SMC&PA Submission 222
Received: 13 May 2013

The Education and Innovation Committee
Parliament House
Brisbane 4000

6th May, 2013

To the committee members,

The inquiry into Senior Mathematics, Chemistry and Physics in schools

I wish to record my strenuous condemnation of the overall assessment methods used with regard to their questionable **validity and reliability (the 3rd Term of Reference)**. My interest stems from career qualifications /experience in marketing /stock appraisals and / merchandising involving statistical analysis. I have also hold and have held positions in science / mathematics based work (including dental prosthetic technician; pharmacy management). Good record-keeping is not only scientific, it is a necessity for accountability in contemporary society.

It appears that the Queensland Studies Authority (QSA) methods, available on the parliamentary website (both the QSA slides presented and transcripts of the QSA briefings to parliament) greatly lack in conventional record-keeping detail.

http://www.parliament.qld.gov.au/work-of-committees/committees/EIC/inquiries/current-inquiries/QldAssessment
click on "Briefing Paper – Queensland Studies Authority – 20th March"
to see pdf document of incomprehensible marking scheme on pages 14, 15 (you will have to count because QSA did not page-number)
http://www.parliament.qld.gov.au/documents/committees/EIC/2013/QldAssessment/bp-20Mar2013.pdf

Even the little data scanned seems to have large gaps. See Tabled Paper (under Public Briefing held on 7th March)
http://www.parliament.qld.gov.au/documents/committees/EIC/2013/QldAssessment/tp-7Mar2013.pdf

- Firstly, there is a lack of a clear system that can be followed in plain language, considered a must for any agency, regardless of complexity. A clear, plain-language system is evident for other school bodies interstate.
- Secondly, therefore, there is no **reliability** in the detail of how students' results are judged, compared, analysed for biases in the system, and ultimately ranked. Without that there is no way of checking on **valid** knowledge.

I will suggest here how the second problem is critical to your determinations on the third term of reference: The ability of assessment processes to support valid and reliable judgments of student outcomes. The overarching conclusions will infer the need to restore a valid and reliable system using numerical marking statistical (i.e., analytical) scoring urgently.

1. **Validity** refers to testing what is supposed to be tested. Two things are clear: a QSA representative stated that
- content knowledge, ie, a "lot of information" (eg rules, equations) is **not** important (Transcript 20th March) and
- that the 'inquiry'-approach is the basis for QSA assessment (Transcript 7th March).

The average person on the street believes young people should be taught and tested on foundational knowledge, yet the QSA is testing them on how well they inquire into the design and research of their own experiments and write them up in detail. My ability to research and synthesise a 1000-word paper was never necessary for my post-school study or jobs involving maths or science. Perhaps, humanities-based subjects at university need sophisticated writing preparation but pharmacy knowledge. for eg, in the form of chemistry has its own language. When hiring employees (regardless of university entry), I would look for people with common sense and fundamental knowledge in maths and science. Writing long essays is irrelevant; and during school this would distract them from learning the basics needed. The issue of cheating also invalidates these projects. The QSA also encourages such computer/ research work to be done in class under supervision. This is still not validly testing knowledge that should have been taught in the first instance.

2. **Reliability** of the marking system overall: In other states, students are told that their assignment is worth 5%, tests worth 45% and the final external exam is worth 50% (similar to NSW example). Currently in Queensland, there is no system to speak of, from an accountable, analytical point of view. This is something that needs to be implemented as soon as possible. A simple announcement that teachers should set a maximum score and keep a numerical score of the items mastered in a test or project is essential to accountability in any layperson's book. Then every test or project that contributes a component to the final subject-grade can be added in a fair and transparent way. Every parent would expect this for their child. The number of proportional marks or percentage allocated to each test paper should be given at the beginning of every term. There can be no central, accountable way of keeping an eye on the data without this. The numerical tallying up and comparing of scores from teacher to teacher, school to school, is not too much to ask.

3. **Reliability** of the letter-scoring 'system': Most significantly, it can be seen in the QSA exemplars, this is evidentially an extremely flawed and deleterious approach. Getting letter-grades, such as Cs an Ds will skew the total result. It will pull the student's result over to the failure-side of the A-E score – even though they got the answers (or part-working) fully correct. Here is an analogy: Imagine, if you will, I rated new employees on basic abilities and **gave them a D** for every time they served a customer ***correctly*** in a simple 'routine' or 'rehearsed' way (that is a typical D 'standard' defined by the QSA)? They would fail their job appraisal. How utterly demoralising. It is far more logical, that I give a score, even if it is only 1 out of 1 mark each time for getting the basics right, then they can see where they got it right.

Their correct performances (scores) can be added up. If there are more important components I can weight them with higher marks. I could give them a letter grade **at the end** (eg, Yes, you're hired – you got 39/40 marks, that's an "A" standard) .However, the point is, in multiple assessments, as in stock-taking or any other large-scale data collections, it is illogical to define each component with an alphabet letter to get an overall result; eye-balling the letters is not reliable. When the QSA forces teachers to mark with letters, and then they have to take a guess at the final overall letter-grade, and then they have to rank between students, the QSA is ignoring statistical mathematics and how it works to make judgments clear, fair and reliable. This has even greater importance when this is apparently a quasi-governing body for all schools around the state (as it is for public and private), especially when needing to compare many sets of data (many students around the state). Whether large-scale market research or social science, data is scored to see the trends.

My humble suggestions are that:

(a) numerical marking be advised to be **mandatory** within days of this inquiry's conclusion - for all teachers - so that common record-keeping can begin and be compared in a reliable manner; and,
(b) Extended essays, research tasks, powerpoints, etc be limited to learning enhancement only. Such largely invalid projects **not** to be used for the high-stakes assessment of students' final grades.
(c) Use one common exam as an 'anchor' to scale other results. Using an exam already in existence that could be agreed upon as an interim measure before a more robust system is started would help.

As an absolute first port of call, numerical marking would be the intuitive and simplest to reinstate. From there, some current tests or assignments, you would think, could still continue if refined. It is through the statistical records of the numerical scores, that pictures of trends emerge. This analysis would surely help with further discussions about the validity of one test or project over another. Test results between schools could also be adjusted up or down depending on the difficulty or standards of the tests. No qualifications are needed to understand that such objective methods of adjusting scores up and down is more transparent calibration than the subjective social moderation going on now.

In particular, recalibrating the results according to one agreed-upon common 'anchor' exam would evidently improve the scaling. And that would be ideal if an agreed-upon external common statewide exam was already in existence (such as used for some schools and distance students already). It could be made available to all students to anchor the results from all schools (provide a yardstick). Cheating in assignments or tests with overinflated results would then show up when comparing the scores from class to class and school to school using statistics. Again, these initial investigations are impossible unless numerical marks are immediately restored in Queensland, to bring it online with the other states.

Until the numerical scoring of tests is implemented in Queensland, there will continue the untenable situation where students are marked on the current alphabet-letter approach, which is both subjective and obviously skewed.
No inquiry is needed to restore numerical marking – it is simply a branch of statistics used everywhere in society.

I believe that at the conclusion of this inquiry, decision-makers in government would care enough to ensure that teachers recommence marking how they used to do until recently. With all respect to teachers who have been forced by the QSA in recent years to use their professional judgment to make 'holistic' judgments about cut-and-dried maths and science, there still remain serious questions about accountability to these students and taxpayers with regard to grades on their CVs and ranking needed for jobs, apprenticeships, TAFE or university. But the past cannot be undone.

However, for the current students in Grades 11 and 12 now, an immediate directive could be issued. It is only by comparing the scores from each school, can decisions begin to be made on an equitable basis. I believe the Australian Constitution states that citizens travelling between states should not be discriminated against by being treated differently in a particular state. Currently Senior Queensland students – unlike students in other states - are not informed at the outset of the composition of their tests and projects towards a final subject-grade which leads to their employment. Surprisingly, this is not a requirement of schools in Queensland (see QSA quotes in syllabuses, sections 6 & 7, about teachers free choice to do 'selective' updating of results). I have not heard of this happening anywhere else. This would obviously be an unreasonable stress on minors who are working so hard for their high-stakes final grades. Their efforts are going to be distorted by such an invalid and unreliable system. They would either be overworked or lose faith in such a system. As a potential employer to school graduates, I believe this is critical to valid and reliable school assessment.

Please see attached appendices.

Yours sincerely,

Noel Dickson and Jennifer Leung

[redacted]

Please withhold my contact details. You may publish names.

APPENDIX A NON-NUMERICAL MARKING – A QUIRK OF THE QSA IN QUEENSLAND

What the QSA has been instructing teachers to do – avoid using numerical records - is the equivalent of striking out a whole chapter on statistics and data-keeping. Ironically, it is ignoring the purpose of teaching mathematics for real-life book-keeping. The QSA is admitting that it is saying: "Please estimate student results by your own professional 'on-balance' intuition, and that is enough for Queensland citizens. The OP-score will tweak your estimations via the QCS test, but we trust you to make the same quality overall decision as every other teacher from all corners of the state using only letters. This is perfectly fine for the all-important pre-university ranking and the future career paths of Queenslanders." Strange? Avoiding numerical marking is denying our citizens good record-keeping and accountability.

TABLE OF CONTENT - Extract from a MATHEMATICS textbook Humprhies M et al (2012) Mathematics for the International Student (IB Diploma): Mathematics Studies SLHaese Mathematics, hasesmathematics.com.au.. Strike-out added to exemplify QSA approach.

QSA forcing Queensland teachers to use letters instead of numerical analysis is akin to pretending that statistics don't exist

APPENDIX B – SCORED TEST PAPERS (NUMERICAL) ENABLE RELIABLE ANALYSES

ON making scaling adjustments using statistics in the marking of scored school papers...

By the director of the largest educational research body in Australia – Australian Council for Educational Research (ACER). Geoff Masters, in a recent 2013 review.

> **Masters, G N., "Reforming Educational Assessment: Imperatives, principles and challenges" (2013). ACER**
> http://research.acer.edu.au/aer/12 Pp 40, 41.

"Objectivity is particularly important when comparisons are to be made across different sets of assessment tasks. To monitor the progress students make over time, it is necessary to compare assessments made on different occasions. In classrooms, these might be assessments made at the beginning of the school year and again towards the end of the year, usually using different assessment tasks"

"The goal of this (statistical scaling) process is to estimate where students are on the same described proficiency scale, even if they attempted different sets of assessment tasks. This statistical approach also is used by national and international assessment programs such as NAPLAN, TIMSS, PIRLS (the Progress in International Reading Literacy Study) and the ICCS (the International Civic and Citizenship Education Study; Schulz, Ainley, Fraillon, Kerr & Losito, 2010) and by commercial assessment resources such as the Progressive Achievement Tests (PAT) (ACER, 2008) and the TORCH Tests of Reading Comprehension (ACER, 2003). In all these assessment processes, students' performances on, or responses to, assessment tasks are analysed statistically and used to infer students' locations on a described proficiency scale."

"Professional test development processes include checks for the possibility of test question 'bias'. Sensitivity reviews are undertaken to inspect questions for features that may place particular groups of students at a disadvantage. Doubtful questions are removed before the test is finalised. Further checks can be conducted after a test has been administered. If a statistical analysis of students' test performances identifies some questions as being unexpectedly difficult (or easy) for particular subgroups of students, and if a substantive reconsideration of those questions raises doubts about their fairness, then responses to those questions can be removed prior to reporting assessment results."