

Towards a common assessment for academic selection

Discussion Paper

February 2017

Introduction

At present the great majority of pupils transfer from primary school at age 11 to Northern Ireland's secondary schools. They may be selected to go to a grammar school. In almost all cases selection is based on an assessment. Up until 2008/09 (almost 10 years after the Belfast Agreement) the selection test was produced officially by the Council for the Curriculum, Examinations and Assessment (CCEA) and it was taken by pupils in their primary school. In 2007, the then Minister for Education's announcement that the selection test was to be dropped created a flurry of activity as the supporters of Grammar schools sought ways forward. The outcome was that about half of the Grammar schools now use an assessment devised and run by the newly created **Association for Quality Education (AQE)** and about half of the schools use an assessment commissioned from GL by the newly created **Post Primary Transfer Consortium (PPTC)**. It is worth noting that a number of grammar schools use both tests. The new arrangements are private and the opt-in assessments are run on Saturdays in Grammar schools.

It is generally recognised that the dual system of two assessments is as a result of the circumstances which applied at the time. The majority of people we spoke to would like to see a single assessment. The Minister requested that the present report be produced as a move toward a single system.

Our perception is that both assessments work well, both are professionally produced and both select pupils who attend Grammar schools that get excellent GCSE results. Although we found that schools were loyal to the particular assessment that they were involved in, the Principals of every school that was visited agreed on the need to work towards a common assessment. There are however some barriers to be overcome.

The major barriers are: the payment method; the number of assessments and how many days should be allocated to take them; and the format of the assessment – should it be based on machine marked multiple choice items or manually marked free response and multiple choice items?

Although previous attempts to move to a single assessment were not successful there is reason to believe that the present attempt may be more fruitful. Although there is differing opinion across political parties regarding the issue, the Minister for Education has been explicit in his stance on selection and has been aware of the importance of the issue to parents since the decision to abolish the selection test was taken in 2007. Many parents have expressed the desire for a single assessment and the views expressed during the production of this report all indicate that a way forward can be found.

This report outlines a purpose and set of principles which may be helpful in considering the way forward and suggests the structure of a single assessment for discussion.

Authors: Peter Tymms, David Galloway and Paul Latham

Key Findings

- 1) There was general agreement that a move to a single selection test would be a good thing. As was said more than once '*children must lie at the heart of any changes*'.
- 2) The existing assessments are effective and have the backing of the schools that use them.
- 3) The major differences of opinion about the two systems centre on costs (who should pay), assessment arrangements (numbers of assessments and number of occasions) and the format of the assessments.
- 4) There are a number of psychometric issues associated with assessments which were not raised but which are important; however time did not allow for these to be analysed in full.
- 5) Strong views were expressed that the assessments have an impact on primary school practice and on families; they cause stress and they cost financially.

A move to a new single assessment would not resolve all the outstanding issues but it has the potential to make improvements. It would have to have the support of the great majority of grammar schools and gain their confidence. Knowing that previous attempts to move towards a single assessment have failed, it must be acknowledged that any move will not be easy. **But there is good will and it does not, to the outsiders who wrote this report, seem that the two groups are very far apart.**

It is our view that any move towards a single assessment must come from the selective schools themselves. They have the organising power, the necessary knowledge and a desire to improve education in Northern Ireland. The backing of the Education Minister will hopefully act as the necessary catalyst for the move.

Proposal for the purpose and principles behind and focus of a common assessment

The existing assessments have widespread support and any new arrangement should build on existing good work to create a single instrument which can be used across Northern Ireland. With that in mind, some very high standards are set out below as aspirations. Before that the purpose of the assessment is defined.

Purpose of the assessment

To provide unbiased data to schools to allow them to select the students who are most likely to benefit from study at their school.

Principles

In order to ensure the highest international standard of assessment, developers and assessment commissioners should refer to the "Standards for Educational and Psychological Testing" developed by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education. (<http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition>)

Development:

- a) The assessment should be trialled outside Northern Ireland prior to use and checked for its psychometric properties including reliability, discrimination and differential item functioning by major sub-groups. The exercise should be used to make adjustments to the proposed assessment.
- b) Every year the new assessment should be considered, against the stated purpose, prior to use, by a group of experts; the variety of schools should be represented on this group as should at least one independent assessment expert.
- c) Three assessments should be prepared together with a contingency assessment (e.g. in the event of unforeseen circumstances) and a supplementary assessment (e.g. due to unavoidable pupil absence).

Predictive validity: The assessment should aim to be strongly predictive of academic success up to GCSE. (This should apply to all relevant sub-groups as well as to the cohort as a whole.) Data should be kept and matched to GCSE results so that lessons can be learned.

Reporting:

- a) The assessment should generate accurate data for all the cut-scores used by selective / partially-selective schools.
- b) A technical report should be made available to the assessment body before the scores are released which, amongst other things, gives precision data (95% confidence intervals) for all scores on the assessment.

Marking:

- a) Any manual marking should be carried out blind. That is to say details of students (names, school, sex, address etc) should not be visible to the markers.

Location:

- a) It was clear from the outset in discussions with various stakeholders that the risks associated with the security and confidentiality of assessment papers can be significantly reduced by continuing to administer the assessments in post-primary schools.

Fairness and openness

Several of the points above will help to ensure that the assessment is as fair as possible but some of the Standards are outlined below to give an idea of the stringent requirement of a assessment which has the respect of the international community. (The wording of the original has been adjusted for a non-American lay readership)

Standard 3.2 The assessment developers are responsible for ensuring that the assessments measure what they intend to measure and are not biased in favour of one group or another because of the wording that is used, the format of the questions or the way in which students are expected to record their responses.

Standard 3.3 The assessment developers should include all relevant groups (both sexes, children with special needs, those for whom English is a second language as well as children from less and more affluent backgrounds) when conducting preliminary studies which are used when preparing the assessment.

Standard 4.2 As well as describing the intended use of the test, the assessment specifications should set the content of the assessment, the proposed length of the assessment, the format of the questions and the order of the sections. It should also set out the desired range of difficulties and discrimination of the questions, as well as the expected reliability the test and other relevant measures of validity.

Proposal for a specification for a “Selection Test”: For discussion

The outline which follows is solely intended to form the basis for initial discussion.

It aims to take the best aspects of the two existing assessments and to suggest ways forward which, with adjustment, will be acceptable to the vast majority of selective schools. It includes some new features which are intended to enhance the procedures using state of the art thinking derived from the "Standards for Educational and Psychological Testing" mentioned earlier.

Number of papers and focus

The assessment should consist of three papers. The first two will be direct, machine marked, multiple choice assessments of the parts of the English and mathematics curriculum which are amenable to such assessment. The third paper will have three parts: an open ended writing section, a mathematical problem solving section and a non-verbal, non-numerical, curriculum free, cognitive assessment. This third paper aims to assess important features of the curriculum which cannot be assessed through multiple choice questions and to provide an insight into students' cognitive functioning which is influenced as little as possible by schooling and home background. Each paper would start with example items.

Nature and number of items

Each paper will last 45 minutes and aim to produce the most accurate picture of pupils possible, in the time available. As a general rule this should involve as many items as possible.

The multiple choice papers should each aim for more than 50 items. This might seem a tall order but the evidence suggests two ways of making better use of the time available to pupils. First, three options in multiple choice items is optimal, cutting down the time needed for four or five options. Second, there are good techniques available which can assess comprehension without students having to read time consuming passages and then answer a few items about the passage.

The third paper will, inevitably involve fewer items but it should remain an aim to generate as much data as is commensurate with the requirements. For example a carefully designed marking scheme for a set of maths problems would be helpful. For the open ended writing section a series of diverse short writing tasks should ensure that a lot of data can be collected in a short space of time.

Number of occasions

The three assessments can all be taken on one day with an option for another sitting to permit special circumstances e.g. bereavement, illness etc that would affect performance on the day in question.

Reporting

Each section (English, mathematics, writing, maths problem solving and non-verbal ability) will be reported on standardised scales with the same mean and Standard Deviation (SD) together with Confidence Intervals. An overall score will be generated by combining the standardised scores, weighting each part according to its importance. An individual's report (profile) might look something like this:

Table: Transfer assessment results for A.N. Student of XYZ primary school: Date

Area	Standardised score	95% Confidence Interval
Mathematics	139	8
English	112	10
Writing	105	15
Maths problem solving	120	15
Non-verbal ability	99	12
Overall ¹	122	5

Reducing bias

The major threats to the fairness of the assessment come from:

- a. Differential preparation by schools, parents and private tuition
- b. Content not being a reasonable representation of the curriculum
- c. The wording of items
- d. The format of items
- e. Manual marking
- d. Parents not opting their children in to the assessment

These threats cannot be removed but they can be reduced and the following steps should be taken to make the assessment as fair as possible.

Differential preparation: Past papers, with answers, should be made available free of charge online so that anyone can access them and use them for practice.

Content: The format and content area of the assessment should be clearly set out and available to all. The draft papers should be independently vetted by an expert panel. The panel should include representatives from the variety of selective schools as well as two independent assessment experts and two representatives from primary schools.

Wording: During assessment development a group of experts should independently vet draft items and the trial data should be analysed for differential item functioning by relevant sub groups. Problematic items should be dropped.

Format: The variety of item formats chosen for the assessment should ensure that bias is minimised. (Speeded multiple choice items tend to favour boys whereas opened ended written items tend to favour girls.)

Marking: Manual marking is potentially biased if information about the students is visible. By ensuring that all marking is blind this problem is circumvented.

Opting in: Parents should be given clear information about the assessments and what they might lead to through literature posted to their homes.

Reliability/precision and confidence intervals

Good selection tests must have good precision at all cut-score boundaries. In other words users should be confident that if a new assessment were taken on the following day the same results, within the stated confidence interval, would be obtained.

¹ The average overall is calculated by taking the weighted average of the standardised scores. The weighting is derived from the "information" within each score which is proportional to the inverse of the squared error. The overall error is estimated.

The first step in ensuring precision is a requirement to report confidence intervals. The second is to make sure that as much high quality data as possible are collected in the time available.

Choice of content

English and mathematics: Well-constructed multiple choice assessments of the curriculum in these two areas are likely to generate the best predictors of GCSE success. The content should be a subset of the statutory curriculum for primary schools in Northern Ireland.

Open ended writing: Some students excel in writing tasks when they do not perform as well in multiple-choice items and this section will allow them to demonstrate their capabilities. Further, the inclusion of a writing element in the assessment is likely to have a positive impact on pedagogy; assessment preparation is likely to be broader.

Problem solving in mathematics: As with open ended writing tasks some students are better at problem solving than at multiple choice assessments and this will allow them to demonstrate this ability. It will also broaden efforts to prepare for the assessment.

Non-verbal, non-numerical, curriculum free, cognitive assessment: The best measure of 'g', the factor underlying intelligence tests, is Raven matrices a non-verbal, non-numerical, curriculum free, cognitive assessment. It cannot be used in the selection test because the students could easily learn the right responses. But a tender could specify that alternatives should be created which correlate well with academic attainment but weakly with socio-economic status.

Validity

Validity is an all-encompassing construct related to the use to which the assessment is put which should be taken seriously by developers and users. Good assessment developers are well aware of the issues, which can be complex. The technical report should deal with some of the key issues but the data should be held for at least five years so that follow up studies can be carried out. These should look, amongst other things at the predictive validity of the assessment and its sub-components.

Ideally anonymised data sets should be made available to independent researchers.

Appendix

Terms of Reference

Aim

1. To make recommendations for the design of a common assessment tool for the purposes of academic selection which (a) builds upon current practice and recent experience; (b) has the confidence and support of stakeholders and (c) draws upon best practice elsewhere.

Objectives

2. Assuming work begins w/c 14 November 2016, the education assessment consultant is commissioned:

- by 16 December 2016:
 - a) to engage with PPTC and AQE to scope current arrangements, and establish points of agreement and difference which need to be addressed;
 - b) to engage with other stakeholders including primary and post-primary schools as appropriate;
 - c) to develop criteria (including in respect of validity, viability and reliability) for a common assessment tool, taking account of agreed international standards for the purpose of selection to selective post-primary schools on academic ability;
 - d) to specify the structure of the assessment including:
 - a. content to be assessed;
 - b. assessment methodology e.g. multiple choice, free response etc; duration of test;
 - c. standardisation methodology – e.g. grading (numbers / letters); awarding, age standardisation etc;
 - e) to recommend arrangements for the administration of the assessment, including implementation, location, timing, funding, marking and reporting;
- to secure, as far as possible, the agreement of PPTC and AQE to the above;
- by 20 January 2017, to have produced a report for DE's consideration which sets out the findings, options and recommendations to be taken forward;
- subsequently, to provide, as far as appropriate, advice in respect of the above.

Support

3. Secretariat support will be provided by the Assessment and Qualifications Team of the Department of Education.

Reporting

4. Throughout this period, and whilst engagement with the PPTC and AQE is undertaken, progress will be reported to the Department on a weekly basis. All discussions and reports to remain in confidence unless agreed by the Department.

5. Progress will be reported to Dr David Hughes, Director of Curriculum, Qualifications and Standards Directorate.

Who was consulted

1. The Minister's Special Adviser
2. Department of Education Officials
3. PPTC
4. AQE members
5. GL Assessment
6. Schools visited where Principals and, in most cases, representatives of the B.of G were present
 - a. Aquinas Diocesan Grammar School
 - b. Ballymena Academy
 - c. Banbridge Academy (4 feeder primary Principals also in attendance)
 - d. Bangor Academy and representatives from Priory College, Strangford Integrated College, Nendrum College, Glastry College, Movilla High School and St Columbanus College
 - e. Foyle College
 - f. Lagan College
 - g. Limavady Grammar School (2 feeder primary Principals also in attendance)
 - h. Our Lady and St Patrick's College
 - i. Our Lady's Grammar School and representatives from The Abbey Christian Brothers Grammar School
 - j. St Michael's College
 - k. Thornhill College (Representatives from Lumen Christi College and two feeder primary schools also in attendance)
 - l. Enniskillen Royal Grammar School
7. Controlled Schools Support Council (CSSC)
8. Council for Catholic Maintained Schools (CCMS)
9. Council for the Curriculum, Examinations and Assessment (CCEA)
10. Education Authority
11. Governing Bodies Association
12. Non-selective schools in North Down area (held at Bangor Academy)
13. Selection of parents hosted by Parenting NI
14. Selection of pupils at Lagan College
15. Unions: UTU & NASUWT (written response from ATL)
16. Workshop for a selection of South Belfast Grammar schools