SMART Assessment for Learning

Kaye Stacey, Beth Price, Vicki Steinle,
Helen Chick, Eugene Gvozdenko

University of Melbourne, Australia

"Specific Mathematics Assessments that Reveal Thinking," which we abbreviate to "smart tests," provide teachers with a quick and easy way to conduct assessment for learning. Using the internet, students in Years 7, 8, and 9 undertake a short test that is focused strongly on a topic selected by their teacher. Students' stages of development are diagnosed, and sent to the teacher immediately. Where available, on-line teaching resources are linked to each diagnosis, to guide teachers in moving students to the next stage. Many smart tests are now being trialled in schools and their impact on students' and teachers' learning is being evaluated. Design issues are discussed.

Models for diagnostic assessment

According to www.moviequotes.com, the film "The Matrix" includes the line "Never send a man to do a machine's job". This paper addresses one of the aspects of teaching that may better be seen as a machine's job than the job of a human teacher: the diagnostic component of 'assessment for learning'. Our "smart tests" research and development project is designing computerized assessment to investigate how the detailed work of assessment for learning might be passed from human to machine and to examine its effect on both student and teacher learning.

The first section of this paper describes assessment for learning, and contrasts current government recommendations with our approach. The middle sections describe the smart test system through two examples. Finally we discuss design principles involved and future directions for research and development.

Assessment for learning (or equivalently formative assessment) occurs when teachers use inferences about students' progress to inform their teaching; especially teaching close in time to that assessment. Assessment for learning is contrasted with assessment of learning (what have students learned to this point of time) and assessment as learning (when students' reflect on information about their progress to understand how to do better). Assessment for learning is not so much concerned with how much students know, but with what they know. It is a matter of intention rather than a particular style of assessment instrument. Our educational authority (Department of Education and Early Childhood Development, Victoria (DEECD)) has enthusiastically adopted the conclusions of Black and Wiliam (1998) that assessment for learning is one of the most effective ways available to raise learning outcomes. DEECD has invested in several tools to help teachers understand their own students' stages of development better. The main tools are two individual interviews which take 30 – 60 minutes per child (HREF3). One is for students in the first years of school and the other, on fractions and decimals, is for students in middle grades. Conducting these interviews is very rewarding for teachers, and greatly adds to the insights into learning that support their instructional planning. However, there are substantial difficulties in finding time to conduct the interviews and as a consequence they tend to be conducted only a couple of times in a child's school career. Another key site of diagnostic assessment is at the beginning of Year 7, the first year of secondary school. Many schools use a bank of written tests (sometimes from a commercial company) to identify the background knowledge of their incoming students and they use this to place students in appropriate groups for their future mathematics learning. Again, this testing tends to happen on entry and not again. Beyond this, assessment for learning is sporadic, mainly depending on the individual teachers' wishes.

Standing in stark contrast with the delivery of the above tools, is the Decimal Comparison Test (DCT), which Stacey and Steinle developed as part of a significant research project. This is available on the CD 'Teaching and Learning about Decimals' (Steinle, Stacey, & Chambers, 2006). The DCT can be administered to a whole class in 5 – 10 minutes and assessed within another half an hour by the teacher (by-hand) or instantaneously on line. The results provide information on students' understanding of decimal notation. The DCT reliably identifies students who hold any of a dozen misconceptions about decimal numbers. The test is not a perfect diagnostic tool, but its limitations are well researched: the most important one is that some students who do well still hold misconceptions but have masked them by accurate rule following (Steinle, 2004). The 'Decimals CD' (Steinle et al, 2006) provides many teaching activities specifically targeting the misconceptions that are revealed. The test can easily be administered again a few weeks later to monitor progress.

The DCT is extremely popular with teachers and we get regular reports of its usefulness. Just today we heard of a school which had recently used it for the first time with all the Year 7 students. Teachers worked together to classify their students. They were amazed to see how a particular misconception was clustered in one class (misconceptions are often due to teaching) and also to see the prevalence of misunderstandings across all classes. At the same time, they were energised by the comparison with the supplied research data that showed that this school was not unusual in the extent of misunderstandings, and that a little attention to the topic could make a big difference. The mathematics coordinator reported genuine enthusiasm amongst the teachers as they selected teaching activities from the Decimals CD to address these fundamental difficulties, and he was pleased and surprised to note conversations continuing in the staff room later in the day. This is assessment for learning working very well. It is our prototype for a 'specific mathematics assessment that reveals thinking'. As a smart test, it is easy to administer, very strongly embedded in research data, focused on conceptual learning that is important for progress, and reveals to the teacher diagnostic information that they can act on. An obstacle is that marking the DCT is complicated, because teachers do not just count how many items student have answered correctly, but have to identify patterns within the correct and incorrect responses. Whilst this is hard for a human, it is easy for a machine, and so it is this diagnostic element that is properly machine work, rather than a teachers' work. Our online system presents items and applies the rules that diagnose understanding.

The smart test project aims to create and study a working system which extends this approach across the curriculum, beginning with the lower secondary school. Information is available at the project website (HREF 1 www.smartvic.com). Fortunately, 3 decades of research around the world into students' understanding of mathematics have created a rich resource of probes that can be used as the basis of smart tests. ISDDE participants Alan Bell and Malcolm Swan, for example, have contributed to the research behind the examples presented in this paper. The project aims to bring this vast resource transformed into a useable format into the hands of teachers.

The smart system and design principles

With all of this in mind, a partnership was set up in 2008 between DEECD, the authors and 7 secondary schools, funded by the Australian Research Council. The brief was to develop a set of diagnostic quizzes or "smart tests" (HREF1) that could give teachers information about the understanding of their individual students in key mathematics topics and suggested teaching responses. These smart tests were to supplement normal assessment.

We now have a set of online tests covering most topics commonly taught in Years 7 to 9. Teachers can read descriptions of the available smart tests, choose one that is appropriate, give their students a password to attempt it, in class, at another time at school or at home. Responses are marked online, and the patterns of results electronically analysed with diagnosis available as soon as the teacher requests the results. A summary of the findings, along with information on the common misconceptions in the topic and relevant links to the Victorian Mathematics Developmental Continuum (HREF2), are visible instantaneously online. Our previous work in writing the *Mathematics Developmental Continuum P-10* (HREF2) had highlighted the necessity of students' understanding certain critical concepts in order to make progress.

Our aim with the smart tests is to target some of these critical concepts, and design

- short and easy to administer on-line tests that can diagnose to what extent students understand the concept;
- informative and prompt feedback to teachers about class and individual performance; and
- targeted teaching suggestions that address the conceptual hurdle
- strong research backing for the diagnosis and for the advice given.

These components, together, highlight the purpose of the smart tests as "assessment for learning". The tests are not designed to give a score, but to diagnose misconceptions and to provide teachers with information that will help them meet students' needs and improve learning outcomes.

By mid 2009, there are three types of tests. Tests of the first type diagnose fundamental conceptual understanding underlying curriculum topics. This is the most important type, and the ones which interest us the most. Parallel versions are provided for testing before and after an intervention. Smart tests of the second type check that prerequisite understanding is in place before a new topic is introduced and the third type check students' knowledge of basic number facts. Smart tests of the second and third types have been created in response to teachers' demand and strategically for us, as a way of introducing teachers to a system that can offer much more.

An example

Mathematical focus and item construction

To further illustrate the purpose, design and components of a typical smart test, we present an item from one of the tests that identifies misconceptions involving multiplication and division with decimals. One of the well-known misconceptions in the area of number operations is that "multiplication makes bigger, and division makes smaller," or MMBDMS for short (Bell, Swan, & Taylor, 1981). As with most misconceptions, MMBDMS arises as a natural consequence of previous learning. When students first learn about multiplication and division, it is with whole numbers, and multiplication does indeed generally make bigger (e.g. $2 \times 5 = 10$ and 10 is greater than both 2 and 5), and division does, indeed, generally make smaller. Strong foundational learning like this is essential for students' progress, but to go further students have to simultaneously build on these concepts and learn how they work in new situations. In the world of fractions and decimals, when multiplying by numbers less than one the formerly useful whole number principle of MMBDMS becomes a misconception. Fortunately, like many other misconceptions, MMBDMS can be readily addressed. Left unaddressed, it can remain to plague students throughout their schooling.

Figure 1 shows one of the screens of items addressing the MMBDMS misconception. Notice that students are not required to do any calculations, but merely to select from three choices of multiplication and division in a drop-down menu. One drop-down menu is shown in the figure as an example. Students doing the smart test would complete the 3-part item in Figure 1, together with some additional items that make it possible to diagnose, with reasonable confidence, whether or not they have the identified misconceptions, or (in other items) are making common errors. One of the consequences of choosing the online format is that items seem to have to be brief and quick for students to complete, without requiring intermediate working. Although students could complete the tests using pencil and paper jottings but only entering answers, their expectations about computer work operate against this.

The diagnosis

As soon as the students submit their responses to the online test, the results are analysed using an algorithm that recognizes patterns of responses corresponding to different types of typical thinking. Teachers then receive feedback on each student's performance almost immediately and they can easily make a class summary. This allows teachers to group students for targeted teaching. The diagnosis includes information for teachers about the different stages of understanding that are revealed by the smart test. An example is shown in Figure 2. Each student's performance on the smart test will allow them to be classified into one or other of the stages of understanding, and the feedback includes a list of all the students and their stage. In some cases a few students will not fit any pattern and so cannot be classified, and if this occurs it is included in the feedback as well. In the context of assessment for learning, this is easily solved – teachers can follow up individually with those students.

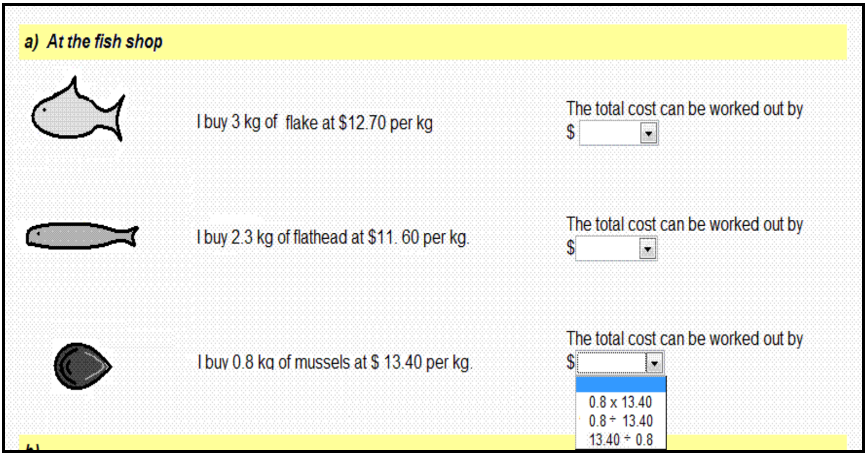
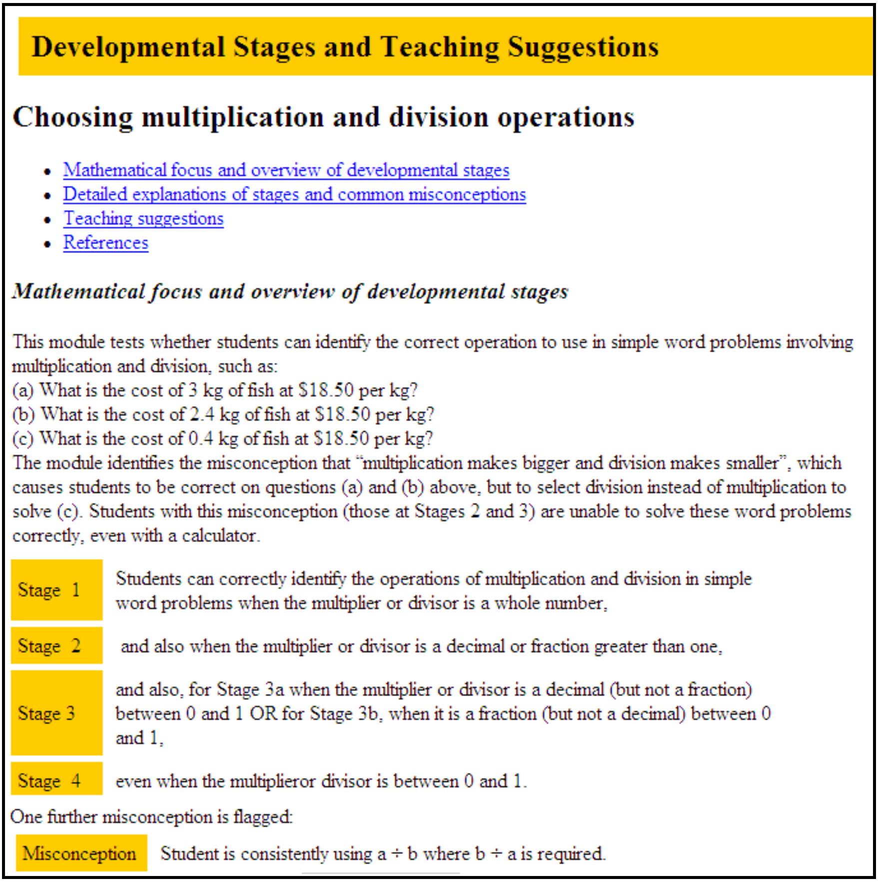


Figure 1. The fish shop item that helps examine misconceptions with multiplication and division operations.

Teaching strategies

Included with the diagnostic information is a set of teaching ideas. We aim to advice for teaching students at the different stages, again with a focus on the particular mathematical concept that is the focus of the smart test. In the case of the fish shop example, the advice for Stage 0 and Stage 1 students is to work on recognising the structure of word problems that involve multiplication or division, moving from repeated addition to multiplication. The repeated multiplication strategy fails at the second problem, when the multiplier is not a whole number. So for them, they need to recognise problem situations where multiplication and division are appropriate, such as equal groups and rates. Relevant progression points and sections in the Mathematics Developmental Continuum (HREF2) are:

- 2.25 Early division ideas
- 2.75 Multiplication from equal groups to arrays
- 3.25 Choosing multiplication and division



Developmental Stages and Teaching Suggestions

Choosing multiplication and division operations

- Mathematical focus and overview of developmental stages
- Detailed explanations of stages and common misconceptions
- Teaching suggestions
- References

Mathematical focus and overview of developmental stages

This module tests whether students can identify the correct operation to use in simple word problems involving multiplication and division, such as:
(a) What is the cost of 3 kg of fish at $18.50 per kg?
(b) What is the cost of 2.4 kg of fish at $18.50 per kg?
(c) What is the cost of 0.4 kg of fish at $18.50 per kg?
The module identifies the misconception that "multiplication makes bigger and division makes smaller", which causes students to be correct on questions (a) and (b) above, but to select division instead of multiplication to solve (c). Students with this misconception (those at Stages 2 and 3) are unable to solve these word problems correctly, even with a calculator.

Stage 1	Students can correctly identify the operations of multiplication and division in simple word problems when the multiplier or divisor is a whole number,
Stage 2	and also when the multiplier or divisor is a decimal or fraction greater than one,
Stage 3	and also, for Stage 3a when the multiplier or divisor is a decimal (but not a fraction) between 0 and 1 OR for Stage 3b, when it is a fraction (but not a decimal) between 0 and 1,
Stage 4	even when the multiplieror divisor is between 0 and 1.

One further misconception is flagged:

Misconception	Student is consistently using a ÷ b where b ÷ a is required.

Figure 2. Diagnostic feedback for teachers on Figure 1 smart test.

Stage 2 and Stage 3 students have the MMBDMS misconception. They may also need to strengthen their recognition of situations (equal groups, rates etc) that involve multiplication and division, and to learn that the type of number in the problem does not change the operation (e.g. small or large whole numbers, fractions, decimals which are smaller or larger than 1). It is for this reason that one of the recommended strategies to find the appropriate operation to solve a word problem is to "consider easier numbers". If we know how to find the cost of 3kg of fish, we can use the same operation to find the cost of 0.4 kg of fish. Once the structure of the problem is understood (and the operation chosen) then this can be used with whatever numbers are provided in the given problem. For ideas on developing concepts of multiplying and dividing by numbers less than 1, see the following indicators in the Mathematics Developmental Continuum (HREF2):

- 5.0 Conceptual obstacles when multiplying and dividing by numbers less than 1
- Number: The meaning of multiplication

Design challenges

There have been some interesting moments in the development so far, and at almost every point we are faced with design issues, the need to explore available research which often results in identifying gaps in research. We have had to work hard to deal with different school computer systems - hardware and software and internet access – and also to design a simple, accessible system. There have been surprising quirks in the data, such as when a whole class did very poorly on a test that involved visual information. Eventually it was realised that their computer screens were of a non-standard aspect ratio and so the visual information was distorted. There are many standard design challenges, such as how to test complex ideas with simple wording and set in situations that are easy to describe. Items need to be attractive, interactive, and robust when the students are using them, as well as give well-founded diagnostic information. Although there is a rich resource of diagnostic items from the research literature, many of these need to be significantly altered or adapted for online delivery. In the future, we hope to extend the nature of the items and responses, so that students' thinking can be probed more by using more sophisticated free response items.

The other aspect of design relates to how assessment for learning can best fit into "normal teaching" and we are collecting evidence of this throughout the life of the project. Early in the project we realized that a smart test to check basic arithmetic facts (e.g. which of their multiplication tables do students know) was popular with teachers, and it introduced them to using the system. This sort of assessment fits well into normal teaching, especially at Year 7 level. Teachers regard basic arithmetic as essential to progress and they are well able to act on the information on students' performance by proving further practice for those students who need it. Similarly, the tests that provide a check on pre-requisite knowledge for a standard curriculum topic, explicitly identified, fit well into normal teaching. For example, a before teaching trigonometry, a teacher may want to know if students (a) can recognize similar triangles and use their properties, and (b) can identify the opposite side, adjacent side and hypotenuse of right angle triangles and (c) can solve equations such as 0.3 / x = 4.6. Smart tests can be readily selected for these components (or pre-assembled as 'readiness for trigonometry') and by using it, the teacher is forewarned about the points that are likely to need more attention in their coming lessons.

Smart tests like those above fit well in the way in which most teachers think about their mathematics curriculum. On the other hand, perhaps surprisingly, understanding of more conceptual aspects of mathematics is not so readily identified as having a home in the curriculum. Our initial impressions are that these tests are less frequently selected for use. We expect that a test of whether students correctly multiply by decimal numbers would be more frequently selected by teachers than the smart test in Figure 1, which they cannot easily place in terms of teaching goals. We will experiment with solutions to this problem, because of the importance that we attach to students' having deep underlying conceptual foundations, and because we believe that using these tests will increase teachers' knowledge for teaching significantly.

Selecting the best grain size is another design issue related to both selection of tests by teachers and practicality. We want smart tests to be short and tightly focused, but we have also found that the overhead of organizing a test can outweigh the information gained from both the teacher's and students' point of view if the test is too narrow. How detailed should a detailed assessment be?

The presentation of diagnostic information for the teacher is the key to one of our major objectives, which is to improve teachers' pedagogical content knowledge. We have had successes here, as evident for example in the way in which teachers are able to adjust their instruction after seeing the results of the DCT, and how their sensitivity to students' thinking is heightened as a result. Our research will explore the conditions that make this likely and the trajectories which teachers' PCK follows.

The future

There is some hard work ahead to ensure that the smart tests become a useful resource available to all teachers of mathematics, with a wide range of tests, accurate and reliable diagnoses of understandings that are key to making progress in mathematics, and linking to helpful teaching suggestions. In the future we will continue to improve each of the smart tests, and refine the power of the diagnostic algorithms. We can also mine our growing data base for other phenomena related to students' mathematical thinking that have not previously been discovered. For example, we organize most feedback in terms of 'stages' expecting students to move from Stage 0 upwards. Which curriculum topics does this model fit? Should students' knowledge of a particular area be best measured or mapped? (Stacey & Steinle, 2006) Currently the feedback goes only to teachers, and we are looking at ways of providing feedback to students as well, again as soon as they have completed the test. The challenge here is that the diagnoses that are supplied to teachers are usually not comprehensible to students. They are written for adults, and some effort, background and technical language is required to understand them. Students are not likely to be able to understand the diagnosis that is being supplied to teachers. Students probably only need an indicator of how well they have done but a simple score on these tests is not meaningful. In addition, teachers are cautious about providing unduly negative feedback to students who have attempted tests at too high a level.

We are optimistic that the smart tests will be a powerful resource for diagnosing students' thinking, easy for schools to use, informative for teachers, and thus an important component of the assessment for learning process. We welcome possible collaborators on the many disparate parts of this project. Our plan is to hand the formative diagnosis over to a machine, so that human teachers can concentrate on using the information to improve the learning of each of their individual students.

Acknowledgements

The smart tests project is supported by Australian Research Council Linkage Projects Program grant LP0882176 in partnership with the Department of Education and Early Childhood Development, Victoria (2008 - 2010).

References

Bell, A., Swan, M., & Taylor, M. (1981) Choice of operation in verbal problems with decimal numbers. *Educational Studies in Mathematics*, 12 (3), 399-420.

Black, P. & Wiliam, D (1998) Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80.

HREF1: http://www.smartvic.com/smart/index.htm, Welcome to smart-tests. SMART research team. Accessed 20 Aug 2009.

HREF2 Stacey, K., Ball, L., Chick, H., Pearn, C., Sullivan, P., Lowe, I. (expert authors) (2006) *Mathematics Developmental Continuum P - 10* http://www.education.vic.gov.au/studentlearning/teachingresources/maths/mathscontinuum/default.htm Department of Education & Early Childhood Development, Victoria. Accessed 20 Aug 2009.

HREF3 Department of Education and Early Childhood Development (n.d.) http://www.education.vic.gov.au/studentlearning/teachingresources/maths/assessment.htm

Stacey, K. & Steinle, V. (2006) A case of the inapplicability of the Rasch Model: Mapping conceptual learning. *Mathematics Education Research Journal,* 18(2), 77 – 92.

Steinle, V. (2004). Changes with age in students' misconceptions of decimal numbers. PhD thesis, Department of Science and Mathematics Education, The University of Melbourne. [http://repository.unimelb.edu.au/10187/686]

Steinle, V., Stacey, K. & Chambers, D. (2006) *Teaching and Learning about Decimals*. (*Version 3.1)* Faculty of Education, University of Melbourne. (CD-ROM). See also http://extranet.edfac.unimelb.edu.au/DSME/decimals