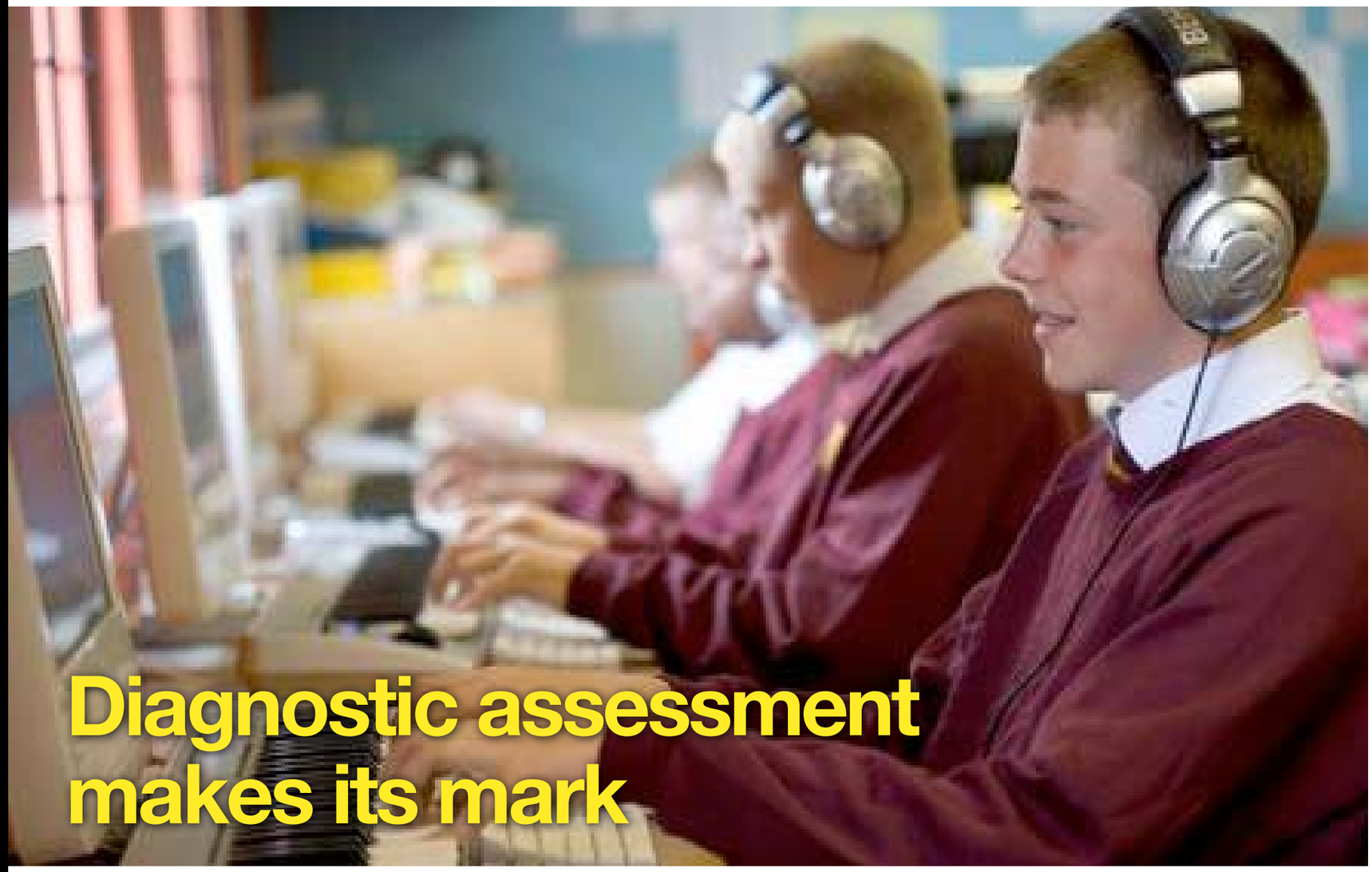

Diagnostic assessment makes its mark

Computer adaptive assessments must be correctly harnessed to identify students' needs and assist them in learning more effectively

Words **Professor Peter Tymms**

The majority of official tests and examinations stamp people with a level, a degree classification, a mark or a pass/fail categorisation. These tests are designed to measure attainment, but there are many other uses for assessment. Among them are the identification of special needs, research and accountability.

The fundamental idea of "assessment for learning" is to use assessments to provide feedback to students on how to improve. The movement has been informed by stringent research in the form of randomised controlled trials, which have demonstrated that feedback from tests that show how to improve performance can have a positive impact, and shown that grades and marks can detract from this positive effect.

Meanwhile, a parallel important and growing body of research has shown how we can diagnose and remediate specific learning problems. For example, if a child has a reading problem, the root difficulty might be diagnosed as: "an obstacle-associated decoding". If the obstacle can be identified, then research-based strategies can be brought to bear and the child's reading improved dramatically.

Efforts must now be directed to link the assessment-for-learning movement with the body of knowledge about learning difficulties; their identification and effective remediation. One important step forward is surely to produce computer adaptive assessments. A teacher should be able to ask a pupil to sit down at a computer and complete an individualised assessment, responding to questions that are tailored to his/her ability so that a diagnostic profile can be generated to identify strengths and weaknesses. The kind of feedback that such an assessment generates is formative and should be linked to research-based advice on what to do about areas that are in need of improvement.

An example from reading should clarify the idea. Competent readers must be able to recognise a great variety of words without a moment's pause, indeed those who read a lot do most of their reading by instantly recognising the words on the page. But they can also decode words that they have never seen before. For example, a skilled reader will have no difficulty reading the invented word "graphile". A competent reader who can recognise and decode words must also be able to make sense of passages of text. There are some children who recognise words and can also decode but who simply "bark" what is written on the page without understanding. Imagine Susan, a seven-year-old, sitting at a computer with headphones on. She is being assessed on decoding skills and hears the computer say "zop", a non-word which is generally easy to decode at her age. She sees several written words that might say "zop" and she chooses the right possibility from the five options. Now she gets a harder word, which she also gets right, and then a harder one. If she gets it wrong she gets an easier one and all the time the computer is estimating her ability to decode. The test takes 10 minutes and she has only been presented with items that she finds easy or almost within her grasp. In other words she was never threatened with questions seriously above her developmental level.

Similar procedures are used for other areas in reading and the teacher gets a comprehensive profile of Susan's strengths and difficulties. It turns out that Susan is good at word recognition for her age but her decoding is two years behind and her ability to cope with reading passages is also impaired. The teacher had an idea that something was not quite right and armed with the diagnostic profile and information on research based strategies for dealing with decoding problems for seven-year-olds she puts a recovery programme into place.

Children have a range of strengths and challenges as their reading develops. Some children, for example, have persistent difficulty in decoding words and *in extremis* may be identified as dyslexic. Others can decode and recognise words but have great difficulty spelling. They both need help but different kinds of help. Fortunately there is now a solid evidence base which can be drawn on to help teachers help their children.

Can we learn from the US?

Although in the UK we are now prevented by law from calling the statutory end of Key Stage assessments Sats, in the United States the original SAT, which has the monopoly on the acronym, continues to be very widely used to select students for universities (colleges).

Unlike A levels it does not assess many areas of the curriculum; it concentrates on verbal and mathematical skills and is taken by more than a million High School students every year. For a long time, students sat down and worked on pencil and paper tests, but gradually that has been moved towards computer adaptive testing (Cat) where students are presented with items which depend on their previous responses. This, of course, means that very large data-banks of items have to be developed together with the difficulties of each item which are established in advance.

The test is very high-stakes and the move to computers has not been trouble-free. For example, there have been attempts to find out which questions are in the databanks. Immediately after taking the assessment, students did a "brain dump", recalling all that they could, which was then passed on to others.

Gradually people have become wise to this and it is much harder to do because the data-banks have become much larger and it is possible to generate items "on the fly". Nevertheless, constant efforts to discover ways of increasing the chance of success, including legal challenges, has caused several u-turns in the use of Cat by the testing bodies and Colleges.

Further, it needs to be stated that the Sat is not a diagnostic assessment. It seeks to give a score but not a profile of strengths and weaknesses.

Paper-free tests boost candidates' experience

More than half of all secondary schools in England are part of the monitoring projects run from the CEM Centre at Durham University (Alis, Yellis and Midyis).

In these projects, schools can opt for a computer-delivered adaptive baseline assessment. Last year 160,000 students took the test, making it the most widely used computer adaptive test outside the US.

Students take this 40-minute test over the internet or via software installed on school networks. The move away from paper has produced a faster and more accurate service with rapid feedback to schools. The experience for pupils is also enhanced as they are no longer faced with self-esteem threatening questions, nor do they spend time on non-challenging easy items.

Research indicates that candidates are more engaged and spend more time answering questions when they are of a difficulty equivalent to, or just above, that of their own ability.

A new generation of children may now be understood by their teachers in detailed ways that were previously only possible with expensive one to one assessments often conducted by an outsider. Personalised assessment and learning is becoming a reality.

Doing the groundwork

So why are these kinds of assessments just appearing now and why are they not more widely available? The short answers are that they are difficult to produce, they require a technical infrastructure and innovations take time to be accepted. The production of a good diagnostic assessment takes a long time, perhaps as long as three years, and it is only in the last decade that the necessary technology has become available. Computers need to be widely accessible in schools and networked with good internet connections. Further, the software must be based on sound statistics and it is only in the last 20 years that the necessary psychometric advances have been made. Much of the statistical groundwork, which started with work on reading by the Danish mathematician George Rasch, has been carried out in Australia and the United States. Britain was at the forefront of such developments but unfortunate decisions nearly 30 years ago left the UK in something of an assessment backwater. Things are changing, but the lack of psychometric expertise has meant the development and adoption of computer adaptive assessments has been slow.

We have extensive knowledge about how to remediate reading problems because of decades of work by top researchers including several in the UK. The same cannot be said of mathematics and other key areas, such as science. Dyscalculia is now being studied widely and advances are being made, but the progress of children towards mathematical competence is not as well understood as it might be.

Similarly, much more research is needed into how scientific competence develops and how challenges can be overcome. But despite our lack of knowledge, it is quite possible to construct computer adaptive assessment in the areas of mathematics and science and they in turn will contribute to the development of the research-based advice on remediation.

Diagnostic assessment is not suitable, as yet, for qualifications such as GCSEs and A levels. These are not intended to be diagnostic and are designed simply to certify attainment levels. The security issue also dictates that new papers must be produced every year in secret. Adaptive tests take years to develop and much trialling of items is needed. They must be used with thousands of children and young people in very different schools. It would be practically impossible to maintain tight security in such circumstances.

The UK examinations tradition involves many open-ended questions and it is unlikely that these exams will be based on adaptive testing without a considerable degree of rethinking.

Adaptive assessments lead to tailored responses

Across all primary schools in Northern Ireland and in many other schools, children now sit down in front of computers wearing with headphones on and are assessed using the InCAS program.

This is a diagnostic adaptive assessment of mathematics and reading for children aged 6-12. Pupils are presented with a series of questions which are selected specifically for them according to their age and their success with questions that they have been asked. The adaptive nature of the assessment means that each child gets a unique assessment. Teachers can choose exactly what they want to assess from a range of possibilities in mathematics and reading as well as in vocabulary, non-verbal ability, spelling and attitudes. After the teacher first indicates the areas of interest, the child sits down, puts a codeword into the computer and the assessment starts.

The data from the test are uploaded via a secure website to the CEM Centre at Durham University, which processes the information providing instantaneous feedback. The teacher can then examine the pupils' profiles and using her professional judgement in conjunction with research-based booklets from the InCAS project can prepare for future action.

It is interesting to note that the adoption of InCAS in Northern Ireland is part of a range of educational changes which include the dropping of the statutory end of key stage assessments in primary schools, the dropping of the transfer test (the 11+) and the encouragement of education and assessment (including InCAS) in Irish.

About the author

Professor Peter Tymms PhD is Director of the CEM Centre at Durham University, which runs projects monitoring the progress and attitudes of a million pupils each year. He devised the PIPS project which monitors the affective and cognitive progress of children in thousands of primary schools.

References

More about INCAS
Merrell, C. and P. Tymms (2006). "Identifying Reading Problems with Computer Adaptive Assessments." Journal of Computer Assisted Learning 22: 1-20.

Black, P. and D. Wiliam (1998). Inside the black box: Raising standards through classroom assessment. London, Kings College London, School of Education.

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model: Fundamental Measurement in the Human Sciences (2nd edition ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Websites
www.incasproject.org
www.cemcentre.org