EDUCATION AND INNOVATION COMMITTEE

Members present:

Mrs RN Menkens MP (Chair)
Mr SA Bennett MP
Mr MA Boothman MP
Mr MR Latter MP
Mrs DC Scott MP
Mr NA Symes MP

Staff present:

Ms B Watson (Research Director)
Mr G Thomson (Principal Research Officer)

FORUM ON THE ABILITY OF ASSESSMENT PROCESSES FOR SENIOR MATHEMATICS, CHEMISTRY AND PHYSICS IN QUEENSLAND TO SUPPORT VALID AND RELIABLE JUDGEMENT OF STUDENT OUTCOMES

TRANSCRIPT OF PROCEEDINGS

WEDNESDAY, 1 MAY 2013

Brisbane

MONDAY, 4 FEBRUARY 2013

_____

Forum commenced at 11.29 pm

BELWARD, Dr Shaun, Head of Discipline, Mathematics, James Cook University

BELLOCCHI, Dr Alberto, Lecturer, Faculty of Education, Queensland University of Technology

FENSHAM, Emeritus Professor Peter, Adjunct Professor, Faculty of Education, Queensland University of Technology

GOOS, Professor Merrilyn, School of Education, University of Queensland

HAEUSLER, Dr Carole, Lecturer in Math, Numeracy and Science Education, Faculty of Education, University of Southern Queensland

JONES, Professor Peter, Head of School—Medicine, Bond University

LEACY, Ms Ailsa, Acting Institute Director, Southern Queensland Institute of TAFE

LINDSAY, Professor Euan, Dean, School of Engineering and Technology, Central Queensland University

MATTERS, Dr Gabrielle, Principal Research Fellow, Australian Council for Educational Research

RIDD, Professor Peter, Head of Discipline, Physics, James Cook University

RUBINSZTEIN-DUNLOP, Professor Halina, Head of School of Mathematics and Physics, University of Queensland

SADLER, Professor Royce, Emeritus Professor, Griffith University

SANG, Professor Robert, Head of School of Science, Griffith University

WALDING, Dr Richard, School of Biomolecular and Physical Sciences, Griffith University

WALTON, Ms Patrea, Chief Executive Officer, Queensland Studies Authority.

CHAIR: Welcome to the Education and Innovation Committee's forum on the ability of assessment processes for senior mathematics, chemistry and physics in Queensland to support valid and reliable judgement of student outcomes. This is one part of a broader inquiry into assessment methods for these subjects at the senior level. I am Rosemary Menkens, the member for Burdekin and chair of the Education and Innovation Committee. The other committee members here today are: Stephen Bennett, the member for Burnett; on my right, Neil Symes, the member for Lytton; Michael Latter, the member for Waterford; and Mark Boothman, the member for Albert. To my left is Mrs Desley Scott, the deputy chair of the committee and the member for Woodridge.

While our inquiry is looking more broadly at the assessment methods used for senior mathematics, chemistry and physics in Queensland than the questions of validity and reliability, it is validity and reliability that we will focus on today. That is why we have invited academic experts and not parents, teachers, students or others with views on the pros and cons of the current assessment system. We are obtaining those views through other means, such as the written submissions and public hearing processes.

We are now commencing a broadcast of the final session of this morning's forum. For the past two hours, forum participants have been discussing the key themes relevant to the question of validity and reliability of assessment methods for these subjects and identifying what evidence exists in support of those themes. In this final hour, the forum will report back to the committee on

what the evidence says and areas of agreement and divergence in respect of the evidence. We hope participants will engage in some debate and discussion as well, and ask questions of each other. Committee members are also free to ask questions.

Hansard reporters are recording this session and so, for their benefit and for the benefit of those watching online, please state your name each time before you speak. As this session is slightly more formal than has been the case so far this morning, I remind all of us that this is a proceeding of parliament and as such rules of privilege and contempt apply. I ask you all to bear that in mind during the discussion we are about to have.

Ms Mary Maher has been engaged to facilitate today's forum. I will now ask Mary to lead us into this final session. Thank you, Mary.

Ms Maher: Welcome everyone to the presentation section of the forum. I will ask table 1, which is covering the topic 'the quality of assessment', to give their presentation. Peter and Gabrielle, you are happy to do it so, we have two presenters; breadth and depth.

Dr Matters: My colleagues are Peter Jones and Ailsa Leacy. I am going to list the criteria or the features that we distilled as indicative of good assessment. Then my colleagues will speak about those. These are our indicators of good assessment: professionalisation of assessment, operability, fitness for purpose, fairness, we must restate validity and reliability probably at a minor level compared to the major over-arching level, and also assessment that has a research base.

Professor Jones: My job is to really tell you about two key features of good assessment. As all of us in academia know, the writing of an assessment item, if it is really going to be a good item that is truly valid and good metrics, is a lot of work. It takes staff an inordinate amount of time to do this job really well. I can give an example from my role as head of a medical school. Bond University hosts the IDEAL Consortia, which is a group of about a dozen medical schools that share assessment items. That way we have a bank of several thousand items of different types—multiple-choice questions, short-answer questions and the like, OSCE stations that can be used in practical examinations—so that different staff members can go in there and write a much better quality assessment in a shorter period because they have access to this. There is the idea that it requires quite a deal of investment to have such access to these high-quality items. I often wonder what a challenge it must be for school teachers who are under pressure to produce high-quality assessments for multiple different classes, several times per year. It is little wonder that they might struggle to have the necessary skills at their disposal or the time to actually write those assessment items. I think one of the recommendations we would like the committee to consider is, really, how much extra training and professionalism we need to put in to the training of our staff and whether this is done on a school basis or in a more centralised fashion. This is work and money to do this and do this well.

The other issue about assessment is really thinking about what is the operability and feasibility of whatever we decide. Under the current system, there is an enormous amount of work done very well, by and large, by the current assessment system in Queensland, but that is not cheap. There is a large investment that occurs through panels trying to standardise across multiple schools and multiple differences. I have a brother who is a schoolteacher. I think that that consumes a large amount of time of any senior schoolteacher across this state and requires a degree of central coordination.

Whatever move there is to try to improve the quality of assessment, there needs to be an acknowledgement that you do not want to necessarily increase assessment, but if you are having a change in the mix of assessment, that change has to be funded if it is actually going to be achieved and implemented. To just say, 'Look, we're going to add something in', all of these decisions have a human cost to them. So they really have to be considered like capital investment. If you say that your staff is going to spend so much more time assessing, that means a reduction in time available for teaching. It is just a simple equation and getting that balance correct is very important.

Ms Leacy: I am going to talk about two aspects of what we defined as 'quality assessment'. The first one is fitness for purpose. It is really about what are we assessing. Part of what I brought up there is from a VET perspective. When you look at the rigour of assessment in VET, it is around the demonstration of knowledge and it needs to be repeated demonstration. So it is actually demonstration over time. Also, a consideration in fitness for purpose is whether they are first-order or second-order skills. Peter made the point also about building foundation knowledge and skills before we move to higher-order skills. Also, one of the issues that we discussed there was around integration and holistic assessment, and also issue to do with, as I said before, the actual repeated

nature, whether it is able to be demonstrated over time. I did speak about the notion of competence that, at times, people see from a VET perspective that that is dumbing down, if you like. But it is actually very rigorous. When you talk to people and industry teachers, they actually have a really high standard that they expect for people going into their industry. So we talked about those notions.

When we talked about fitness for purpose, we discussed that really it depends on what the outcome of the assessment or the actual course is. You could look at it from three different perspectives, and we are not making a judgement about any of these three. If you look at it that students are undertaking education for an individual purpose, then the assessment would be based on what does that individual actually know and are they able to do that thing. If it is for university, obviously that actually would inform the purpose of the assessment. Again, from a VET perspective, our assessment is very much on employment outcomes, so is it that the assessment in chemistry, physics and maths should be for university, or should it actually be for employment purposes? These are the things that would inform fitness for purpose.

The next aspect is fairness. We talked about plagiarism. We talked about timing. Again from a VET perspective, I was describing that people really do have—not forever, but they do not have to demonstrate competence within a set period. They can have multiple opportunities. If they are not competent, they can have another go. I was comparing that potentially with assessment items at school where you have a set period. We talked about anonymity. We talked about fairness in relation to styles of learning, labelling. So appropriate higher-order thinking skills in there as well is what we are testing and when.

Dr Matters: I will just round off. Probably I will take one of the three topics that I have left and that is that good assessment or assessment that is described as good can only be described as good if there is an empirical basis for that statement. Two aspects or two questions that will be asked in order to answer the two questions that lead to the conclusion that the assessment is good, the assessment is of high quality, need to be grounded in fact, that people do understand the aspects of the system that they have opinions about; and No. 2, that there is an empirical base for answering the question, 'Is the Queensland system functioning well?'

Ms Maher: Thank you, group 1. The second topic is 'the marking and judging involved in assessment'.

Professor Lindsay: Our group focused on the physics syllabus side of things. We believe that we did that without a loss of generality, because of the similarity across them. Our focus was essentially on how teachers take the work that the students produce and then turn those into grades; how do they actually decide what a student has achieved and at what level. Hopefully, this grid is somewhere in the notes. There are a number of different criteria and the teachers assess the piece of work against those criteria. They place ticks in a box, that is actually the way that it works, and they look at them. There are five different levels of achievement, A through E, each with descriptors saying 'this is what an A is', 'this is what an E is', and so on. It is all very easy if a student gets all As. It is all very easy if a student gets all Es. Where we got into the meat of our discussion is what happens if a student is spread across multiple columns, that some of the criteria are being achieved well and some of them are not being achieved well. One of the key sticking points for our group was the fact that the system as used at the moment is not explicitly quantitative.

The physicist engineers around the table working in higher education are used to examples where you have: criterion 1 is worth 10 per cent, criterion 2 is worth three per cent, criterion 4 is worth 20 per cent. So you can put a number: A equals five, B equals four, et cetera. At the end of it, you crunch the handle, out comes a 72 and that must be a B.

The system that is used in the high schools at the moment is not like that. There are deliberate educational reasons for that. There are issues about, is a B really four where an A is five? Is A 25 per cent better than a B? Is there an equal gap between them? Should all of the criteria be weighted equally? This came to the root of one of our bigger discussions that I shall return back to at the end.

As we flowed through the process where each teacher marks each piece of assessment, they put the ticks there, they look at a portfolio at the end of the year of all of the work that the student has done, see where the bulk of the ticks are, and based on—there is an objective panel, but the dominant mode, so if they were getting mostly Bs they will get a B, if they are getting mostly Cs they will get a C. But there is a caveat that there is a minimum tick. So if there is a tick down in the Es, you cannot give the student more than a D. So there are mechanisms there to guide the teacher in how they make this averaging.

This was one of our concerns, certainly from a reliability process. If it is going to be reliable, it has to give you the same outcome regardless of which teacher does it. While we have to rely on the judgement of teachers to put the ticks in the right place, further, we need to rely on them to average them correctly. But it seems that there are guidelines to say, you must average in this way, use your discretion within this box.

Further to that, there is a process where samples are taken from each school and external panels compare samples. They look at 10 from this school, 10 from that school, 10 from this school et cetera and, in particular, borderline cases—this one is on the line between a B and a C; this one is on the line between an A and a B. They get examined. The panels moderate. They make recommendations back to the schools. So, while all this is happening in a qualitative way, rather than the perhaps more brutal quantitative numerical method that I would have been more comfortable with, it does seem that there is a mechanism to ensure that this gets back. There was a discussion that it works better with experienced teachers rather than novice teachers. But, then again, my understanding is that most of the education system works that way as well.

One of the big discussions we had was what 'sound' means. This is where you have that issue of, if the average is a C, does this mean that the student receives Cs for everything they did or did that average out of some Ds, some Cs, some Bs and some As? The meaning of 'sound', the choice of the word, is something you can rely upon: it is sound. That is fine if they have received all Cs. If they have received some Bs and some Ds, it means that on the whole they are sound but they are unsound at some things and quite good at some other things. This led to the main tension we had around our table—that is, the difference between senior chemistry, physics and mathematics being seen as high school leaver representation. So when you say, 'This student is sound,' that person is someone we would expect to have completed mathematics at high school versus that person is someone who is sound to enter university and complete mathematics.

The consensus that we had around the table was that of the criteria that there are—there are nine the different criteria for physics broken into three categories—the universities do not value each of them equally for entry. So, if the 'sound' that you have is an average of quite good, sound and not sound but you happen to be not sound in the bit that the university is valuing, then there is a dissonance between the assessment criteria we are producing, the statements we are making about our high school leavers and what society perceives of them versus what the universities perceive of them. This is how we get to situations where deans of schools make comments such as 'I get these students and 20 per cent of them cannot add up.' So there was some suggestion of does this mean we need two separate streams? Do we need senior physics for people going to university and senior physics for people who are not going to university? And I would leave that to the committee rather than dare answer that right now.

So we discussed the concept of what does being rated as 'sound' actually mean. Does this statement mean that they are okay at everything or does it mean that they are good at some things and not good at others? And that gets back to the validity. What is it that we are actually saying? What is it that we are actually measuring? Are we measuring their fitness to leave high school or their fitness to enter university or some other measure? This was particularly in the context of open questions versus closed questions. It seems the curriculum has moved to answering open questions. When you look at the criteria that are set out, there are criteria that deal very well with closed questions and criteria that deal very well with open questions, but it seems to be the closed questions that we at the university level see a deficiency in. So from our validity of what we would like the high school senior curriculum to say when we get someone who is 'sound' is that we mean them to be sound in that first set, but if they have come through unsound there but good in other places it might mask that.

Certainly with the A level students, when they are all As it is very easy to aggregate those ticks. But there is a qualitative system. It does seem to produce exactly the same thing as you would get if you did it quantitatively. But because the people who are relying on it at a university level are used to crunching the numbers rather than being guided by the rubric, there is perhaps a certain lack of trust in the process even though it does give the same outcomes. Have I missed anything, team? Thank you.

Ms MAHER: Thank you, group 2. Now can we have Merrilyn speaking for group 3. Group 3's topic was the links between curriculum, learning and assessment.

Prof. GOOS: That was the topic that we were meant to look at it and, even though we wandered quite a way around that topic, I think it is indeed what we addressed. It does not make any sense to talk about assessment in isolation from curriculum and teaching. Those three things must obviously go together. When we looked at alignment between assessment and curriculum, I think we took it for granted that this was important. Assessment must be able to provide evidence of the extent to which students have learned the things that the curriculum says they should learn. But, even so, we then went off into a discussion of why is it then that it seems when students come to university and start to learn, particularly university mathematics—that was one of the main areas that we discussed—they do not know the things that we would like them to know? They even seem to be struggling with what seems to be lower secondary knowledge of mathematics.

But I think we agreed it was important to realise that this transition difficulty that students experience is not unique to Queensland. It is not even unique to Australia. It is an international phenomenon, no matter what kind of education or assessment system exists—which should tell us that maybe there are some other reasons for these sorts of problems happening. So we talked about things like the amount of time that students spend learning mathematics in schools now is less than it was, say, 40 years ago. There have been a lot of changes in university entry requirements over the last couple of decades in terms of changing or removing some prerequisites, which send the unfortunate message to schools and students and parents that you do not really need to do maths C, for example, if you are going to be an engineer when actually it is really, really important that you do know that kind of mathematics.

We also acknowledge that young people are different now from the way they were 20 years ago. The world is a different place. Technology has changed the way that we access information, and of course information is not the same as knowledge. So we know that we need to change the way that we teach students where 'we' mean teachers in schools and in universities. So we came back to what we think is a fairly key question about assessment: does it provide us with confidence that students when they leave school know enough mathematics, chemistry and physics to be able to succeed at university? So we tried to unpack that a little bit more. We do not think it is possible to guarantee that every student leaving school with the same apparent level of achievement knows exactly the same amount of stuff or to the same depth or the same quality, and that has never been the case. But we would like to be confident that school leavers when they get to university have some common foundational knowledge upon which we can build and they are able to acquire new knowledge at university.

That brought us to what, I think, is a really important piece of information which has already been emphasised before, and that is that the quality of teaching that students experience is absolutely crucial. There is a large amount of rigorously conducted research around that demonstrates that teachers do make a difference to student learning. So it is important that teachers know how to design assessment tasks, that they know how to teach students how to respond appropriately to those tasks and that they know how to make sound judgements of students' learning based on those tasks. So we ended up by asking: what do teachers need in order to be able to do those things? Firstly, they need time. They also need resources, advice, examples, professional development. The experience of social moderation where they sit down with other teachers to examine students' work and come to a consensus professional judgement about that work is also extremely valuable preparation and training for teachers in making good assessment decisions. Do I need to say anything else, team?

Ms MAHER: Thanks, group 3, and Merrilyn. We will go to group 4 and the coverage of the specific needs looking at the subjects of senior maths, chemistry and physics. It is over to Carole.

Dr HAEUSLER: There were two main points that we debated in our group. The first one related to the purpose of senior science and maths education and the second one was the role of written work. There was a difference of opinion regarding the purpose of senior science and maths education. One was that it is for university. The other opinion was that it had a wider application—that is because education now expects most people to go on to year 12 and, of those that take science, not all of them will go to university and, of those who go to university, not all of them will do science at university. Therefore, we argued that perhaps science education was for developing a scientific literate community which then would have implications for the nature and content of science education.

One of our members, Peter, expressed a concern that in physics and maths there was in some cases written work that exceeded 3,000 and even perhaps 10,000 words and the argument was that boys who take physics are unable to handle this amount of writing. Peter Fensham reported that research shows that written work has in fact encouraged girls into physics and other areas. He has also done a study of chemistry across four different states, and it has shown that in

Victoria they have more chemistry topics but they have less focus on higher order learning than South Australia and Queensland, who have less content but a greater focus on higher order learning, and this is demonstrated primarily through written work.

The advice that we have is that perhaps QSA needs to provide clearer advice on the length and nature of written work such that excessively long assessments are not required of people in physics, chemistry and mathematics. We also suggested that perhaps teachers should be given professional development on how to develop questions that target higher order understanding in more brief written work. Perhaps the training of teachers needs to be addressed in this area. I think that is all we are going to say in our group. We argued over that for quite a long time. Do either of the Peters want to say something?

Prof. FENSHAM: I would just like to add to what Carole said. I am a voluntary tutor at a homework centre for refugee kids and I will be going there this afternoon. From some high schools I get students coming with a 3,000-word physics task and their question to me is: 'How can I turn this into 1,500 words?' That does not seem to be a physics task; it is an English task. On the other hand, we have students coming from other schools with a task of writing 1,500 words but that is broken down into five lots of 200 words which are very clearly targeted by the teachers to indicate the difference skills and aspects of the content knowledge the student is to provide evidence of understanding. So I think the same sort of task needs to be monitored rather carefully by what was mentioned by one other group. But moderation by groups of teachers across schools is, we know, an extremely effective way of both professional teacher development but also of ensuring comparability and reliability.

Ms MAHER: I would like to thank the four groups for working so hard on those topics. I now open it up for discussion. You have table microphones and the committee members also have microphones. So it is an all-forum discussion based on the points that have been raised or additional points. We know that we parked some items. They may resurface in this open discussion. So let's throw it open. There were quite a number of questions that the groups raised and they are marked.

The criteria of quality group, group 1, did some very good work on the criteria of quality and the importance of investing to get that quality. The marking and judging group, group 2, talked about qualitative and quantitative. They gave an example of looking at the 'sound' rating and talking about what that means and looking at where the marking and judging feeds into, from school into what—what is the point? What are we measuring?

The links between curriculum, learning and assessment group, group 3, had a discussion about the fact that people do not arrive at universities, for instance, or where the schools feed into, knowing enough. There is a multiple set of causes for that, but there is the need for assessment to give confidence that students know enough to succeed. So succeeding at uni would mean a common foundation knowledge plus the ability to acquire knowledge. Teachers were rated as a fairly critical set of players in the assessment links question.

Then we had the specific needs in physics and chemistry group, group 4. They looked at what the purpose is of assessment in those subjects. Is it to have a wider application of that knowledge or is it for university? And then they looked at the question of the amount of writing and how the design of the assessment items needs investment and monitoring.

So they are some of the issues. Given that we have had a good look at some aspects, what are the powerful questions moving forward if we were going to make a difference? Let's play with some.

CHAIR: I would like to throw the direct question to any and as many members here who would like to answer it: are you happy with the current assessment system in our maths and science subjects in senior? As I say, it is an open question. I really congratulate you also for the wonderful comments that we have heard, but this is a simple straightforward question and answer. Where do we go from there?

Prof. JONES: One of the challenges here is that they have adopted a single criterion based assessment for all things. There is no doubt that there is room for having competencies of being able to describe effectively what your knowledge base is, but I do think that there is room for some type of quantitative examinations which look at whether the students can do the relevant calculations or formula that they might need to understand. Testing those things is difficult, so not only should there be sharing amongst the panels which look back at the decisions that teachers have made to try to make them as fair as possible—and Queensland does that—but how could they best share items to allow that type of accurate quantitative analysis to occur within schools? I think

in a lot of ways the question that was mentioned by my colleague—that is, 'Gee, what does sound mean? Does sound mean an average of soft and loud?', which is what it appears to mean at the moment—means that you are therefore guilty of averaging things and having mixed results blended into the one bucket at the end. You need to take apart those things which are things best assessed with numbers, facts and figures and those things which are assessed as a demonstrated competency. That way there might be enough information so that we know that they have some knowledge, but we also know that they can do these certain sums so that they will be all right when they turn up at university.

Prof. FENSHAM: I think we need to get away from the notion that a quantitative score has a greater reliability to tell us that a student can do a range of things than a qualitative score, because you can take a common test and get 60 per cent and it can be made up of all sorts of different combinations from the various questions that are asked. So the 60 per cent may seem to be quite a good score, but it will mask the fact that on some questions you have a very low score and other questions you have a high score. So I think what Peter is suggesting is that we need a clear profile of what the student can do—whether it is a quantitative measure or a qualitative measure—but the notion of a profile as distinct from a single score seems to me to be an important thing that solves quite a lot of problems that we have talked about.

Prof. SANG: I guess the question I would throw up is this: what are you assuming as a successful high school graduate out of one of those disciplines—chemistry, physics and maths? After our discussion around the table I see that there are differences of opinion what that should be, whether someone is going to university or whether they are just going out into regular employment straight after that. Surely your assessment system has to be based around what your idea is of what a graduate of the discipline is. I actually do not understand what the current assessment system is trying to produce—who are they trying to produce? For example, for somebody who has completed physics, what is that for? Are they competent to enter a university and study the discipline of physics or science or STEM engineering? So that would be my question about your assessment system: who is it designed to produce in terms of graduates?

Ms MAHER: A powerful question. Would anyone like to comment on the powerful question that Robert is putting?

Ms LEACY: I just want to add one little consideration; it is not really an answer. I would again just refer to the VET context in that assessment is driven by industry, so the question I would raise is: what actual employment, and if it is university employment that is a fine outcome too? I am not suggesting that research is not a good outcome, but what are the range of people who will be employing those graduates in the future and should they not have some role in informing what the purpose is?

Ms MAHER: Thank you, Ailsa. There is a lot of thinking going on. What is the system designed to produce and can we have a mix of quantitative and qualitative? They are becoming some of the headline topics. Would anyone like to take up either of those? If you were designing a system, what would you be designing it to produce? Robert, you asked the question. Give us a hypothesis.

Prof. SANG: Unless I am wrong about this—I am not a high school teacher—I would imagine that most students who study the enabling sciences and mathematics have an inherent interest in that. I would think that, with regard to the graduates of that—and I do not know what the numbers are, the percentages—a lot of them would be thinking about going on to further study at university. If they are going to study the hard sciences, I think that is the general profile of that type of graduate. The question I would then design around that is this: what knowledge do I need to impart to these students so they can succeed later on? That is what I would be thinking about—that is, in terms of what I need to teach, what is the general knowledge and skill bases that they need in order to succeed in those disciplines?

Prof. JONES: This is speaking with my medical educator hat on. Most of the skills, as my experience has been, are reasonably generic in that if the student is a highly performing student they will adapt to their learning environment that they come from. For instance, in medicine whether I have taught in undergraduate or graduate schools there has always been a substantial number of people who have come from a graduate background but not necessarily from a science graduate background. All the evidence for medical education at least is after about one semester it does not matter what the background of the student was and it is really the general aptitude of the student that determines their performance at medical school at university. So the question really for this assessment is: how do we give the students the general aptitude skills via physics, chemistry and maths to be able to do things?

One of my concerns about the system is that, yes, it is very important to be able to write written answers, but at times as the pressure is on to try and improve and demonstrate that you are better—and I know everyone hates the idea that it is all about ranking, but I know from the way students try and climb over one another to get into medical school and get that OP1 or that 98 or 99 mark—it is often all about trying to get that ranking. We push students to try and demonstrate very high order thinking skills before they have got foundation skills in place. I think there should be a greater emphasis on making sure that some of those foundation skills are really well and truly taught, whether that is numerical, chemistry or the like. I think all of us who are academic know those basic hard skills that seem to somehow be able to compensate for the ones with regard to the noise that Euan was talking about.

Ms MAHER: So, Peter, you are saying that the feed-in process for you would work better if it had some of these characteristics—that it mixed the quantitative and the qualitative but that you are not saying the system should be geared to university entrance because universities have become more flexible and entry—

Prof. JONES: What I am saying is that the people who succeed at uni are those students who are the best performers. What should happen for those students who perform really well is that they should have really good foundation skills and sometimes in an effort to meet the assessment criteria that is so dense with what the students have to produce. This is where I am an illegal parent, because I am on my third child out of five through the school system. When assignment time hits for kids, they might have three big assignments—one in history, one in English and one in maths—all coming up in the next fortnight that they are just worried about demonstrating and getting the necessary number of ticks to get this job done and there might be an inability of time to really focus on that opportunity to make sure that they have the foundation skill. So they can still get enough ticks to get those 3,000 word assignments written and they will pass all of them, but having done all of that work we still do not know if they know how to do early secondary school maths because they have been doing all the noise.

CHAIR: I was actually going to direct a specific question to Professor Peter Ridd. Taking it right back to our current assessment system, if there were to be changes in the current assessment system, do you believe that it could be tweaked at the edges or that there should be a total change?

Prof. RIDD: I think it needs a total change. I think the problems with not being able to rank the students properly using the letter grades are so massive that there is no way to get around it and go to a numerical system. That is not the biggest problem. The biggest problem is the overuse of written material. That is a tweak because it could be done very easily. You could stipulate much shorter EEIs. You could even enforce the rules properly, because that would be a useful thing. So that is a tweak, but the letter grade thing I think just has to go. In order to ensure comparability across schools there is no choice but to have a small external examination as well, but that is not a tweak.

Prof. LINDSAY: If I might respond to that, I think part of the challenge that we have here is that we are using the same assessment regime for two purposes—the high school exit and the university entry. I am very supportive—I think it came up with table 3—of producing a scientifically literate society. I think that is a very important goal, because if you as engineers have to convince people to drink recycled wastewater the more people who have done year 12 chemistry and understand that the water is pure and it is all good and you can drink it the better. It is not just about then putting them into chemical engineering degrees. With regard to the issue that Peter has referred to—the issue of being able to rank them—you do not need to rank people when you are deciding whether or not they have completed high school; it is whether they have achieved the things that a high school graduate should do. Where the ranking comes in is in our ability to test that—that is, the validity of the assessment for us to predict whether or not they will succeed if they go into further or higher education.

So the problem is at the moment the same tool is being used to produce both—that is, we are trying on the one hand to see whether they have passed high school and whether they are ready to be citizens and be trusted to vote, and on the other hand we are using the same tools to decide whether or not these people will succeed if we then put them into the four, five, six years for a medical degree. In terms of how you then take that qualitative As, Bs or ticks and turn that into an OP ranking so we can distinguish who gets an OP8 and who gets an OP9 and who gets let into my engineering program and who does not, I think that is part of where the problem is. So when we say, 'Are we happy with the assessment regime? Is it valid?', it may be a valid tool for deciding who has passed high school. But whether the current system is producing a valid predictor of success at university is a very separate question. They have different demands and they do not overlap. How

you would decouple them I think would be very difficult, because then you are making students do one set of tasks for one purpose and then do you have them do a different set for another one? But at the moment the same scheme where we are trying to ask, 'Is it reliable? Is it valid?', we are actually looking for it to serve multiple purposes and to be valid for both.

Ms MAHER: Peter has a comment.

Prof. FENSHAM: I have a couple of comments. The chairperson asked us whether we were satisfied. I think as educators we are never satisfied with the status quo, but what the committee has to decide is whether there are any advantages in radically different systems or, as you put it, can we treat this system to improve it? I think the committee has to ask itself: what are the advantages and what would be lost if you really radically change the system?

The second point is that the committee has to realise we are all meeting together because these three areas of the school curriculum are ones where there is a shortage of students and particularly a shortage of students at university. So we have to ask ourselves the question: does the form of assessment hinder or encourage participation in the subject? That is not a topic which falls within the validity-reliability question but I think it is fundamental. It may be that the system we have is more encouraging than some of the other systems, but I would suggest the committee has to think about whether the system—whatever system you use—encourages students to take these subjects.

Prof. RIDD: With regard to the question of whether we should go to a radically different system, we actually have a radically different system now. We are radically different from every other state except for the ACT. I do not think there is any doubt that in some regard we are encouraging more girls into physics. I have had a lot of teachers tell me that and it would be interesting to know what the numbers from the QSA say. That is part of this business about them being more comfortable with the written side of things. Now that is a good thing. No doubt that is a good thing. But, on the other hand, are they doing the physics which is valued by universities or is it the stuff that is valued by society? We seem to keep coming back to that. It is what Euan was saying about the assessment trying to do two things and maybe it is doing one better than the other. From my perspective it is certainly no good for university entrance.

My feeling is, yes, we may be getting more participation by girls but what is happening to the boys? I think the boys are a segment of the community that are doing very badly at the moment. They are doing the hard yards at school. It is partly their own fault because teenage boys are mostly completely hopeless but we should not hold that against them. They will turn into decent people eventually and we should not make it so they are badly disadvantaged in those formative years and they make the wrong decisions as to what they do at university perhaps because of a bad experience at school.

Prof. GOOS: I would like to go back to some of the really important questions that were being asked earlier about what is it we want our school leavers to be able to do. I am glad those questions were raised because they are fundamental to curriculum development. They are not primarily assessment questions; they are curriculum questions. I would like to remind people about that. If we are pondering on what kind of people we want to produce, we should not look for the answer in our assessment system as the first place to look. We need to consider what are our curriculum aims and objectives. Whatever those aims and objectives are, the way we assess must be able to provide evidence as to whether students are able to do those things. So we are kind of taking a step back here, and I think that is quite a reasonable thing to do in asking the question: what are the things that we want students to be able to do?

Mrs SCOTT: I wonder whether we should also be asking what the drop-out percentage at university is or the percentage of students changing courses—in other words, those who select a certain course and find that they are not adequately prepared and they then change to a course which may be a little less rigorous. I wonder whether that might inform whether or not we are preparing students and assessing them for university entrance.

Dr BELWARD: I can answer the question about drop-out rates in first-year maths subjects. They are quite high and sufficiently high that it causes concern amongst the hierarchy at the university. Why do you want to fail 40 per cent of your class? That is an entirely reasonable question. We should not be having a situation where students are misled as to their readiness for university study on the basis of grades that are not accurate in terms of suggesting how they might succeed at university. So there are issues around all of that. I note Alberto's comments before that when he was in first-year maths at JCU 20-odd years ago the failure rate was also around 50 per cent. So in some sense things have not changed. It is just that there tends to be more of a focus on

failure rates these days than there was in the past. Certainly understanding the decisions students make about study at university based on their high school grades and what they perceive those grades permit them to do is something that we need to get a handle on. I think the university sector has got some way to go in terms of unpacking that. We are probably not astute in understanding what the grade system from the secondary school system is saying. Having said that, we think they come with a certain skill set but when you read the curriculum documents there are certainly flaws in terms of some of the skills they are supposed to have.

Dr BELLOCCHI: I would like to follow up on Shaun's invitation to speak. I would like to throw the question back to those of you who are academic scientists as to what changes you believe you have seen in your student cohorts from the past until the commencement of the criterion reference system. There is an underlying assumption in the hearing that there is a problem with the current system. How is it different from the problems of the past which Shaun alluded to with the comment about mathematics?

Prof. RIDD: Firstly, I think there is no doubt that their algebra and their calculus are worse. They come in supposedly having done physics but they have not done vectors and they have not done basic mechanics. They may be better at writing—I am not sure about that, because we probably do not test a lot of that. They come in and they have not done a whole lot of basic stuff which we have definitely noticed. We are going to have to yet again change our first-year physics offering for next year or the year after to take that into account.

Dr BELLOCCHI: To follow up, Peter, do you have a sense of the level of achievement that these students have gained in high school and whether there is a difference between students who have achieved higher levels or lower levels? Are you talking about your entire cohort or is there a difference within that cohort between people coming in with an A level of achievement, a B level and a C level in physics?

Prof. RIDD: In the stuff that Shaun did on maths, there is no statistical difference between the level of achievement as to whether they get a high or a sound, for instance. That is saying that the QSA assessment has not been reliable—well, at least it has not been accurate because you would expect the people who got a high to do better than the people who got a sound. Irrespective of the fact that both of them failed, there is no doubt the people who got very highs did better than the rest, but the assessment was coming back unreliable. In terms of their abilities it is very mixed. People are coming from some schools and they have tremendous skills and they may only have a high or even a sound. Because in the physics syllabus there is almost no content that is actually stipulated as being the basic stuff that you have to teach, schools can do almost anything. So some kids will come from schools having everything that we want and others with very good grades might have very little of what we want in terms of background skills.

Dr HAEUSLER: I think there are two wider issues that need to be addressed. One is that the drop-off in interest in science and mathematics is an international problem. It is not just Queensland; it is everywhere. Secondly, with regard to the issues that you raise about the quality of high school graduates entering physics, is it any different from the comments that are made in Melbourne, Adelaide, Sydney, Perth et cetera?

Prof. RIDD: This is an extremely good question and it is something the QSA has not done, which is to benchmark us against the other systems. We have an extreme system here. I cannot tell you whether the system is better or worse in the other states in terms of our assessment and all the rest of it, but neither can the QSA. Considering we have such a different system, why has the QSA not looked at this and benchmarked our students against other students at the end of year 12 to see whether they are better or worse? The onus should not be on somebody such as me to say, 'It is better in Victoria.' The onus is on the QSA to demonstrate that our system performs at least as well as the others and that is something that they have not done.

Ms WALTON: There are a number of things I would like to talk about. Some of the things that have been said here this afternoon need a bit of clarification. There has been a lot of discussion about writing in senior mathematics, chemistry and physics. I want to make it very clear that in year 12 students are required or mandated in physics to do one extended experimental investigation out of the assessment in year 12. That is one assignment if we want to use that language. Of about six assessment pieces that are usually required in the year 12 physics assessment program, one is an assignment and the school can make the decision around the other assessment items. I have to tell you that typically the other five are exams. It is an exam environment that is used—exams, tests and one assignment. I just want to make clear that all of the assessment pieces in physics are not assignments.

Secondly, with regard to the physics content, I have the physics syllabus here in front of me and the physics syllabus does describe areas of learning. Can I also say that the physics syllabus has standards and those standards are not all that dissimilar to the standards that you will see in the Australian curriculum. There was an announcement of 7 December last year where the ministers agreed to 14 senior secondary subjects for the Australian curriculum. They, too, surprisingly have standards and very detailed standards, even more detailed than the Queensland QSA syllabuses.

In regards to students' preparation for university, I have been talking to Shaun about some of the work that he has been doing around students' performance when they get to university and certainly we would like to see more of that because that is of great interest in the education community. But there are other factors around student performance when they get to university. It is a big step leaving high school and going to university, just like it is a big step going from primary school to secondary school. It is a big step today just like it was when I went to university. It was a big step years ago. Students perform well in semester 1 for a whole range of reasons. They may not perform in semester 1 for a whole range of reasons. They may not go to lectures, for starters. They may not go to tutorials, for starters. I am sure that will be in the research, Shaun, and I am keen to see that. But students' performance in semester 1 at university cannot wholly and solely be attributed to their performance in one subject when they are in year 12. That will do me.

Prof. SADLER: The discussion today has not really gotten close to the validity and reliability questions that were meant to be the focus. One of the difficulties is our use of words. For example, if we are talking about predictive validity it is the ability of high school examinations to predict performance at university. It is not a reliability issue. Given that, the figures have shown historically never to be very high for reasons which have just been outlined by Patrea.

In the days when universities had elitist student populations—and I do not mean that in a nasty sense—the predictive validities were limited because the selection fraction was very small and that automatically lowered your predictive validity. There are a whole lot of factors like that. Our context has changed. I really think that the issue that has been at the forefront of the discussion today has not been the validity and reliability issues, neither of which have been shown to be major problems, but the whole question of the purpose to be served by studies in sciences and mathematics at high school level. What I think we have got is something we have always had. It is trying to serve dual purposes and it is not going to serve either of them brilliantly. But the point is we want to serve both of them well enough.

At that point I think we really have to say that that is about as far as the discussion can go. We do need to keep an eye on the curriculum. We do need to make sure that when we say sound achievement it agrees with what people think is sound in the normal sense of the word. So those kinds of things we really must attend to so that misconceptions do not develop. Apart from continuing to be vigilant and make sure that the integrity of the moderation system is not compromised by some untoward behaviour in some places and there are various checks like that—I am glad to see that the QSA engages in the random sampling of student work; they just do not look at the results they look at the actual student work—these are very positive things that contribute towards the very properties that we want the assessments to have.

Ms MAHER: Thank you Royce, could I hand to the committee chair to make the final words.

CHAIR: Thank you, Mary. Ladies and gentlemen, thank you all so very much. There have been some excellent points brought forth today. We really appreciate the time you have taken out of your very busy schedules to be here today.

The committee will now take quite some time to digest your contributions. We greatly appreciate your input into an issue that is of significant consequence to Queensland not just to students, parents and teachers but also to the broader community and economy. The committee is due to report to parliament on 16 August. It will be taking written submissions until close of business on Monday, 13 May. I would invite all of today's participants, as well as other people who have an interest in this area, to provide a written submission to the committee. We will then invite further witnesses to attend public hearings during May and June.

I urge all who have an interest in this inquiry or the committee's broad areas of responsibility to subscribe to the committee's email update list via the parliamentary website to ensure that you are kept up to date about inquiry and committee activity. May I thank you all again. I now declare this forum closed.

Forum adjourned at 12.33 pm