Using Online Assessment to Inform Teaching and Learning in Primary and Secondary Schools

Paper presented at the 2005 ACER Research Conference on Monday 8 August 2005 by Professor Jim Tognolini of the Australian Council for Educational Research (ACER).

Introduction

I intend to present this paper in four distinct, but interrelated parts. In the first part I will present a contextual framework for developing assessments and meaningfully interpreting and using feedback to inform teaching and learning; in the second part I will introduce two closely related applications of one of ACER's online testing programs, focusing upon how the programs provide reports that are consistent with the theory that underpins the construction of the tests; the third part will focus upon how the information might be used to inform teaching and learning; and, the final part will provide some ideas regarding future research and development in the area of assessment and use of data.

The first and one of the most important steps in any assessment exercise is that the tests and tasks that we construct are reliable and produce valid results. In addition we are also interested in ensuring that the results that are generated from the assessment activities that we conduct have meaning in the context of student learning and the performance standards that benchmark student learning.

When we start with this basic principle that our tests must produce information that can be referenced to student learning or standards, then we effectively shift the focus in assessment from notions of rank ordering students to those of monitoring growth or progress and MEASUREMENT. Of course the two are not mutually exclusive as when monitoring growth it is still possible to rank order students.

PART 1: Constructing and Using Measurement Scales

Constructing measurement scales in the physical sciences

In order to monitor growth, or measure something we must first develop the measurement scale. This is true for measuring attributes such as height, weight, and heat or mathematics achievement. Once you have developed the scale, you can then use it to measure the thing you want to measure (that is, you can find out HOW MUCH [a measurement question] of the property - height, weight, heat or mathematics knowledge and skill – the object has).

It is quite interesting to examine some of the history of measurement to see how we (the human race) have developed different measures over time. In the case of the measurement of length history shows that the human has been the "the measure" and the "measurer".

In an article in the Journal of Outcome Measurement (1998) Mark Stone highlighted how "man" had accommodated the need to measure by referencing the unit of measurement. That is, he showed how "man" had to develop measurement units in order to be able to measure. He illustrated the process using an example of how the "foot" was derived to be the unit of measurement for length, or height. He said,

> An etching made in the 16th century shows 16 men standing in a line outside the entrance to a church. Each man has placed his left foot heel-to-toe with another. Observing the scene are three monitors. The accompanying narrative to the etching describes the method being employed. The narration indicates that monitors were stationed at the door of the church following the weekly service to

> "Bid the sixteen men to stop, tall ones and short ones, as they happened to come out … their left feet one behind the other … gives …the right and lowful rood … [the sixteenth part of which is] … the right and lawful foot" (Nicholson, 1912, page 47)
>
> (Stone, 1998, page 27)

It is interesting to note how this simple, ingenious method of defining a unit of measurement has taken account of the variation that exists in human physiology. Early metrology recognised variation and devised ways to address it (Stone, 1998, page 27).

Once the basic unit has been defined, other units are derived; scales calibrated in terms of the unit of measurement are then constructed to enable measurement to occur. The object to be measured is then referenced to the scale to obtain an estimate of how much of the property the object has. Once the measurement has taken place, decisions regarding the adequacy of the current situation and what to do next, evolve. In other words the results of measurement provide evidence that is used to inform further action.

Figure 1 shows how a ruler can be used to measure height or length.

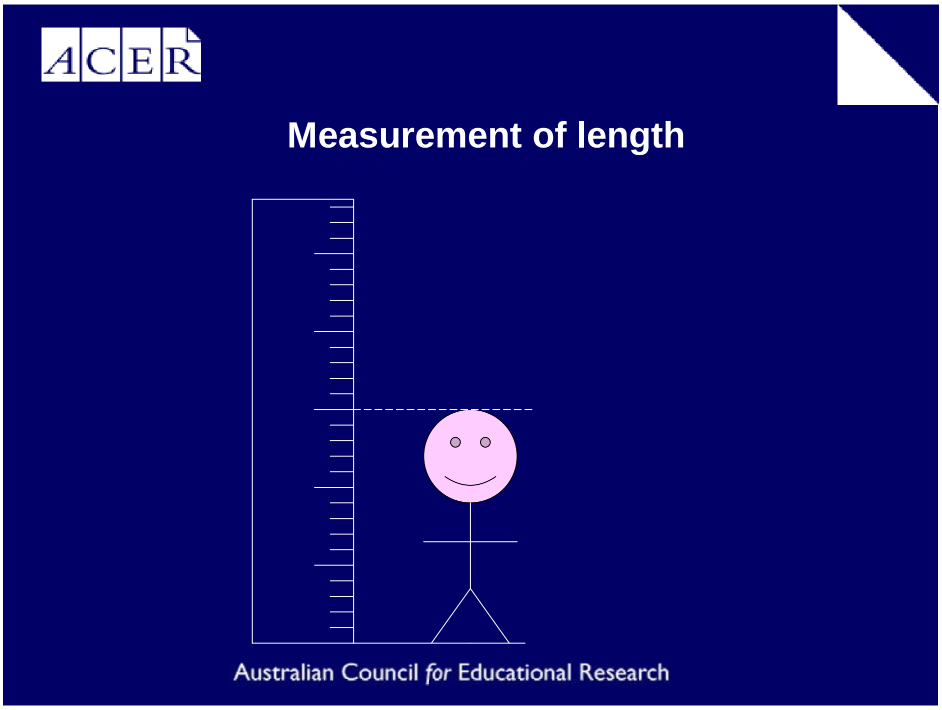

Figure 1: Measurement of length or height

When we are asked to measure properties or constructs (such as, mathematics knowledge and skill or leadership potential) in the social sciences, we are being asked to solve a measurement question. We can get some ideas of how to do this from the measurement experiences in the physical sciences.

Lots of people say that it is much easier to construct measurement scales and measure properties in the physical sciences because you can actually see height and weight. However, this is not always the case. For example, we cannot see heat; yet we can measure it. This is done by making a link between heat and the manifestation of heat (increased movement of molecules causing expansion of mercury up a capillary tube) that is calibrated in degrees Centigrade or Fahrenheit to form a temperature scale (thermometer). Once the scale has been constructed measurement is possible using the scale.

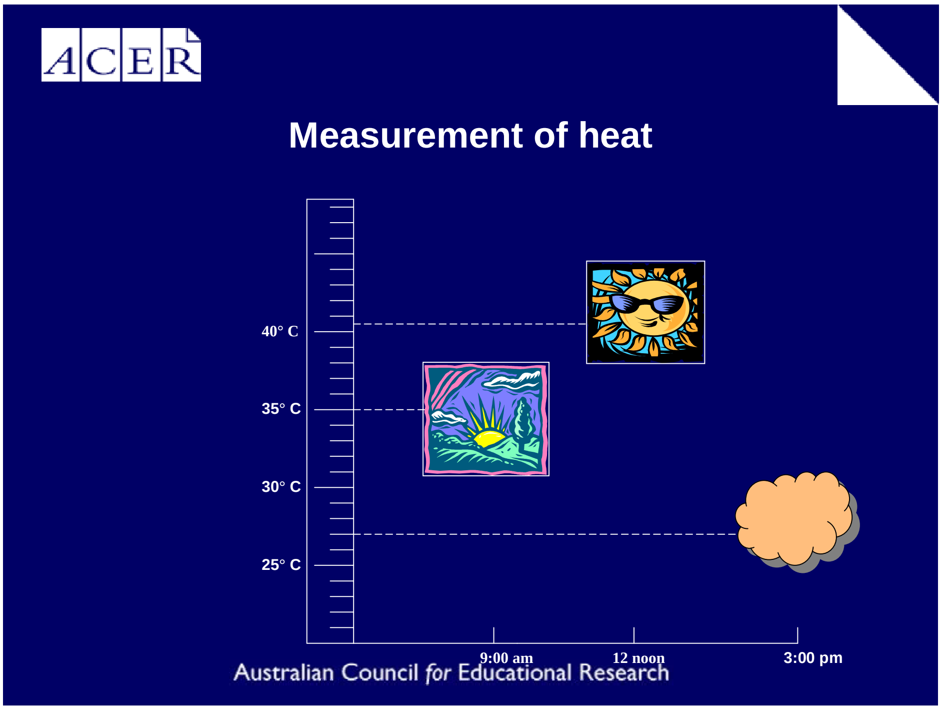

Figure 2: Measurement of heat

One of the advantages of measurement in the physical sciences is that it is generally objective and independent of the location in which it is used. This means, for example, that temperature can be interpreted without reference to the particular thermometer used for the measurement. It also means that different people can use it in different locations and on different occasions and the measures are comparable.

A second advantage is that measurement is always referenced directly to the scale and as a consequence the results can be more meaningfully interpreted in relation to what is measured by the scale. There is a direct relationship between the amount of the "thing" being assessed and the scale. For example, the further the mercury is along the scale, the more heat energy there is.

These features are highly valued in measurement in the physical sciences and would be just as highly valued (although rarely present) when measuring in the social sciences.

The challenge for the social scientists (which includes teachers in this case) is to develop scales that enable measurement of the constructs that they need to measure. While the challenge is demanding the principles underpinning the development of measurement scales are the same and the advantages that accrue from using such scales for measurement in the social sciences make the efforts well worthwhile.

Constructing measurement scales in the social sciences

If the same principles are applied to constructs that we are interested in as teachers, educators and social science researchers then the first stage in measuring student achievement or potential must be the construction of an appropriate measurement scale.

In most, if not all, the constructs or subjects that are of interest in the social sciences, it is not possible to see the actual construct or observe directly HOW MUCH of the construct a student has – just like heat. So it is necessary to create ways in which the students can provide explicit evidence about how MUCH mathematics knowledge and skill they have. When test items or assessment tasks are constructed they provide the students with the opportunity to demonstrate what it is they know and can do in the subject. The items and tasks are the manifestations of the construct and as such define it.

Figure 3 shows how a measurement scale might be conceptualised in much the same way as measurement scales are in the physical sciences.

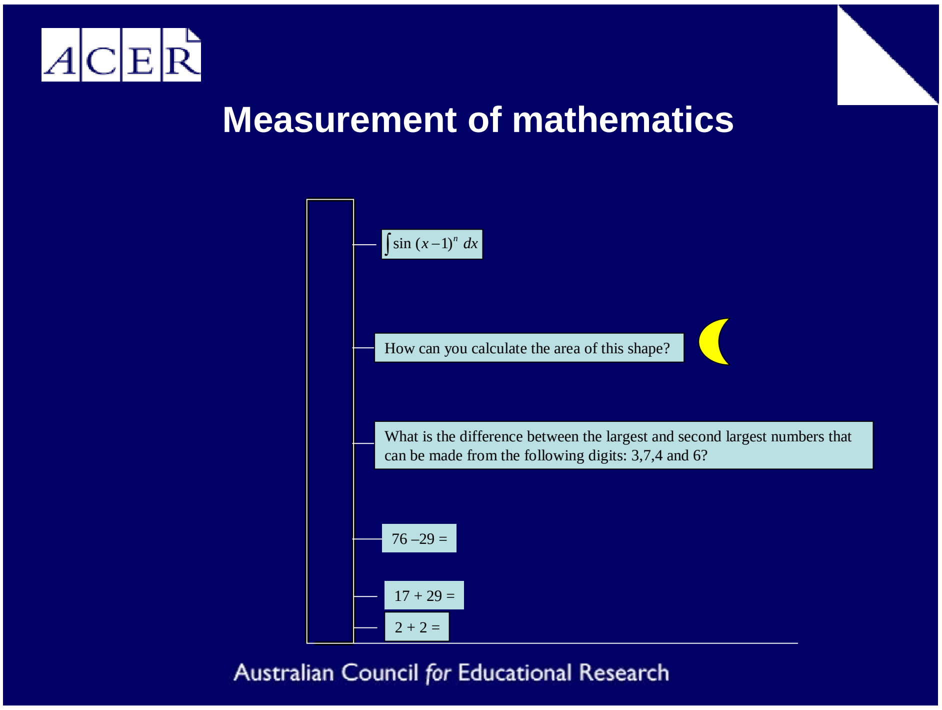

Figure 3: Measurement of mathematics

In this case the items themselves are the calibrations on the scale and the location of the items is determined by HOW MUCH mathematics knowledge and skill is required to correctly answer the question imbedded in the item. Hard items that challenge the most able persons are located towards the upper end of the line. Conversely, those items that are relatively easy, and can be answered by students with small amounts (and large amounts) of mathematics knowledge and skill are located towards the bottom end of the line.

In order to develop such scales teachers, test constructors and examiners are challenged by the need to define items that enable students at different stages along the scale to demonstrate that they have enough of the property to correctly answer the item. They must ensure that the items are assessing the outcomes for the particular location on the scale. They must also ensure that when the items are written, the ones that are intended to be located further to the top on the scale are, in fact, harder then those that are located to the left on the line. In addition, they must ensure that the reason that the items are harder is a function of the construct that is being measured and not a function of some extraneous feature (validity).

As a consequence of developing a scale like this the items towards the top are the ones that challenge the most able students and indicate GROWTH of the student.

Figure 4 summarises the challenges facing test constructors when developing scales for measuring student performance and GROWTH.

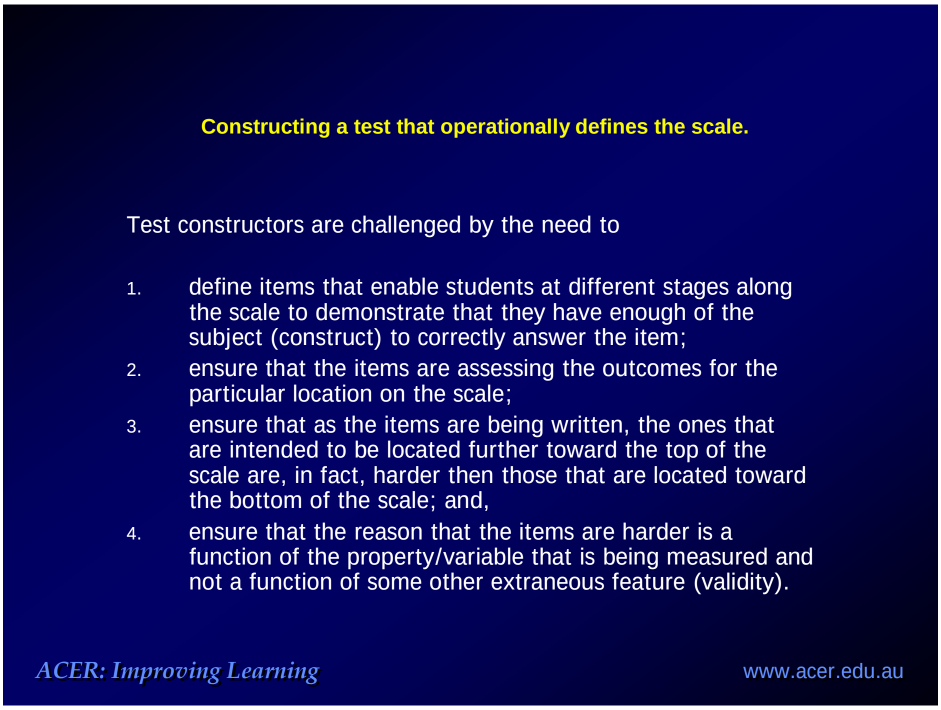

Figure 4: Developing the test

Once the items have been written then they are given to the students in order to determine just how difficult they are. The items that most students get correct are the easiest; those that few get correct are the most difficult. These item difficulties are then converted into calibrations on a scale using a measurement model. The easiest items will be located towards the bottom of the scale and the most difficult will be located towards the top.

Figure 5 shows the items for a mathematics scale where the items assess outcomes in Year 6.

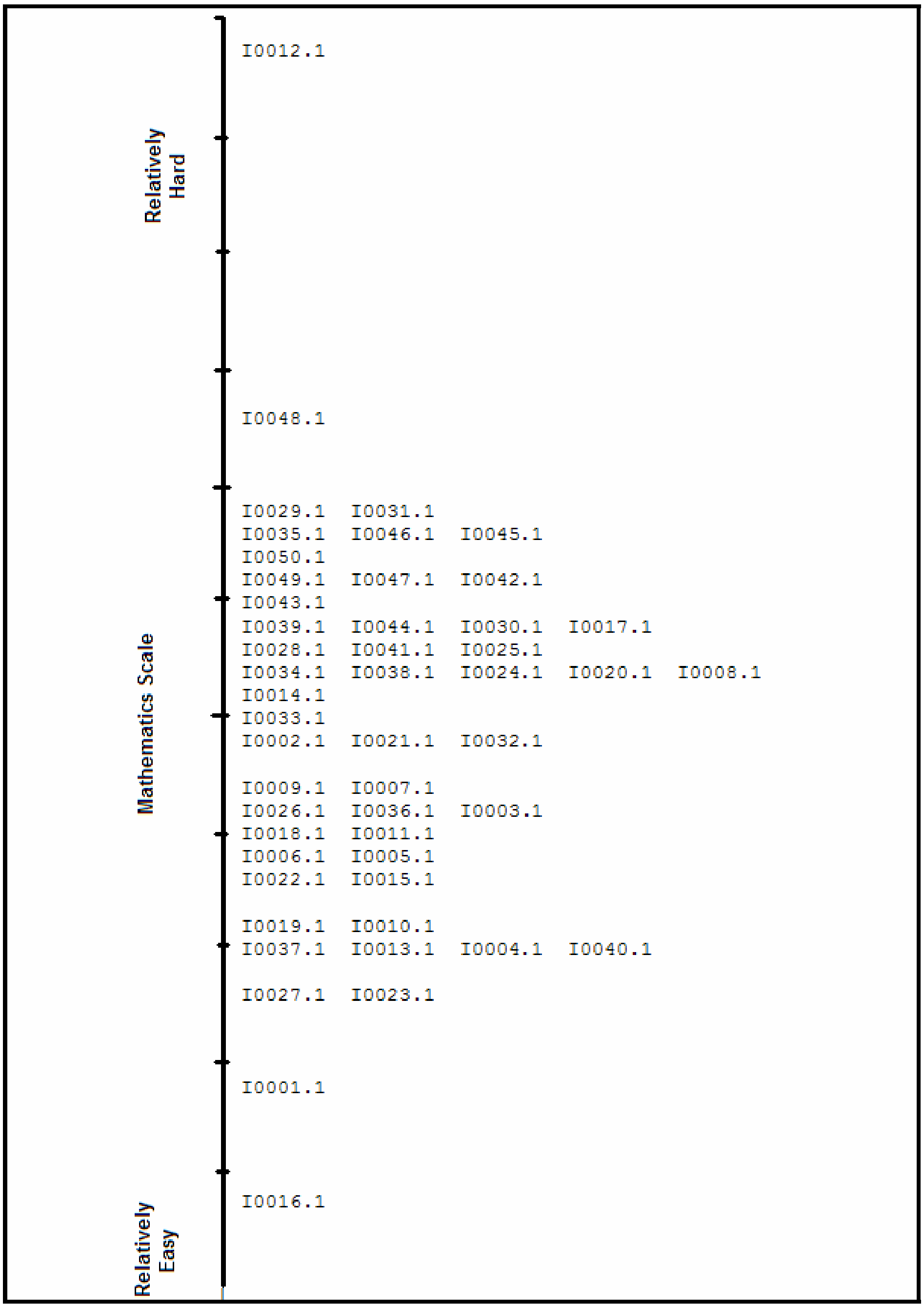

Figure 5: Calibrated mathematics scale for Year 6

It can be seen in this particular example, the easiest item in the test was Item Number 16; the next easiest was Item Number 1; and the hardest was Item Number 12.

Figure 6 shows Item Number 16, the easiest item on the test.

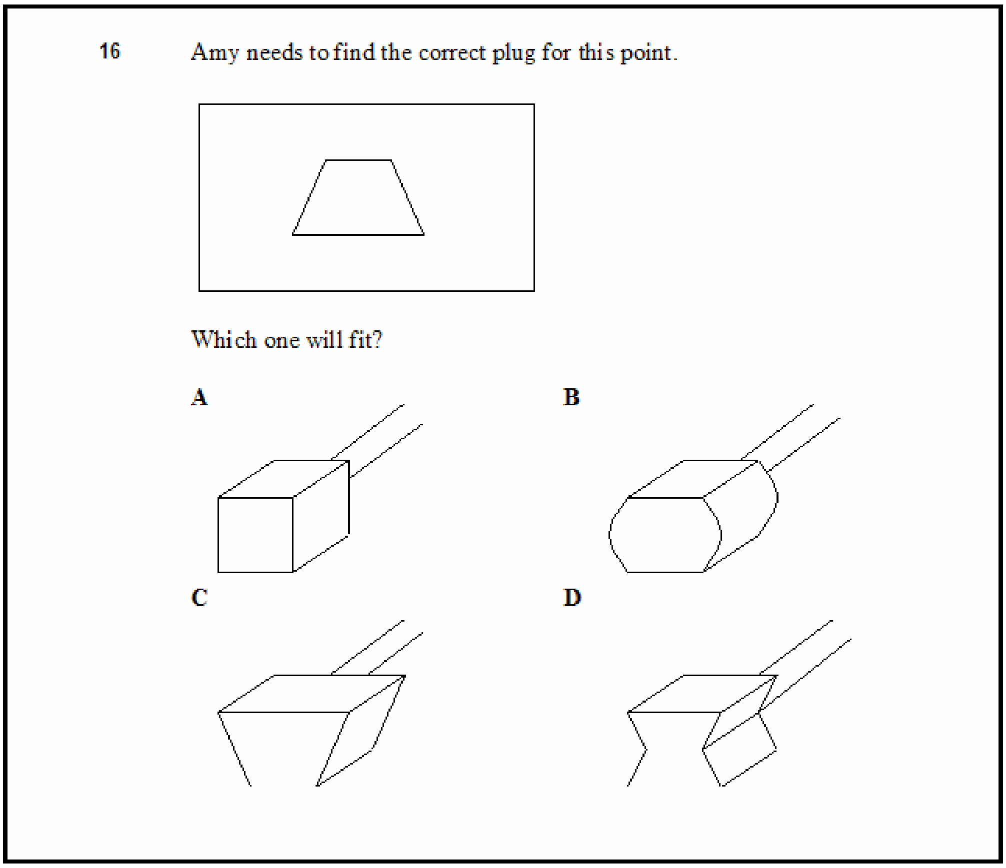

Figure 6: The easiest item on the Year 6 Mathematics Test

That is the item that required most mathematics knowledge and skill to answer it was Item Number 12.

Figure 7 shows Item Number 12 which is the hardest item on the test.

12 $? \times 7 = 378$

Which number sentence could be used to find the missing number?

A $7 \div 378 =$ **B** $378 \times 7 =$
C $7 + 378 =$ **D** $378 \div 7 =$

Figure 7: The hardest item on the Year 6 Mathematics Test

Now that the mathematics scale has been constructed, student scores can be converted into measures so that they can be referenced to the scale.

Converting student scores to measures on the scale

Once the tests have been marked students' scores are calculated. These scores are then converted into measures on the same scale using the same measurement model. Once calculated they can be located on the scale as shown in Figure 8.

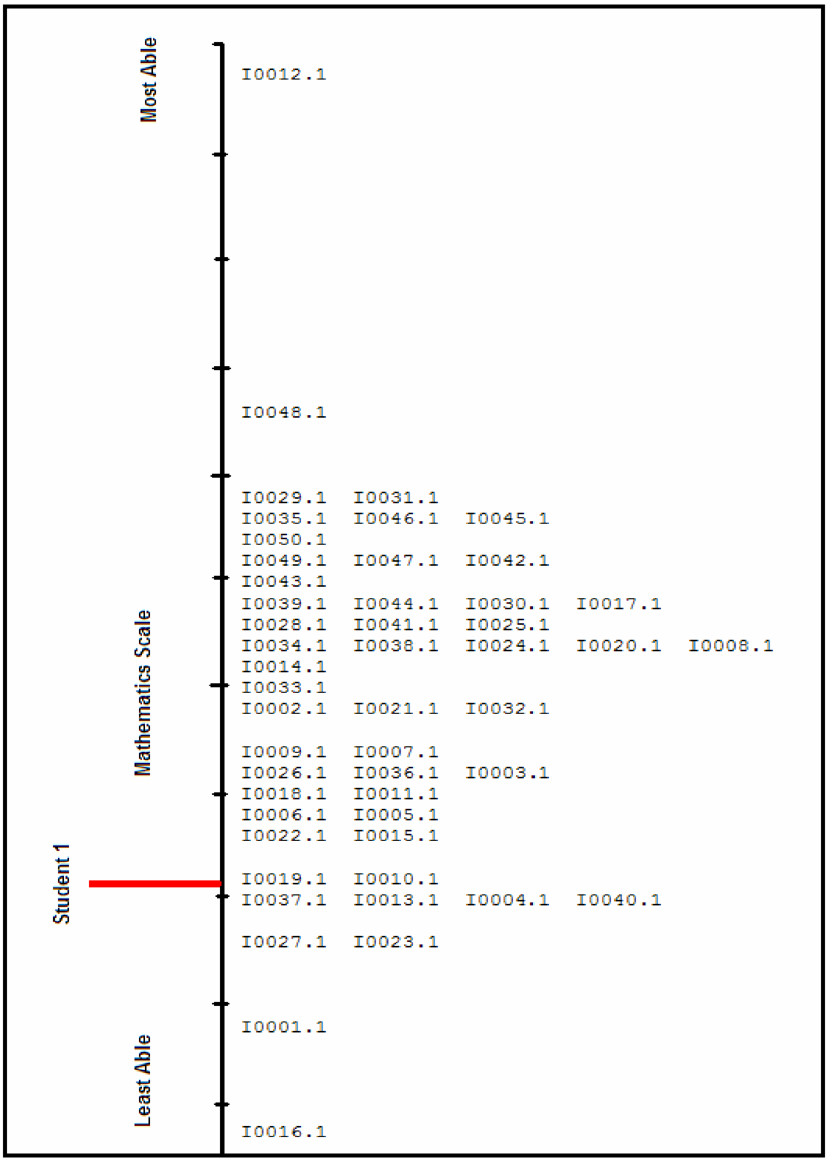

Figure 8: Student located on mathematics scale

It can be seen from Figure 9 where Student 1 is located on the mathematics scale. The advantage of being able to locate students and items on the same scale is that the performance of the students can be directly referenced to the items, achievement progress maps (that we value) and standards. More importantly the whole process of constructing a scale and representing results in this way adds a new dimension to the type of feedback that emerges and the way that assessment data can be incorporated into teaching and learning from such a system.

For example, it can be seen from Figure 9 that Student 1 has enough mathematics knowledge and skill to generally be able to do items lower on the scale (e.g. item numbers 16, 1, 27 and 23) but not be a stage where they can more often than not do items higher on the scale (e.g. item numbers 12, 48, 5 and 6). Not surprisingly students do not always follow this deterministic model and they get some of the easier items wrong and the harder items correct. In these response patterns therefore there is feedback information for the students and the teachers, which if used effectively, can contribute to improved teaching and learning.

The Australian Council for Educational Research (ACER) values this way of collecting information and presenting the results of its testing programs. It has used the measurement principles outlined above in one of its latest assessment programs that also takes advantage of the online environment to deliver tests to students and provide immediate feedback to students, parent and teachers.

The remainder of this presentation examines the program (called OPI [school version] and *iAchieve at home* [home version]) and shows how the information that emerges from such a program can be used inform learning and teaching.

PART 2: Two Applications of One of ACER's Online Assessment Programs

An overview of the Online Placement Instruments (OPI) and *iAchieve at home* Programs

The OPI and *iAchieve at home* offer online multiple choice tests in Mathematics and English for students from Years 3 to 10. One of the driving factors that led to this innovation was the desire to bring together curriculum, teaching, assessment and learning into a package in an online environment that would benefit children. A second desire was to better focus learning opportunities in the home so that a) parents could become more involved in supporting the learning process; and b) children can become more empowered in learning.

Key features of the program

Key features of the program include:

- tests being taken at school or at home in a class.
- items in the tests assessing the key generic skills that underpin the mathematics and English areas of learning. They do not just test content learnt within school.
- tests being tailored to the individual ability of each and every child.
- different tests being available at the beginning and the end of the year to monitor students' progress over the year and across years.
- comprehensive, state-of-the-art feedback being provided instantaneously. Results are also referenced to national and international benchmarks.
- helpful hints being provided on a way forward for students and for concerned parents on those areas that might be proving a problem.

The program aims to support learning in the home and within schools. It also aims to provide schools with a powerful assessment system that enables them to monitor the performance of students through areas of learning. In addition it aims to provide parents with a mechanism with which they can become more effectively involved in supporting student learning in the home.

PART 3: Using Feedback to Inform Learning

I must admit that I am not a strong advocate of making a distinction between assessment for learning and assessment of learning. In my opinion assessment is the collection of information for a purpose. I would not see a case, even in high stakes examinations where the assessment information that is collected should not be used to inform teaching and learning. Consequently, all tests should provide data at different levels of generality to inform the teaching and learning process. More importantly, I believe that students, teachers and parents should be taught how to interpret data themselves. As I have talked about feedback and the use of data around the world, teachers generally say to me “How am I expected to use the reports to give feedback to 40, 50 or 60 students in my class?” I am always perplexed by this question. Surely rather than go through each of the reports with the students it is more appropriate and efficient to teach the students themselves to analyse their own performance and tell you, the teacher, what they did wrong on those items they might have been expected to get right and how they might do in future to ensure that they get such items correct. In this way the learner is involved in the process. However, in order to ensure that this process works effectively the students need to be provided with a context or reference frame within which they can interpret their performance.

It is this context that the feedback for the online testing programs has been developed.

Features of the feedback

There are four main reports that accompany the OPI program (the iAchieve at home program only has three because it applies to only the individual. As such, there is no class-level or school-level report).

The first report is an individual report that shows for each student the questions he or she got correct and those she or he got wrong. One of the advantages of online reporting is that students have access to their results as soon as they finish and submit the tests. There is no doubt that this is when they are at their most receptive and the feedback is most likely to influence their future learning. (See the “Individual Report for Students”). I am reminded here of when I taught Year 12 mathematics and physics. As the students walked out of their examinations I had the solutions available for them. While they didn’t really embrace the practice they all very quickly delved into the solutions to look at the questions that they had grappled with during the course of the actual examination. They then talked about this with their fellow students or challenged me. This was very effective learning.

The second report is designed to tailor the results so that they and the individual item results can be more meaningfully interpreted. In this report the results are referenced to what it is students know and can do and what it is they have to be able to do next to move through an area of learning in a key subject area (See the “Where You’re At Report”).

Thirdly, the measurement model also means that students, parents and teachers can track individual progress across years (See Growth Report).

Fourthly, at the school level reports are produced that provide the teacher with indicators on an item-by-item basis with areas of potential strength and weakness for the class. Such tables can be used by a teacher to evaluate general areas of class weakness (particularly if the items are clustered into strands); and, diagnose student and class weakness.

Care must be taken when interpreting such data to make sure that the teacher and the students do not over interpret the results. That is, the individual item results should be taken as just one indicator of the performance of the students on the skill. It should be taken interpreted in the context of other information obtained from other sources. For example, if the results show that the class is not performing well on problem solving in mathematics in the measurement strand, then it would be appropriate for the teacher to ask if there is any other evidence that would suggest this to be the case. If not, then more information might be sought before treatments are introduced to overcome the perceived problem.

The most meaningful class or school level report is one that presents the data in a Guttman Scale. This scale takes advantage of the measurement theory outlined earlier. More specifically it presents the data in a table in which the items are ordered from easiest to hardest (based upon the item difficulties obtained from across the country [in this case] as part of the calibration exercise for OPI and iAchieve at home) and the students from most able to least able. Anomalous performance across the set of items becomes apparent for the school or class by the pattern of item difficulties for the school. Individual student patterns can also be identified and interpreted meaningfully.

One of the particular features of the feedback associated with the online programs is that students can access individual item feedback designed to help them understand what it is they might have done wrong on the item.

PART 4: Future Research and Development

Once again as I venture to countries like Indonesia, India and the Middle East the key question that teachers and principals have is now that my students have shown that they are poor at this aspect of their work, what can we do about it? Online technology enables this question to be answered. The individual item feedback is very preliminary at this stage. There needs to be more research and development work done on linking the distractors of the items to incorrect strategies and then building up learning strategies that help the student move from the incorrect strategy to the correct one. Worksheets can be added so that students can then practice the skill that they have just acquired.

Lectures, teaching activities, books and articles can also be added to the database of activities available online.

The items for this program have been written to accommodate school and home situations that have minimum bandwidth. There needs to be more research and development work done on developing item types that are more creative and take cognisance of the many software (multi media) programs that are available at this time. At the same time technical research has to be done on how to minimise the requirements for bandwidth for such items.

Conclusion

The purposes of this paper were to

1. introduce some basic ideas of measurement which might provide a meaningful context for teachers to generate assessments and for students, teachers and parents to interpret results in such a way that they can positively inform the teaching and learning process; and,
2. show how a particular assessment program has been constructed in accord with the measurement theory, as a consequence of which, the assessment information that emerges from the program can be used to inform teaching and learning.

The beneficiaries of teachers and students being able to meaningfully interpret and use of data in this way is undoubtedly the students.