



EDUCATION AND INNOVATION COMMITTEE

Members present:

Mrs RN Menkens MP (Chair)
Mr MA Boothman MP
Mrs YM D'Ath MP
Mr MR Latter MP
Mr NA Symes MP

Staff present:

Ms B Watson (Research Director)
Ms R Stacey (Principal Research Officer)

PUBLIC BRIEFING—AUSTRALIAN COUNCIL FOR EDUCATIONAL RESEARCH

TRANSCRIPT OF PROCEEDINGS

WEDNESDAY, 7 MAY 2014

Brisbane

WEDNESDAY, 7 MAY 2014

Committee met at 10.34 am

MATTERS, Dr Gabrielle, Australian Council for Educational Research

CHAIR: Welcome to all. Before we begin, I remind everyone present in the room to turn off their mobile phones or set them to 'silent'. I also remind members of the media who might be recording these proceedings to adhere to the committee's endorsed media guidelines. Staff can provide a copy of the guidelines should you require one.

As many of you are aware, last year the Education and Innovation Committee undertook an inquiry into the assessment methods used in senior mathematics, chemistry and physics in Queensland schools. The committee tabled a report on its findings in October 2013, and the government's response was tabled in January. We noted in our report that some of our recommendations would impinge on the considerations of the independent Review of Senior Assessment and Tertiary Entrance that has been commissioned by the government. As we pointed out, a parliamentary inquiry is not undertaken by content matter experts but by representatives of the public and considers evidence presented to it by stakeholders, from published research and submitted directly by experts. The review being conducted by ACER is being undertaken by those with technical expertise. Naturally, that review comes from a different perspective.

The government agreed with our view about the respective roles of a parliamentary committee inquiry as an independent review undertaken by content matter experts and, as part of its response, referred several of the committee's recommendations to the independent Review of Senior Assessment and Tertiary Entrance. Those recommendations had implications beyond the senior mathematics, chemistry and physics focus of our inquiry and also fell squarely within the terms of reference for that review. The review being undertaken by ACER will present proposals to the Queensland Minister for Education, Training and Employment by the end of July 2014. The government will consider those proposals after that. The government has also indicated that if it decides to make changes to the senior assessment and tertiary entrance systems, it will carefully consider the timing so that students will be able to plan their senior schooling and tertiary entrance pathways.

The Education and Innovation Committee has invited Dr Gabrielle Matters, one of the lead reviewers of the Australian Council for Educational Research, to give us an update on the review today. I welcome you, Dr Matters, this morning to this hearing. I will just briefly introduce the members of the Education and Innovation Committee. I am Rosemary Menkens, member for Burdekin and chair of this committee. With me are the committee members: Mr Michael Latter, member for Waterford; Ms Yvette D'Ath, member for Redcliffe; Mr Neil Symes, member for Lytton; and Mr Mark Boothman, member for Albert.

This session is a formal process of the parliament, and parliamentary privilege applies to all evidence presented. Any person intentionally misleading the committee is committing a serious offence. While these proceedings are public, you may request through me, as chair, that any material or information you provide be kept private, and you can object to particular questions that a committee member might ask you. You might also wish to take questions on notice if you do not have information at hand. The rules we are operating under in respect of witnesses are contained in schedule 3 of the parliament's standing orders, 'Instructions to Committees Regarding Witnesses'. Today's proceedings are being broadcast via the internet and we are also being recorded by Hansard. The transcript will be published on the committee's web page once it is available.

Before you start, Dr Matters, would you prefer committee members to ask questions throughout your briefing, or would you prefer to leave all of the questions to the end?

Dr Matters: I think there is a stage, if I could signal that to you, after probably about 10 minutes that might be an appropriate time to ask questions.

CHAIR: Thank you so much, Dr Matters. For the benefit of Hansard, I ask everyone to please state their names the first time they speak. Dr Matters, over to you, thank you.

Dr Matters: Thank you, and thank you to the committee for inviting me to speak with you. We are nine months into a 12-month review. What I am talking about now is ACER's position three-quarters of the way through the review. Professor Geoff Masters and I understand completely that the government will seek its own responses to the proposals we put, so what we are giving to you is our position at the moment. This position has not yet been adapted, if it needs to be adapted, in the light of submissions that have come in—we are analysing those submissions now—nor in the light of a very important piece of research done by Professor Claire Wyatt-Smith on the operation of panels, but our position is informed by our own reading, observations, knowledge of other places and analysis of data, and we are most grateful to QTAC and QSA for making all data available to us. So that is where we are at the present.

We, of course, do adapt our position in response to consultations. We have had a wide consultation process that culminated last Tuesday night on a meeting of 120 key stakeholders at which I hope I gave a clear description and explanation of our position. Our consultation process has been of two types: we have had a survey; we have met and spoken with the 27 key stakeholder organisations or representatives of them. We have invited ourselves to key meetings, and we have accepted invitations beyond the consultation process to have discussions with any group who wish to have a conversation with us.

After nine months, our position is that we believe that there should be—I'm sorry to use the word 'should'. I am not quite sure how to discuss what we think might happen. Our position is that a remodelling of the secondary tertiary interface is the direction that Queensland should go in. By the 'secondary tertiary interface' I mean that zone where senior schooling butts up against university entry, so it is that zone where selection decisions are made. The review's terms of reference were wide ranging, as you would well know, but we take one of the particular terms of reference as defining our task, and that was to consider the effectiveness of the systems of senior assessment and tertiary entrance and identify ways to improve, revitalise or reform them.

I will just start at the end of the story as it exists right now. In July 2014, as Mrs Menkens said, we will deliver our report to the minister. In July 2013, at the beginning of the review, we were constantly asked two questions that seemed naïve to us at the time but, as it turns out, are the questions that are in the news right now across the country. I will just digress for a moment. Today's *Australian* has a story about the ATAR in it. The early questions we were asked were: are we going to have external exams?; is Queensland going to the ATAR?

Right now our response to the question, 'Is Queensland going to have external exams?', meaning are we going to recommend external exams, is that we would recommend that school based assessment be preserved on the proviso that it be revitalised, that we should introduce a new species of external assessment, and that that external assessment is not the three-hour written external exam that characterises the HSC. To the second question, 'Is Queensland going to the ATAR?', our statement would be that the OP is not sustainable and that the ATAR is not the long-term solution here or anywhere else in the country.

Now going back to the beginning of the story, the terms of reference for senior assessment related to these four facets of the system: comparability, moderation, assessment instruments and exit levels of achievement.

For those we looked at evidence of comparability in school assessments. We looked at the nature of the moderation process, realising that comparability and moderation are, in fact, related via standards. We looked at assessment instruments and assessment programs and we looked at exit levels and achievement in terms of how they are arrived at, which in turn requires that we look at the way exit standards are described and we looked at the current debate about the use of numerical marks.

The terms of reference for the tertiary entrance system relate in summary to the OP, to the QTAC rank, to the ATAR and the QCS test and to common curriculum elements. For that we looked very closely at the OP system—and I will return to that—to the QTAC rank, and I will take a moment to explain this to you. I just need to explain also that when we look at tertiary entrance, we are looking at the procedures by which year 12 completers are selected for university entrance. We are not looking at the whole landscape of tertiary entrance in Queensland; we are looking at what happens to year 12 completers. Going back to the QTAC rank, students at the end of year 12 are OP eligible or OP ineligible. Both sets of students might apply for university selection; most of them do. The OP eligible students are those who receive an OP obviously and are considered on the basis of that instrument which was designed precisely for that: to make selection decisions. There are students, though, who for various reasons are not eligible for an OP or who have been rendered ineligible for an OP who also want to be considered for university entrance. For those students,

QTAC produces a ranking. That ranking is based on a simple addition of results in various types of subjects from authority subjects to authority registered subjects, VET and so on. That particular process was an exception to the rule, if you like; that was for students who were not OP eligible but sought university entrance. I will show you a graph later that demonstrates that there is almost the same percentage of both now; the percentage of OP eligible students has fallen and is heading for 50 per cent and the percentage of OP ineligible students is rising and is about to hit 50 per cent. The point I am making here—there are two actually. The first one is that the exception is greater than the rule or is equal at least to the rule. Those who are seeking entry by the rule, if you like the classic way, are almost the same number as the OP ineligible. The universities would say—and I agree with them—that this is an unfair binary system because the processes for arriving at the OP are through a scaling mechanism through the QTAC rank by simple addition of scores.

The ATAR itself—I will just take a minute to explain again the ATAR. The ATAR, as you know, is a national ranking of students who are eligible for university entrance. It is reported in four digits such as 99.05 or 99.50. I labour that point to say it has got four digits; it is a very, very fine scale and all states can map their own ranking onto that scale. I would like your indulgence also to make a comment about this myth that Queensland is different or is so different. Every state in the country produces an ATAR, including Queensland. The difference between Queensland and the others is that the ATAR for Queensland students is not known to them, although it can be accessed at their request by QSA and used by the universities to make difficult decisions within OP bands.

The other interesting thing about the ATAR that is not well known is that the Queensland way of getting to the ATAR at the moment is via the OP. The mechanism in other states for getting to the ATAR is not the same mechanism. Queensland is different from the other states but they are, in fact, different from each other. The process in New South Wales, for example, has different eligibility requirements from the process here or in Western Australia. The black box that is used to scale school assessments or HSC results or whatever is different from state to state. I will use some technical terms not to bore you silly but to actually give you some labels about how they are different. New South Wales, for example, uses something called intersubject scaling within the black box. Queensland uses scaling to an omnibus test in their black box. Tasmania uses item response modelling within their black box. I labour the point that every state is different; it is not Queensland being different from the rest. The point that is interesting is that Queensland is different from the rest in that the ATAR is not known to students. There is a reason historically for this in that the OP predated the ATAR and it was decided in Queensland that the four digits used for the ATAR scale were not acceptable in statistical terms as in the error of measurement on that fine scale meant that the students were, in fact, often not in the rank order or they were actually separated out in a way that could not be sustained by the mathematics of it.

In relation to the QCS test—I am not sure I should use these adjectives in this context; please correct me if I should not be—we have found that one of the big surprises has been the almost rejection of the QCS test in the education community not on the basis of its content, not on the basis of its role in scaling—although scaling of course is not well understood and this fact that people do not understand it is almost treated as a virtue—but more because the preparation for this QCS test has taken over, a lot of money is being spent by schools in coaching programs and a lot of time is directed into this process from teachers.

As well as looking at those facets of senior assessment and tertiary entrance we were also, as indicated by Mrs Menkens, asked to look at certain recommendations from the maths and science inquiry that were referred to us. Before we started our deliberations, before we got tied up in the nitty-gritty, as it were, we set for ourselves some principles—the principles that would underpin all of our investigations and our decisions. Some of them seemed to be the obvious ones. I will list the principles and then expand on how we used them. I will just give you an example of each, although there are many things we looked at and then examined according to these principles. I am working back to front in describing these in listing them and then giving an example of an issue that we had to look at and then consider according to these principles.

The first two are the hallmarks of any good assessment or assessment system: validity and reliability. So in the present system of senior assessment we looked at the validity of the assessment instruments individually and the validity of the assessment program. Validity is simply a measure of the fidelity of the assessments—how faithful they are to the official curriculum, if you like. In doing that, we came to the conclusion that the validity of assessment programs is not as high as it could be because the range and balance of assessments perhaps does not deliver coverage of the desired curriculum.

Reliability we take to mean consistency of judgements. In this system it would be comparability or consistency of teacher judgements. We came to the conclusion that there is not any empirical evidence—no studies done—in our opinion of comparability in the system. So we really cannot report on the level of comparability because there is no evidence. Panels can tell us there is high comparability. A report by Geoff Masters in 1994 that says there is more comparability here than in external exams I think is a bit dated. There is a process called random sampling, a post-hoc process as part of the moderation process and we are not persuaded that that provides the evidence that we are looking for. This is not to say there is not any comparability; it is simply to say we cannot prove the levels of comparability. We are prepared to say that the comparability for the five levels of achievement in the present system is probably adequate for what these levels of achievement are required for. But any system that I am coming to which requires finer divisions on a scale requires high reliability.

In relation to the futures orientation, we set the principle for ourselves that we should not be introducing into a new system or recommending introduction into a new system any processes that are part of the halcyon days of yesteryear or—I am thinking about the early days of the QCS test perhaps—that are deemed to be retro not just by us but also by educators and researchers around the world. We also took the futures orientation to mean that anything that we suggest should be desirable not just for tomorrow but for 20 years into the future at this stage. In that sense, we looked at issues like the whole digital aspect of education in both assessment—students doing assessments, assessments being marked—and we looked at the fraught issue of 21st century skills.

I jump forward a step to say that we had the notion of some cross-curriculum testing occurring of key cross-curriculum competencies because ACARA itself is looking for a definition of these competencies for assessment. This is one place where our opinion or our direction was completely changed by reactions from the key stakeholders, which were very cold about the notion of such testing. We understand that having three types of assessment in a system—school assessment, external assessments and such testing—would be a load on students and on the public purse; we understand that. We are fascinated though by the fact that Queensland was looking at this notion of generic skills at the end of the last century and giving papers at international conferences on the case of testing generic skills. Without sounding provincial, I have to say that the rest of the country and, in fact, the world is now talking about testing high order thinking skills and we have decided that this is perhaps the time not to have such tests. Our compromised position and one that has been suggested from the field—and we are grateful to receive it—is that maybe we can have testing of high order thinking skills but within subjects and prescribed, say, in school assessments or external assessments. The difference there is that those assessments would be within the disciplines and not across them, and that is a very different type of assessment. I give you that example for two reasons: one, to actually let you know what we were thinking and, two, as an example of where we were prepared to change direction in terms of the arguments that were put from the professionals.

I have now mentioned three of our principles and gone off on tangents. The fourth one was the centrality of student learning. It is often a throwaway line, but we take that one seriously. We rejected any notion that, in our opinion, takes away from student learning or teaching time or teaching energy. We, therefore, reject the QCS test on the basis of this staggering amount of time and energy that is taken away from teaching and learning for QCS preparation.

We reject subject achievement indicators as the input into the OP, which, for other reasons, we think is no longer sustainable, because of the time and energy that is spent by teachers and schools at the end of the year ranking students within subjects in what should be a very simple process but is turned into—usually I say a Russian novel, but that is probably not appropriate either—a process that requires a document five centimetres thick about how to do it and which turns into many conversations from schools that can only be described as gaming. I use the word 'gaming' with some sort of regret, but the evidence is not from us looking at data but having been told directly by schools about themselves or of course about other schools. Obviously any system that replaces the present one is going to eventually change equilibrium and it will be gamed. The New South Wales system is gamed at the moment—you have probably read about it in the press—whereby students are choosing the easier subjects. That is a strange thing to do in terms of one's future, but they are doing that because the New South Wales scaling process results in the top student in every subject, whatever the subject, having the same top score to go into the scaling process. So if you could choose between higher level maths and basic mathematics and you knew that the top score was going to be the same in both of those, you might be tempted to choose the

lower level maths. It is pretty stupid to do that of course if you wanted to study engineering, but nevertheless that is an example of gaming the system and an example of it happening in other places.

With regard to transparency, I have given an example already about what I consider to be a lack of transparency at present and it relates to a selection procedure. I will just give one example. I could be here today and tomorrow giving you examples that we have looked at and examined according to these principles, but the ones I am using really are not necessarily the most important ones but they are examples that really come with a good story, if you like, and they are easy to understand. When the students do the QCS test in September of their senior year, they go into that test hopefully well informed by their school and believing that the individual results do not affect their own OP. Again, parked at the side of this badge of honour that is worn by people who should know better is that that is not the case. I will return a lot of times to this theme that we are staggered by the lack of understanding of the system, by people telling you they understand the system and by people swearing black and blue that a certain statement is in fact true when factually it is not true. That is the hallmark of the discussion at the moment. That is something new, I believe, in this state where it is not considered to be someone's professional right—professional duty—to understand the system in which they operate, and that is at all levels in the system.

Going back to this student who sits for the QCS test in September fully believing the fact that individual result will not count for the OP, in the selection process that occurs in December a university—any university, and this has happened—can find it difficult to make in-band decisions. They have so many places to fill and they have filled X of them by the time they get to the end of OP3, for example. They have some places left over. They do not want to overfill to go to OP4. They say, 'What information can we use? We've got the OP. We've got the FPs,' and there is another story about why they are no longer useful. They say, 'We've got the QCS grade, but it's only five levels. We've got the levels of achievement. They're only five levels too. How are we going to do this? We can get the ATAR, we can get the student to ask for her ATAR or we can actually get the student to ask for her raw QCS score report on percentile band.' So I can now find out from QSA my QCS raw score on a percentile band and give it to universities. I am not talking about QCS grade; I am talking about the score only. So that is an example of a lack of transparency. That is not known to the student. I will not labour that one; it is just such a surprise that they do continue to use it as a transparency issue.

With regard to operability, obviously feasibility and practicability are principles that underpin the design of any system. We realise that in framing our recommendations we need to be conscious of how feasible they are, and I will give you an example. We are persuaded that some of the universities for their high-status courses should consider the reintroduction of prerequisite subjects—not all courses, not all places—and the simple example we give is for engineering. I have actually had some really interesting comments on the Twittersphere that have been directed to me about the way I use the term 'maths 2'. We actually talk about maths A, maths B and maths C in our system and our highest level maths is maths C. I presume that was about the psychology of students wanting to make sure the maths A students did not feel bad about doing lower level maths. The students of course see through this completely because students are quite happy to talk about easy, medium and hard subjects. Nevertheless, I find 'maths C' a very hard term to use. Maths 2 of course is a thing of the past, so I will choose my words and say highest level mathematics.

In this state and in this country it is quite scandalous about the number of students—the proportion of students—who actually study high-level maths. That is a value judgement, so I will leave that one to the side. The engineering faculty will often get students from our schools and say, 'They come here into engineering and they can't do mathematics.' Of course they cannot because they have not been asked to do higher level mathematics. So for engineering we would think that a possibility would be for students to have three subjects prescribed, including physics and higher level mathematics. In actually signalling to you the three subjects, we are also of the opinion that we do not need to have five subjects as inputs into tertiary selection. You might be fascinated to know something completely irrelevant, but our timetables in schools here and in the country have six subjects. Students can do five, but there are six subjects. Where did that tradition come from and what was the rationale for it? Last century—probably the middle of it—somebody asked how many subjects in the senior curriculum and they looked for the answer in Scotland and the answer was six, and people doing school timetables have been tortured ever since in actually having six subjects online.

In terms of the OP, why five subjects? There is a theoretical problem—and I use the word 'theoretical' on purpose—about the number of inputs you need to get a rank order list that does not suffer some sort of perturbation just by tiny little butterfly effects in one place, and five subjects does

that. Four subjects would probably be enough to calculate a rank order list should it be necessary. We realise there are backwash effects of those sorts of decisions on the curriculum. That is a big discussion for elsewhere, but I use that example of engineering as prerequisites as probably a good place to have prerequisites. Another issue about operability though is one that comes from a decision such as that one. If universities were to consider prerequisites such as that one for engineering, then the question of fairness comes into play. What if high-level maths were a prerequisite for engineering? Do we have enough teachers of high-level maths to cover this diverse state? So would students outside the metropolitan area be disadvantaged by that? So there are both sides of the coin and all of our principles are not surprises.

We do introduce another one that often is not found in such a list, and that is privileging subject specific achievement. The notion of ranking in this country based on overall achievement is peculiar in the sense that nobody else in the world does it or has considered it. If you think about it, we take subject results—results in any five subjects out of a total of about 50 authority subjects that are offered—we scale results from the subjects, we add them together, we scale results from the school and we add them together, and then we chunk them into OP bands and that OP is a measure of a student's position on an overall achievement scale where the subjects have lost their identity, if you like. As I say, no other country in the world does that. Why does it happen in Australia? Revisiting rationales has always been interesting here. It happens in Australia because, again, last century it was necessary to allocate Commonwealth scholarships and that was the basis for allocating Commonwealth scholarships. We think it is time to challenge the rationale of overall achievement that masks subject performance as the main indicator of university selection, and that is why later on I will describe to you how we see that subject achievement being measured and reported.

All of those led us to something that we call separation of responsibilities where we think that it is time to challenge the notion of the schooling sector having to calculate the ranking of students for use by the universities. Again, that is a thing that is peculiar to Australia—that is, that the schooling sector does the job for the universities. It is probably easily done here because all of the states and all of the universities want the rank order based on overall achievement, but nevertheless that is calculated within the secondary sector. So the review's current position on senior assessment is revitalise school assessment—all of these verbs we have considered very carefully—revamp moderation and introduce external assessment. There are some more, but I will just stop there. I said revitalise school assessment. We are deliberately not using the term 'school based assessment'. We just think school assessment says what we mean—assessments where the teachers in the school devise them, mark them, grade them and where the accountability for those assessments rests with the school. With regard to revamp moderation, we believe that the current processes of quality assurance for standards have probably reached a stage where we need to completely rethink the process—going right back to the literature as to what are the options and what are the ways in which moderation can be done. Our present model of consensus using panels is one way. There are arguments for why that probably is not the most efficient way of delivering the best comparability data in the system right now. That is nobody's fault; it is what actually happens when a system goes on for a long time and has adaptations over a long period of time. But we could put forward a criticism of the way the moderation model has evolved. These things do not necessarily mean more work for teachers. These things do not necessarily mean more expense; they just mean different ways of doing things. Again, I use the term 'revisiting the rationale'. Why moderation this way?

In terms of introducing external assessment, again I could speak for ages on how we see that external assessment. We do see it though as a different type of assessment from subject to subject, depending on the nature of the subject, and we do see it as having a different weighting from subject to subject. I would predict—though it is not my business to recommend—that the physics community would probably want to have 50 per cent external assessment and would include in that assessment some, I hope, sophisticated multiple choice testing. That does not mean to say that there cannot be really good constructive responses or other mechanisms; I am just using that as an example. The English community might say, 'Okay. We're happy to have some external assessment.' We think that should be a common essay on the same day under the same conditions and—let us have a fantasy here—that it might be on a Shakespearean play for example. I am just giving you the nature of two very different types. So I think there will be some subjects where 50 per cent will be appropriate, some subjects where close to zero would be appropriate and some subjects like biology and geography perhaps where 30 per cent might be sensible. We are not in the position of recommending how much but are recommending the process by which those decisions should be made and that those decisions, we believe, should be made in a boiler room—

not over a long period of time but in an intense environment where we have university disciplines, teachers who are experts in the discipline, assessment experts and curriculum experts coming to that particular conclusion. As I say, that is enough said there. It is a big, big issue and then we move on to combining results in school assessment and external assessment.

I keep using the word 'theoretical'. I just use that so it does not sound like the simple process of adding A plus B, the percentage in external assessment and the percentage in internal assessment. We reject those notions of percentages completely, but that those subject results should be combined in a way that does not weight one thing more than the other if it is not supposed to be weighted more than the other which can often happen by adding all scores.

I probably should have said before that that we are of the opinion, and we think this is possible, that subject results can be reported on a much finer scale than at present. For the precursors it would be necessary to get those subject results valid and reliable and that we could go as high as a 15-point scale. That would require a different quality assurance process leading up to the grades, a 15-point scale, but in conjunction with the enhanced school assessment, the new external assessment, an appropriate way for combining results, reporting those results on a 15-point scale—we believe, by the way, that that scale should be derived empirically, not just on the basis of verbal descriptors of standards—that the outcome would be valid, reliable, credible in the education community and the broader community, stand-alone subject results, that those subject results would be good enough, excellent enough, credible enough for anybody to use them, including, of course, universities, that the output from schools would be subject results on a 15-point scale. The universities could use those—could choose to use those—in their selection processes, might choose those as the basis of ranking students, calculating their own ATAR if that is what they want, but that that calculation could be done without a QCS test for scaling and without the use of subject achievement indicators from schools. That really indicates the way in which the OP could be no longer used.

I will move on to the OP quickly now in terms of the review's current position on tertiary entrance. The OP certainly has served Queensland well, but it is no longer sustainable. The reasons relate to this big change in the OP eligible population and the unfair binary system, also to the fact that with the proliferation of new subjects in schools there are more and more small groups and the scaling model does not apply to them so again the exception is getting almost as big as the rule. There are other reasons that are not of a theoretical or technical nature such as the marketing of schools according to their OP1s. There are a lot of other things that are more societal than theoretical and these are listed in the record of our meeting last Tuesday. Our current position on tertiary entrance is that the OP is no longer sustainable. The ATAR's integrity has been questioned across the country. I draw your attention to an article in the *Australian* today, one of many. There are alternatives to the OP and ATAR.

I think I should stop there. I think that that really is the crux of the situation. There are arguments more detailed about the ATAR, about the OP, about implications and consequences but, Mrs Menkens, I think that probably gives a view of where we are at, as it were, and it might be the place to start with questions if that is all right with you.

CHAIR: Dr Matters, thank you so much for that very interesting overview. We understand the huge challenges that your committee is facing. To start the questioning, because I know there will be lots of questions in the time that we have left, students are assessed against standards rather than against each other. Might your review make any proposal to change that standards-base approach to assessment?

Dr Matters: We believe that it is appropriate to make decisions on the performance level of a student based on external standards and that the student is assessed against them and not against each other. However, an outcome of that is that a student who gets a VHA in this subject and a student who gets an HA in the same subject are, in fact, competing with each other, if you like, for that space, that excellence space. Where we think there needs to be a radical change, and it is something actually mentioned in our submission to the maths-physics inquiry, is that the process to be followed by teachers in combining, aggregating, making overall judgments about how these students' marks—I am going to use an old fashioned term 'marks'—on individual assessment instruments are sucked up, as it were, to make an overall judgment could be based on numbers where numbers are codes for achievement, we can see no value in the current profiling of students with A, A plus plus, B minus minus. We suspect if they were numerically coded you would get the same result, but that the standards, the five levels—the five grades really—are defined by an empirical process that actually involves scaling, not the scaling we are used to now, but a process of calibrating assessment instruments so that they are all on the same scale and then can be added

together and the standards are related to those cut-off points. I am sorry, I cannot do justice due to that white board, but we think there is quite a radical way of arriving at a final level—we will not call it level, a grade—that has a numerical process that certainly is not that naive one. Some people are saying why can't we have 80 per cent. Well, I have to reject that and say quite frankly to you that is just purely stupid. Anybody who knows anything about mathematics knows you would not say that. So when I say numbers I do not mean naive percentages, but I mean a way of assigning grades, marks, quality indicators, to assessments in which they can use numbers but that when those numbers are sucked up into the overall there is a process that is very, very sophisticated—it can be done with a little computer black box, not sophisticated as in it will take a whole lot of time—and that derives the results against standards that way and not just by judgement. I have not done justice to that question. I will try to write that up more clearly. But, yes, we see a radical change in the way that the overall grade is arrived at.

CHAIR: But no doubt one of your focuses would be to try to relieve the stress that certainly we discovered within our investigations. There is a high level of stress amongst the teaching fraternity in arriving at the final results.

Dr Matters: Yes. We think perhaps there is energy that is used that perhaps does not need to be used in that particular type of decision. It does not mean to say some teachers are not doing it brilliantly, it just means it is really not an efficient process and we are not sure about the validity or reliability of the outcome.

Ms D'ATH: Thank you, Dr Matters, and thank you for that comprehensive overview this morning. I just wanted to go to your comments on transparency. I am going to probably put it way too simplistic in summarising what you said here, but if I understand correctly you were saying that the students do not understand the system because those teaching them do not understand the system. I am interested to know will your report actually deal with these issues, because no matter what your recommendations are and what is adopted into the future, we have to deal with why this problem is occurring and how we might address it under future systems.

Dr Matters: First of all, it is not just the teachers. I don't say that for one moment. There are people in the system who, even more so than teachers, should understand their own system, if you can understand what I am saying. Myths abound and the myth breakers, if you like, do not seem to be having an impact. This not an ageist thing, but the teachers who have lived and breathed this system for some time can explain it well. I think there is perhaps a trend where someone tells you and you tell the next person and it all gets passed on, like the oral tradition, and errors are introduced there. There is also an incredible sort of behaviour where you will say to people that is not so and they will tell you yes it is so. But factually it is not so. it is not an opinion. So there are all these things which I call a badge of honour to say it is impossible to understand. There is a lack of acceptance of factual knowledge, opinion, and I wonder if that direct communication is a function of the tradition we have—the education community, QSA, QTAC—of trying to explain all things to all people. I think this is probably a Queensland obsession. We think everybody has the ability and the willingness to understand every detail. I think we should reconsider the way we explain any system. If we are going to resort to journalese we might have to get the help of journalists actually to simplify it, but to have the main principles of it understood and then behind the curtain have all this other stuff, like if the scaling were to occur, you know, the mathematics of it, examples of it, available to people of course if they want to know but not try to even dream of having everybody understand it because once you start trying to make everybody understand it I think that is where the myths start to happen.

Ms D'ATH: But you agree that there has to be some level of understanding for the students.

Dr Matters: That's right.

Ms D'ATH: I just wanted to get to that final point that I am hoping that you are going to be raising these issues and addressing them and maybe making some recommendations around that in your report.

Dr Matters: Yes. I haven't actually covered the whole of the terms of reference. There are more. They are about communication, costing and governance and I have touched on those today.

Ms D'ATH: Thank you.

Mr BOOTHMAN: What would be the purpose of an external examination in your model? Would it have the function to moderate the school based assessment in some way?

Dr Matters: The principle that we are applying to our decision to recommend an external assessment is not that there is an uprising in the community to have it, nor that everybody else has it and so should we. Our rationale for introducing external assessment is that we do not believe it is

possible to get coverage of the curriculum using one assessment regime only. That is the first thing. We believe by having those types of assessments we will have more validity across an assessment program. In other words, all of the outcomes, if you like, all the concepts, all the content of the curriculum is more likely to be covered. That is it.

In terms of the question you asked, it is true that in other states in Australia—not in the rest of the world—that have internal-external assessment that the assessments are combined in a two-stage process. First of all, the internal assessment is scaled to the external assessment, then the adjusted internal assessments are added to the external assessments to create the score—in terms of the HSC score, for example. Scaling of internal assessment to external assessment of itself privileges one type of assessment. It makes one look like it is the right thing, one like it is the wrong one. That is the first thing. Scaling that way, which is called statistical moderation, is an expensive process and one that probably is not worth the expense. You have to have both external exams full blown and internal assessment and scale one against the other. We believe we can actually get a combination of internal and external without going down that particular path. As well as Geoff Masters and I conducting this review, we also have an international consultant, Dr Peter Hill, and we test all these ideas against him. He has worked in Victoria and understands that process completely, and in Hong Kong, and is working with Michael Barber in the UK at the moment. So we test some of these more provocative, if you like, recommendations against somebody like that. You will hear people say, people in universities, people from schools, people who look to other states, why aren't we having statistical moderation. That has been queried, by the way, as a mechanism in other states. I don't think I have answered your question properly except to say it is the first question people ask and it is a good question. We would reject it completely as something that is not needed. It has passed its use by date in other places and is peculiar in terms of the rest of the planet.

CHAIR: Our time is coming to a close. Dr Matters, this has been extremely interesting. It has given us a very good understanding of the challenges that you and your committee are facing. I congratulate you on the work that you are doing. It is an exciting time for education. We, as are all of Queensland, are waiting expectantly to see the direction this goes in. As I say, it is a changing time in education. You are certainly addressing the challenging changes that we have with students who are attending right through to year 12 and, of course, the technological era that we are confronted with. Thank you so much for your time this morning, Dr Matters. We will be very interested in hearing the final proposals made as a result of the ACER review and in due course the government's response to these. As usual I urge anyone with an interest in the work of the Education and Innovation Committee to subscribe to the committee's email subscription list via the Queensland parliament's website. I now declare this briefing and the committee's proceedings for today closed.

Committee adjourned at 11.31 am