

The Large-Scale Application of a Computer-Delivered Adaptive Assessment for Children in Primary Schools

Paper Presented at AEA conference

Stockholm November 2007

Peter Tymms*, Christine Merrell*, Richard Hanna#, Brian Henderson* & Stephen Albone*

*** CEM Centre, Durham University**

CCEA Northern Ireland

Draft please to not cite without the authors' permission

Abstract

The recent experience of conducting large-scale trials with a view to using a computer adaptive instrument as a statutory assessment across a whole country has raised some important issues. In addition to the obvious hardware requirements within schools, the system requires a complex secure system to store and process data and a website through which schools can access the system. A national database of unique pupil identifiers to monitor progress over time is also essential. The professional development of teachers in relation to the assessment is a huge expensive undertaking. Teachers need training on the mechanics of using the program, on interpreting pupils' results and there have been political issues involving the use to which the data might be put in the background.

1 Introduction

It has been clear for years that the future of assessment lies in electronic forms and not in the paper-based delivery. But changes have been slow to happen. Perhaps the reason for this is that statutory assessments require security with new tests every year and this is not a system that lends itself to computer adaptive assessments or to delivery over the internet. What is described here is a first; a first for a statutory assessment using computer adaptive diagnostic assessments in which schools are required to do the assessment. They do the same assessment every year and the intention is that the information is to be used by schools to identify pupils' strengths and areas for improvement as well as to provide parents with good quality assessment information for their dialogue with class teachers. Formative assessment is already an important part of classroom practice in Northern Ireland, but the introduction of standardized statutory computer-based assessment is new. A lot of issues had to be solved in developing the system. One was the development of the assessment itself; how that would work and the kind of information to be given to schools. As with any systemic change, there were issues to address such as credibility of the assessment tool, comparability of outcomes with existing tools, teacher awareness and training. The system had to be embedded and at the same time the constant concern that the kind of feedback that is given to schools would have a positive impact. A decision has been taken that no data would be collated centrally across Northern Ireland so that no league tables could be produced, so that one school could not be set against another. This, of course, means that one needs to have an alternative way to monitor standards over time. In the paper what is dealt with is the way that the statutory system was set up, the way that the problems are being solved and some of the issues are set out in detail.

2 The Context

The statutory school curriculum and assessment arrangements in Northern Ireland schools have been revised under the terms of recent legislation; The Education (Northern Ireland) Order 2006. The introduction of the revised curriculum as well as new assessment and reporting arrangements is phased, beginning in September 2007 (The Northern Ireland Curriculum: Primary, NICCEA). Details can be found at <http://www.nicurriculum.org.uk> and <http://ccea.org.uk>

The revised curriculum is focuses on the development of skills and capabilities for life-long learning and for contributing effectively to society. These whole-curriculum skills and capabilities consist of Cross-Curricular skills, Thinking Skills and Personal Capabilities. The skills are embedded, providing opportunities for pupils to develop and demonstrate them in all areas of the curriculum.

A minimum statutory requirement for content is specified but schools are empowered to design programmes which best meet the abilities and interests of their pupils. This is in contrast to the prescriptive curriculum being superseded, characterised by compulsory content and lack of flexibility for schools.

New statutory computer-based assessment arrangements have been introduced in the Primary sector which focus on identifying pupils' strengths and areas for improvement

early in the school year. These arrangements replace statutory summative assessments at two key milestones, end of Key Stage 1 and end of Key Stage 2, in the Primary phase. The computer-based assessment tool initially specified by the Department of Education is InCAS.

Pupils with severe learning difficulties are not required to take the assessments and those with statements of special education needs which may specify moderate learning difficulties may be exempted by their school if this is considered appropriate.

All schools are required to offer parents an opportunity to meet with their child's teacher early in the school year to share the outcomes of these assessments, together with plans for the year ahead and discuss ways in which they can help at home. In this way, assessments are used as a context to inform teaching and learning in partnership with parents. In addition to the assessments, schools are required to provide parents with a written report of their child's educational and other achievements before the end of the school year.

3 The INCAS software

The InCAS software comprises a suite of independently developed assessments or modules. These can be broadly categorised into three main types as follows; measures of academic achievement, measures of academic potential, and attitudinal measures. The measures of academic achievement consist of reading, mathematics and mental arithmetic. Reading itself is further divided into word recognition, word decoding, comprehension and spelling. The measures of academic potential consist of English vocabulary and non-verbal ability, the latter taking the form of a task known as problems of position. For each of these cognitive measures the software uses an adaptive algorithm to match the difficulty of the assessment items to the ability of the individual pupil. The attitudinal measures are somewhat different in that the choice of items presented is dependent only on the age of the pupil. There are separate attitudinal measures for reading, mathematics and school in general. Details of the form of these assessments and the adaptive algorithms used are published elsewhere (Merrell & Tymms, 2007).

From the outset, the development of InCAS has been informed by rigid criteria that have had a profound influence on its design and implementation. Foremost amongst these criteria is that it should be easy to use by both pupils and teachers, and that the data generated should be of the highest possible quality. Something will be said about each of these in what follows.

InCAS has been developed for use by children as young as five years of age. The interface is designed in such a way that no prior experience of using a computer is assumed. All input into the computer by the child is conducted via a point-and-click device such as a mouse. The keyboard is used only under adult supervision to input passwords that give access to the assessment modules. This system gives teachers the necessary means to coordinate the activity of the pupils. A simple and intuitive interface allows pupils indicate their responses by clicking objects on the screen. Typically the choice of objects is limited to between three and five. All instructions for completing a particular task are given aurally and visually, and children are provided with practice examples before attempting the task proper.

An important consideration in developing a computer based assessment system such as this is that most primary schools do not have the advantage of permanent on site technical assistance. In addition, and for very good reasons, school computer networks tend to be tightly regulated to restrict both internal access permissions and communication with the outside world. It was therefore decided at an early stage that the web based functionality of the system be restricted to uploading completed assessment data, and downloading feedback. This is done via a secure central server located at the CEM Centre in Durham. In order to facilitate the day-to-day business of carrying out the assessments a copy of the software is installed separately on each client machine on the school network. Each client machine accesses a single database held at a central location on the network. The database contains biographical data of the pupils, which are copied from the schools management software, together with the assessment modules to be completed by each pupil. An IT technician sets all of this up in advance and generates a set of password lists. All that remains for the teacher to do is administer the assessments using those passwords. Additional pupils and assessment sessions can be added later if required using an intuitive user interface without the need for specialist technical support. This approach ensures a system that is robust, flexible and easy to use.

The ease of use of the system for teachers does not stop at administering the assessments. The use of adaptive algorithms, and the analysis of the data generated using Rasch theory, has allowed the development of feedback that is much more intuitive than that generated by traditional standardised tests. The Rasch approach allows both the assessment items and the pupils to be measured on the same equal interval scale. By gathering data from pupils over a broad age range it is a simple matter to equate that equal interval scale to one measured in units of time. The difficulty of an assessment item can be expressed in terms of the age at which a typically achieving child would have an equal chance of answering correctly or incorrectly. Pupil measures may be similarly expressed in terms of age-equivalent scores which can be compared directly with their chronological age. For the average primary school teacher the knowledge that a child is operating at a level one year below his actual age is of much more immediate and practical benefit than knowing he is one standard deviation below average. In addition it provides a very straightforward way to monitor the progress made over time.

The adaptive approach also allows us to monitor the quality of the assessment data collected at the level of the individual pupil, something that is not offered by traditional standardised assessments. With InCAS, individual errors of measurement are reported back to teachers in the feedback. The reporting of confidence intervals, together with a little training in what they mean, allows teachers to view statistics as best estimates rather than precise measurements. Not only does this make the data less threatening, it encourages reflection as they reconcile the assessment results with their other knowledge of each child.

From a practical test development perspective, targeting assessment items using adaptive algorithms increases the quality of the assessment because it allows all of the children to be measured with equal precision, independent of their ability. Considering the errors of measurement allows the quantification of the trade off between length of the assessment and its reliability. In general adaptive assessments are more reliable than standardised assessments of the same length because of the reduction in poorly targeted items. This

gives test developers the option of using the slack to shorten the assessment or to increase its precision.

Finally, consider the children's experience of using InCAS. On the whole, children seem to enjoy taking part in the assessments, although it is not clear how much of this is down to the novelty value of the exercise. Unlike traditional paper tests where all of the children receive a similar experience in terms of the questions asked, with InCAS they all get a similar experience in terms of the number of items they are able to answer correctly. Weaker children tend to report this as a positive experience, which is perhaps not surprising since they have been given the opportunity to demonstrate what they can do as much as what they cannot. By contrast, the more able children tend to report finding InCAS challenging.

4 Issues and strategies

The testing software is installed on a school network or on stand alone machines. This can be done either via a CD or by download. The school ensures that pupils do the required assessments and then the data are uploaded to the University where the data are processed and feedback can be downloaded.

4.1 Hardware requirements

Schools use a wide variety of different hardware therefore the InCAS software has been designed to run on as many systems as possible. As schools have limited budgets for technology many are still using ten year old computers. Although the software will run on a stand alone machine it is recommended that a network configuration is used where the software is installed on each client machine then a file server, where each machine can access, is used to store the results.

The minimum requirements for InCAS are a Pentium processor or better, 32MBs of Ram, Windows 95 or later, 100MBs of hard disk space, an internet connection, SVGA Monitor and Headphones are recommended ensuring that the software will run on the vast majority of Windows based PCs manufactured in the last ten years.

4.2 Secure data transfer

In order to set and improve the standardisation of items each year it is necessary to collect the data from the participating schools. As the assessment results are confidential it is also important to have a secure method of returning the data. To facilitate this process a secure web site has been setup where the results can be uploaded with the data being encrypted over the internet. This method also allows the schools to receive near instant feedback from the secure web site; again the feedback is encrypted to ensure complete confidentiality.

4.3 Data storage

It was decided at an early stage in the design to store the assessment software and results in a database on a local machine rather than over the internet. While the internet has

proved a very useful tool it could not be relied upon during assessments to deliver questions which require much data quickly and keep a connection live over the whole assessment period especially with such a large number of schools assessing concurrently. There can also be problems configuring firewalls when assessing over the internet which can cause a burden on any technical support.

The software is installed onto computer suites by technicians before any assessments take place. Through working closely with SIMS a system has been developed where pupil biographical details are imported from the School Management Software to save schools reentering the data which can be time consuming.

4.4 Pupil identifiers

A useful feature of the InCAS software is the ability to track pupils as they move through school. Simply using the pupil name would not be sufficient as the name may have spelling mistakes, a pupil may change have a name change or they may move to a different school. The Unique Pupil Number (UPN) which is assigned to every pupil in Northern Ireland is therefore used for this process as it is already used by schools and is readily available through the link with the SIMS School Management software.

The software generates and stores a unique identifier based on the school identifier (DENI number) and the academic year of a pupil starting school. A check character is included to ensure the validity of the UPN.

4.5 Professional development

The first year of using InCAS in Northern Ireland involved a single year-group of pupils in approximately 900 schools. Training was required for approximately 1800 school teachers in the first year with further year groups each year. For the assessment its feedback to be successful and useful, it was necessary for it to be incorporated into the education system, which required training and development. There were several areas of professional development for different tiers of professionals. The first area involved understanding the rationale behind the content of the assessment. The second involved being able to use the system; import pupils' biographical data into the program, set up assessment sessions, oversee pupil assessments, upload the assessment data to the secure website and download the feedback. The third area was to understand and interpret both pupil and school-level feedback, and to be able to use the information in to plan of appropriate work for individuals and classes. The different tiers of personnel were education officers working at national and regional level, head teachers and classroom teachers working at school and pupil level.

Changing teachers' ways of working through professional development is not easy and takes a significant amount of time (Chinn and Brewer, 1993, Desforges, 2000). The provision of in-service training, peer observation and coaching by professionals can be effective techniques (see for example Adey et al., 2004, Sparks, 1986) although Adey et al. suggested that 30 hours were required to change a teacher's practice, and a parallel length of time might be expected to learn and use InCAS and the feedback that it offers. Joyce and Showers (1980) reviewed the literature on the acquisition of new skills and

strategies by teachers. They classified the outcomes into different levels; awareness of a concept, acquisition of concepts and knowledge, and the ability to apply that knowledge to different situations. Raising awareness and then acquiring new concepts and knowledge can be achieved through direct teaching and lecturing but this does not automatically lead to the application of new skills and knowledge in the classroom. Application was found to be most successfully achieved by following the awareness and acquisition stages with modelling or demonstration, practice in simulated classroom settings, structured feedback and coaching. Joyce and Showers suggested that the demonstration of a new concept accompanied with discussion and then followed up with practice, observation and structured feedback leads to many teachers transferring new skills into their everyday practice. However, direct coaching on how to apply skills appears to be necessary for some teachers. This involves individual classroom support over a sustained period. The coach observes, encourages, models and corrects with the specific aim of transferring skills.

The way in which different tiers of professionals have been trained in the use of InCAS shares common features with Joyce and Showers' model. Firstly, Education Officers attended sessions to learn about the assessment program, the secure website where assessment data were uploaded and feedback downloaded, the various feedback formats and the remediation strategies for helping children with specific problems. The sessions included lectures, hands-on practice with the software and the opportunity for discussion. Following these sessions, the Officers contributed to the production of training material for the school staff and they were then required to teach the school staff, modeling the process, thus consolidating their understanding and ability to answer queries.

The first training sessions for school staff took place at the beginning of the academic year just before the assessments of pupils were completed. The Officers presented information about the rationale for the content of the assessment, the mechanics of how to use the software to carry out assessments and how to upload the assessment data for processing and feedback. Teachers practiced the procedures as part of their training before returning to their respective schools to complete the assessments with their pupils. Later in the term, after completing the assessments, the school staff were invited to a second session to learn about how to interpret and use the assessment feedback. This session consisted of learning through lectures and then discussing the performance of example pupils with colleagues. They were also given the opportunity to analyse data from their own pupils, thus applying theory to practice.

Each training session lasted one day and therefore all school staff completed two full days of training. Further guidance and support was offered through on-line tutorials.

4.6 Political issues

Children in Northern Ireland attend government funded Primary schools until age 11. The post-Primary phase consists of similarly funded selective and non-selective schools. The curriculum in both school types is the same. Entry to a post-Primary school is based on a range of criteria published by the school but only selective schools can use performance in the transfer procedure test as a criterion.

This transfer test is based on English, Mathematics, Science and Technology aspects of the Primary school curriculum. Taking the test is not compulsory but it is considered by many as being high stakes because of its role in the process of academic selection.

The process of transfer between Primary and post-Primary education in Northern Ireland is a matter of Government policy and currently the subject of political debate.

The assessments described in this paper are designed to inform learning and teaching and play no part in the process of selection on the basis of ability.

5 Conclusions

The implementation programme is at an early stage, but the signs are that the systems seem to be working well. The key two major areas are technical and human. Under the technical heading, we need to be confident that the system is operating well in schools; that it gives rapid feedback; that the pupils who transfer between schools can be accommodated into the system; that the schools can upload and download data well; that the headphones and the sound systems and the visuals are good; that we deal with those children with special needs; that we acknowledge there are some groups such as the profoundly deaf for whom this system does not work; and that all that technical background needs to have the confidence of people and to be generating useful information. Also on the technical side is the need to get the psychometrics right; to ensure that the data is sufficiently accurate; to produce feedback which is useful but not so time-consuming that it disrupts the life in the schools; and there is a tension to be had between the time taken to collect information and the accuracy of that information. At the moment, the system takes an hour and it will be a matter of judgment as to whether this is being got right as time progresses.

The second major area relates to pupils, teachers, parents, and Government policy. For the pupils, the aim is to have an experience which is enjoyable, which commands their attention sufficiently so that they are motivated to treat the tests seriously so that it produces good data; The assessment, being adaptive, should present items which are in the area that they find challenging, that is the zone of proximal development. To be successful teachers must consider the assessment useful and manageable. The data is intended to be used to support professional judgments about what to do if a child has decoding problems, what to do if the child has a text understanding issue, even though they can decode and word recognize.

The decision to use this assessment tool across the region was based on three trials, the outcomes of which were used to modify the tool. The implementation programme beginning in autumn 2007 is the subject of an evaluation which will explore a range of issues such as how useful teachers find the information informing learning and teaching. The revised Northern Ireland Curriculum focuses on assessment for learning and the InCAS tool plays a key role in supporting this policy.

School Principals expressed concerns about how the data might be centrally collated and used for other purposes such as comparing performance between schools. Such practice could result in pupils being coached or practicing assessments in order to improve

performance which would undermine the effectiveness as a diagnostic tool. The Department of Education has made it clear that data will not be used for such purpose. Alternative methods will be used for measuring the effectiveness of government policy to improve literacy and numeracy.

Parents will receive assessment outcomes at a meeting with their child's teacher and this will inform a discussion about plans for the year ahead and how parents might help at home. Details can be found at http://www.deni.gov.uk/information_for_parents.pdf

The voluntary use of standardized pencil and paper assessments for formative purposes is common in Primary schools in Northern Ireland and the use of a range of tools is encouraged by the Department of Education. However, the decision to specify an annual assessment to be used in all schools and for that to be computer-based is new. As such, there is much to learn about its effectiveness to support teachers' professional judgments and its role in improving standards.

6 References

Adey, P., Hewitt, G., Hewitt, J. and Landau, N. (2004). *The Professional Development of Teachers: Practice and Theory*. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Gardner, J. (2007). Review of "Devolution and pluralism in education in Northern Ireland" by Caitlin Donnelly, Penny KcKeown and Robert Osborne, (2005) Manchester: Manchester University Press. *British Educational Research Journal* **33**(1): 125-126.

Joyce, B. and Showers, B. (1980). *Improving in-service training: the messages of research*. Educational Leadership Feb □.

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

Showers, B., Joyce, B. and Bennett, B. (1987). *Synthesis of Research on Staff development: a framework for future study and a state-of-the-art analysis*. Educational Leadership.

Sparks, G.M. (1986). The Effectiveness of Alternative Training Activities in Changing Teaching Practices, *American Educational Research Journal*, 23(2) p217 – 225.