



NFER Thinks

What the evidence tells us

**Evidence for
Excellence in
Education**

Delivering a reception baseline assessment

Why a standardised administration is fairer than an observational approach



NFER has been developing high-quality, robust, reliable and informative assessments for 70 years, from classroom assessments to large-scale national and international testing programmes that support system-wide improvement.

In this *NFER Thinks*, we argue the case for having a standardised approach to the administration and scoring of the reception baseline assessment so that it can be delivered consistently across all primary schools.

The Department for Education (DfE) has now published its response to the primary assessment consultation. This includes a plan to introduce a school-entry or reception assessment to act as a baseline for measuring progress across the primary phase. Not much detail on the proposed assessment has been given as yet, but the DfE has clearly stated that it does ‘not intend this to be an observational assessment which is carried out over time’ (DfE, 2017). As this baseline measure is to be used as part of the primary accountability system, NFER believes the introduction of an assessment that is administered in a standardised way is the best way to enable *fair and consistent comparisons across schools*.

Support for measuring progress rather than attainment

The introduction of a reception baseline assessment (RBA) will enable progress to be measured across seven years of schooling rather than across four years as is currently the case.

A greater emphasis on progress is likely to be welcomed by schools. Almost 90 per cent of primary senior leaders and classroom teachers participating in our [Teacher Voice omnibus survey](#) considered the measurement of progress, rather than absolute attainment, to be a better way to assess the effectiveness of schools (NFER, 2017). The proposal to introduce an RBA has also been welcomed by organisations such as the National Association of Head Teachers (NAHT, 2017).

What do we mean by a standardised administration approach?

A standardised or uniform administration approach is one in which every child or pupil is assessed in the same way. The tasks or questions, any resources used or permitted and what is credited as a correct response are the same for each pupil. For example, key stage tests and written examinations such as GCSEs are administered in a uniform way, with each pupil attempting the same questions (or equivalent questions in the case of some GCSEs).

For young children, where a written test would be inappropriate, a standardised administration approach involves the teacher carrying out an assessment on a one-to-one basis, using exactly the same instructions or ‘script’ for each child. This ensures there is little or no variation in the way the assessment is conducted across all children in a school and across all schools.

An observational approach is the most common type of non-standardised assessment used. Children are assessed over time during normal classroom activities, using a set of assessment criteria. Although children are assessed using the same criteria, the contexts and the opportunities to demonstrate what they can do, know or understand will be different for each child. Depending on the specificity of the criteria, a greater degree of interpretation of children’s responses may be required, with teachers having to make subjective judgements as to whether children have met the criteria. With an observational approach, achieving consistency within and across schools depends on high levels of ongoing training and moderation and remains challenging.

Why a standardised administration approach is crucial to the success of a reception baseline assessment

The need for fair and consistent comparisons.

The purpose of an assessment should be made explicit. This stated purpose then guides decisions about the design of the assessment. The proposal by the DfE to reintroduce a baseline assessment is intended not as a starting point for measuring individual performance over time but for measuring the performance of a whole cohort. If this school-entry assessment is ultimately to be used to hold schools to account (as part of a high-stakes accountability system) it is essential that the assessment provides accurate and reliable outcomes, and that it enables fair and consistent comparisons across schools.

The importance of validity and reliability. Issues of validity and reliability arise in relation to any assessment – not just assessments of young children.

Reliability. An assessment is reliable if the outcomes produced can be shown to be accurate and consistent (for example, the outcome for each child would be the same or very similar if the assessment was repeated or carried out by a different practitioner).

Validity. The validity of any assessment is linked to the purpose for which the results are to be used. It represents the extent to which the score or outcomes of the assessment allow valid inferences to be made about the skills, understanding and knowledge of the child (or in this case the cohort). In research terms, validation is a process that takes place during the development of an assessment and continues during its use; threats to the validity of the assessment are identified and evidence is sought to demonstrate that the threats have been avoided or minimised.

A standardised administration approach is often perceived as more reliable than an observational one, whereas an observational approach is often perceived as more valid than a standardised one. Both forms of assessment have their advantages and disadvantages *but the issue is which of these approaches is the more appropriate to act as a baseline for measuring progress*.

Although it is often claimed that observational assessments during normal classroom activities are more valid due to their authenticity, a serious threat to the validity of such observational judgements is that there is too much variety in the way the assessments are conducted and scored.

If judgements of children are made based on observations in different circumstances, the danger is that the outcome may reflect the context of the assessment rather than the abilities of the child. Some circumstances may offer more support or contextual cues than others, making the task being assessed more (or less) accessible for the child. Controlling the circumstances in which the observations take place in order to make the judgements fairer can reduce this particular threat to validity but may be very time consuming to set up and result in something akin to a standardised assessment.

Reliability and mitigating the effects of external variables.

It is often assumed that because their performance can vary from day to day, it is not possible to get an accurate picture of children's abilities from a one-off assessment. In other words, the results of formal assessments of children will not be reliable.

In fact, research evidence demonstrates that high levels of 'test re-test' reliability (over the assessment as a whole) can be achieved with young children. In other words, they tend to achieve very similar scores when assessed on two separate occasions. On a sample of children assessed with the NFER RBA, an overall 'test re-test' correlation of 0.96 was achieved when the same children were assessed a second time within a week of the first administration. The outcomes were also very robust to different practitioners carrying out the assessments (i.e. high inter-rater reliability).

Authenticity and familiarity. Although uniformly administered, one-off standardised assessments of young children should be carried out sensitively by trained professionals and with sensible and age-appropriate administration instructions and guidance. Possible threats to validity can be minimised by using familiar practical resources and by including some familiarisation or introductory activities. In many cases young children will not even realise they are being assessed.

Critics of the use of the standardised assessments of young children have pointed to the lack of authenticity or the adverse impact of unfamiliar environments or unfamiliar tasks. And yet most young children in Reception experience a very common type of one-to-one 'formal assessment' – that of a familiar adult listening to them read or sharing a picture book to assess their familiarity with books.

In a child's eyes, a one-to-one baseline assessment may therefore appear no stranger (or possibly less strange) than several other activities they are asked to do in school that they have never done before, such as lining up, assemblies, or doing the register.

Equality and fairness. A key advantage of assessments administered in a standardised way is that every child in every school across the country is given the same opportunity to demonstrate what they know, what they understand and what they can do. The tasks, the resources and the way they are administered are the same for every child, reducing sources of irrelevant variation in scores. Further, the clearly defined yes/no criteria do not require any subjective interpretation and make it easy for teachers to reach consistent judgements within and across schools

Sensitivity to the age group. Standardising the tasks and the administration instructions does not mean assessments are automated or sterile. This underestimates the professionalism both of the assessment developers and the practitioners carrying out the assessments.

Well-designed assessments provide guidance on how to put the child at their ease at the start. Such assessments will also often employ discontinuation or routing rules to ensure children are able to attempt any tasks that they may have a possibility of completing successfully but are not faced with tasks that would be much too difficult for them.

The dangers of inaccurate media representation

Inaccurate media representations of standardised assessments of young children (for example, photographs suggesting children will be assessed by means of written tests, sitting in rows of desks) have helped to create a very negative image of such assessments – often referred to as 'tests' by critics to further foster this unhelpful representation. However, many children will enjoy the opportunity of working on a one-to-one basis with the teacher and having his or her undivided attention, if only for 20 or 30 minutes.

We would encourage the DfE to retain the flexibility for teachers to be able to choose when and where they administer the assessment, meaning it does not have to be carried out on a particular day or at a particular time (as with the current Baseline Assessment). By considering carefully when to assess each child, ensuring they are sufficiently settled in school, choosing an optimal time of day, etc., it should be possible for teachers to minimise the impact of factors such as the mood of the child or whether the child is hungry or tired.

Aggregation of cohort data. The DfE's proposed baseline measure will be at the cohort level, so individual variability will be 'smoothed out' over the cohort as a whole. Although some children may perform slightly better than expected on the assessment and some may perform slightly worse, the aggregated outcome should be sufficiently representative of the intake as a whole.

Of course, gaming by schools could distort the extent to which the baseline is representative of the cohort but this distortion would occur whatever the type of assessment. And there are wider issues about how the progress of an individual cohort, which may be small in a primary school, should be treated in accountability measures (for example, Allen, 2016).

Objectivity and practicality. Observational assessments are often perceived as more valid than standardised one-off assessments. The argument made is that you can build a more accurate picture of the child's abilities by observing them in more than one context over a period of time.

However, in practical terms it is very difficult to carry out multiple observations against every criterion if the assessment window is relatively short. There may be difficulties ensuring the activities observed provide sufficiently challenging opportunities for some children to demonstrate the full range of their abilities and there is no structured means by which the different contexts in which the children are being assessed and the impact of the actions of other children can be taken into account. The criteria used in such assessments are often open to some level of subjective interpretation.

Avoidance of 'localism'. Although training can help to moderate judgments in observational assessments and help build consistency within schools, as with any system based on subjective judgements, localism (local standards) can evolve and it can be very difficult to develop and maintain consistency across different schools.

It is also the case that because the observations may give contradictory evidence, teachers may err on the side of caution, which when multiplied over the assessment as a whole may give an inaccurate picture of the child (and of course an inaccurate picture of the cohort).

Summary and conclusions

The purpose of the DfE's proposed baseline is to assess the starting point of the cohort. Having measured the progress made between school entry and the end of Key Stage 2, schools will then be credited for the value they add by comparing the progress made by schools with similar intakes.

All assessment data should be treated with some caution but, if schools are to be held to account, NFER's judgement is that a standardised administration approach should be adopted for the new baseline assessment. Carefully developed so that threats to validity and reliability can be minimised, such an approach will be fairer and ensure more consistency between teachers and schools than an observational approach that will require excessive levels of initial and ongoing training and moderation in order to limit subjective interpretation.

References

Allen R. (2016). *Education Datalab's View on Primary Assessment*. Available: <http://educationdatalab.org.uk/2016/12/education-datalabs-view-on-primary-assessment/> [26 September 2017].

Department for Education (2017). *Primary Assessment in England. Government Consultation Response* [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644871/Primary_assessment_consultation_response.pdf [10 October 2017].

NAHT (2017). *Report of the Assessment Review Group: Redressing the Balance*. Available: www.naht.org.uk/welcome/news-and-media/key-topics/assessment/assessment-review-group-publishes-report-on-the-future-of-assessment/ [26 September 2017].

Wespieser, K. (2017). *Teacher views on baseline assessment*. Slough: NFER. Available: www.nfer.ac.uk/publications/IMTV01/ [26 September 2017].

Published October 2017

© National Foundation for Educational Research
Registered charity no 313392

978-1-911039-61-7

How to cite this publication:

Kirkup, C. (2017-P). *Delivering a reception baseline assessment. Why a standardised administration is fairer than an observational approach* (NFER Thinks: What the evidence tells us). Slough: NFER.

National Foundation for Educational Research
The Mere, Upton Park
Slough, Berks SL1 2DQ

T: 01753 574123
F: 01753 691632
E: enquiries@nfer.ac.uk
www.nfer.ac.uk

NFER

NFER is a not-for-profit organisation with a worldwide reputation for providing independent and robust evidence for excellence in education. Our authoritative insights inform policy and practice, offering a unique perspective on today's educational challenges.

Any surplus generated is reinvested in research projects to continue our work to improve outcomes for children and young people.

NFER's RBA

The NFER Reception Baseline Assessment, launched in 2015, uses a standardised administration approach. Throughout its development, researchers, early years experts and practitioners reviewed the tasks and questions (and all the associated trialling data) to ensure they would provide a valid assessment: examining evidence of the match of the tasks and checklists to the content domain, of the suitability of the assessment for the target age group and of the avoidance of bias.



Catherine Kirkup

Catherine has extensive experience managing and directing test development projects and other assessment research. She is a Practitioner of the Association for Educational Assessment-Europe and a Fellow of the Chartered Institute of Educational Assessors.

