USING CAT FOR 11-PLUS TESTING IN NORTHERN IRELAND:

WHAT ARE THE ISSUES?

Dr Pamela Cowan

Using CAT For 11-Plus Testing In Northern Ireland: What are the Issues?

Dr Pamela Cowan School of Education Queen's University Belfast p.cowan@qub.ac.uk

Abstract

This paper discusses the current concerns surrounding the psychometric properties of the Transfer Test used in Northern Ireland to select pupils aged 10-11 years old for a grammar school education. It highlights the lack of validity and reliability in the current selection system and offers computerised adaptive testing as the viable alternative for academic selection which reduces inequities associated with coaching and meets the international standards of validity and reliability.

Introduction

Since 1947 the majority of schools in Northern Ireland (NI) have been operating a two-tier system of selective secondary education, commonly referred to as secondary and grammar schools. Places in the grammar schools are awarded on the basis of a Transfer Test, also known as the '11plus test', taken at two unique times in the P7 year (final year of primary schooling). These tests are supposed to measure one or more of 'ability'. 'achievement' or the 'potential to benefit from a grammar school education'. Gardner and Cowan (2005) completed a detailed psychometric analysis of these tests and revealed that they were lacking in validity and reliability as defined by the American Educational Research Association's Standards for Educational and Psychological Testing (AERA, APA and NCME, 1999) and that only 18 marks out of a possible 150 marks spanned the top grade (Grade A which secures a grammar school place) and the bottom grade (Grade D, commonly known as a 'Fail'). This paper outlines the concerns raised with the current Transfer Test and offers an analysis of the variety of alternative selection mechanisms currently under review by the educational bodies in NI including the use of Computerised Adaptive Testing for primary school pupils.

The problems with the current 11-plus test

The Transfer Test comprises items addressing the NI Curriculum requirements in mathematics, English and science with the overall proportion

of marks awarded being in the ratio 26: 26: 23 respectively for these subjects. Since the Test determines the next stage in a child's education, they are viewed as 'high stakes' by parents, teachers and pupils. Although three subject areas are assessed, only one final grade (A, B1, B2, C1, C2 or D) is awarded summarising the scores in all three areas across both Tests. The technical fidelity of the Transfer Test was investigated using the following research questions:

- Are the tests unidimensional and therefore capable of differentiating pupils on a single construct, ability?
- How do children perform in the test?
- Are the tests successful in grading children accurately?

A random stratified sample of 52 primary schools of various sizes and school management types was used resulting in 1288 Test 1 scripts, 1270 Test 2 scripts and 623 Supplementary papers being returned. The Supplementary paper is only used if a pupil is absent from either Test 1 or Test 2. Confirmatory Factor Analysis, CFA, was used to test the null hypothesis that the proposed one construct model fits the observed data. The results were deemed to be 'safe' as the sample size was in excess of 200 (Boomsman, 1987). Using Joreskog and Sorbom (1989) the following limits were defined:

- $\chi^2/df < 2.0$
- RMR < 0.05
- AGFI > 0.8

The one construct model, namely that the 'Test' captured 'the pupils' ability to benefit from a grammar school education', was tested for:

- the whole sample
- boys only
- girls only

and this model failed on the $\chi^2/df < 2.0$ criterion in every case.

Comparable tests for these three categories (whole sample, boys only and girls only) were conducted with the 3-construct model, that the test measured pupil performance in 'Mathematics', 'English' and 'Science/Technology', and for all three categories the 'goodness of fit' criteria listed above were met.

Nonetheless the disattenuated correlation coefficient between each of the 3 constructs indicates high levels of correlation (>0.8) as shown in Table 1.

Test/subject	Maths &	Maths &	English &
area	English	Science/Technology	Science/Technology
Test 1	0.853	0.915	0.898
Test 2	0.872	0.870	0.899
Supplementary	0.816	0.837	0.940
Table 1	Disattenuated correlation coefficients		

A possible explanation for these high correlations is the phenomenon known as the Positive Manifold Effect whereby pupils no longer view the subjects as separate entities but as one interconnected unit called the 'Test'. As Primary School children are being taught by the same teacher, with the same teaching style and equal emphasis on each of the subject areas, the pupils find it difficult to distinguish between these three subjects resulting in a blurring of the boundaries. Messick (1989) advocates not collapsing different constructs or domains into a single measure even if they are highly correlated. Consequently, the results of the CFA show that the Transfer Test does <u>not</u> measure a single construct and treatment of the test score as a single measure, combining scores in the three subject areas, is open to question. As a result there is no evidence that the scores from the Test can be used to infer 'ability' or 'the potential to benefit from a grammar school education'.

Test performance and grades

Frequency analysis of the Test scores showed that over 65% of pupils answered over 70% of the test items correctly yet this score equates to a Grade D ('Fail'). Clearly the 'easiness' of the Test lulls the pupils into a false sense of security in which they feel they have done well. On average pupils are correctly answering 70% of English and maths items and 83% of science items. This aspect of the test design is not acceptable for 'high stakes assessment'.

In terms of the grade allocations, the top 25% of pupils get a grade A which should secure them a place in a grammar school, the next 5% of pupils' scores are each awarded grades B1, B2, C1 and C2. The remaining 55% of pupils are awarded a Grade D (generally viewed as a 'Fail'). Due to the perceived 'easiness' of the test and the clustering of scores above 70%, the actual number of correct items required for a grade A in this sample was 123 out of 150, grade B1 ranged from 119 to 122 inclusive, while B2 was 116 to 118, C1 was 112 to 115 and C2 was 106 to 111. Scores of 105 or below out of a possible 150 marks were awarded a Grade D. With only 18 marks separating the highest and lowest grades it is imperative that the test is measuring accurately and the reliability is high. The Standard Error of Measure (SEM) of the test provides an indication of the precision with which the observed (raw) score reflects the pupils' 'true' score. With a SEM of 4.75 for the combined Test 1 and Test 2 scores, the 95% confidence interval reveals that the pupils' true score may be between 9 and 10 marks above or below their actual scores. Due to the close proximity of the grade boundaries, the potential misclassification for the Transfer Test is up to 3 grades and could effect over 30% of pupils (Please, 1971).

To date, the Transfer Test is not underpinned by any published standards of practice or technical fidelity. If international standards for educational and psychological measurement (AERA, APA and NCME, 1999) were applied to this test, two standards would raise particular concern:

Standard 1.2 (on validity)

The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intending to assess should be clearly described.

Standard 2.1 (on reliability)

For each total score, sub-score or combination of scores that is to be interpreted, estimates or relevant reliabilities and standard errors of measurement or test information functions should be reported.

Given the 'high stakes' nature of the Transfer Test in NI, issues of technical fidelity should be addressed. It is clear from this research study that serious concerns exist around the grading system embedded in the current Transfer Test. At present, debate is raging on whether or not academic selection should be retained and if it is, how can these issues of technical fidelity be addressed for future young people in NI.

What are the solutions?

If academic selection is removed then parental preference will prevail. Parents will be encouraged to choose a school which best suits the needs of their child. This informed choice will be dictated by the Pupil Profile which is to be completed for every pupil in the primary school. The teachers will be required to complete and award levels in Communication (language and literacy), in Using Mathematics (mathematics and numeracy) and in Using ICT for each pupil. To supplement this record of academic achievements, the teachers will also be commenting on the pupil's Thinking Skills and Personal Capabilities, The Arts, Personal Development and Mutual Understanding, Physical Education, The World Around Us (Science, Geography and History), Religious Education, other interests and strengths and any other comments.

Although a database of pre-defined comments and phrases has been created, teachers are concerned about the time needed to complete each profile and also the subjective nature of the comments. Parents have been consulted and raised issues about the purpose and role of the pupil profile and how it could take into account the child's development over time – as soon as the profile is completed it is effectively out of date as the child will have moved on in his or her educational development. Parents viewed the pupil profile as guiding their decision-making regarding the 'best school' for their child however many parents found it difficult to interpret the content of the pupil profile and so training is needed for the parents.

If academic selection is retained or if schools are free to use academic selection if they wish then the key issues requiring attention are the international standards of validity and reliability of any tool used to select pupils into a post-primary school. However additional criteria have also been uncovered such as the need to accommodate pupils of all social and ethnic

backgrounds, the need to remove the pressure and anxiety associated with the Transfer Test as expressed by pupils, parents and teachers (Sutherland, 2000 and Save the Children, 2001), to allow parents to monitor their child's progress over the final years of primary education and to minimise the impact of coaching to ensure equity for all to all.

Some of the options under consideration include the use of NFER standardised tests in English and mathematics which will be administered on a specific day to all pupils, that is examination days like the current Transfer Test arrangement. This will not address the issues of validity for socially disadvantaged pupils and a new set of test papers would have to be created each year to prevent coaching. The stress on pupils, parents and teachers would remain as all tests would have to take place on a specific day.

A second alternative under consideration is the 'test when ready' facility of an Instructional Database Management System (IDMS). Unlike the NFER tests, IDMS has a large existing databank of test items for English and mathematics so pupils can have multiple attempts at the test. This option would reduce the feelings of pressure experienced on the 'test day' however it will not address validity issues for the socially disadvantaged child who has less access to coaching.

The University of Durham have been involved in the creation and use of an InCAS system (Interactive Computerised Assessment system) for English and mathematics which could be used to supplement the qualitative information provided in the Pupil Profile and to verify teacher assigned levels. Although the test is computerised, the system is not classified as a computer adaptive test and the website warns against making high stakes decisions based on the outcome of the assessments. This system would be a viable option for formative assessment and would assist the pupils, parents and teachers in determining a child's progress over time.

The final option under consideration is the use of computerised adaptive testing (CAT) which meets the standards of technical fidelity and removes the option of coaching to the test thereby ensuring equity to all pupils. Using Item Response Theory the test can be tailored to each pupil and so feelings of pressure and anxiety will be minimised as difficult questions are not administered. Also the pupils obtain instant feedback and due to the reduced assessment time, pupils can take the test when they are ready and repeat it as often as they wish. It is advocated that the CAT option could be taken at regular intervals in the final three years of primary schooling so parents will also have information about their child's progress over time and the results can be used for formative assessment prior to the P7 year. Although primary schools in NI are connected to the C2K network which offers secure 24/7 access via the external portal.

Discussion

The use of computerised adaptive testing (CAT) for assessing pupils' mathematical attainment against the levels of the NI Curriculum for end-of-Key Stage assessment has already been demonstrated (Cowan, 1997) however its use for province-wide assessment leading to **high stakes** decision-making has yet to be piloted. International research in the use of CAT has focused on its role for admissions to US graduate programmes (GMAT and GRE), and formative and summative assessment in UK Higher Education courses however there appears to be a deficit of research into the role of 'high stakes' CAT with school age children and in particular with primary school pupils. If UK examination bodies are aiming to include on-screen assessment for all new qualifications and at GCSE and A level "*by 2009, e-assessment should certainly be normal, if not the norm, for thousands of students each year*" (Boston, 2004) then perhaps educators should consider preparing these students at an early age for high stakes assessment of this nature.

Worldwide, the pupil's age may vary for transferring from the first school to a second school however all pupils have the right to be assessed with validity and reliability as they make the transition to the next stage in their education. The name 'Transfer Test' may be synonymous to the NI context, however all pupils undergo some form of assessment as they move from one school to another whether it is an entrance exam or a decision made by parents or informed by teachers. Consequently the concept of using CAT to provide all pupils with the opportunity to demonstrate their ability with a high level of technical fidelity and consequently to be considered for a school which best matches their needs rather than based on assumptions from other adults, seems to be a child's human right.

Conclusions

In terms of addressing the concerns raised in the Gardner and Cowan (2005) paper, CAT appears to be the only option meeting the call for international standards of validity and reliability. At the same time the nature of the CAT process facilitates a 'test when ready' approach which minimises stress and pressure on pupils and teachers alike while also minimising the impact of The content domain for the items will not distort the primary coaching. curriculum as teaching to the test is almost impossible. Pupils with special educational needs can be accommodated via the use of voice-overs and teacher time is not wasted creating lengthy pupil profiles with limited use to parents and post-primary schools. Since the tests are delivered online there is no need for extensive in-service training and moderation, a simple explanation of the -3 to +3 range for the scores is all that is needed for teachers and parents to interpret the pupil's test scores. Areas of strengths and weaknesses will be evident from the tracking facility within the software and detailed feedback against categories of test items can be provided for formative purposes. This system could be used over the final few years of primary education and weighted scores could be calculated to summarise

pupils' achievement over time. By making it an integral part of the primary school assessment system, pupils from all social backgrounds would have the opportunity to gain a place in the grammar school as 'opting out' would not be an option! So what is stopping the introduction of this CAT system as a means of academic selection in NI? At present it is 'cultural obstacles' (Hambrick, 2002) and politicians!

References

- AERA, APA, NCME (1999) Standards for educational and psychological measurement. Washington, DC: AERA with APA and NCME.
- Boomsman, A. (1987) The robustness of maximum likelihood estimation in structural equations models. In P Cuttance & R Ecobs (Eds) *Structural modelling by example: applications in educational, sociological and behavioural research* (New York: Cambridge University Press).
- Cowan, P. (1997) Using information technology to assess and report mathematical attainment against a scale of standards. Unpublished doctoral thesis.
- Gardner, J. and Cowan, P. (2005) The fallibility of high stakes '11-plus' testing in Northern Ireland. *Assessment in Education*, 12, 2, 145-165.
- Hambrick, K A (2002) Critical Issues in Online, Large-scale Assessment: An Exploratory Study to Identify the Issues. Capella University
- Joreskog, K G and Sorbom, D (1989) *Lisrel 7: a guide to the program and applications.* Chicago, IL: SPSS Inc.
- Messick, S (1989) Validity. In R L Linn (Ed) *Educational measurement*, New York: American Council on Education. 13-103.
- Please, N W (1971) Estimation of the proportion of examination candidates who are wrongly graded. *British Journal of Mathematical and Statistical Psychology*, 24, 2, 230-238.
- Save the Children (2001) *Thoughts on the 11-plus: a research report examining children's experiences of the Transfer Test.* Belfast: Save the Children Fund.
- Sutherland, A (2000) Interviews with groups of Year 8 pupils. In T Gallagher and A Smith (Eds) *The effects of the selective system of secondary education in Northern Irelan*d. Bangor: Department of Education.