

EDUCATION AND SKILLS COMMITTEE: Scottish National Standardised Assessments**Submission from the School of Education, University of Glasgow.****Executive Summary**

The University of Glasgow welcomes the opportunity to respond to the Education and Skills Committee on the assessment issues under consideration. Assessment is one of the main areas of research of the University's School of Education, which has a long history of involvement in assessment policy, practice and research in Scotland and in assessment systems internationally.

Curriculum for Excellence and the values that lie behind it define what matters in the education of young people who are educated citizens in Scotland.

- Curriculum (what matters in learning),
- Pedagogy (how learning takes place) and
- Assessment (showing how much and how well learners are learning)

are inextricably linked. Within this context, assessment can be viewed from two perspectives. It can focus on the use of evidence to provide feedback to inform next steps in learning or on judging, comparing or categorising learners. To improve an education system, the focus has to be on learning.

There is a persistent myth that Scotland does not record sufficient data to support improvement. The Scottish Education system is replete with data but these data are not always well targeted, as dependable as we might wish them to be or used to best effect.

Key Points

1. The assessment system in Scotland, the National Improvement Framework, should be a means to provide key participants, including learners themselves, with dependable evidence that they can use to improve learning. Much of the system is in place but we may need to review key aspects. Crucially, we need to build assessment capacity across the system.

2. Traditionally, assessment systems serve three main purposes, to inform learning, to sum up learning over time and to hold people to account . Assessment information gathered about the past is only helpful if it informs future action that leads to improvement. Children are not data – they are not numbers. They are people with lives and futures that depend on successful learning relationships in and beyond the classroom. Learning must be our principal concern.

3. In an assessment system every action taken has consequences. Whole-cohort tests and sample surveys are simply different ways to collect evidence. Each approach has advantages, disadvantages and consequences for learners and for learning. Scotland's assessment system should promote only consequences that are positive for both.

4. Assessment policy in Scotland - the National Improvement Framework - exists in a wider social context. The context within which policy emerges will influence how

assessment policy is translated into practice. We can learn from our own previous experience in Scotland with Assessment 5-14.

The University of Glasgow is a founder member of IEAN, an International Educational Assessment Network of researchers and policy makers in twelve nations/states who share their insights and experiences of assessment. Members of this international network who are tackling similar issues can contribute to our thinking and to the deliberations of this Committee.

Submission from the School of Education, University of Glasgow.

1. The assessment system in Scotland, the National Improvement Framework, should be a means to provide key participants, including learners themselves, with dependable evidence that they can use to improve learning.

In Scotland, information on progress in what matters in learning (assessment evidence) is generated at a number of levels, eg,

- *in classrooms* – to determine how individual learners are progressing, using evidence collected to inform next steps in learning and to provide information for parents/carers;
- the *department/faculty/school* to indicate how groups of learners are progressing, using evidence collected to inform better planning and classroom practice;
- *in a local authority/regional improvement collaborative*, using evidence to inform and share better understanding and expectations about learning across schools and action to improve it
- *for the nation* to indicate how much and how well young people are learning in relation to the curriculum nationally, using evidence to identify trends, evaluate overall provision and inform action to promote better learning nationally (or for specific groups);
- *in society more widely* assessment evidence, principally from the Scottish Qualifications Authority, is used to provide access to limited resources, eg, College or University, through a process of selection that seeks to be fair and transparent;
- *internationally* -how comparative evidence from other countries can inform thinking about improving learning in Scotland, eg, through OECD's international assessment survey (PISA) or through inviting OECD to review the Scottish education system.

Evidence provides information for all of the above in our national assessment system (National Improvement Framework). This assessment evidence should serve as a major driver to ensure that Curriculum for Excellence promotes the best possible life chances for all of Scotland's citizens, ie, an education system that is more socially just. However, the collection of evidence itself does not lead to improvement. Improvements in learning come from targeted action informed by high quality evidence from assessment approaches that are fit for purpose.

Much of the framework to provide the information Scotland needs is already in place and is consistent with practice that is regarded as effective, eg, *Synergies for Better Learning* OECD (2015). There are, however, tasks to be undertaken to make sure that our current systems remain fit for purpose. For example, we may need to

- review aspects of Curriculum for Excellence in the light of recent thinking about progression in learning

- reflect on current assessment practices to ensure that all the data we gather matches a clear purpose
- ensure that our system is efficient, so that different parts of the system are not duplicating the collection of information, we are not gathering more evidence that is required for our core purposes, and we are not gathering evidence where more time is spent in collecting information than in using it.

If information gathered is not used to inform improvements in learning, it is not worth gathering. Persuading people to stop gathering data that they have traditionally collected is a very difficult task.

Future investment in assessment should pay particular attention to investing in people. If assessment is to support learning rather than superficial compliance with curriculum statements, all involved, including policy makers, practitioners, parents and young people, need to have a deep understanding of assessment purposes and practices. For example, if the whole curriculum matters, teachers' professional learning may need to focus on the design of tasks that require the knowledge, understanding and skills specified across the curriculum. Creating such tasks supported by professional learning opportunities related to monitoring pupils' progress will develop and improve teachers' professionalism. Although progress has been made in moderating teachers' professional judgement, for those judgements to play their intended central role in Scotland's assessment system, time and focus are required to ensure that professional judgement is consistently dependable. The confidence that develops from depth of understanding is a necessary part of developing the assessment culture that will consistently support the aspirations of Curriculum for Excellence.

2. Traditionally, assessment systems serve three main purposes, to inform learning, to sum up learning over time and to hold people to account. Assessment information gathered about the past is only helpful if it informs future action that leads to improvement. Children are not data – they are not numbers. They are people with lives and futures that depend on successful learning relationships. Learning must be our principal concern.

Assessment systems in education are complex and seek to serve a range of purposes. The Assessment Reform Group, an internationally recognized group of experts in educational assessment, identified three main assessment purposes in their publication *Assessment in Schools – Fit for Purpose? (Mansell, James et al, 2010)**.

- The use of assessment to help build pupils' understanding, within day-to-day lessons.
- The use of assessment to provide information on pupils' achievements to those outside the pupil teacher relationship, eg, to parents (on the basis of in-class judgments by teachers and of test and examination results), and to further and higher education institutions and employers (through test and examination results).
- The use of assessment data to hold individuals and institutions to account.

These three main purposes interact in any national assessment system. Any action taken in one area will have an impact on the other areas. This phenomenon is commonly referred to as washback.

Assessment systems provide information and influence what people do. A National Improvement Framework influences the actions of those who work within it, policy makers, practitioners and researchers. A current challenge internationally is how to make that influence positive, ie, enhance learning (Hayward, 2015). Too many current national

performance frameworks have not had a positive influence (Mons, 2009); there is powerful, consistent evidence that high stakes test-based monitoring systems lead to undesirable effects. Washback effects commonly include teaching to narrowly defined tests, narrowing the curriculum, teaching test behaviours, demotivating more vulnerable pupils and reducing levels of teachers' confidence in their professional judgement and in their wider professionalism. The decision of the Scottish Government not to collect data from standardised assessment separately from evidence from teachers' professional judgement was a welcome attempt to reduce the stakes of standardised assessment in Scotland.

Any decision about how to collect evidence at a national level has to consider the potential for washback and those responsible should seek to avoid predictable undesirable consequences and design a system where washback is positive, ie, leads to improvement in learning. Above all, it is crucial to remember that behind every number or letter or comment lies a young person. Motivation matters and assessment information should encourage learners to make progress not label or categorise them in ways that make them less likely to want to learn.

3 In an assessment system every action taken has potential consequences for other parts of the system. Standardised whole-cohort tests and sample surveys are simply different ways to collect evidence for particular purposes. Each approach has advantages, disadvantages and consequences for learners and for learning. Scotland's assessment system should promote consequences that are positive.

Surveys and standardised tests are two approaches that can be used to collect evidence to inform national decision making.

Standardised assessments have advantages, eg,

- They are designed by experts who have developed procedures to make them reliable.
- They allow the performance of individuals and groups on the test to be compared with the average performance of the whole population that has been used to standardise the test; and they can indicate changes in performance over time.
- The SNSA are diagnostic, in that they are linked to proposals for action based on each young person's performance.
- SNSA evidence for a group of young people may point to areas of the curriculum where the teacher needs to place future emphasis

Standardised assessments have disadvantages, eg,

- In general, the advantage that standardised tests have in respect of *reliability* often reduces their *validity* as means of assessing progress in the planned curriculum, because many aspects of this cannot be assessed through test items that match *statistical* reliability criteria
- While they can provide information on a limited number of aspects of Curriculum for Excellence, they cannot do so on all that matters; eg, it is easier to measure a child's phonic awareness than their motivation to read; yet both matter. Coverage of the curriculum is limited to those aspects that are readily measured, and by keeping tests to an acceptable length.
- The use or misuse of standardised assessment data can lead to the washback effects described in the previous section.
- There are many practical issues that prevent standardised assessment being enacted as intended – lack of access to technology, perceived pressure, alternative understandings of the nature of the assessment, perceptions that data will be used

for purposes of comparison.

- Learning in schools and classrooms is based on dialogue and standardised assessments based on the performance of the child do not recognise this

Sample Surveys have advantages, eg,

- They can provide dependable national level data without identifying individual schools and teachers and thus avoiding the washback effects of whole-cohort approaches.
- They can monitor national levels of performance in learning over time, providing information about the impact of *Curriculum for Excellence*.
- They can be designed to sample a wider range of aspects of the curriculum than is possible in a whole-cohort test, including aspects that are harder to measure, for example through observation of tasks and questionnaire enquiries, as was the case in a previous survey, the Scottish Survey of Achievement (SSA).
- They can over time provide evidence on different areas of the curriculum beyond English and Mathematics (the SSA provided information on Literacy and Numeracy in Science and Social Subjects as well as in English and Mathematics).
- The survey sample can be adapted for different purposes, eg, a boosted sample can give a local authority or a regional collaborative information specific to that context or information can be generated relating to a specific population for a specific purpose – eg, boys' performance in reading.
- As surveys in Scotland were designed in partnership with teachers, involvement in the process helped to build capacity.

Surveys have disadvantages, eg,

- Commonly, they do not offer information on every pupil.
- If the survey sample is too small the evidence emerging can be compromised.
- Surveys are often poorly understood and they come under attack for not addressing purposes they were never designed to serve.

It is difficult to be certain why the decision was taken to move away from the use of the SSLN survey, but there were a number of contributory factors which led to a negative perception in Scotland of the survey method of monitoring achievement, in contrast to the very positive international view of this approach as it had been applied in Scotland. Possible factors were:

- A misinterpretation of the recommendations of the OECD report. A view emerged that the OECD had recommended the introduction of standardised assessment. The OECD report recommended an 'integrated framework for assessment'. The final paragraph on p.161 is clear:

'Currently, however, the way national assessment is constructed in Scotland does not provide sufficiently robust information at all levels of the system, including LAs or an individual school or across important domains of CfE for learners and their teachers. This problem does not mean that everyone must be tested at particular year levels in order to have this information. An alternative, for example, could involve sample testing of a range of learners within each school on rich tasks which can then be used to benchmark the achievement of other learners on the curriculum.'

- Problems that existed in the SSLN survey methodology, as opposed to that of the previous more extensive survey in Scotland - the Scottish Survey of Achievement (SSA).
- A concern to have nationally available information on the progress of every child and

perhaps a lack of awareness that surveys can be designed to allow such data to be generated.

- Insufficient involvement of all key education authority stakeholders in the design of the survey.
- A strong commitment in some education authorities to other forms of data collection, principally standardised testing. (Although the understanding was that when SNSA was introduced Local Authorities would cease to use other standardised tests, this has not happened).

4. Assessment policy, in Scotland - the National Improvement Framework - exists in a wider social context. The context within which policy emerges will influence how assessment policy is translated into practice.

No matter the intention underpinning it, all policy develops within a cultural context. There are patterns of activity in assessment over time in Scotland from which it is important to learn to make sure that mistakes from the past are not repeated, now or in the future. Although current circumstances are different in some aspects, our own history offers a salutary reminder of how tricky it is to keep the focus on teachers' professional judgement when standardised assessment is part of the system.

In the early stages of Education 5-14, assessment policy emphasised the importance of teachers' professional judgement. To build a shared understanding of national standards, light touch national tests which covered limited aspects of the reading, writing and number curricula would be used to moderate teachers' professional judgement. The policy explicitly stated that if a difference emerged between a teacher's professional judgement and the national test, the teacher's judgement would be the result reported (although the school should review such cases to try to understand the reasons for the difference). What happened in practice was that teachers used the tests to discern whether or not a child had 'achieved' a level. Although the policy asked teachers to test when the child was ready, ie, when the evidence suggested that (s)he was ready to move on to the next level, practice in schools across the country was very different. Some schools had 'testing weeks' when every child took a test and children were reported as 'passing' or 'failing' the test. Some schools sent home award certificates. A few schools took all of the children into the school hall and ran national tests as if they were national examinations. Ironically, although both teachers and government stated an intention that tests should be low stakes, in practice they became high stakes for children, teachers, schools and local authorities. Teachers and schools responded to the culture within which they perceived the tests to have emerged rather than acting in ways that were consistent with stated policy.

Patterns of behaviour such as this have emerged in countries internationally (Mons, 2009). Onora O'Neill in the Reith Lectures of 2002, 'A Question of Trust' (<https://www.bbc.co.uk/radio4/reith2002/>), reflected on international trends in accountability and argued that countries needed to pay greater attention to culture and values if they were to design more 'intelligent' accountability systems - systems that placed greater emphasis on professional judgement.

Any assessment method designed by 'assessment experts' is likely to be attributed a level of significance that will impact on the value placed on teachers' professional judgment. No assessment method is perfect and the dependability of any one is a matter of the appropriate balance between validity of curriculum coverage and reliability of the interpretations of the

assessment evidence., Throughout the world people tend to overestimate the dependability of tests and examinations and underestimate the dependability of teachers' professional judgement. However, it is the use or misuse of data that leads to distortions in education systems. League tables and other ways of comparing teachers, schools, authorities or nations have left a deep scar on professionals' consciousness. Even if data are not collected and published nationally, if there is a perception that data might be used to gauge performance in classrooms, schools, local authorities or nations, distortions are likely.

It is difficult to be certain what is actually happening in schools in Scotland as they begin to incorporate national assessments into their assessment practices. Different sources of evidence are providing different pictures of the impact of the standardised assessments in schools and classrooms. The plan to investigate current practices will provide crucial evidence to help the system understand what is actually happening in schools and, most importantly, to identify factors that are driving practices.

Learning with others: international insights. In partnership with members of the Policy Division in Scottish Government, the University of Glasgow founded an international network of assessment experts, both researchers and policy makers from each nation or state. Twelve nations/states are members of the International Educational Assessment Network (IEAN): Scotland, Wales, Ireland, Norway, Iceland, Denmark, Slovenia, Switzerland, Singapore, Canada (Ontario), Australia (Queensland) and New Zealand. Through this IEAN network we would be happy to contribute international evidence on areas of interest to the Committee.