InCAS (Interactive Computerised Assessment System): Using Individual Diagnostic Profiles In Assessment For Learning

Christine Merrell and Peter Tymms
CEM Centre,
Durham University

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

Paper Presented at EARLI Conference, Nicosia, Cyprus
August 2005

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 specific strengths and weaknesses.

This paper describes the development of an adaptive assessment known as InCAS (Interactive Computerised Assessment System) that is aimed at children of a wide age- and ability-range to identify specific reading problems. The rationale for the structure and content of InCAS is discussed and then the way that Rasch measurement has been used to create the equal interval scales that form each part of the assessment will be explained. Finally, examples of the feedback for teachers and the subsequent strategies for remediation will be shown.

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 full class of pupils it may become 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 identifies 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.

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 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 Cohen 1980). 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, this paper will now 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).

When starting on the development of InCAS, the aim was to develop a single computer program that could assess several areas of development and the attitudes of children from age 5 years to the end of primary school (age 11) in England and in international settings. Feedback would give a detailed profile of each child that would identify problems and provide a basis for planning appropriate, personalised, programmes of study. Initially InCAS had a particular focus 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 a text but to comprehend it 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 pointed out the importance of the use of background knowledge, ability to understand verbal communication and task persistence within fluent reading. The process of reading is an interactive-compensatory process (Stanovich, 1980). The time allocated to the processes of word recognition and comprehension is variable, according to the needs of the reader. For good readers, the process of word recognition is automatic much of the time. Not being a conscious action it takes up very little of the brain's processing capacity, which leaves plenty of resources for 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 means an awareness of the skills, strategies and resources necessary to succeed, and self-regulation strategies, which is 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, the assessment 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 they could opt to miss some sections if they wished. Therefore the program had to be modular. The program also had to be modular in order to provide feedback 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 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/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.

Spelling The child hears a word and then a sentence, putting 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 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 an equal interval scale for each of the InCAS modules. This involved extensive testing of many items some of which were derived from the PIPS assessments (www.pipsproject.org) and some of which were written specifically for InCAS. 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 program for item selection in all modules except for reading comprehension and attitudes, which will be discussed later.

The algorithm that the 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 a first item that is quite easy. The program then presents harder and harder items as the child answers items correctly. After the child had given a number 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 and then a third.

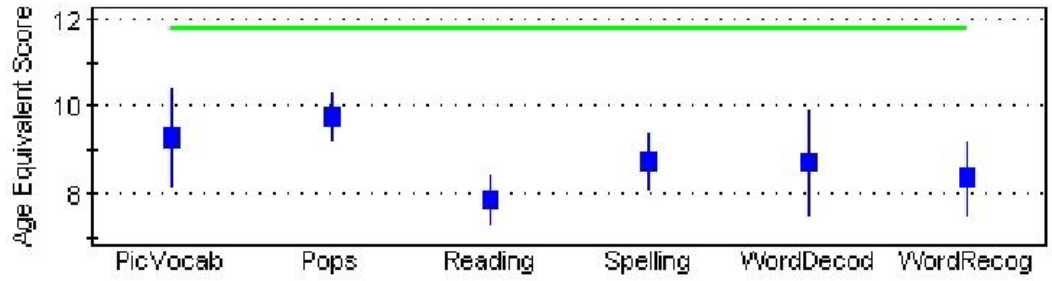
The reading comprehension module followed a slightly different set of rules. One passage of appropriate difficulty is selected on the basis of the child's word recognition and decoding scores. The child works through the whole passage regardless of their responses, the reason being that it would make a very disjointed assessment if the passage disappeared in response to a number of wrong answers if the child was half way through it. For the attitude 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 was available for teachers to download and print. This feedback gives age-equivalent scores for each section on which a child had been assessed. Teachers can see strengths and weaknesses of each pupil and focus on areas that require remediation. A booklet that linked 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 of Black and his colleagues.

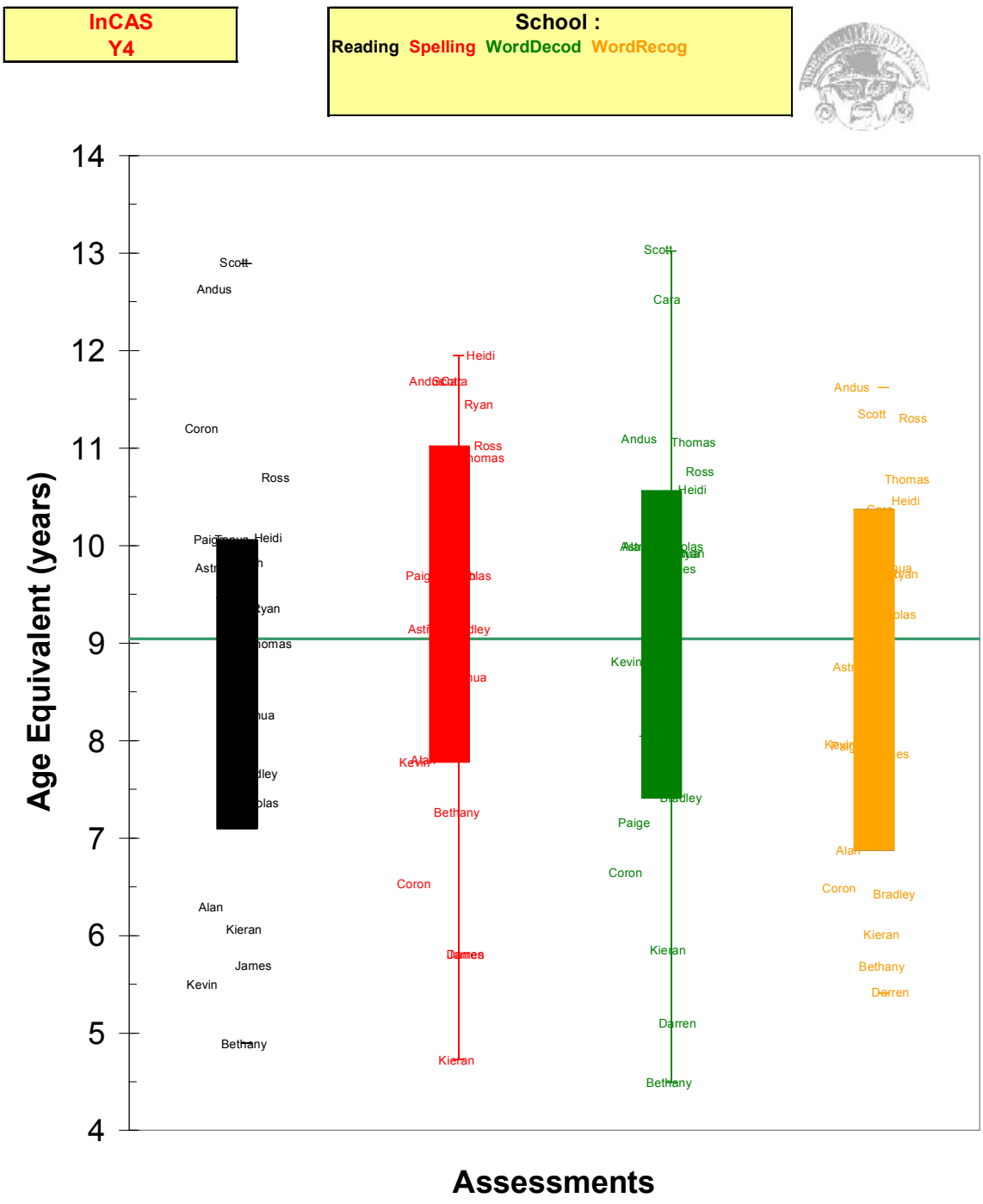
Although InCAS is a diagnostic tool for teachers, from a management perspective it can provide feedback in a convenient format for comparing cohorts and monitoring

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



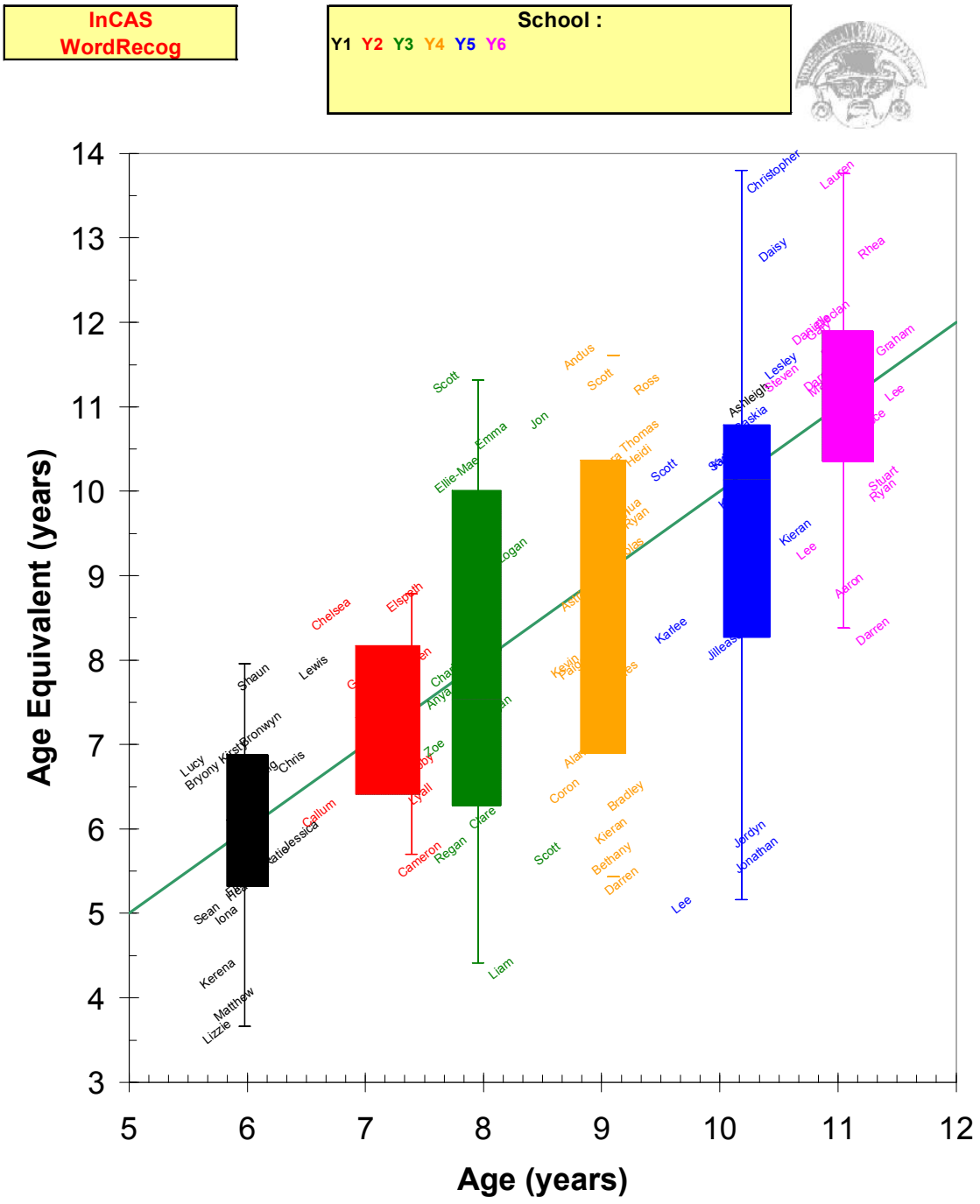
The blue squares represent the child's age equivalent scores for each module and the lines above and below are confidence intervals. The green line shows the age of the child when they did InCAS. In the example, the child was almost 12 years old when s/he was assessed and her/his age-equivalent scores are all significantly lower than their chronological age. The accompanying guidance to teachers includes strategies and programmes for helping all aspects of reading. It is also interesting to see that both the English vocabulary and non-verbal ability are also significantly lower than expected.

Below is an example of class-level feedback:



These box and whisker plots show the range of scores of pupils in a class for each module. The green line shows the mean age of the pupils at the time of assessment. The chart gives a comparison of modules. In the example above, there is a wide range of scores for each module and within that single class, differences of several years in the age equivalent scores of its pupils.

Below is an example of school-level feedback for all year groups in a primary school for one of the InCAS modules.



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. There are no particular issues for any year group but it is interesting to see that the age-equivalent scores of some of the older pupils in Year 6 are the same as some of the pupils in the youngest year group (Year 1).

Future Developments

The modular structure of InCAS means that further assessments can be added to the existing suite and a diagnostic mathematics assessment is currently under development.

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 Wiliam, 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

Cohen, P. A. (1980). Effectiveness Of Student-Rating Feedback For Improving College Instruction: A Meta-analysis Of Findings. *Research in Higher Education, 13*(4), 321-341.

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).

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.