

Primary Assessment

Written evidence from the *Department of Learning and Leadership, UCL Institute of Education*

This submission was prepared by Dr. Jake Anders, Dr. Melanie Ehren, Dr. Guy Roberts-Holmes and Prof. Dominic Wyse. It draws on their knowledge of some key research in the curriculum and assessment field, and on their own previous research projects and scholarly work.

Introduction

Primary assessment has many purposes. One useful way of separating out its role is into formative, summative, and evaluative assessment (Harlen, 2016). Formative assessment, sometimes called assessment for learning, is used throughout the teaching process to help pupils' learning. Summative assessment, sometimes called assessment of learning, provides a report on the learning that has happened. Evaluative assessment is used to monitor performance at school or national levels. As such, primary assessment, in a broad sense, is vital to the teaching and learning process, with teachers using it interactively to adjust their approach and help pupils to reach their potential. However, it can also be used in ways that have negative consequences for pupils, teachers and the education system as a whole (Wyse, McCreery & Torrance, 2009; Klenowski & Carter, 2016). This submission provides a brief review of the importance, uses and challenges of primary assessment.

Uses for primary assessment

In relation to primary assessment most people think first of standardised assessment tests, particularly those at age 11 at the end of Key Stage 2. However, assessment has a much broader role than this. At its most fundamental level, teachers are constantly assessing pupils' understanding of concepts and adjusting their methods and instruction as a result. This kind of informal, formative assessment is extremely important to successful learning but researchers have noted the tension that can form between summative and formative assessment (Black & Wiliam, 2006, p. 19; Harlen & James, 2006). Nevertheless, this external summative assessment is seen by many to have an important role for the purposes of reporting to pupils and parents.

Problems with current primary assessment arrangements

National and international research on standardised assessment indicate a number of problems that are likely to follow from its use (Hamilton et al, 2002). For example, its consequences for teaching and learning have been highlighted recently by Ehren's work on the end of Key Stage 2 mathematics test, which has revealed some implications for pupils' motivation and confidence, and for the processes of taking tests (Ehren, 2016a). We discuss four key concepts relevant to understanding potential issues with the current system.

1. *Narrowing the curriculum*

The standardised primary assessment system in England has focused mainly on the three 'core' subjects of English, mathematics and science (Wyse & Torrence, 2009), and more recently mainly on English and mathematics. This results in a narrowing of the curriculum that teachers provide, due to their awareness of what they and their pupils will be judged on (Boyle & Bragg, 2004). Even within these core areas, assessment is not currently addressing important elements of learning, such as creativity, which may be neglected as a result (Wyse and Ferrari, 2015). In addition, the existence of standardised summative outcomes have been used to judge the accountability of teachers and schools through trends in performance across the country, rather than addressing accountability through a separate mechanism as originally recommended nearly 30 years ago in the Task Group on Assessment and Testing report.

2. *Reductionist models of teaching and learning*

National standardised assessment systems used for accountability have a tendency for reductionism to enable numerical comparison across schools and systems (Ozga, 2009). This can contribute to a deficit model of teaching and learning (Ball, 2014), focussed on passing and failing. This potentially sets up low expectations, particularly for disadvantaged and SEN children (Bradbury and Roberts-Holmes, 2016).

3. *'Teaching to the test'*

High stakes assessments lead to a tendency for schools to 'teach to the test' in order to avoid the consequences of not meeting standards. This, in turn, produces artificial test-score gains (i.e. test-score inflation). Scores on high-stakes tests- often become severely inflated in being far larger than true gains in students' learning. According to Koretz (2008), this inflation is highly variable and unpredictable, so one cannot tell which school's scores are inflated and which are legitimate (Koretz, 2008, p. 131).

4. 'Educational triage'

Another consequence of high stakes assessment is that it often provides incentives to focus teaching resource on specific students on the margin of key test thresholds. This has been referred to as 'educational triage', where particular children are identified, who, with increased teaching will achieve a particular standard and raise the school's metrics. This leads to other children, particular those with SEN who are deemed as not potentially making the standard, being allocated less teaching resource (Roberts-Holmes, 2014; Hutchings, 2015; Marks, 2016).

National trends in performance

At present, the results of primary assessment tests, specifically the proportion of pupils exceeding a specified score, are used as a way of tracking performance over time. However, concerns have been expressed that the combination of these two functions leads to a downward pressure on the standard required to be judged to reach the specified level, so as to provide evidence of "improvement" in educational standards over time. Separating these two functions by introducing a system of national sampling (Sahlberg, 2011; Harlen, 2014) where assessments do not have consequences for individuals or schools would be a more appropriate way of judging such trends, although there are still challenges to implementing this in a way that is truly meaningful (Johnson, 2016). Potential advantages to such a system is that such sampling could be linked to internationally comparative tests such as TIMSS and PIRLS; they could also widen the focus by covering a wider range of the curriculum on a rotating basis.

Value Added

Even if these concerns are surmountable, it is generally accepted that good school accountability measures, of the type that Key Stage 2 tests are current used as part of, should be "value added" measures. Rather than simply giving a school a score based on the current performance of its pupils, value added measures attempt to estimate the average progress that pupils have made during their time in the school. While the statistical methods used are more complex than this, the basic intuition is to adjust a pupil's absolute performance in a test at the point of leaving school by their attainment at the point they entered the school. As such, for a value added measure to work well it is important to have good estimates of pupils' performance as close as possible to the time they enter the school. This can be achieved with at least some reliability at secondary school, using tests at age 11.

However, Bradbury and Roberts-Holmes (2016) discuss how it is much harder for primary schooling because it is difficult to generate reliable and valid baseline statistics with young children. The potential stress that testing places on pupils, along with lower reliability of testing pupils at this age, means teacher assessment at the end of Key Stage 1 is used as a baseline. Yet, the knowledge that these form the baseline for end of Key Stage 2 value added measures may give teachers perverse incentives to depress scores in these tests in order to boost the value added measure on which their school will be judged (Education Datalab, 2016). This raises fundamental concerns about the utility of primary assessment for accountability purposes.

Conclusions

This submission has highlighted some of the key issues relevant to primary assessment, particularly around the use of standardised, summative assessment for purposes beyond their direct reporting role. We have reflected on problems resulting from the use of standardised assessments increasingly being used for accountability purposes and to measure national trends in performance. However, the process of conducting these types of assessment and, especially, using them for these purposes may cause unintended consequences, including narrowing of the curriculum and gaming of such measures. These issues should in our view be at the forefront of the Committee's mind as it considers the overarching purpose of primary assessment.

References

- Black, P. & Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5:1, 7-74, DOI: 10.1080/0969595980050102
- Boyle, B. & Bragg, J. (2006) A curriculum without foundation. *British Educational Research Journal* 32:4, 569–82.
- Bradbury, A. & Roberts-Holmes, G. (2016). 'They are Children...Not Robots, Not Machines': *The Introduction of Reception Baseline Assessment*. National Union of Teachers/Association of Teachers and Lecturers.
- Education Datalab (2016) *Seven things you might not know about our schools*.
<http://www.educationdatalab.org.uk/wp-content/uploads/2016/02/EduDataLab-7things.pdf>
- Ehren, M. (2016a). Life after levels: is the new Year 6 maths test changing the way teachers teach? IOE London Blog. Available at: <https://ioelondonblog.wordpress.com/2016/07/19/life-after-levels-is-the-new-year-6-maths-test-changing-the-way-teachers-teach/>
- Ehren, M. (2016b). Word problems in standardised maths tests: how fair are the Key Stage 2 SATs? IOE London Blog. Available at: <https://ioelondonblog.wordpress.com/2016/10/25/word-problems-in-standardised-maths-tests-how-fair-are-the-key-stage-2-sats/>
- Harlen, W. & James, M. (1997) Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4:1, 365-379, DOI: 10.1080/0969594970040304
- Harlen, W. (2014) *Assessment, Standards and Quality of Learning in Primary Education*. Cambridge Primary Review Trust.
- Hutchings, M. (2016) *Exam Factories? The impact of accountability measures on children and young people*. National Union of Teachers (NUT). Available at <https://www.teachers.org.uk/files/exam-factories.pdf>
- Johnson, S. (2016) *National Assessment and Intelligent Accountability* in Wyse, D., Hayward, L., & Pandya, J. (eds.) *The SAGE Handbook of Curriculum, Pedagogy and Assessment*, vol. 2, ch. 49, pp. 790-804. SAGE: London.
- Johnson, S. (2016) *Curriculum Reform in Testing and Accountability Contexts* in Wyse, D., Hayward, L., & Pandya, J. (eds.) *The SAGE Handbook of Curriculum, Pedagogy and Assessment*, vol. 2, ch. 53, pp. 853-870. SAGE: London.
- Koretz, D. M. (2008). *Measuring up*. Harvard University Press.
- Marks, R. (2016) *Ability Grouping in Primary Schools: Case studies and Critical Debates Northwich*, Critical Publishing
- Ozga, J. (2009) Governing Education through Data in England from regulation to self-evaluation. *Journal of Education Policy*, 24 (2) pp. 149-162. doi:10.1080/02680930902733121
- Roberts-Holmes, G. (2015). The 'datafication' of early years pedagogy: 'if the teaching is good, the data should be good and if there's bad teaching, there is bad data'. *Journal of Education Policy*, 30:3, 302-315. doi:10.1080/02680939.2014.924561
- Sahlberg, P. (2011) *Finnish Lessons: What can the world learn from educational change in Finland?* Teachers College, Columbia University.
- Schagen, I., Sainsbury, M. & Strand, S. (1999) Statistical Aspects of Baseline Assessment and its Relationship to End of Key Stage One Assessment. *Oxford Review of Education*, 25:3, 359-367, DOI: 10.1080/030549899104035
- Wyse, D., McCreery, E. & Torrance, H. (2008) *The Trajectory and Impact of National Reform: curriculum and assessment in English primary schools* (Primary Review Research Survey 3/2), Cambridge: University of Cambridge Faculty of Education.
- Wyse, D. & Torrance, H. (2009) The development and consequences of national curriculum assessment for primary education in England, *Educational Research*, 51:2, 213-228, DOI: 10.1080/00131880902891479
- Wyse, D. and Ferrari, A. (2015) Creativity and Education: Comparing the national curricula of the states of the European Union with the United Kingdom. *British Educational Research Journal*. Vol. 41(1), 30–47.