# E-assessment for improving learning

**Chris Whetton and Marian Sainsbury**

Paper presented at the 33rd International Association for Educational Assessment Conference, 16 – 21 September 2007, Baku, Azerbaijan.

**Presenter: Chris Whetton**

Assistant Director,

National Foundation for Educational Research

The Mere, Upton Park Slough SL1 2DQ

United Kingdom

**c.whetton@nfer.ac.uk**

# 1 Background

## Policy context

For most of the 20th century, part of the assessment world was dominated by the continuing promise of psychology and psychometrics. These developing sciences held out the possibility of precise and accurate measurement of mental attributes, in parallel to the precision and accuracy of physical science. There was an emphasis on objectivity, and a technology of test development grew up, heavily dependent upon statistical analyses. Alongside this with its own emphases, generally in Europe, an alternative approach, based on written examinations prospered, which has some of the same characteristics, but differed in the styles of questions and scoring processes. In the psychometric tradition, the test-taker was the object being measured and the test outcome took the form of a battery of numbers – raw scores, standardised scores, percentiles, confidence intervals, correlation coefficients – which related performance to established norms, or to other tests. Test items were frequently multiple-choice, and were selected mainly for their statistical properties. Most current paper-based tests are still located in this psychometric tradition. Their claim to rigour lies in large-scale representative sampling and statistical analyses following established methods. These principles are now frequently also applied to current computer-based tests. Some are simply paper tests adapted for the screen, with statistical equating exercises to relate the scores from the two methods. Others are adaptive, with statistical test-development technology being extended to deal with the situation where each test-taker takes a different set of items.

Towards the end of the 20th century, in some countries, a further strong imperative arose in the assessment world. Governments increasingly required a testing programme for the purpose of evaluating the success of schools, teachers and the education system. For this, it was necessary to broaden the scope of the tests to match the depth and richness of the subjects as defined by the curriculum. At first, this led to tests which had high validity, but were less manageable than their psychometric predecessors, cumbersome to mark, with a resulting loss of 'objectivity'. However, as the requirements for accountability grew, the tests themselves moved back to a psychometric model, with a strong emphasis on reliability, security and the comparability of results.

- In the United Kingdom, or specifically in England, recent policy announcements have suggested that Government thinking on assessment is moving into a new phase. Added to its existing monitoring and accountability functions are new demands for tracking pupil progress and giving teachers the evidence they need to provide personalised learning. To meet the new requirements, assessment will have to deliver new types of data and become more sharply focused.

A consultation document entitled *Making Good Progress* set out this ambitious vision. It expresses concern about the discrepancies in progress made by individual children in the course of their school careers. Based on this, it suggests that a

combination of high expectations and targeted interventions could be mobilised to address the problem. For assessment to play its part in this process, it should take on a new focus on progress, with progress targets supplementing the existing targets for absolute attainment. Alongside this, the integration of effective assessment into ongoing teaching and learning is seen as central. A major implication of another government commissioned report, *2020 Vision,* is that teachers will be expected to make systematic use of data analysis, regular assessment of curriculum topics and techniques such as pupil peer and self assessment. This will lead to a clearer understanding in the minds of both teacher and pupil about the pupil's existing understanding and help to formulate sharper and more achievable targets for continuing progress.

However, this picture of a coherent, constructive assessment system makes enormous demands on the teacher. Schools will have to continue to manage national tests and stringent accountability targets, but there is now to be a new requirement that each child has a personalised curriculum planned in the light of comprehensive assessment evidence. There is a risk that these demands will become overwhelming, and indeed there are indications that many teachers are already finding them so. There are also reports that pupils find assessment stressful. It is essential to ensure that assessment is a coherent, manageable and useful process for teachers so that pupils benefit from it.

At the same time, with strong Government backing, digital technology is playing an unprecedented part in school life. ICT is now an integral part of school activity, both in the classroom and in school administration. ICT makes its mark on teaching through the use of interactive whiteboards, teaching packages to aid pupil learning and word processors and presentation software that enable pupils to present their work. There is now one computer for every 6.7 pupils in UK primary schools (up from one for every 12.6 pupils in 2000) with most having the required resources for large group use – and this trend can only increase with time. Computers should surely be able to make a positive contribution to solving the teacher's assessment administrative burden. Yet so far this provision has had little impact on the assessment process, other than in analysing results.

# 2 A New Approach

## E-assessment

Introducing e-assessment for important, high stakes national tests has the potential for great improvements, but also brings with it significant problems. In England, The Qualifications and Curriculum Authority announced a policy in 2005 for the gradual introduction of more computerised testing, with GCSEs, AS and A2 examinations having an on-screen option within five years. In time, this could also apply to the proposed national curriculum progress tests. However, in moving towards this situation the confidentiality of high-stakes test materials must not be compromised, and the chances of candidates cannot be blighted by technical failures.

Less attention, however, has been paid to the potential of e-assessment in low-stakes contexts. It is clear that teachers are required to focus on the understanding and attainment of individual pupils in order to develop effective plans for personalised learning. This will involve the management of a great deal of assessment evidence for planning teaching, in the form of test data and information on progress through the ongoing curriculum. Making sense of this mass of evidence requires teachers to discern patterns, interpret their meaning and use the results to formulate targets and specific differentiated teaching plans. Traditionally, this has been done informally, based on the teacher's personal knowledge of each pupil. With a requirement for more systematic assessment of progress and recording of targets, however, e-assessment can occupy a central role, first in gathering detailed information about the nature of individual pupils' understanding and attainment, and then in collating and analysing this data. Rather than supplanting the teacher's role in relation to the child, it could supplement it, reducing the marking and recording workload while increasing and easing the flow of genuinely useful information.

Using e-assessment for low-stakes, formative purposes would seem to offer some attractive opportunities:

- A bank of assessments could focus in depth on individual curriculum topics, rather than attempting to cover an entire subject, leading to richer data on individual pupil progress.

- Printed test questions can assess only limited aspects of the curriculum, whereas the dynamic and interactive capacity of the computer allows for a wider range of question types and styles of assessment. E-assessment also helps pupils to demonstrate their visual and kinaesthetic understanding more effectively.

- More frequent assessments can give formative data before a subject has been taught, and summative (as well as further formative) data afterwards. Tests can be used to inform individual teaching plans as well as to assess achievement. By contrast, many conventional tests are primarily summative, with only limited diagnostic or formative information.

- Instead of being taken at distinct points in the school year, formative e-assessments can be integrated into ongoing classroom teaching. In this way, assessment becomes an integral part of the normal teaching and learning cycle for teachers and pupils.

- E-assessment can build a profile of strengths, weaknesses, partial understandings and misconceptions. Going far beyond scores and standard outputs, these reports can be used to improve teaching as well as communicating effectively to various interest groups.

- Because it is low-stakes, e-assessment can be a positive part of the classroom experience. Administering tests on computer and using engaging tasks can make them more motivating. Boys, in particular, may engage more readily with computers.

There are corresponding challenges, however. Where pupils' responses are to be computer-marked, computer-based assessments are usually limited to the use of questions with closed answers, rather than allowing pupils to express their responses in their own ways. To be effective, these questions must be carefully devised to capture as much as possible of the full richness of the curriculum, and to be thought-provoking rather than superficial. Computers have a potential for dynamic and interactive activities that are not available on paper, and this potential must be exploited to create effective, innovative digital question types. Computers can effortlessly capture a vast quantity of data about each pupil's responses, but much work has to be done before this becomes a useful, meaningful and manageable tool for the teacher.

Previously the main purpose of assessment was the provision of information in numerical form to be used *outside* the classroom to summarise and compare pupils' performance. What is now required is an unprecedented quality of information to be used *inside* the classroom, by teachers and, to some extent, by pupils too. This is reflected in some of the recent educational and assessment trends.

*1. Assessment for learning*

Assessment for learning, or formative assessment, highlights the gains to be made by providing carefully focused ***feedback*** resulting from an assessment. This enables teachers to make precise teaching plans and helps pupils themselves to understand how to improve. A seminal paper by Black and Wiliam[i] drew attention to evidence that improving formative assessment in the classroom leads to higher achievement in students. Since then, these ideas have been further elaborated in the course of several action research projects. In policy terms, a self appointed pressure group known as the Assessment Reform Group have met with some success in urging the adoption of these principles[ii], and assessment for learning forms an integral part of the 'personalised learning' agenda of the current English and other UK governments[iii]. It is strongly supported by assessment advisers in local education authorities.

*2. Thinking skills*[iv]

There is some overlap between thinking skills and assessment for learning, in the sense that both attach importance to pupils' conscious monitoring of their own cognitive processes as they think and learn. Thinking skills research has investigated ways in which students can be taught to think more effectively, and has given rise to a number of teaching programmes[v]. The questions in the assessments are designed to elicit a variety of thinking skills that can be loosely grouped into the following categories:

- thinking for clarification
- thinking for probing reasons and evidence
- thinking creatively to explore further or different ways and views
- thinking to test implications and consequences
- thinking through to solutions and linking to situations beyond the immediate.

*3. Engagement*

Once again, there are areas of overlap between engagement theory, formative assessment and thinking skills, in the notion of the learner as proactive and thoughtful. Engagement theory highlights the positive emotions and growth of knowledge that result when the learner is strategically involved in his or her own learning. This active process involves a number of cognitive strategies, for example: activating prior knowledge, questioning, searching, summarising, monitoring comprehension, organising knowledge and making inferences. Guthrie, a leading researcher in this area, summarises engaged learning with 'five Cs': connections, competence, choices, comprehension, collaboration[vi]. Guthrie's work has mainly been in the field of reading comprehension, but the ideas are applicable to all learning and also to the strategies involved in answering test questions in any subject.

## A new generation of assessments

Building upon all of this, the time is now right for assessments which offer an extra dimension not available in traditional tests. The pupils are no longer the 'objects being measured' but instead become active, engaged participants in the process.

The experience of trialling the questions with groups of children provides evidence that pupils find demanding open-ended questions genuinely interesting. Extended discussions take place, with different pupils spontaneously providing evidence and reasoning to support their points of view. Children are animated in their opinions, but also show willingness to adapt their views in the light of reasoning to the contrary.

Such thought-provoking questions should underpin high-quality information that is available to teachers. They allow in-depth probing of pupils' knowledge and

reasoning, uncovering misconceptions and pinpointing the limits of understanding. This information should be available to teachers for each individual pupil if required. However, adding further value will be the provision feedback so that teachers can confidently go straight from the test results to detailed and tightly focused teaching plans. In addition, the 'child-speak' versions of the reports can make it easy for teachers to involve their pupils in discussing progress and agreeing targets.

These types of requirements can be addressed through computerised testing, but this will require a new type of test with different approaches to structure and, crucially, to the reporting mechanisms.

# 3 A New Type of Computerised Test

## Assessment information

NFER has been in the process of developing a new approach to computerised assessment that meets the needs for a low stakes assessment, that fits the philosophies of personalised learning and assessment for learning. At present, this is mainly in the primary stages of education, but the principles involved can be adapted to all phases of education.

The structure of the package differs from a standard test, in that it attempts to cover the curriculum as a whole, but as a series of units, each relating to a curriculum area. The units can be taken individually to provide diagnostic information to aid teaching and learning. In addition, the information from the units can be combined to give an overall measure of progress in the curriculum. This overall measure will have greater or lesser certainty according to the number of units which have been taken, and this level of certainty will be part of the reporting to the teacher.

There will be a pair of tests in each unit, addressing a topic, or 'chunk' of curriculum content, called here a programme. Teachers can choose to use one or both, at the beginning, the end, or during a programme of teaching. However, the central purpose of the design is to 'wrap around' the programme of teaching, giving sharply focused formative information at the beginning and both formative and summative information at the end.

The assessment information that will be available from each unit will have a variety of forms, which are listed below. The information towards the beginning of the list is designed to be more useful at the beginning of a programme of teaching, and that further down the list is designed to be more useful at the end.

Reports will indicate:

- which pupils have mastered which aspects of the curriculum content of the teaching programme;
- suggestions for grouping pupils according to the extent and quality of their understanding;
- diagnostic reports on understanding and misconceptions for each pupil;
- National Curriculum level with probability estimates which will become more precise the more tests the pupil has taken;
- age standardised scores, with confidence limits, since teachers continue to find such summative information useful in reporting to parents and others..

Most of these types of report will be available for individuals, groups, whole classes and larger groups (eg year groups).

In any given class, pupils will be working at a range of levels, and it is not realistic to expect a short manageable test to give equally detailed information for the full range of ability. The tests will address a range of attainment within and around the most relevant levels for the age group. The approach to difficulty levels is likely to replicate the teacher's curriculum planning, where lessons are aimed at most of the pupils in the middle of the range, with some special arrangements for very high and very low attainers.

The topics of the teaching programmes to which the tests are matched are intended to be chosen by teachers in accordance with their usual planning procedures, and flexibility is therefore built into the assessment structure. Each teaching programme might take anything from a couple of weeks to a whole term. This structure allows for the content to be age- and level- appropriate, whilst not constraining teachers to a particular approach to the curriculum for a single term or year. The content for set of assessment units will, in total, cover all the major areas of the programmes of study, but will not attempt to be completely exhaustive..

## Test structure

Within the overall structure outlined above, each test has 12-16 items, to be taken by an individual pupil working at a computer, with a timing of approximately 30-40 minutes but no strict time limit.

The item types are mainly standard format closed items, but there are some innovative item types which might include dynamic operations, a more complex choice of answers, or help screens to give hints as to how to proceed. For younger children, clicking and dragging operations have been found to be best, as their keyboard skills are undeveloped. However, for the older pupils, some free-text answers are included in order to allow them to formulate their own answers.

Each item has a mark scheme reflecting its complexity and giving credit where appropriate for partial understanding and speed or process of response as well as correctness. As a result, items can carry between one and three or four marks. The total mark allocation for a test is 20-30 marks.

# 4 The Development Process

This paper describes work in progress. As such some conclusions, methods and products are still in a state of change and this description reflects that stage of development. The development process for these new products, whilst drawing on existing expertise and experience within NFER, is nevertheless be innovative in many ways. The underpinning vision, as always with a research and development process, is to design a methodology that will deliver the required assessment information with high levels of validity and reliability for its intended purpose. To do this, a process of generation, selection and refinement through review and trialling is necessary. Alongside the requirement for a product that meets quality standards is the need for a short, streamlined and focused development process that meets cost and time constraints.

§ Because this is an innovative product, the specification of the tests becomes an important first step in the process. The curriculum topic or teaching programme for each unit is the first stage of this specification. Within each of the specified areas, it is necessary to set out the exact content in terms of the skills, understanding and knowledge that the tests are to address. This is being done more exhaustively than in the usual test development process, as these definitions will underpin the precise formative information to be derived from the test. These specifications may take the form of a series of statements, including in some cases a concept map or other diagrammatic information. The information in the test specifications will directly inform the teacher reports generated by the final product.

§ Once the specifications are developed, the item writing process will follow. The aim at this stage for each unit was to generate two parallel tests containing about 20% more items than will eventually be needed – 16-20 items. Both simple and complex item types were included, with rather more simple than complex. Item writers were guided to make full use of the diagnostic strategies envisaged for these tests:

- Timing analysis, where the speed of response contributes particular information about the pupil's understanding (for example, in knowledge of number facts);
- Misconception/partial understanding analysis, where incorrect answers are analysed for evidence of systematic misconceptions or partial understandings;
- Feedback to pupils on their incorrect or partial answers;
- Help screens giving hints or teaching points;
- Supplementary questions to probe understanding further.

§ The informal trialling stage is important as it is the single opportunity to obtain developmentally useful information on the functioning of the items. This will

contribute to the validity of the units, in that the trials will reveal central features of pupils' thought processes as they go about answering the questions. The trials will be designed so that each unit (ie pair of tests with 2 x 15-20 items) is thoroughly trialled with a small group pupils in a number of different schools. The findings will directly inform the formative and diagnostic information yielded by the test.

To structure the informal trials, protocols have been designed consisting of the test items in paper-based form, together with probes for researchers to explore pupils' reasoning and underlying understanding at certain points in the tests. A think-aloud procedure is being used to explore pupil understanding. All informal trialling sessions is being audio recorded to provide evidence in addition to the pupils' written responses.

During the visits, teachers in the schools are asked to comment upon the materials, including curriculum coverage and suitability for age and levels, thus contributing to the validation of the assessment.

At the conclusion of the informal trialling phase, the development team should know:

- typical responses to each item from pupils of different levels of understanding;
- something of the reasoning underlying these responses, including partial understanding and misconceptions;
- the approximate difficulty of the items in relation to one another and to pupils' typical attainment at the target ages and levels;
- the accessibility, clarity and informativeness of the items in measurement terms;
- teachers' overall view of the appropriateness of the materials.

On the basis of all this information, the items for each unit are decided. They are prepared in paper form together with their mark schemes and subjected to further team review for validity, accessibility and appropriateness, followed by final revisions before the programming stage.

- The process of programming the tests has changed over the development period. The technology partners in this project programmed the initial tests themselves, working to detailed specifications provided by researchers. Using these specifications as the basis, an authoring tool has now been developed, refined and handed over. It is now possible for NFER researchers to enter their own questions and publish their own e-assessments using the authoring tool.
- In this system, each question is built by selecting a template from an existing bank and populating it with a variety of 'assets', a term used to encompass every textual, visual or multimedia component. Textual assets are produced in Word and may also be designed in QuarkExpress. Illustrations must be available in digital

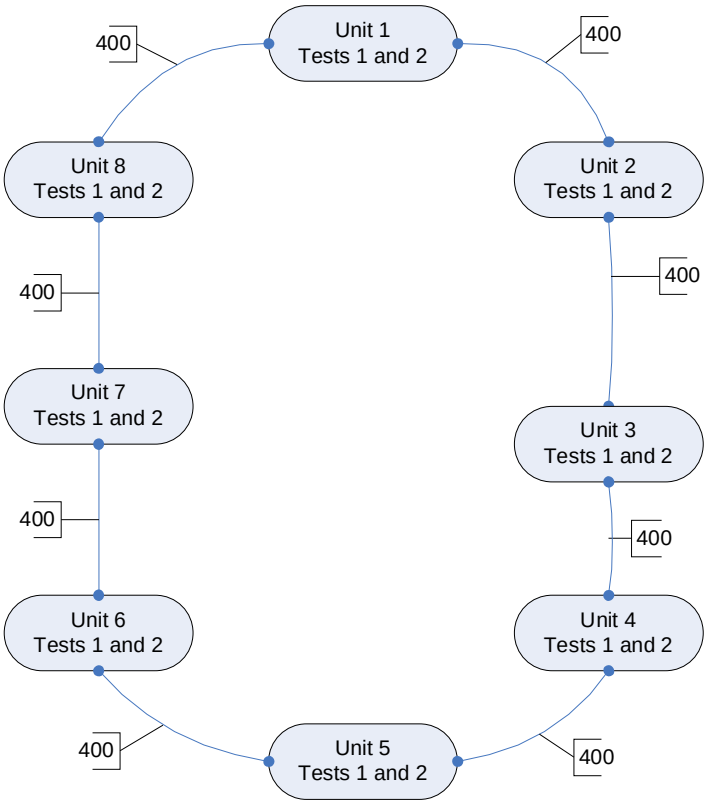
form, either by electronic creation or scanning. Digital video and photographs may also be included. Animations and interactive assets need to be specified and programmed. For this purpose, vector graphics are used in the creation of the visual elements.

The formal trials are being based as far as possible on the finished product, trialling complete tests with design and item functionality in their final form. The order of items in these tests is fixed rather than random, as their purpose is to build a profile of a pupil's understanding and the order of presentation will affect the reliability of this. The trials are intended to establish the technical functioning of the new tests, both in terms of the delivery system and the measurement properties of the assessment.

A test delivery system has been developed and progressively refined over the formal trialling phases. It involves schools installing software from CDs to the school network and to each individual computer. Pupil information can be imported from the school management system. Test results and pupil data are fed back to a central server automatically and collected for analysis. Further pupil data (teacher assessment levels in reading, mathematics or science; special educational needs; English as an additional language) is collected via an Excel spreadsheet directly from the schools.

## Trialling design

This sample design is based upon an 8-unit assessment, each unit consisting of two parallel tests.




In this diagram, the 800 pupils taking each unit are split so that 400 link with one other unit and the other 400 with another. The cartwheel design ensures that results from all the tests can be linked together for analysis.

Initial analysis involves a basic descriptive analysis for each whole test, giving score distribution, mean, median and standard deviation. This provides information about the overall suitability of the tests for the target pupils. Mean total scores are broken down by boy/girl and by teacher assessment level. Correlations are calculated between the pairs of tests in each unit; since the tests are intended to work in parallel, these correlations should be positive and high. These initial analyses indicate whether the level of difficulty of the tests is appropriate and reasonably consistent across all the tests in the assessment. If this proves not to be the case, some adjustment, in terms of removing a small number of items, remains possible.

The initial item analysis use classical test theory and yield information about the facility of each item (the percentage correct), omission rate and discrimination. Distractor analyses, showing the percentage choosing each incorrect option or coded response, are also conducted. A mean response time for each item is calculated at this stage, together with any other item-level information such as incorrect attempts and use of help screens. These are all used in the profiling analysis described below.

All the items in an assessment are then scaled using item response theory (IRT). This provides a basis for the national curriculum levels, age standardised scores and the profile analysis. It also gives further information about the difficulty of the items, distractor analysis and item analysis by boy/girl and level.

The relationship between teacher assessed levels and scale scores is used to derive for any pupil with a given scale score the probability distribution for a national curriculum level in the given subject. Bayesian statistics provides a methodology for updating these probability distributions in the light of new data.

Age standardised scores in three-month bands are calculated for each test, with confidence limits. These, too, will be updated so that the confidence intervals narrow the more tests are taken.

To provide focused formative information, analyses seek out 'profiles' of understanding, identifying patterns of response that distinguish particular types of pupil understanding (see the next section). A profile will be built from whole test scores, scores on individual items and incorrect or partial responses on particular items. This will use latent class analysis to explore patterns of performance, with reference to those suggested by the original test content specification. It underpins the suggested groupings of pupils and suggestions for next steps in learning.

All of the above analyses play their part in establishing a solid research basis for the claims made to validity, reliability and fitness for purpose by the new assessments. They will also provide a wealth of data to inform the reports from the tests, including:

- the pupil raw scores, levels and age standardised scores for individual, group and class reports;
- a nationally representative mean score on each test, which can be compared with the performance of a school or class;
- nationally representative facility and pattern of performance on each item, which can be compared with the performance of a school or class.

The Bayesian analyses, updating of probabilities, updating of confidence intervals and analysis of profiles are all innovations without any precedent in previous test development research.

## Finalisation

Once the analyses are complete, they are being used as a basis for the programming of the reports. The team is working with the software developers in designing and generating tables and graphs which embed necessary algorithms into the software. This will take the form of a 'black box' program, provided by NFER and built into the analysis and reporting engine. Once again, this will include innovative elements that require new approaches to test design and reporting.

# 5 Some Early Results

As described above, the main objective was to develop a statistical methodology for providing teachers with formative information about their pupils in an innovative computer-based test. In order to do this, two statistical methods have been combined to provide researchers and teachers with classifications of pupils based on their abilities and response patterns. This work therefore utilises state of the art statistical methods to produce outcomes which can feed directly into enhanced teaching and learning.

It is well known that tem response theory (IRT) provides an estimate of pupil ability as well as various item characteristics such as difficulty and discrimination (in the case of a two parameter model). This is useful because it allows a researcher to rank pupil performances based on item responses and characteristics rather than on just their total test score.

In this project, we have combined this with the statistical technique known as Latent class analysis (LCA). This provides a probability of class membership based on homogeneous response patterns. Class is considered to be an unobserved latent variable that is imperfectly measured. As such the technique calculates a probability class membership for each student. This differs from conventional cluster analysis which assumes the student either is or is not a member of the class or cluster.

In this study an exploratory approach was taken to find a solution that was interpretable, fit the data and tried to minimize 'inaccurate' class membership. The aim of analysis was to classify pupils based on e-assessment results to aid teaching and learning. The classifications needed to be sufficiently fine-grained and detailed to plan teaching and group children.

A latent variable model was defined for each test such that a pupil's chance of correctly answering an item was related to their ability and their latent class. All analyses were based on responses to at least 20 items. Ability was defined as an observed variable derived from a preliminary IRT analysis. Latent classes were determined such that the difficulty of each item could vary within classes. Latent classes were also allowed to have differing average levels of ability. Models were run for up to five latent classes and the most appropriate number of classes was decided upon using a combination of information criteria.

Models have been tested for 2, 3, 4 and in some cases, 5 classes. A variety of model fit statistics were examined to determine the most appropriate statistical models before curriculum and assessment experts began to investigate interpretation. The following provides an example of the constrained model for a four latent class solution for pupils taking the tests of reading.

Such analyses have been undertaken for a test of reading for ten year-olds, and this has identified four groups of readers with different patterns of performance, even though their raw scores could be the same:

- Reluctant reader
- Developing reader
- Reasoning reader
- Involved reader

As an example, ***reluctant readers*** are those who are beginning to get the gist of what they have read but are unlikely to be 'very interested' in what they read. They answer questions very quickly and have undeveloped inference and analysis skills. For the teacher, the next steps are to build enthusiasm by finding topics of interest

This can be contrasted with ***developing readers*** who may have the same mean score but generally get the gist of what they have read. They are also unlikely to be 'very interested' in what they read but are more persistent, spending some time considering questions. Their inference and analysis skills are also undeveloped but for them the teacher's next steps involve lots of reading and discussion of texts at an appropriate level and topics of interest.

Similar profiles can be given for reasoning readers and involved readers, each with their own next steps for teachers.

This analysis presents an innovative use of IRT and LCA methodology to provide formative assessment in a computer-based test. These analyses contribute to online feedback to pupils and teachers. E-assessment holds the potential to collect data other than item scores which could, in the future, contribute to the derivation of latent classes. Examples include time taken on each item, options selected, requests for help etc.

The potential of this methodology, combined with e-assessment technology, for providing robust yet accessible information to contribute to improved teaching and learning cannot be over-emphasized. It is essential that there are robust statistical methods available to underpin the assessment goals of the new forms of formative feedback to support personalised learning and teaching.

---

[i] Black, P. & Wiliam, D. (1998). *Inside the black box: raising standards through classroom assessment* (London: School of Education, King's College).

[ii] Assessment Reform Group (2002) *Assessment for learning: 10 principles. Research based principles to guide classroom practice* [online]. Available: http://arg.educ.cam.ac.uk/CIE3.pdf [7 June, 2005].

[iii] Department for Education and Skills (2004). *A National Conversation about Personalised Learning* [online]. (Available: www.standards.dfes.gov.uk/personalisedlearning)

Department for Education and Skills (2006). *2020 Vision: The Report of the Teaching and Learning in 2020 Review Group.* [The Gilbert Report] [online] (Available: http://publications.teachernet.gov.uk/default.aspx?PageFunction=productdetails&PageMode=publications&ProductId=DFES-04255-2006&)

[iv] This section draws heavily on the work of Hilary Cox and Juliet Sizmur, members of the NFER e-assessment development team.

[v] For example, Fisher's critical thinking, Lipman's philosophy for children and McGuinness's infusion methodology.

[vi] Guthrie, J. (2006). Engagement and expertise in reading: students' needs and classroom practices. Keynote address to the UKLA Reading Focus Day, University of Nottingham.