

CORRECTED VERSION

EDUCATION AND TRAINING COMMITTEE

Inquiry into the education of gifted and talented students

Melbourne — 25 July 2011

Members

Mr P. Crisp
Mr N. Elasmarr
Ms E. Miller

Mr D. Southwick
Ms G. Tierney

Chair: Mr D. Southwick
Deputy Chair: Ms G. Tierney

Staff

Executive Officer: Ms K. Riseley
Research Officer: Ms M. Scott, Ms A. Madden
Administrative Officer: Ms N. Tyler

Witnesses

Professor G. Masters, Chief Executive Officer, and
Mr R. Saubern, Director, Assessment Services, Australian Council for Educational Research.

The CHAIR — I would like to welcome ACER to the public hearing of the Education and Training Committee's inquiry into the education of gifted and talented students. All evidence today at this hearing is protected by parliamentary privilege — the same privilege we are protected by — but any comments made outside this hearing may not be afforded such privilege. Hansard is recording today's hearing, so we will provide a proof transcript for you to have a look at. If there are any errors, you will have the opportunity to correct them. We have 45 minutes. We will have a brief opening statement from you. We have received your submission, we have read it and we have a whole host of questions for you, so we want to have as much time as we can to go into the questions and answers, so if we can keep the opening statement brief, we will get straight into the rest of it. Welcome to both of you.

Mr SAUBERN — I would like to start by thanking the committee for the opportunity to make a submission and to give evidence here today. We are here to answer your questions about our submission and more broadly about the work of ACER. Before we hand over to you to ask some of those questions, I thought we might very briefly point to the main issues we raised in our paper. I note firstly that whilst your inquiry is rather broadly constituted, quite rightly, around the areas of gifted education in Victoria, our submission was constituted quite narrowly. We focused on issues in terms of the identification of students using objective assessment measures.

In that we noted that there is considerable discussion and debate about the nature and definition of giftedness in students, about how to identify such students and indeed about how to teach such students once they have been identified. ACER's work focuses less on the idea of giftedness and more on the idea of high academic achievement or the potential for high academic achievement, and our assessment instruments are designed and developed in order to identify those students who have higher order reasoning skills and abilities in those contexts. ACER test instruments are underpinned by a particular statistical approach that ensures a level of reliability and validity. It is based around the definition of a construct, can be measured and measured fairly from one student to another and can be compared fairly from student to student and from time to time.

It is our submission that such research-based effective assessment is an important part of whatever processes are used at the school or system level to identify gifted students for particular programs or streams. That is probably what we wanted to draw attention to in the first place, and we are very happy to take any questions the committee has.

The CHAIR — Your submission refers to the decline in performance of high-achieving Australian students by international students. What do you think is responsible for this decline?

Prof. MASTERS — Nobody knows exactly what is responsible for the decline. Various hypotheses have been proposed, but what we do know is that over the past decade there has been a decline at the top end in reading and mathematics. One of the hypotheses is that, because we are now so focused on making sure that all students achieve at least minimally acceptable levels, schools have taken their eyes off extending very high-achieving students. This is a phenomenon that has been observed in other parts of the world. When you set minimum standards what teachers tend to do, quite logically, is focus their attention on students who have a chance of achieving the standards and they put their energy into students who are just below the cut-off point, often at the expense of students who are well below and are unlikely to meet the standard and certainly at the expense of students who have already met the standard. That is one hypothesis.

Another suggestion that has been made is that students — and these are 15-year-olds we are talking about in the OECD study — may just be reading less extended text than they once were. We see that the reading levels have come down. Now with modern technologies and other ways of communicating one possibility is that fewer students are reading extended text and so students are less able than they once were to make connections across extended text and to extract deeper meaning from text. However, these are just hypotheses; we actually do not know the reason.

The CHAIR — In terms of the testing, the committee is aware that ACER tests are used by some SEAL schools — for example, Mill Park Secondary College. How many SEAL schools in Victoria use ACER tests, and which are the ACER tests that are used in Victorian schools for entering into programs for gifted students?

Mr SAUBERN — As far as we can identify, and this may change from year to year, about half of the SEAL schools use ACER test instruments or programs. About 15 have in the last 12 months used one or other of our programs. Those include particular assessment services that we operate. One is called the Higher Ability Skills

Test (HAST), which is used by 10 or 12 of the SEAL schools at the year 6 level or at the level of entry into year 7. There are also other SEAL schools that are using published assessments, which are widely available to schools for a variety of purposes, including identifying ability levels of students, and two of them are the ACER General Ability Test (AGAT) and the Middle Years Ability Test (MYAT). Those are the ones that we can clearly see are being used by SEAL schools.

Ms TIERNEY — In terms of the ACER tests, is it correct that they identify giftedness only in academic areas?

Prof. MASTERS — Our most commonly used tests identify academic ability, but it depends on how broadly you define ‘academic’. A lot of our tests focus on quantitative reasoning and verbal reasoning, but we also have tests that include written expression, so the ability to make arguments. There are assessments that look at interpersonal skills as well, and there are assessments of spatial abilities. It is academic, yes, but it is a question of how broadly one defines ‘academic’.

Ms TIERNEY — Does it potentially cover leadership and music and ICT?

Mr SAUBERN — Potentially, yes. ACER will create a test which is based on a construct, so if part of the construct is leadership skills, we can develop a test which measures that construct. As far as I am aware, none of our school-level assessments include constructs for all things such as leadership skills or particular talent such as musical ability or ability in language learning. Some of our assessments that have been designed for university level or for adult use do include such constructs.

Ms TIERNEY — You say in your submission that you give very high regard to the importance of an objective assessment for potentially gifted students, and it notes the biases that may arise if teachers are administering it. What is the role of teachers then in identifying gifted students, and what tools do teachers need to help them identify gifted students?

Mr SAUBERN — I think the point that we would like to make is that teachers, schools and systems all have an important role in identifying gifted students. Students in general, teachers, schools and systems have a role in identifying the achievement and ability of all students. Teachers should know a lot about the achievement and the ability of the students in their class, and so should schools. Certainly one element of that is their knowledge about those students in the high-achievement area. The teachers should have the professional ability, the capacity, to be able to identify or help with the identification of those students, not simply for the purpose of identifying them to be taken into a separate stream but because teachers need to teach to students at the level that is relevant for that particular student or group of students. Teachers certainly have a very important role, and that role is enhanced by the addition of objective measurements as well as teacher judgements.

Prof. MASTERS — I think the difficulty with teacher judgements often is that teachers know their own students very well but they are not always in a good position to compare how their students are performing with students at other schools and students of the same age. I think that is where an objective test often has real value in that it allows that comparison to be made. It is not a reflection on teachers; teachers can make very good judgements but they usually have limited access to a range of abilities that we will see in a particular year level.

Ms MILLER — We have received many submissions in regard to this area. It has been suggested that all Victorian students should be screened for giftedness. Do you agree? And if so, at what age should children be assessed? If not, what other mechanisms should be used to ensure that all gifted children are identified?

Prof. MASTERS — There is an assumption that underlies that question, and that is the assumption that there are children who can be identified as gifted and others who are not gifted and so should not be identified as gifted. The principle that underpins the work that we do is that every child needs to be given learning opportunities appropriate to their current needs as learners. They need to be challenged, and that is true of every child. Our approach, certainly our assessments, is not based on the idea that out there somewhere there are gifted children who need to be identified and there is a clear cut-off between gifted children and children who are not gifted. It is more a matter of identifying children who are very advanced in their development and their learning and who would benefit from additional challenges.

Ralph made the comment earlier that our approach is not one that is driven by the notion of giftedness. It is a recognition that there are children who are highly intelligent and very able and who need to be challenged and

need different kinds of learning experiences. Our assessments are primarily designed to ensure a better matching of students to the kinds of learning opportunities that are going to be appropriate for them.

Ms MILLER — Do you think every child should be screened, though?

Prof. MASTERS — I think every child should be assessed in an attempt to establish where they are currently at in their learning and development, and out of that we can draw an inference about the kinds of learning opportunities that are going to be most appropriate for them. That will include identifying children who have significant potential and who are very high achievers.

Ms MILLER — In terms of identifying gifted students from disadvantaged backgrounds, the evidence received by the committee suggests that giftedness occurs in groups but that some children are less likely to be identified as gifted, such as those from indigenous, culturally diverse and low socioeconomic backgrounds. How can we increase identification in these groups, and how do ACER tests ensure that all gifted children are identified regardless of their background?

Prof. MASTERS — I might pass to Ralph in a minute. We are well aware of the international research around the possibility that tests disadvantage particular groups of students. There has been quite a debate in the United States, for example, around the cultural sensitivity of IQ tests and the extent to which traditional kinds of IQ tests are appropriate for different cultural groups. Our approach is to attempt to ensure that the tests that we develop do not include — and you can only do this up to a certain point, obviously, to a certain extent — cultural bias. We go to a lot of trouble to try to ensure that our assessments are equally appropriate for girls and boys and indigenous and non-indigenous students and so on. That is the approach that we take. We recognise that often children's backgrounds and their language backgrounds can be an impediment to identifying exactly where they are at in terms of what they are capable of, and that is why we tend to use a mix of assessments, including sometimes spatial assessments, not just verbal assessments in particular because they can sometimes put particular groups of students at a disadvantage. Ralph, I do not know whether you want to elaborate on that.

Mr SAUBERN — I think the only thing I would add is to say that our assessment is not intended to identify "giftedness". We are describing performance on a construct, so our efforts are put into ensuring that that construct, mathematical ability or numerical ability or musical ability, is measured in a fair and unbiased way. What we are measuring is the performance of the student on what we hope is as fair and unbiased an instrument as we can develop. To that end, for example, knowing that a particular group might have differing levels of English language ability if we are measuring numerical ability, we will work very hard to reduce the language load of an assessment to ensure that we are assessing numerical ability and not something else. That would be an example of trying to achieve fairness, which I guess is something of a construct. If, on the other hand, we are attempting to measure verbal ability in English, we will measure verbal ability in English and not something else. If a student from a non-English-speaking background performs less well on that test, that would not be unusual because that is the construct that we are attempting to measure.

Ms MILLER — You mentioned about testing for non-bias, like girls versus boys. Can you give me an example of one of those tests?

Prof. MASTERS — This is routine across our assessments. We have a fairly rigorous process, a test development process, in which we have all the test items as they are developed subjected to rigorous review. They are reviewed from various perspectives — do the items include content that might not be easily accessed by a particular group of students? This is the question we are asking constantly, and we have people involved who are very experienced in looking at items for possible cultural bias or even geographical bias.

We would be cautious about writing an item that requires knowledge of the ocean, because many of our students live very far from the ocean and may not see it, as an example. In the test development process there is that rigorous review. The other thing we do, though, after we have results, is that we undertake what is called a differential item analysis — a technical term. What we do is look at how individual items are performing for different groups and if a group performs unusually well or unusually poorly on an item, given their performances on the other items, then we will look at that and see if we can see why that might be the case. Then we will remove the item and not include it in the final test that we are developing.

It is a two-pronged approach that we have used in the test development process, and then after we have developed the materials, we use a statistical analysis to see if we can identify items that are biased, if you like, that are favouring one group over another.

Mr SAUBERN — That becomes quite an elaborate process. You can imagine that description, between identifying something that should be measured, developing the content to measure that, reviewing it in a panelling process with many eyes looking at it, trialling it, looking for those biases, then going to a live instrument and again looking at those biases, and feeding that back into the next round of assessment development so that that instrument can become, over time, a better and better instrument. You can see that there is an elaborate process there, more so than just putting together a piece of paper with questions on it.

Ms MILLER — By doing that it actually enables the socioeconomically advantaged and disadvantaged students; is that what I am to understand?

Prof. MASTERS — It tries to identify items that might give a special advantage to a particular group. It is not trying to wash out the fact that, on average, students from lower socioeconomic backgrounds may perform less well. It is not trying to make them perform as well as students from higher socioeconomic backgrounds. Not that we would do it, but if you wrote an item about a haberdashery, for example, then there would be certain groups that would not have access to that word; they would not know what it was. That is the kind of thing we are doing: identifying specific test items — questions — that may be biased.

The CHAIR — Do you need a decent sample size to be able to do that; if you are dealing with a small group, how does it work?

Prof. MASTERS — No, you need a good sample size. You need at least hundreds of respondents, and that is what we do. We have a trial test, and then there is what Ralph called the panelling process, the peer review process, where we go through the materials looking at them from the point of view of their possible cultural bias. Then there is a trial testing phase where we go out and test the items on, usually, several hundred people, if we can get them. Then we analyse how those candidates have performed. If we wanted to look carefully at indigenous versus non-indigenous people, for example, we would need to sample in a way that provided us with adequate sample sizes to do that.

Mr ELASMAR — It has been suggested to the committee that the Coolabah Dynamic Assessment is effective for identifying gifted indigenous students. Are you aware of this tool, and what is your view of its efficiency? Are there any other tools you can recommend for identifying giftedness in the indigenous student?

Mr SAUBERN — We have not been invited to look in detail at any of the results of the various pilots that have been undertaken using the Coolabah Dynamic Assessment technology, so we are not very well informed about that process. From what we can see, Dr Chaffey's approach is not so much a different approach to assessment. As you heard earlier, in his thesis he used a very old assessment called the Raven's progressive matrices, which is a non-language-based, pattern-based assessment which people have taken as a way of, effectively, measuring intelligence — learn-ability, but intelligence, effectively.

He has used that assessment as it is, but interspersed it with a fairly intensive one-on-one coaching process. He has asserted that at the end of that process, retesting children can result in very significantly different results. He brings this back to a notion of self-efficacy. That is, in a straightforward way, some children, because of their social, cultural, personal, psychological background, are not being picked up by assessment because they do not want to be seen as standing out, and the process of coaching helps them understand that that is not a bad thing and that they can be identified using standard assessment at the end of that process more effectively.

As I said, we have not been invited to look at any of that data, and the pilots that I am aware of being conducted in various places have not been completed. That is something to observe. One thing that one might note is that such a one-on-one approach is obviously potentially very labour intensive compared to a system application level of an assessment, but it does probably draw attention to the important role of teachers in classrooms and to the point that if teachers well understand these dynamics, they are well placed to be able to play a role in the development of individual students' attitudes towards achievement, both in their teaching and learning.

Mr ELASMAR — On identifying gifted children at a young age, your submission refers to a version of the higher ability selection test which has recently been developed for use in primary schools. Which schools are

using this test, and what special considerations need to be taken into account in developing and administering the test for preschool and primary school students?

Mr SAUBERN — There are a number of Victorian schools using ACER tests at primary-age levels. As far as we understand, most of these schools are using our assessments at middle and upper primary rather than at lower primary. One thing to note is that the assessment of students at a very low level in schools, in the very early years — preschool, kindergarten, prep level — is different to the testing of students at middle primary. At middle primary students are familiar with NAPLAN-style tests, multiple-choice questions, some open-ended questions and so on. This is a very different prospect from testing students at a very early age.

At a very early age we tend to use different forms of assessment which are more supportive of individual students, which have lower requirements in terms of reading load and are often administered one on one by the teacher in some kind of structured interview arrangement. By the end of year 2 many students have the literacy and other abilities to be able to undertake a standardised assessment. Does that answer your question?

Mr ELASMAR — That was good, but can you tell me more about which schools are using this?

Prof. MASTERS — Which use HAST?

Mr SAUBERN — I do not have a list of the exact schools using ACER instruments, but we have identified about 20 schools that are using scholarship testing services at about middle primary, grade 4, level. These are, by and large, independent schools in Victoria. About 20 schools are using the published tests that I mentioned earlier, the ACER general ability test and the middle years ability test, at grade 4 or grade 5. There are at least five schools that are using the HAST assessment service. Three of them are Catholic schools and two are government schools. They are using them at the middle to upper primary level.

Mr ELASMAR — Is it easier for you to provide the committee with a list of the schools that are using this test?

Mr SAUBERN — I imagine it would be easy for us to send a list, but I am not sure that those particular schools necessarily would want their school names published in an open forum. We could ask them if they wanted to be; but we can certainly identify them for the committee.

Mr ELASMAR — Thank you.

Mr SAUBERN — I do not know, Geoff, if you want to add anything on our tests of very young students?

Prof. MASTERS — No, I think you have covered it. Clearly it is a different challenge. Often we will do things like read to the children. Certainly you are talking about one on one. It is very often observations of children at that level, including observations that these sorts of assessments can get into, looking at psychomotor skills as well: observations of their ability to perform simple psychomotor tasks.

Mr CRISP — I have a question on online testing. Does ACER administer online tests, and what are the advantages and disadvantages of this?

Prof. MASTERS — We certainly do.

Mr SAUBERN — ACER has developed tests for use on online platforms or, I should say more generally, on computer-based platforms — some of them are online and some of them are not — for many years and for particular purposes. There certainly are some advantages and some disadvantages. Usually the advantage is not the one that people think is the advantage, which is that it will be cheaper and easier. Our experience is that using computer-based assessments is rarely cheaper or easier, but there are advantages. For example, computer-based testing has the ability or the potential to deliver a very fast turnaround of results for individual students and in general for reporting.

Electronic test instruments can make use of more sophisticated stimulus materials and more contextualised stimulus materials, such as audiovisual content and increasingly potentially more sophisticated interactive test items, where we are actually asking students to do more than simply answer a question; we are asking them to effectively either explore a data set or some particular context, make some judgements and show us what the

judgements they are making are as they are making them. It has the potential to use more sophisticated interactive materials.

The move to touch screen computing platforms, which is something that we looked at several years ago but it has now obviously become more possible with different kinds of technology that has become very widely available, certainly may offer improved ability to assess younger students and students with lower computing skills.

Finally, there is the potential to use computer-based systems for computer-adaptive models of testing which have the potential in some contexts to shorten the length of assessments — that is, the amount of questions that need to be asked to arrive at a valid and reliable answer as to the student's performance on a particular test.

Prof. MASTERS — Or maybe even more important than shortening the test is targeting the test on where the student is. In a computer-adaptive test what one is doing is getting the computer to select test questions from a bank of questions, and the computer monitors how the candidate is performing and selects test questions appropriate to the candidate's performance. You can have different candidates being administered questions of different levels of difficulty based on their ability level.

Mr SAUBERN — Whereas in a paper-based format those judgements are usually made by the assessors, the teachers or the test developers before the child begins the assessment.

Prof. MASTERS — And every child takes exactly the same set of questions, whereas you can get more precise measures of the abilities of students by better targeting their ability levels and giving questions that are more appropriate to their ability.

Mr SAUBERN — Some of the disadvantages include the fact that certainly not all students are equally confident with the electronic media. That may be an issue that is disappearing at a rate of knots for most students, but still it is relevant for many students. What is probably more relevant is that moving to computer-based models highlights the difficulties of standardised uses across the system, for example. The capacity to do something in a standard way across the system is difficult given the dependence on local conditions. For example, when testing in schools we must consider what are the facilities and capacities within a local school to deliver that assessment on a computer-based platform and what support is available to that local school in a remote regional area or in a small school compared with a large school, for example.

There is some unevenness that we need to look at. There are also some new and different kinds of issues that online or computer-based testing platforms introduce. We are always looking at issues around ensuring that there is a standardised way that a test is delivered. Going to an online or computer-based platform introduces a whole lot of extra things. For example, we need to consider whether a particular computer has particular things on it that a student can use that another student cannot, such as spellcheckers and access to the internet. These are just practical issues that we need to manage.

Finally, in our experience, online or computer-based testing platforms tend to be more expensive to develop and deliver, for a variety of very practical reasons — it is at least as expensive as any other method. Perhaps I will say one more thing. We should note that when we are talking about computer-based or online testing we are often talking about taking an existing test on paper and putting it online. I just sound a note of caution there: taking a test in one administration mode, on a piece of paper, and putting it into a computer on a screen which may come in a variety of sizes is not the same test. Many elements of it are the same but one cannot assume that once it is put on a computer screen or into some other electronic medium the test will perform exactly as it performs on paper in relation to students. In order to be confident about the continuity of information you are getting in moving from paper-based to computer-based testing you need to generate data and analyse that to understand what is happening with shifts of delivery modes.

The CHAIR — I just have a few questions. There are a number of examples where students may have been coached to get a certain outcome in testing; there has been a lot of talk about NAPLAN tests and preparing students to get an end result. How do you ensure that that does not occur in this area, particularly in those middle years where students are looking at going off into specific select schools and what have you in terms of some of the products you are putting out there?

Prof. MASTERS — We do not pretend that coaching will have no effect at all. Some people make claims for their tests that it is impossible to improve students' results through any kind of preparatory activities. We would not make that claim. Certainly just familiarising students with a test often helps. Particularly with students who have had no or very little contact with tests, just teaching them some basic test-taking strategies is often quite helpful. If you are talking about curriculum-based tests, then it is a slightly different matter, but in these kinds of tests the way that we address that issue is by really trying to assess higher order thinking, higher order skills, reasoning skills and things that are quite different from teaching facts and procedures that you can memorise and then go in and apply. That is what we are trying to do through these tests: we are trying to capture students' critical thinking and reasoning abilities.

Mr SAUBERN — Obviously it would be difficult to study the effect of coaching in a systematic way. What we have done is some internal analysis which looked at students who have re-sat examinations and looked at improvements. That research, which we did for our own benefit, shows that improvements that occur on these higher order thinking tests tend to be in that middle range. With assessments which are really designed to identify the top range of academic potential or achievement those students are not really moving much by re-sitting the test or being coached in particular ways. But you can see movements around that middle range. As Geoff said, that may be more to do with familiarisation with test conditions and ways that you might move through the test in an effective way. But if you are not in that top band of these tests, it appears that it is not that easy to coach someone into the top band.

The CHAIR — Do Victorian selective entry schools use ACER instruments?

Mr SAUBERN — Not currently. They have in the past, but not currently.

The CHAIR — Is that something that you are looking at doing into the future?

Prof. MASTERS — We are available to do that, and we have developed tests that have been used by Melbourne High and MacRob in the past. They are not currently using ACER tests, but rather purpose-built ones.

Mr SAUBERN — You may not be aware but we have currently tendered for the provision of selection assessments for the four selective entry high schools. That is currently under review now by the VCAA and the department

The CHAIR — Could you comment on the use of specialised assessment tools to identify giftedness in particular groups — for example, indigenous students — as opposed to using general tools that pick up giftedness in all groups including indigenous students?

Prof. MASTERS — I think we started to address that a little bit earlier in that our approach is to try to develop assessment instruments that are going to be appropriate across the board by minimising features of the tests that are likely to put particular groups at a disadvantage. That is our general approach. Beyond that I guess what we are trying to do is to use breadth as well, so we are looking at not just verbal ability or quantitative reasoning ability but at spatial ability, where there is no language as such involved in the test. It is a two-pronged approach. We are doing what we can to ensure that our tests are what is sometimes called 'culture fair' or fair to all groups. By using a breadth of assessments addressing a range of constructs we are trying to ensure that the tests are useful for identifying students of high ability in different demographic groups.

The CHAIR — We would like to thank you for coming along today and for your contribution both with the submission that you have given us and your appearance. If there is anything you feel that we have not covered on which you want to make a short 1 or 2-minute additional contribution, we are happy to take that now.

Mr SAUBERN — Geoff has recently written one of what we call in ACER an occasional essay which broadly covers the issues that he began to describe earlier in our evidence around expectations of individual students. This is not particularly about giftedness or about students in that higher ability range but more generally talking around the issues of identifying individual students and where they are at and expecting all students to be able to achieve at a high level, and the dangers of the opposite, which is expecting students to achieve the way their demographic suggests they should achieve. You often succeed if you do make that assumption. This essay looks at what we should be doing in terms of the opposite. We should be expecting all students to achieve irrespective of their demographic, irrespective of the description you might make of them as

indigenous, low SES, girl, boy, whatever it might be. It is a short essay, and I think it is something that the committee might like to look at as additional evidence in its thinking around this area of gifted education.

The CHAIR — We will certainly include that as part of the evidence.

Prof. MASTERS — Thank you. I think the only thing that I would add is that what has happened over recent decades is that our understanding of learning itself has considerably increased, and it has been through developments in neuroscience and cognitive science. There was a time 100 years ago or less when we believed that people's capacity to learn was markedly different from one person to another and that those things were pretty fixed. I think one of the things that we are starting to understand through neuroscience and other research is that pretty much every individual is capable of learning if they are motivated and they are given appropriate learning opportunities.

Our approach has been to move away from labelling people saying, 'You are talented, you are gifted', or, 'You're not', towards saying that everybody is at some point in their learning. They are at different points. They are progressing at different rates, but they are all capable of making progress if motivated and provided with appropriate learning opportunities. That is not to say that there are not some students who are at very high levels of achievement and who have high levels of ability. We need to recognise those students and provide them with appropriate opportunity learning opportunities as well.

The CHAIR — Perfect. Thank you very much for coming.

Witnesses withdrew.