new syllabi (especially the criteria for grading) actually produced a better result (even increased participation) for students and teachers - but they have not, instead the change has been great collateral damage, increasing teachers' angst and working-hours (and long assignments for students - since the best time to do most of these in Terms 2 & 3 the Science-Maths "A" students could be doing 10 exams and up to 50,000 words of assignments in total across 5 subjects within about 18 weeks).

Physics (and for Chemistry since it has the same syllabus)

The problem is that in producing a syllabus to meet a template, the QSA created a syllabus which:

- had assessment-criteria that had nothing to do with Physics, nor with Physics-education (presumably due to lack of resources for proper research and trials);
- had assessment criteria that had not been trialed (a Trial Pilot had different criteria);
- and had assessment-tasks that drove (and drives) the whole teaching (without substantial trials to show how much work would be required in implementing the assessment-techniques, on behalf of students and teachers, and particularly in the effect on students of having 5 subjects demanding such large assignments).

Even so, there would not be such difficulty (as a result of the assessment driving the teaching) if the assessment-criteria were certain, clear, and had a plain-meaning to both teachers and students.

Even now, as experienced at the current (2013) round of Monitoring, teachers, experienced in teaching or at physics, consummate professionals, start teaching Physics, read the syllabus, take the criteria at face-value and are in retrospect (after a full year of teaching Grade 11 in the following February, or after 3 terms of teaching Grade 12) told that is not what the criteria are taken to mean.

Any time the position is sought to be clarified, or a common meaning created, the QSA response is to tell DRPCs not to do so, and told expressly during negotiation not to put anything in writing by fax or email – thus some operate in contravention of the instructions – for Verification (end of Term 3) or Exit (end of Term 4) changes; and in explanation of Monitoring-comments; and told not to talk to Heads of Departments but only to the teacher in the subject and ensure the discussion is not recorded); yet the QSA obviously does not have the resources to do so.

We work in a subject(s) in which words have precise meanings – and we spend a lot of time seeking clarification of the meaning, and teaching the students how to understand that – and yet on coming in to teach it, teachers are given a document that is full of words that together have no clear meaning (or rather words that could have many meanings). For example even when one might allow that late changes to the syllabus are necessary as problems are revealed during implementation (since no trial was held) , the QSA tried (2010) to fix that by creating a second “glossary” but has no way of ensuring every teacher new to the subject has the glossary or knows that the glossary exists. A teacher new to the subject has no reason to go looking for a subsequent glossary.

Even at more bureaucratic levels, there are shifts in meaning which never completely flow to all teachers – as QSA refuses to try to contact each subject-teacher directly, and forbids those DRPCs who might attempt to maintain a district list and semi-annual circular) to tell everyone whatever is any new “interpretation” until the (retrospective) comments appear on the comments to schools after Verification (end of Term 3 Grade 12) or Monitoring (February following Grade 11).

For example – previously “fullest and latest” has been used to justify a late change based on a perceived trend in results of later assessment pieces (e.g. those in Semester 2 of Grade 12 compared to those in Semester One of Grade 12) provided the pieces cover the three General Objectives, but this year by words spoken at a few meetings, and then in semi-informal pieces such as an email to a single teacher (in 2012), the “interpretation” is now being said to be that it is merely the reasoning for making Gr 12 results summative in place of the Gr 11 results, no more than that. There is no system for distributing the news to every one of us. And since the system works in retrospect (teachers receiving advice on the R6 at Verification after 3 terms of work) that is a further guarantee of inconsistency and unreliability.

Physics Teachers Discussion List 5 August 2012 – still discussing word-length

On looking back through the Physics-Teachers Discussion-list I saw that these issues were well-debated in July-October 2010, because of the Memo released at the beginning of the year [March 2010] (as the Memo nominally reduced the Word-Length).

The official response was that the “word-length” was not to be enforced as a “limit” as one of the panellists wrote on 25 August 2010:

‘At panel training the other day we were clearly told that students were not to be marked down for exceeding the word-limit. It came up because the “A” exemplar provided by QSA was about 5000 words when the word-limit stated 1500 words. There was some debate but the conclusion was if the document contains valid discussion and it exceeds the limit (by any amount) the student was not to be penalised at all. If however the document contained waffle then the student should be. Hence word count is only a “suggested limit”.’

And the official “Form R6 advice” for Panels for 2010 listed as “Inappropriate wording”:

“The panel strongly recommends that students be penalised for length of response ..”, and the “Rationale” [for not enforcing the word-length as a limit] given was “Advice must be based on the evidence available and the requirements of the syllabus.”

Nothing has appeared since officially to contradict that advice to panels.

In response to a newspaper report

Wed 8 August 2012 Courier-Mail.

In regard to the points in the article with respect to the Memo of July 2012, in which the QSA sought to present the “facts” about the syllabi.

The issues arise from issuing of the syllabus in 2007 without the field-testing of the Criteria and Standards that are supposed to tell us what is expected of an “A” student. There had been trials of a syllabus– but the Standards in the final version that was issued were very different from the ones used in the trial. And the words in the Standards often did not carry their usual meaning – and worse, in combination of words, Physics-teachers could not easily work out what was expected. [I’ll insert a few examples at the bottom **in the Appendices**].

Then it became apparent that the "A" grade maths-science students who may enrol in at least two of the three Senior sciences, and at least one Maths, could be writing 3 to 5 assignments in one year, and very often (because of timing within schools, and to give students time to adjust to what was expected each year) the majority would be due within Term 2 or 3.

Since the Standard for what is "A" seems very open-ended, students over-wrote. After some complaints apparently, the QSA Board of Governors issued in March 2010 a Memo ordering a reduction in the "Word-length" of Written Reports, and also a Glossary of the words used in the Standards.

The Memo-writer stated that the word-lengths in the syllabus 'till then had been "recommended", but the word "recommended" did not appear in the original syllabus, except in the sense of increasing the word-length from Grade 11 to Grade 12. (One can infer that the memo-writer was not au fait with the syllabus, a short-coming presumably again a consequence of lack of resources.)

In reference to the QSA Memo of July 2012:

First – Marks

The QSA says that it has never disallowed the use of marks.

But – in all training-sessions up to 2010 Marks were explicitly made taboo. In other advice to teachers in implementing the syllabus the advice was explicit.

At least two documents of the QSA make it EXPLICIT that marks were wrong and had no place in the new syllabi in Physics and Chemistry as issued in 2007.

I refer to these two in particular:

1. Notes to Panellists reviewing work programs:

Information for Physics Work Programs based on the 2007 Syllabus:

About marks and marking schemes:

Remember there is NO clear, transparent link between marks, general objectives and the Exit criteria& standards matrix...we need to squash the idea that marks can be used to grade tasks, keep referring people back to ROSBA discussion paper #21

2. Chemistry and Physics – Advice for teachers

The importance of instrument-specific criteria and standards

"Decisions about levels of achievement need to be made in the context of a folio of evidence and the extent to which that evidence matches with the syllabus standards at a certain level. The syllabus requires each assessment instrument to be accompanied by a task-specific criteria sheet that is derived from, and is consistent with, the syllabus exit criteria and standards. **Attempts to encode these standards in numerical marks are not consistent with the syllabus intent.**" [emphasis added]

<http://www.qsa.qld.edu.au/10-12/5713.html> listing for "Chemistry and Physics – The importance of instrument-specific criteria and standards" www.qsa.qld.edu.au/.../10-12/snr_chem07_phys07_instrument-specific.pdf

Also the syllabus requires that at Verification (the final Moderation of the work of students in October of Grade 12), Grades have to be awarded and if any other system is used then an explanation of the conversion of marks to grades has to be supplied – and no-one has ever worked out what that would look like.

Second

Individual questions do not have to be graded. Could be correct in principle -

but impractical on several counts.

A. If schools have a lot (say, 50) of students doing Physics– and these days most who would fail change subjects half-way through Grade 11, so these are going to have Grades C- to A+. The school has to issue to the QSA a final ranking of these students on a 10 point scale in each of A, B and C grades – that is from C1 to A10. So those 50 students have to be ranked on a ladder of about 30 rungs. Since the schools overwhelmingly follow the advice not to use marks (for the reasons above), then the school has to have some way of making a very fine judgement between the students – and that amounts to ensuring that in each exam-paper each individual question **is** graded and recorded.

B.

B. The students have to be graded in Three Criteria – and it is difficult in the time allowed for an exam to have a range of questions from C grade to A grade for each of the three criteria (in fact it is more difficult than that as each criterion has three sub-criteria). There are many types of questions that are said to be merely C-grade type questions (i.e. most of the factual recall and simple calculations). So the school has to set an exam with a good number of C-grade questions, and then try to squeeze in a few “A” grade questions. So certainly each of the questions HAS to be graded as there may only be one, at most three, A-grade questions of EACH criteria on an exam. (Sometimes the assessment of one of the criteria is left to the Written Reports, as the QSA says that is the “best opportunity” to assess that criterion).

Third

Word-Length has been reduced.

In the Memo [Appendix C] of 2010 the QSA Board of Governors declared that the Word-Length of Reports was to be reduced to 1,000 – 1,500

Yet through-out 2010, the QSA repeatedly told teachers that that was only a “recommended” Word-Length NOT a limit. This occurred in the August 2010 meetings of Panels – for which there is no written record, but one can find written contemporaneous recollections by teachers shortly afterwards in an on-line Discussion-List, and in one place that advice not to enforce the Word-Length did appear in the formal set of wording (**Form R6 advice**) given to the District Panels, for use in assessing the work of schools in October 2010. The explicit statement is:

“Form R6 advice” for Panels for 2010 listed as **“Inappropriate wording”**: “The panel strongly recommends that students be penalised for length of response ..”, and the “Rationale” [for not enforcing the word-length as a limit] given was “Advice [about grades] must be based on the evidence available and the requirements of the syllabus.”

To show how inconsistent is the system – also note that in the March 2010 Memo “The word length for the EEI and ERT reports **includes** discussion / evaluation / conclusions / recommendations.” [my emphasis].

But when the QSA released in June 2010 an updated version of the whole syllabus, it reverted to stating that the Word-Length applied [only] to the “discussion/conclusions/evaluation/recommendations” .

No explanation was ever given as to who authorised that increase (from Word Length including those section, in the Memo, to Word Length applies to those sections, in the updated syllabus) or why [other than that I was telephoned in response to my query (as I

had followed the Memo and instructed a class to write the total length to 1,500 words) and told "that was always the intent"].

In Conclusion

Consequently the teachers have been (and remain) under great stress, both in setting assessment-tasks, in creating grading schemes, and in trying to allot an overall grade. The best that some experienced teachers can say is that: "[w]hile I agree with the sentiments about taking time to work out what each criteria means and what the standards are, it happens with the implementation of each new syllabus document" and in that case, referring to the implementation of the previous, 1995, Syllabus "Complex reasoning took about 3 years for people to reach a consensus about standards and difficulty".

As someone from a business-background, I would ask how can a professional, educational organisation make that same faulty implementation twice in succession, at great cost of time and effort (and consequent loss of teaching-time)? That is, not having improved in the process of implementation in 12 years. One can only infer that the organisation has been severely under-funded in that period?

Part of the statement in each new of the syllabi is "[s]tudents will be challenged to ... communicate effectively" and teachers are encouraged at all times to model appropriate learning-behaviour and required to support student-learning and student-wellbeing. Without adequate funding the QSA cannot be modeling the same clarity of communication, and effectiveness, that it (and the Government's College of Teachers) expects of teachers.

Otherwise, confusion reigns, for example, teachers and the public being told, as in 2010, that DRPCs should have not been telling teachers that marks could not be used when explicit instructions were given to DRPCs to the contrary, and the QSA saying now (and in 2012) that it ordered that the word-length be reduced when in practice QSA (from 2010 to 2012, since the 2010 syllabi-change) told everyone not to enforce it.

One must infer that organizational mis-communication such as that could be bettered with better funding of specialist staff.

Appendix A

To show how the syllabi of 2007 (Physics and Chemistry) were, and are, incapable of being implemented **as written**.

The standards that describe grades of A and B use jargon that is: (a) not of any physics/chemistry meaning; nor (b) of any physics-education (chemistry-education) meaning.

ONE –To the extent that in mid-2010 the QSA in Memo 030/10 announced that a glossary of meanings would be released **June 2010** – for **2007** syllabi.

TWO – teachers through repeated attempts have had to try to create a workable set of phrases – still continuing (as in “systematic analysis” and “relationship between patterns” (see in table below))

Criterion	A –Grade As written in the Syllabus	As we have had to re-word it, or been told to re-word it, in practice or how it has been implemented in examples on the QSA web-site
<i>Knowledge and conceptual understanding</i> KCU	The student work has the following characteristics: <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 	<ul style="list-style-type: none"> • reproduction and interpretation concepts, theories and principles in complex and challenging situations • comparison and explanation concepts, processes and phenomena in complex situations
<i>Investigative processes</i> IP	<ul style="list-style-type: none"> • systematic analysis of primary and secondary data to identify relationships between patterns, trends, errors and anomalies. 	<ul style="list-style-type: none"> • systematic analysis of primary or secondary data to identify relationship between patterns, trends, errors and anomalies.

THREE – a Definition of Complex & Challenging was released on **July 2008 (for a 2007 syllabus)** (and stands on the web-site in a new format dated April 2010)

It shows that a “C” grade is to be awarded for successful completion of questions that were somewhat complex and a little challenging (that is, with only a few concepts)

Whereas the official standard for a “C” is:

*“application of algorithms, principles, theories and schema to find solutions in **simple** situations.”*
(and certainly nothing about *concepts*)

FOUR - Objective labeled Knowledge and Conceptual Understanding (KCU) 3rd point

The A & B Grades are for completion of questions that are complex and challenging, whereas a “C” is for “simple” questions

– so that a “C” student may not even be able to start a question of standard “A” or “B”

- so then the result on that A-Grade or B-Grade question of a student who is about “C” on the “C” questions can be ignored if there is no attempt or something completely wrong –but not if the students manages to do a reasonable amount of the questions of standard “A” or “B”.

For the “B” grade descriptors in KCU 3rd point

B Grade requires “linking concepts” – so explicitly, more than one concept is required to be used. Immediately then that invokes “challenging” according to the definition of Complex & Challenging released in July 2008.

Yet the next phrase in the B-descriptor is “or challenging” – yet by the first part “or challenging” cannot be an option – it is by nature already “challenging” by reason of being required to have two (or more) concepts in use.

But that means the “or challenging” is redundant

So the B-Descriptor is then simply “linking & ... of ... in complex situations”

which means that it is then the same as an A-grade descriptor – for it is already by necessity “challenging” and it is also “complex”.

Further notes on Complex & Challenging – July 2008 Memo v 2010 Glossary

Debate between teachers (including some panelists):

Panelist A

Just rereading part of your document

B Grade requires “linking concepts” – so explicitly, more than one concept is required to be used. Immediately then that invokes “challenging” according to the definition of Complex & Challenging released in July 2008.

Yet the next phrase in the B-descriptor is “or challenging” – yet by the first part “or challenging” cannot be an option – it is by nature already “challenging” by reason of being required to have two (or more) concepts in use.

But that means the “or challenging” is redundant

It appears you have mis-read the glossary definition of “challenging”. The “linking concepts” invokes more closely the “complex” definition:

Complex - Relationships or interactions that have a number of elements, components or steps

Whereas “challenging” relates to the inherent nature of the concept or task:

Challenging - Demanding and thought-provoking; usually involving less familiar or unfamiliar elements, high levels of synthesis and greater abstraction

Therefore, the distinction between “A” and “B” is that “B” may be rehearsed routine multistep OR less familiar or inherently abstract or difficult to apply. “A” should be BOTH multistep/multiconcept AND inherently abstract or difficult to apply. HOWEVER this is a definition that is really NOT appropriate for students. Individual questions need not be BOTH complex and challenging, rather the student must successfully complete challenging questions AND successfully complete complex questions.

However, this provides an entirely different understanding than that which I have used, and implies that there is no need for an “A” *question*, just different types of “B” questions! This interpretation brings on a whole new minefield upon which I am not yet willing to tread! But it is one that further invalidates the reliability of the QSA assessment and moderation process.)

Reply – by Gary

It is an issue that arises because the **2010 Glossary** offers so many options:

e.g.

I tend to focus on **Challenging** as being abstract, or unseen, or using several concepts (i.e necessary if there is a high level of synthesis)

and in **Complex** see that as necessarily then "*Complex - Relationships or interactions that have a number of elements, components or steps*"

The definition is not talking about number of concepts, but rather the "number of elements, components or steps"

(Complex can't mean number of concepts as that is covered in Challenging (by necessity) and in fact I see that QSA has updated the Complex & Challenging Memo and the pdf version stills lists, as it did in July 2008,

that:

"Complexity may relate to the number of steps involved in applying knowledge to the situation or task. Tasks requiring many steps may be viewed as more complex than tasks that require fewer steps. The level of scaffolding may also influence complexity, for example, where scaffolding steps students through the task rather than requiring them to independently complete the steps. "

Next reply:

Sorry to be argumentative Gary, but your definition of challenging differs from the QSA glossary. Increased number of elements would imply increased number of concepts (Otherwise it would just be repetitive application of a single step).

I would also note the contradiction in the memo that says "level of scaffolding may influence complexity" - actually, increased scaffolding influences the challenge, not the complexity. It doesn't change the number of elements or steps required in order to get to a final response!

Next reply by Gary

I am but merely trying to make **the July 2010 Glossary** match the **July 2008 [re-issued in 2010] "Challenging & Complex" Memo** definitions

You have to try to read them to make them consistent, (that is, use parts that are consistent) not to highlight the inconsistencies

-otherwise you have "number of concepts" in both "Challenging" and in "Complex"

From another panellist (B)

Are we stuck on the definition of complex and challenging 5 years on???

1. Complex involves many components or steps which, by its very nature involves many concepts that are concrete, familiar, bog-standard and can be applied with a low level of synthesis. Few steps = C, many steps = B.
2. Challenging would appear to me to be those questions that need the 'aha' realisation. There would be many ways to achieve this (multiple concepts, obscured concepts, simplification of problem, etc.). Once that is found, the rest is elementary. However if there are still many steps to a solution then it would be classes as both complex and challenging I suppose.
3. ... scaffolding can water down both the level of complexity or challenge depending on the nature of the scaffolding given.

It would appear that Complexity is more a subset of Challenge because **all** questions must draw on aspects of familiarity, concepts learned and abstraction and must therefore have some degree of Challenge. According to the memos, the only thing that makes a Challenging task an A standard is the number of steps.

Next reply by Gary

the real challenge is trying to define what "Complex" means and "Challenging" given the differing meanings in the glossary of 2010 from that of the "C&C" Memo of 2008.

From Panellist B

Interestingly the C&C Memo refers to just steps whereas the glossary refers to *elements* and *components* as well as *steps*. However there is no strict definition in the glossary of what an *element* or a *component* refers to in relation to a Physics context??

Does '*component*' or '*element*' refer to the concepts employed and if so it would be already covered in the Challenging criteria?

More questions to pose to the QSA.

Next reply by Gary

That is my point - the shift between 2008 and 2010. Did the 2010 Glossary-writer read the 2008 Memo?

Appendix B.

Memo of the amendment of March 2010 to the Syllabus.

Word-Length 1,000 to 1,500 words

"The word length for the EEI and ERT reports *includes* discussion / evaluation / conclusions / recommendations." [my emphasis].

Previously the Word-Length had been 2,000 to 2,500 words - and that range applied ONLY to the "discussion/conclusions/evaluation/recommendations".

That is, the Report could be any length - and often were very long, more than 6,000 words (including the sample that was on the QSA web-site until the beginning of last month, July 2012) - so as to cover the Three Criteria and show each Criteria to the A-level.

Appendix C.

There are other pieces that add to the general confusion, and the work-load and uncertainty for teachers and students - something that should not be happening if the Syllabi had been READY for Implementation when released in 2007 (even after a version being on trial since 2001).

These pieces are more debatable and less clear, but the fact of the debate still occurring is itself evidence of the confusion created by the flawed implementation of these imprecise, patched-up syllabi.

1. Extended Experimental Reports

The matter at issue in July 2010 was:

the effect of the amendment of March 2010 to the Syllabus.

The memo changes the word-length, and consequently changes anything that depends on that word-length.

In the original syllabus, there had to be assessment of at least one part of each of the three General Objectives (GOs).

Aspects of each of the three criteria should be evident in the investigation.
(for an EEI] - (page 22)

BUT that was in the section in which the range was set as 2,000 to 2,500 words – & that range ONLY applied to the “discussion/conclusions/evaluation/recommendations” .
That is, the Report could be any length – and often were very long, more than 5,000 words – so as to cover the Three GOs and show each GO to the A-level.

But now the limit is 1,000 to 1,500 words –

The question is: now what could be assessed in the range 1,000 to 1,500 words and to what level?

If the original syllabus-writer had an expectation that the standard was A-level in Three General Objectives in 2,000 to 2,500 words of discussion etc (with a lot of other words in other sections not counted), then maintaining the same standard would mean that not so much of the GOs can be assessed in a maximum of 1,500 words.

The samples published on the QSA web-site as being examples in which all three of the GOs can be assessed to A-standard.

That A-Standard sample contained **6,323 words** (NOT including tables and figures)!

[During the years, at workshops etc, it has been emphasised that certain parts of the GOs are best assessed via specific types of Assessment Instruments (eg IP in EEI or EC in ERT)]

2. Summary of EC 3rd point

In the third sub-part of Evaluating & Concluding (EC3) about "innovative use of a range of formats" No-one could address the issue of “innovative use” – see for example the annotations on Samples on the QSA website.

So the Glossary re-defined a meaning of “innovative” to include the option of merely being “appropriate”!

Appendix D.

From: Panellist B

I thought there was one omission that could be included.

The use of the word *typical* when arriving at a LoA (Level of Achievement).

According to advice from the QSA, the word *typical* cannot be based strictly on any one common statistical measure (mean, median, mode) and is largely based on the individual teacher's opinion. This has effectively reduced the role of the Physics Panelist to one of a rubber stamp because there is no quick way to check the validity of the LoA unless the panelist effectively re-marks each item and then (by chance) uses exactly the same subset of assessment results (the subset does not have to be the same for all students according to the syllabus) to arrive at a decision for a LoA that is *typical* (not based on any one defined statistical measure) for that student.

In order to arrive at a final judgement for a LoA we have:

1. An opinion by the class teacher on how well the student's response addresses the Assessment Criteria in each Assessment Task.
2. An opinion by the class teacher on which responses over the course of study make up the final LoA for each of the three Exit Criteria.
3. An opinion by the class teacher on the Exit LoA based on the opinions for the three Exit Criteria found in Step 2 above.

The Exit LoA is basically an opinion of an opinion of an opinion. Needless to say the waters are very murky by the time Verification comes around and the task of splitting the students amongst the 30 rungs from VHA10 to SA1 is a farce and akin to trying to thread a camel through the eye of a needle!

As a result, Panelists are only able to clearly discern between bands and threshold placements between bands are difficult to justify one way or the other. This opens up the whole assessment process to the possibility corruption as the classroom teacher's opinions (and any associated bias) cannot be definitively isolated from the assessment and verification process.

It also defies logic how the QSA then gathers 5 or 6 similarly derived LoAs and converts them back to a definitive number called an Overall Position!

Reply by Gary

There is an additional piece we could add -
to that piece about rungs -

what struck me was then recently hearing of teachers trying to separate four students within a single rung [e.g. VHA2] for the purpose of allotting an SAI. The only way is to minutely dissect the responses to individual questions in exams over the course of a year.

APPENDIX E – Formulae Sheets – April 2013

Despite the 2007 syllabus differing from the 1995 syllabus, the reasoning for supplying a Formulae-sheet (which was stated on 02 March 2013 that we should) seems to be that the 1995 (!) Syllabus State Panel said teachers should.

Yet the official advice (2013) is that:

“The Physics syllabus refers to

*Recall and interpret concepts, theories and principles of Physics – this includes the abilities to remember, reproduce and interpret subject matter such as facts, definitions, **formulas**.....*

.....

It is appropriate to give a generic formulae sheet in Supervised assessments.”

If the Criterion is to “recall” a formula, how can it be “appropriate” to supply a formulae-sheet?

Physics teachers might identify this advice as a non-sequitur, but in any case is an example of the contradictions between the plain wording (i.e. “recall”) of Criteria in the syllabus, and the official interpretation (“give a formulae sheet”) – and if the latter is to over-ride the plain-meaning, then such a special meaning should be included in the syllabus-document so that all teachers are reading the same document..