Diagnosing and Remediating Literacy Problems Using INCAS Software: Identifying Reading and Spelling Difficulties And Providing Help

Christine Merrell and Peter Tymms

CEM Centre,

Durham University

Contact: Christine.Merrell@cem.dur.ac.uk

CEM Centre,

Mountjoy Research Centre 4,

University of Durham,

Stockton Road,

Durham,

DH1 3UZ.

Tel. 0191 3344226

Fax. 0191 3344180

Running Title: **Computerised Assessment Of Reading**

Identifying Reading Problems With Computer-Adaptive Assessments

Abstract

Widespread access to computers and advances in psychometric theory have led to the development of adaptive assessments. These are computer-delivered assessments in which the program selects items appropriate to the ability of each student. The processing of the student's responses can be automated to provide rapid feedback on their performance including diagnostic information on specific strengths and weaknesses.

This paper describes the development of an adaptive assessment called InCAS (Interactive Computerised Assessment System) that is aimed at children of a wide age- and ability-range to identify specific reading problems. Rasch measurement has been used to create the equal interval scales that form each part of the assessment. The rationale for the structure and content of InCAS is discussed and then different formats of feedback supplied to teachers will be shown. This feedback is accompanied by research-based strategies for remediation, following the principle of 'assessment for learning', which promotes the provision of feedback on how to improve as well as test grades.

Key words and phrases: Adaptive Test, Computerised, Reading, Rasch, Remediation

Introduction

Traditional group assessments, administered on paper, tend to be limited in the range of item difficulties that can be included and yet frequently they are considered to be a suitable method for assessing children with a wide range of abilities. Typically they will have a few questions that are very difficult and a few questions that are very easy for the majority of the target group. This means that a child of low ability can find the experience damaging to their self-esteem and yet a gifted child will only be challenged by a small number, if any, of the questions. The amount of information provided about children at those extremes of the population is limited and their scores have higher margins of error – a point which is often missed by proprietary tests (Tymms 1998). By contrast, adaptive assessments present children with a higher number of items that are appropriate to their ability and far fewer questions that are either too easy or too difficult, thus having the potential to provide more accurate and reliable information in a shorter time. This method of assessment is difficult to deliver in paper format but could be achieved from an interview conducted on an individual basis. The problem with individual interviews is the time taken by the tester. If the tester happens to be a teacher with a whole class of pupils it may be impossible and so adaptive assessments are ideally suited to computer-delivery. A computer-delivered assessment applies pre-determined algorithms to select questions on the basis of children's previous performance to give a more individualised assessment (see for example Thissen and Mislevy 1990). Gradually the computer focuses in on the child's zone of proximal development and estimates the child's ability level - just the place for subsequent teaching. The child's experience is very different from a traditional test – it is challenging but within reach, and shorter.

Computer-adaptive assessments are not likely to be used for statutory assessment, nor high stakes tests such as GCSE and A-levels in England because their development requires a large number of items and every item from the easiest to the most difficult must be extensively tested in order to get accurate item characteristic information. The task of collecting this information requires far more effort than producing a traditional test, which already stretches the resources of examination bodies on an annual basis. If, however, the purpose of a computer-adaptive assessment is to provide information for internal use by teachers and not to hold schools to account then the investment required for its development is justified because it is not discarded after a single use. Computer-adaptive assessments fit perfectly into a philosophy that seeks to teach each student as an individual and which wants them to feel successful and yet sees the need to diagnose difficulties.

If the information from a child's assessment is to lead to an improvement in attainment, the teacher needs be able to interpret the information and then apply appropriate strategies. It has been found that giving students a grade or test mark without further information did not automatically lead to an improvement in their attainment whereas providing feedback on how to improve did translate into higher attainment (Black and William, 1998, Black et al. 2003). This theory could equally be applied to teachers. Providing them with scores from an assessment for each of their pupils does not necessarily provide a blue-print for improvement (see for example Penny and Coe 2004). Teachers need to know what kinds of strategies are required for particular children to help them improve and ideally these strategies should be linked to assessment results.

The Development of a Computer Adaptive Assessment

Having outlined some of the issues associated with traditional assessments and merits of computer-adaptive assessments, we turn to describe the methods used for the development of such a computer-adaptive assessment system; Interactive Computerised Assessment System (InCAS), developed by the by the CEM Centre at Durham University, UK (www.incasproject.org and www.cemcentre.org respectively).

The long term aim of InCAS is to develop a single computer program that could assess several developmental areas as well as the attitudes of children from age 5 years to the end of primary school (age 11) in England and in English-speaking international settings. Feedback will give a detailed profile of each child that would identify specific levels as well as problems and provide a basis for planning appropriate, personalised, programmes of study. Initially the development has focused on diagnosing reading problems and so the starting point for the assessment framework was to consider the elements required for fluent reading.

Readers need to be able to decode unknown words. Familiar words should be rapidly recognised without conscious decoding (see for example Harrison, 2004). A combination of these strategies enables the reader to decipher text but comprehension requires additional contributory factors (see for example Gersten et al. 2001). There is a strong relationship between vocabulary knowledge and reading comprehension and

this is reciprocal in that vocabulary knowledge contributes to reading comprehension and it grows through reading experiences. Gersten et al. also identified other factors that were important for fluent reading, namely the use of background knowledge, an ability to understand verbal communication and task persistence. The process of reading is an interactive-compensatory process (Stanovich, 1980) in that the time allocated to the processes of word recognition and comprehension is variable, according to the needs of the reader. The process of word recognition is a low-level cognitive activity and for good readers it is automatic much of the time. Not being a conscious action it takes up very little of the brain's processing capacity, which leaves the reader able to devote resources to the higher-level task of interpreting the meaning of a text. The process is 'compensatory' because more processing capacity can be allocated to decoding and word recognition when required, but at the expense of the interpretation of the meaning of the text. So, less fluent readers have to devote more processing capacity to word recognition and decoding, thus limiting resources for understanding the meaning of what they are reading. When readers become more fluent and able to recognise or decode words easily, they then use cognitive and metacognitive strategies to comprehend more demanding text. Cognitive strategies include re-reading difficult phrases or paragraphs, utilising background knowledge and adjusting reading speed when a difficult word or phrase is encountered. Metacognitive strategies are self-monitoring activities, which require an awareness of the skills, strategies and resources necessary to succeed, and self-regulation strategies - the control of those skills, strategies and resources so that effective performance is achieved. A further important element in the acquisition of literacy is learning to spell words correctly. Some children find it difficult to read and spell. Other children are good at reading but have problems with spelling. Either way, this slows them down in their written work, which means that they perform badly in written assessments. They have to work out how to express their ideas using words that are easy to spell. Older pupils who have spelling problems are at risk of taking inadequate notes, which also impacts on exam performance. When looking at a full profile of a child's strengths and weaknesses it is useful to include a measure of non-verbal ability, particularly if the child's first language is different to English. Perhaps poor reading is due to poor English acquisition or perhaps it is associated with a wider spectrum of special needs. Moseley (1976) developed the Problems of Position (POPS) test, which is a measure of non-verbal ability but it is also an indicator of potential spelling problems.

Although attitude to reading is not strongly related to attainment or progress in the primary years (Tymms, 1999), it is important for children to be able to express their feelings and it links to motivation, which leads to reading.

This understanding of the processes involved in literacy acquisition provided the structure for InCAS. From a management point of view, if the assessment was to be useful for assessing whole classes fairly rapidly, it needed to be presented in a child-friendly way that required a minimum level of adult supervision. Teachers had to be able to select particular sections for individual children so that younger children could do the assessment in several short sessions, and so that teachers could opt to miss some sections if they wished which meant that the program had to be modular. A modular program also meant that feedback could be provided on specific areas of cognitive development and reading processes so that appropriate teaching and remediation of problems could follow. InCAS was developed to include modules to assess English

picture vocabulary, non-verbal ability, word recognition, word decoding, reading comprehension, spelling and attitudes.

Below is a brief description of the format each of these modules:

English picture vocabulary The child hears a word and sees five pictures. They are instructed to select the picture that best describes the word.

Word recognition The child hears a word, and then a sentence, putting the word into context. They are instructed to select the target word from a choice of five presented on-screen.

Word decoding The child hears a nonsense or unfamiliar word. They are instructed to select the target word from a choice of five presented on-screen.

Reading comprehension The child reads a passage of text. Approximately every fifth word is presented as a choice of three. The child has to select the word from that choice of three that best fits within the overall meaning of the sentence. This requires children to use both cognitive and metacognitive strategies. If the child chooses the wrong answer this is sometimes associated with not understanding the meaning of the text and sometimes associated with not understanding the grammar or spelling. Having just three options in a multiple-choice format might seem less discriminating than if more are used, however a meta-analysis conducted by Rodriguez (2005) has shown three options to be the optimum number of answers for multiple choice questions. More three option items can be administered than 4 or 5 option items per testing time without having a detrimental effect on the psychometric quality of the test scores.

Spelling The child hears a word and then a sentence that puts the word into context. They spell the word by selecting the appropriate letters from an on-screen keyboard.

Attitudes The child hears a series of statements intended to assess their attitude to reading, and also to mathematics and school-life. They indicate a negative through to positive response on a sliding scale.

Non-verbal ability This is the Problems of Position (POPS) test, developed by David Moseley, that was described earlier. The child sees a pattern of dots and then has to find the same pattern within a more complicated arrangement of dots.

How InCAS works

Rasch scaling (see for example Bond and Fox, 2001) was used to create equal interval scales for each of the InCAS modules. This involved extensive testing of many items. Some items were taken from the PIPS assessments (www.pipsproject.org), also produced by the Curriculum, Evaluation and Management Centre and which had already been used extensively in pencil and paper tests whilst others were written specifically. Age-equivalent values were calculated for all items corresponding to the age at which pupils had a 50:50 chance of getting an item right. These values were used by the InCAS program for initial item selection in all modules except for Reading comprehension and Attitudes, which will be discussed later.

The algorithm that the InCAS program uses for several of the modules is to take the age of a child as its starting point and select an item that a child three years younger would have a 50% chance of answering correctly. In other words, it selects an easy question.

The program then presents progressively more difficult items so long as the child answers them correctly. After a predetermined pattern of incorrect answers, the program estimates the age-equivalent ability of the child on that module and then a second group of items is presented to refine the first ability estimate, followed by a third group of items.

The reading comprehension module follows a slightly different set of rules. Rather than selecting a passage for text comprehension on the basis of the child's age on assessment, the child's word recognition and decoding scores are used to select one passage of appropriate difficulty. The child works through the whole passage regardless of their responses, the reason being that it would make a very disjointed assessment if the passage disappeared in response to a number of wrong answers when the child was half way through it.

For the attitudes module, the child sees a fixed number of age-determined items regardless of their responses.

Each module typically takes 10 minutes to complete, varying slightly according to the ability of the child.

Analysis of Results and Feedback to Teachers

InCAS can be administered at any time during the academic year. Once children have been assessed, their data are uploaded to a secure website where they are processed by a fully automated system and immediate age-standardised feedback is available for teachers to download and print. It would be possible for the program to process data locally but there are several reasons for uploading the data. One is to facilitate regular updates to the standardisation. Items can change their characteristics over time and can become easier or more difficult. For example, several years ago the word 'cash' was quite a difficult item in the English picture vocabulary test for young children but with the increasingly widespread use of cash dispensing machines, it has become easier. The uploading of data also provides the opportunity to gather data on new items. If several new items are included in a module and the program randomly select one per child during the assessment, information is collected easily. The item bank is then gradually extended without the need for extensive trialling.

The feedback for teachers gives age-equivalent scores for each section on which a child had been assessed, except Attitudes. Teachers can see strengths and weaknesses of each pupil and focus on areas that require remediation. A booklet that links research-based strategies for the remediation of specific difficulties to InCAS profiles is included with the assessment package. The complete InCAS package provides teachers with children's scores and strategies for helping them to improve, in line with the recommendations for assessment for learning (Black et al (2003).

Although InCAS is a diagnostic tool for teachers, from a management perspective it can provide feedback in a convenient format for comparing cohorts and monitoring

standards over time. It also has potential as a research tool for evaluating the impact of interventions on reading. Figure 1 is an example of pupil-level feedback for teachers, generated from InCAS data.

Insert Figure 1 here

The blue squares represent the child's age equivalent scores for each module and the lines above and below are confidence intervals. The green line shows the age of the child when s/he did InCAS. In the example, the boy was almost 12 years old when he was assessed and his age-equivalent scores were all significantly lower than his chronological age. His spelling, word decoding and word recognition scores were more typical of a child of between eight and nine years of age, and his reading (comprehension) was slightly but not significantly lower. It is also interesting to see that both the English vocabulary and non-verbal ability were also significantly lower than would be expected for a child of his age. The accompanying guidance to teachers includes strategies and programmes for helping all aspects of reading.

Figure 2 is an example of class-level feedback.

Insert Figure 2 here

Box and whisker plots are used to present the range of scores of pupils in a single class for each module. The middle 50% of pupils are included in the box and the whiskers usually extend to the highest and lowest pupils. The line in the middle of the box represents the position of the median pupil of the class. The green line shows the mean age of the pupils at the time of assessment. The chart gives a comparison of the modules assessed. In Figure 2, there is a wide range of age-equivalent scores for each module and within that single class, differences of several years in the age equivalent scores of its pupils. If a school is located in an area which includes a high proportion of families for whom English is an additional language, it is possible that the pupils have low English picture vocabulary scores for their age and correspondingly low word recognition, word decoding, reading and spelling scores but perhaps higher non-verbal ability (POPS) scores.

Figure 3 is an example of school-level feedback for all year groups in a primary school for one of the InCAS modules such as English picture vocabulary.

Insert Figure 3 here

This chart shows the age-equivalent scores for a single module (word recognition in this case) for each year group in the school. Now the green line indicates how the chronological age would map onto the age equivalent score with a perfect match and the box and whisker plots show how far the pupils in each class deviate from that. The vertical representation gives the age-equivalent scores of the pupils. The horizontal

representation, in other words the width of the box, covers the chronological age-range of the middle 50% of pupils in the class. Take the youngest class, Year 1, whose chronological mean age is 6 years. The age-equivalent word recognition scores of those pupils ranges from 3½ to 8 years. Although the scores of each year group are distributed above and below the green line and each median score tends to be close to the green line, it is interesting to see that the age-equivalent score of some of the least-able pupil in Year 6 is very similar to the most-able pupil in the youngest year group.

As pupils move through the primary school and re-take InCAS on an annual basis, feedback is provided on their progress. This is in the form of a pupil-level longitudinal chart that plots the age-equivalent score for each module assessed against the chronological age over time.

Discussion

This paper has described a newly developed, computer-delivered, adaptive assessment that can be used to identify the strengths and weaknesses of children of a wide age and ability range which is currently available to schools in England and other English-speaking countries. The program is capable of delivering individualised assessments in schools to multiple computers at one time and the web-based feedback system is capable of handling data from tens of thousands of pupils. This system has advantages over a traditional pencil and paper group assessment in terms of the ease and time for administration, the ability to provide items appropriate to the ability of individual pupils, which gives more reliable results for pupils at the extremes of the normal population, and a reduction in administration, marking and analysis for teachers. However, if an assessment is to lead to an improvement in the performance of pupils, regardless of how sophisticated its method of delivery becomes, teachers need to know how to interpret the feedback and implement appropriate strategies. InCAS has incorporated that further dimension but there is still room for further development. It is intended that pupil profiles will be more detailed with an analysis of the errors made by children and correspondingly more detailed advice on how to help. Modules that assess aspects of mathematics are currently under development and there is no reason why other aspects of development for which research-based remediation interventions exist can't be included, for example an assessment of inattentive, hyperactive and impulsive behaviour.

References

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

Black, P., Harrison, C., Lee, C., Marshall, B. and William, D. (2003) *Assessment for Learning: Putting it into practice*, Open University Press: Maidenhead.

Bond, T. and Fox, C. (2001) *Applying the Rasch Model: Fundamental Measurement in theHuman Sciences*, Mahwah, NJ; Lawrence Erlbaum Associates

Gersten, R., Fuchs, L.S., Williams, J.P. and Baker, S. (2001) Teaching Reading Comprehension Strategies to Students with Learning Disabilities: A Review of Research, *Review of Educational Research*, **71**(2) 279 – 320.

Harrison, C. (2004) *Understanding Reading Development*, Sage Publications: London.

Moseley, D. (1976) *Helping with learning difficulties*, Course E201 OU: Block 10, (Milton Keynes: Open University).

Penny , A. R. and R. Coe (2004). Effectiveness of Consultation on Student Ratings Feedback: A Meta-Analysis. *Review of Educational Research* **74**(2): 215-153.

Rodriguez, M.C. (2005) Three Options Are Optimal For Multiple-Choice Items: A Meta-Analysis Of 80 Years Of Research, *Educational Measurement: Issues and Practice*, 24(2) p3 – 13.

Stanovich, K. (1980) Toward an interactive-compensatory model of individual differences in the development of reading fluency, *Reading Research Quarterly*, 16, p32 – 71.

Thissen, D., & Mislevy, R. J. (1990). Testing Algorithms. In H. Wainer (Ed.), *Computerized Adaptive Testing: A Primer* (pp. 300). Hillsdale, Hove and London: Lawrence Erlbaum Associates.

Tymms, P. B. (1998). Opening a can of worms: a critical examination of age-standardised scores. *British Journal of Curriculum & Assessment, 8*(3), 21-25.

Tymms, P. (1999) *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*, David Fulton Publishers: London.

Figure 1 Pupil Profile

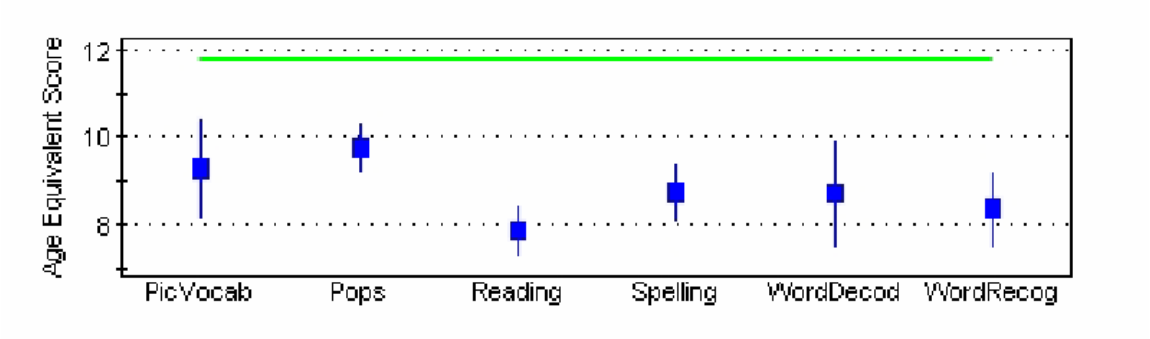

Figure 2 Class-Level Feedback

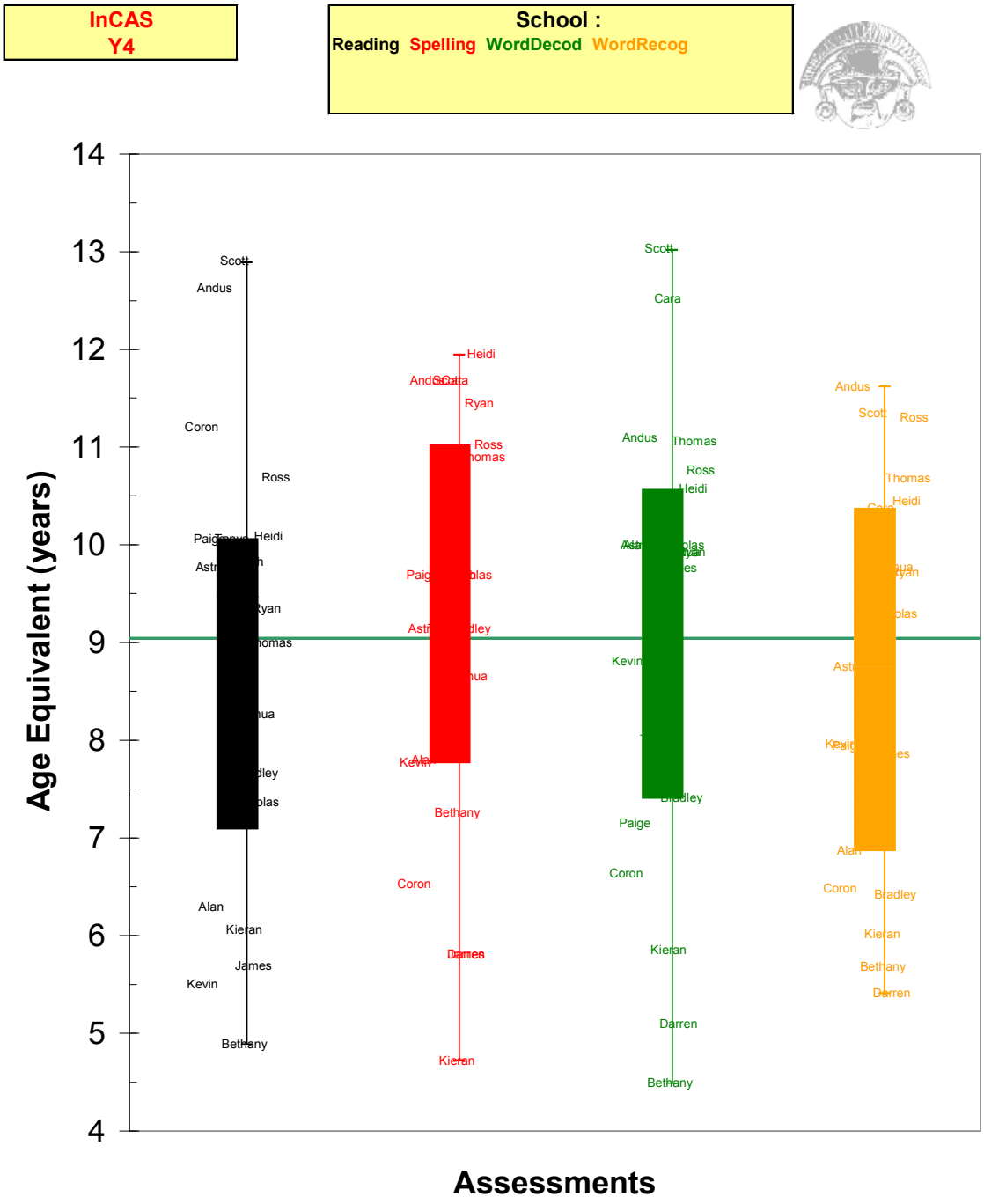

Figure 3 School-Level Feedback