

## **Young children as guinea pigs: the Reception Baseline Assessment Framework**

Primary schools in England are now being invited to volunteer to take part in a pilot of baseline assessment, by administering a full baseline assessment test to all children entering primary school in September 2019. This follows trials of specific test items in the current academic year. The government's Standards and Testing Agency (STA) has released documentation on the Assessment framework along with a brief supporting document.<sup>i</sup> The government intends to make the baseline assessment statutory for all schools beginning in 2020, in exchange for removing the statutory tests at the end of Year 2.

The STA document reasserts the purpose of the assessment, which is highly contested. It includes some clarification of content and administration of the test, as well as the intended use of the data generated. The framework addresses none of the substantive problems with baseline assessment that have been repeatedly identified by education experts, teachers and parents, which leaves young children entering school as unwitting subjects in an experiment, which is bound to fail, in generating and using data.

Objections to the use of baseline assessment have been clearly articulated elsewhere including detailed argument and evidence against the plans.<sup>ii</sup> A damning conclusion was reached by an expert assessment panel of the British Educational Research Association who asked, '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 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.'<sup>iii</sup>

Given such existing full examinations of the baseline proposals, this critique will be limited to considering the new documents released by the STA.

### **Purpose**

The STA framework states that the purpose is to create data as the starting point for a cohort-level progress measure for children from the start of their primary education to the end of Key Stage 2. Leaving aside for a moment the clear difficulties in developing a valid measure – that is, one that is measuring things that are in fact related to later progress – the statements of what the baseline is **not** intended for are also of interest. The baseline, STA explains, is not to provide on-going formative information for practitioners, nor to provide detailed diagnostic information about pupils' areas for development. In other words, it is very clearly not for the benefit of

supporting children's learning, but is simply a data collection exercise as the starting point for making judgments on schools seven years later.

Strangely, however, 'a series of narrative statements to describe how each pupil performed on the different content domains' will be provided to the school. It is difficult to understand how this decision has been reached and what the purpose of such information could be. Is it intended to help teachers find out about their pupils, in order to teach them more effectively? If so, then this is a formative purpose which was explicitly ruled out. Given the woefully inadequate nature of a brief test to uncover accurate and useful information about what 4-year-olds know and can do, this does not appear to be a useful adjunct to the more in-depth and appropriate ways teachers already assess young children on entry to school. Indeed, in the previous pilot of baseline assessment which included a very similar format from the same provider, only 8% of teachers said it was a fair and accurate way of assessing children, while 68% thought it was not a helpful way to assess children with English as an additional language, 71% thought it was not useful in identifying children with special needs, 68% said it increased their workloads, and 92% said they already had on-entry assessment systems in place.<sup>iv</sup>

Further difficulties arise in relation to the shifting population in schools, caused by a significant proportion of children moving around the country over the seven years of their primary education. Teachers and school leaders also change schools, adding to the illogicality of judging schools on data obtained seven years previously.

### **Reporting results**

Presumably the government has been sensitive to criticism about the dangers of early labelling of young children, which can result in lowered expectations for some children and ability grouping which maintains and exacerbates inequalities. Therefore the plan is that numerical results for individual children will be 'black-boxed' for seven years and not made available to schools, but will be held in the national pupil database.

Two major issues arise from this. The risk remains that children will be labelled on the basis of an inadequate test, since in administering the test teachers will see how each child is responding to the test items, as well as receiving the 'narrative summaries'. In addition, there are serious questions regarding the rights of parents to the data held on their child – and to give permission for the data to be generated in the first place. This is data produced and held without parental consent or oversight, regarding children who are not even yet of statutory school age.

### **Administering the test**

The time needed to administer the test is intended to be around 20 minutes. Since the trial phase did not involve any children sitting the entire set of tasks but only selected items, it is not clear how this estimate will relate to reality. It is unlikely that the time stated includes the necessary surrounding time for organising the context, orienting the child and ensuring they are reasonably comfortable with the unfamiliar situation before beginning. Even if 20 minutes is estimated, that still requires over 10 hours for a class of 30 children – time which would be much more valuably spent by staff in the early weeks in helping the children as a group to settle in to classroom routines and to build relationships, and getting to know the children in authentic contexts.

The STA explains that the test will include 'carefully designed routing', so that if children do not give correct answers in a section the test will move on to another area

so children are not 'presented with too many activities in which they are unlikely to be successful'. While it is understandable that children should not be presented with inappropriate demands, potential problems in accuracy arise. In a brief test designed to cover four areas in each of mathematics and language/literacy, the items selected can be only a sample of the skills and knowledge in these areas. Routing children out of an area following incorrect responses assumes that the items represent progressive knowledge, and that children attain knowledge in a linear way. Learning, however, is non-linear and what young children know and can do will not necessarily follow a predictable pattern.

### **Content domain**

The test is, according to STA, 'an age-appropriate assessment of mathematics and literacy, communication and language, that will be delivered in English'. These areas were chosen as 'skills which can be reliably assessed and which correlate with attainment in English and mathematics at the end of key stage 2, most noticeably early literacy and numeracy.' Again, the mistaken view of linear learning holds sway here, with the assumption that there will be a straight line of progress and therefore that measuring solely these two areas has a predictive capacity. There is no sound evidence that testing in these narrow areas is predictive of later attainment.

Unlike statutory tests at the end of other phases of education, which assess children's attainment in areas which have been taught, the reception baseline tests children not in what has been the holistic curriculum of their early years education – including the foundational prime areas of personal, social and emotional development, physical development, and communication and language, as well as the Characteristics of Effective Learning – but in the narrow areas where previous experience will have varied widely, depending on home background and early years settings. Just because a child has not been taught letters or numbers does not mean that they will not learn them rapidly when they are taught at the appropriate time.

The trial phase of the test development included exploration of whether it could include self-regulation, which is recognised as a significant factor in children's learning and a strong predictor of progress. Although the areas trialled were in any case an inadequate subset of self-regulation, looking at only limited areas of executive function, the STA explains that the Department for Education (DfE) has decided not to include self-regulation in the measure. As many critics predicted, the trials showed that it is not possible to assess self-regulation through test activities; as well as giving unclear results, the items took too long to administer and so couldn't be included within the 20-minute test. Therefore the decision has been to count what is easily counted, but not what actually counts.

All content in the baseline test is to be delivered in English. Although STA quotes an Ofqual requirement that 'A pupil should not be disadvantaged by factors that do not relate to what is being tested' and reports that the baseline assessment was subjected to a cultural review, it is inconceivable that children in the early stages of learning English will face a level playing field in responding to either language/literacy or maths tasks. Assessing children with special educational needs also requires particular skills and well-founded familiarity with the child's preferred communication style; it is not easy to see how these considerations will be fulfilled in the early weeks of attending an unfamiliar setting. Delays with documentation are not uncommon, so staff will not necessarily know at this stage in the term which children have special educational needs.

## **Cognitive domains and complexity**

The STA describes the test tasks as 'designed to maximise active interaction between the pupil and resources', with manipulation of physical resources possibly enabling some pupils 'to better demonstrate their skills and level of understanding to the practitioner'. Referring to young children as 'generally active and engaged pupils who learn through exploration and possess a natural curiosity', it is presumed that the hands-on tasks will stimulate children's thinking skills and intellectual processes. The test package includes physical materials such as plastic shapes and picture cards, used in activities in which children are required to respond orally (30-40%), by pointing (25-35%) or through ordering or moving objects (25-35%).

Children are indeed active learners, who demonstrate what they know and understand as they interact with other people and their environment. Early years assessment approaches are grounded in understanding that children show their embedded knowledge as they apply their skills and understanding across a range of experiences, and that therefore observation of children over time and in a range of contexts is the most reliable way to understand their level of learning and development. Children employ their highest cognitive abilities in play, where they ask their own questions, set their own challenges, and solve problems.

Closed assessment tasks presented in an unfamiliar and structured situation will in no way elicit children's highest functioning, and minimal handling of a few plastic materials in order to respond in limited ways does not begin to replicate the rich exploratory experiences through which children make links, express their understanding and find creative solutions. Although there may be 'reliability' in terms of replicating the same results regardless of who is administering the assessment, this does not mean that the test offers a valid picture of children's thinking and learning.

The STA reports that the tasks are designed to involve a range of cognitive complexity, from recall through to conceptual understanding and application. This is accomplished by moving from concrete activities involving physical resources to abstract activities, with a roughly 50-50 split between concrete and abstract activities in both maths and language/literacy. While STA describes this as 'appropriate for this age group', it is in fact inappropriate to expect such young children to engage with abstract items taken out of contexts that are meaningful for the child.<sup>v</sup>

## **Data**

Each child will be assigned a raw numerical score, having been offered between 26 and 45 possible points dependent on routing of the tasks. The data is not age standardised, so that children who may be a year apart in age will be ranked without regard to the enormous difference in maturity and experience of a quarter of their lives. The resulting differences in test scores could completely invalidate accountability measures for schools that have unbalanced proportions of younger children. But more importantly it can begin a negative cycle for the youngest children; they are unlikely to be ready to be judged against same standards as children nearly a year older. This can lead to them being taught in 'lower' groups which cement slower progress and cause disaffection with learning, leading to a high proportion of summer-born children being labelled as having special educational needs.<sup>viii</sup>

## Guinea pigs

There is no evidence to show – and much reason to doubt – that data from a baseline assessment can be of any use in judging schools’ performance seven years later. Young children are to be subjected to this experiment in accountability, using an untried assessment for no benefit whatsoever to the children themselves. Schools are currently being asked to take part in a pilot of the tests in September 2019, with the final version to be adjusted according to what is learned about how it works or does not work in practice. The government would like a significant number of schools to take part, because the results of the pilot year will be compared with the key stage one SATs three years later, to see whether in fact it is possible to construct a progress measure that links the two tests.

Assessment should be for the benefit of children, enabling teachers to understand their learning and development, and more widely their well-being and ways of interacting with the world around them – all to the end of meeting children’s needs more effectively. Baseline assessment offers the opposite, taking teachers away from authentic assessment while demanding a focus on a narrow range of learning that does not support the needs and potential of young children to grow as confident, engaged learners.

Children at the very start of their school experience should not be used in an experiment in an ill-advised accountability system.

---

i

Standards & Testing Agency (Feb 2019), Assessment framework – Reception Baseline Assessment, 978-1-78957-247-6; Reception baseline assessment framework: supporting document, 978-1-78957-248-3

ii

More Than a Score (2018), Baseline Assessment: Why it doesn’t add up, <https://morethanascorecampaign.files.wordpress.com/2018/02/neu352-baseline-a4-16pp-crop.pdf>

iii

BERA (2018) A Baseline Without Basis: The validity and utility of the proposed reception baseline assessment in England, <https://www.bera.ac.uk/researchers-resources/publications/a-baseline-without-basis>

iv

Roberts-Holmes, G. and Bradbury, A. (2016) “They are children... not robots, not machines”: The Introduction of Reception Baseline Assessment, <http://www.teachers.org.uk/files/baseline-assessment--final-10404.pdf>

<sup>v</sup> Wood, E. A. (2013). Free choice and free play in early childhood education: troubling the discourse. *International Journal of Early Years Education*, 22(1), 4–18.

Zosh et al (2017) Learning Through Play: a review of the evidence [https://www.legofoundation.com/media/1063/learning-through-play\\_web.pdf](https://www.legofoundation.com/media/1063/learning-through-play_web.pdf)

Ingrid Pramling Samuelsson & Maj Asplund Carlsson (2008) The Playing Learning Child: Towards a pedagogy of early childhood, *Scandinavian Journal of Educational Research*, 52:6, 623-641, DOI: 10.1080/00313830802497265

<sup>vi</sup> Crawford, C, Dearden, L, Greaves, E (2013) When you are born matters: Evidence for England, IFS Report R80

<sup>vii</sup> Bradbury, A. (2011) Rethinking assessment and inequality: the production of disparities in attainment in early years education, *Journal of Education Policy*, 26:5, 655-676, DOI: 10.1080/02680939.2011.569572