

A BASELINE WITHOUT BASIS

THE VALIDITY AND UTILITY OF THE PROPOSED RECEPTION BASELINE ASSESSMENT IN ENGLAND

SUMMARY REPORT

Authors: Harvey Goldstein, Gemma Moss, Pamela Sammons, Gwen Sinnott and Gordon Stobart

Full report: <https://www.bera.ac.uk/researchers-resources/publications/a-baseline-without-basis>

This document summarises the case against the government’s proposal to use a baseline assessment test of pupils in reception to hold schools in England to account for the progress that pupils have made at the end of key stage 2.

When the government’s plans were published in 2018, BERA convened an expert panel to consider whether the evidence from the assessment literature could justify such a test being used for this purpose. The conclusion of the expert panel is that it cannot. This report is intended to inform public debate by providing an accessible account of the reasons why the proposals are flawed.

In the panel’s view the proposed baseline assessment will not lead to accurate comparisons being made between schools, as policymakers assume. Perhaps most importantly, they will not work in the best interests of children and their parents.

The panel has drawn the following conclusions on the basis of the evidence set out in the body of this report.

1. Under these proposals, children will be exposed to tests that will offer no formative help in establishing their needs and/or in developing teaching strategies capable of meeting them.
2. Any value-added calculations that will be used to hold schools to account will be highly unreliable.
3. This is an untried experiment that cannot be properly evaluated until at least 2027, when the first cohort tested at reception has taken key stage 2 tests.

The panel arrived at their view by seeking answers to the following six questions, summarised below.

1. Is it legitimate to use baseline assessment for school accountability purposes seven years hence?

It is both ethically and methodologically questionable to use reception baseline assessment (RBA) for such a purpose. As currently proposed, RBA is likely to produce results with little predictive power and dubious validity.

The assessment of very young children may be ethically justifiable when used to support a child's learning, in which case they stand to benefit directly. However, the government's RBA will be used solely for school accountability, a purpose for which the test is not fit.

The research evidence demonstrates that any early-years assessment system will have little predictive power. Aggregating scores in a proposed 20-minute test, covering the three domains of literacy, numeracy and self-regulation to produce a single number, is misguided. Besides its inherent unreliability, it would ignore the fact that children may perform differently in each domain, and that some domains are better predictors of progress in different areas of the curriculum than others. Generalising from a cohort to the school would be unwise given the limited sample size in each primary school. Furthermore, no proposals have been made regarding how predictive validity will be investigated and reported across different years.

For accountability purposes, it remains unclear whether the reception baseline tests are intended to align, in terms of method and content, with the relatively narrow formal testing at key stage 2 against which pupils' progress in the intervening years will be measured. Insisting on a close alignment may result in a narrowing of the early-years and primary curricula.

2. Will the proposed tests be accurate or fair?

There is good reason to question the reliability of the data that the test will produce, and the ways in which that data will be interpreted and acted upon.

The panel expects the baseline tests to show low levels of reliability because, firstly, no indications have been given that age effects will be controlled for at both the initial baseline test and the outcome tests at key stage 2 – yet this is essential if the data is to be used for school accountability purposes, for two reasons.

- Just a few month's difference in age has been shown to produce pronounced developmental differences at reception age. Autumn-born children have demonstrated a strong advantage in attainment over their younger, summer-born peers in assessments similar to the one proposed.

- Pupil cohorts within primary schools are statistically small, and often have uneven distributions of younger and older children. Schools serving more children who are young for their year of entry may appear to have less favourable effects on children's later attainment than those that serve children who are old for their year, unless age and season of birth are accounted for with sufficient precision.

Secondly, pupil mobility poses a problem if the RBA is intended as a measure of pupil progress seven years hence: either mobile pupils will have to be taken out of the progress measure in all schools, resulting in varying numbers of children being 'missing' from the accountability measure; or baseline assessment results will 'follow' pupils between schools, resulting in schools being held accountable for pupils' progress despite being unaware of their starting points, and having been responsible for only part of pupils' school lives. Teacher turnover, and the likelihood of a change in head teacher over a seven-year period, will also muddy the issue of accountability.

3. What recognition is being given to contextual factors in the interpretation of the data?

It is generally recognised that the only proper way to make comparisons between schools is to make adjustments for the prior attainments of their pupils when they enter those schools, and to control for other relevant characteristics of pupil intakes such as parents' educational levels, family income and having English as an additional language. Such adjustments lead to what are known as 'value added' comparisons. There is strong evidence that these characteristics affect both attainment and relative attainment in value-added measures. However, under the government's current proposals, school-level attainment at year 6 will be adjusted for using the reception baseline assessments alone, and *without* controlling for any contextual factors. This approach cannot lead to fair comparisons.

4. Will this form of accountability lead to useful comparisons of schools?

The available evidence suggests not.

Little research has been carried out on the efficacy of using pupil progress measures to hold schools to account at the primary level. However, research at the *secondary* level has found that, when ranking schools in this way:

- value-added scores suffer from considerable statistical uncertainty due to low sample sizes
- data used to inform parents' choice of school is extrapolated from the results of students who entered school several years earlier, and is thus significantly 'out of date'

- the fact that specific sets of value-added school effects will prevail at any given time means that predictions for new cohorts based on this data will be very weak.

The analogous uncertainties will almost certainly be greater for the reception baseline tests, because:

- the time-gap between reception and year 6 is greater than between year 6 and year 11
- baseline tests of very young children will be prone to greater inherent measurement error.

5. What is the likely impact of these accountability measures on pupils and schools?

The results themselves will do little to help secure positive outcomes for pupils, teachers or parents in either the short or longer term.

The government intends to hold the baseline test data until the cohort reaches key stage 2. It is not yet known whether it will release a limited set of data to schools during the test year. Certainly, publishing the data at the point at which it is collected in reception could encourage the production of statistically worthless ranked league tables of school performance. Conversely, while non-disclosure of the data may prevent the over-interpretation of individuals' and schools' results from a potentially unreliable test, it is likely to frustrate teachers and parents, who may well ask, 'Why administer a test that doesn't help teaching and learning?'

While the assessment is not intended to have any diagnostic value for schools and individual children, teachers administering the test will see children's scores. This could mean that some children – particularly the summer-born, those with English as an additional language and those with special educational needs – could be unnecessarily labelled as low-ability at the very beginning of their education, with the risk that premature judgements about their abilities may then become 'self-fulfilling'.

6. Are there better alternatives to baseline testing?

Baseline testing reflects a more general trend in public services towards using 'performance indicators chosen for ease of measurement and control rather than because they measure quality of performance accurately'.¹ However, there are alternatives. An 'intelligent accountability' approach would allow practitioners to use their professional judgement more fully in the assessment process – gathering deeper and more meaningful data that can

1 O'Neil O (2002) *A Question of Trust: The BBC Reith Lectures 2002*, Cambridge: Cambridge University Press: 54

take account of contextual factors, help to support individual pupils, and inform improvement planning both within and between schools.

Both the Surrey value-added initiative and the London Education Research Network, among other examples, have demonstrated that principles of intelligent accountability can readily be adopted and put into practice to support school improvement and spread good practice in the use of data.

The panel believe that the government's proposals for the reception baseline assessment are flawed, unjustified and wholly unfit for purpose. They would be detrimental to children, parents, teachers and the wider education system in England. We publish this report in the hope of informing public debate by offering an accessible and thorough account of why these proposals must be comprehensively rethought.

The report summarised in this document presents the views of an expert panel established by BERA in early 2018 to consider the evidence behind the government's proposal to use a baseline assessment test of pupils in reception to hold schools in England to account for the progress that pupils have made at the end of key stage 2.

Full report

Goldstein H, Moss G, Sammons P, Sinnott G and Stobart G (2018) *A baseline without basis: The validity and utility of the proposed reception baseline assessment in England: Summary*, London: British Educational Research Association. <https://www.bera.ac.uk/researchers-resources/publications/a-baseline-without-basis>

Permission to share

This document is published under a creative commons licence:

Attribution-NonCommercial-NoDerivs 2.0 UK
<http://creativecommons.org/licenses/by-nc-nd/2.0/uk/>

For commercial use, please contact publications@bera.ac.uk.

British Educational Research Association (BERA)
9–11 Endsleigh Gardens
London WC1H 0EH
www.bera.ac.uk | @BERANews

