

EDUCATION
POLICY
INSTITUTE



Response to 'Primary Assessments in England' Government Consultation

June 2017

The aims of assessment in primary school

Summarising the aims and requirements set out in the consultation document, the Education Policy Institute has arrived at a list of 14 features of an effective primary assessment system:

1. Resilient to change
2. Rigorous, reliable and trusted
3. Differentiates between pupils with sufficient granularity
4. Provides useful information to schools for benchmarking and identifying problems
5. Suitable for identifying pupils who need additional support
6. Inclusive, taking account of pupils with SEND working below age expectations
7. Suitable for measuring school performance
8. Captures all of progress made during primary
9. But is also suitable for infant/junior/middle schools
10. Keeps stakes low, especially in early primary
11. Rounded, with proper reflection of teachers' work
12. Does not require changes to teaching practice
13. Proportionate and minimises burdens on schools
14. Avoids perverse incentives for schools

The consultation document states that the purpose of the national assessments would be for school accountability, which is reflected in most of the above points. The exception to this is point 5, which recognises that a useful function of the assessments would be to assist in identifying pupils with additional support needs.

Below, we outline a coherent system of assessment for primary schools which aims to promote these objectives within the constraints of measuring the attainment of young children. This is intended as an outline model for consideration and further development.

Baseline assessments

The consultation document proposes baseline assessments to provide a starting point for measuring pupil progress as a way of assessing school effectiveness. These assessments do not need to be modelled directly on the taught curriculum in order to capture the starting points of young children. Indeed, they are more reliable and less prone to manipulation or perverse incentives if they are more broad-based in nature, provided they capture important aspects of cognitive development. However, they must be reliable and granular, so as to ensure that progress can be accurately measured.

What could this mean in practice?

I. A single suite of validated assessments

A single suite of assessments based on tasks and questionnaires with standardised scoring, implemented individually adult-to-child, and with scope to repeat the assessment at a different time, or separate it into components to allow for the variable response typical of young children at different times.

There already are existing assessment tools that have been validated and used in longitudinal studies, such as the British Ability Scales test used by the Millennium Cohort Study and both the Wechsler Preschool and Primary Scale of Intelligence and Griffiths Mental Development Scales used by the Avon Longitudinal Study of Parents and Children.¹ These typically take around an hour per child in total to administer, less time than completing the EYSFP.

The use of established validated assessments would enable:

- Much better reliability than current arrangements, plus existing datasets for comparison at different time points or with different populations or age groups.
- Implementation of tests that have passed research ethics approvals and are not stressful for children as they involve supported one-to-one tasks and observations.
- Inclusion of extra assessment outcomes such as vocabulary development, self-regulation, and wellbeing to create a rounded assessment and provide objective initial screening for unidentified learning difficulties and/or emotional / behavioural / mental health problems. Emphasising the use of different outcomes will maximise the explanatory (predictive) power of progress at the key stage 2 results. Whilst this may add a layer of complexity to the measure, the alternative (simply averaging a range of outcomes) may result in a baseline that has insufficient explanatory power.

¹ More information on the assessments here: British Ability Scales <http://www.psychometrics.cam.ac.uk/services/psychometric-tests/gl-assessment>; Wechsler Preschool and Primary Scale of Intelligence <http://www.psychometrics.cam.ac.uk/services/psychometric-tests/wppsi-iii>; Griffiths Mental Development Scales <http://www.hogrefe.co.uk/gmds-er-2-8.html>

- Substantially reduced assessment, recording and moderation burdens for schools; even if repeated annually for three consecutive years, the burden would be lower than the current assessment regime for reception and key stage 1.
- The possibility of outsourcing the assessments to trained professionals such as educational psychologists to provide additional reliability and time-efficiency.

II. Age-standardised and repeated during reception and key stage 1

An ideal suite of assessments would be designed to be age-standardised, for use with children from age three through to age seven; and for most children, implemented three times – in reception, year 1 and year 2. The advantages of this repetition structure would be that:

- Three data points will enable more robust statistical baselining for all-through primaries, where each child has up to three progress estimates, enabling erroneous baselines to be smoothed down in the data. It would also lower the stakes of the assessments compared with a single assessment point.
- Children working below age expectations will be able to access the assessment by age seven, in some cases where they could not by age five. Especially if it is designed to be suitable from age three, early years practitioners can also use it to check development.
- Scores can be converted into a single comparable metric so everyone is using a coherent system throughout nursery and early primary education.
- Results from reception, year 1 and year 2 can be compared on a like-for-like basis. They can be averaged to reduce measurement errors, or substituted for one another where children missed initial assessments or were not proficient enough in English for it to be the standard baseline. The baseline for progress measures could then be based on a statistically designed model, which would select and combine the three data points to construct the best estimate for each child, taking into account the availability and plausibility of their results and the age range of the school in question, and place all children with results on a comparable footing.
- Children who arrive after reception but before the end of key stage 1, or whose English proficiency is insufficient to access the assessment at age five, can be picked up by age seven and included in progress measures and funding models. Currently, the number of EAL children with missing data leads to progress measures with incomplete coverage, and those with low prior attainment not being recognised as such, preventing the attraction of intended additional funding under the proposed national funding formula. However, an age-standardised suite of assessments would bring more children into the mainstream assessment system, so that their progress can be measured more accurately and robustly even if they do not fit within the expected range in their early years.
- Infant and junior schools could be included within the same framework, with infant schools using assessment at age five as the baseline where possible and junior schools in age seven. This would, in effect, allow progress models to be considered ‘progress per year’ due to the

age-standardisation. However, middle schools would remain an unsolved issue in accountability.

- We currently have no official assessment of English proficiency that is rigorous and consistent, but the language components of an age-standardised test may be re-usable with some older newly-arrived children learning English as an additional language to assess their proficiency in a way that is efficient and consistent with mainstream assessment.
- Standardised assessments can be implemented by other trained professionals such as education psychologists, who would then have an opportunity to screen children for unidentified SEND. This would ensure that these needs are identified early, and such professionals could provide ongoing specialised support to these children as part of a wider job role. They would provide expert judgement in addition to better national assessments.

A proportionate assessment system

The requirements of pupil-level diagnostic assessment, formative assessment, summative assessment and assessment for school accountability are likely to be different and should not be conflated. The stated purpose of assessment in the consultation is school accountability, with a principle of not collecting data that are not needed to provide this robustly.

There is no need to cover the whole curriculum to achieve this, and teacher assessments at key stages 1 and 2 should be made non-statutory to reduce workload and unnecessary tasks designed around tests rather than the learning needs of children. For example, outcomes from current pupil assessments show a strong correlation between reading and writing scores. Thus, collecting teacher assessments on student writing in addition to test data does not provide much more information about pupil performance in literacy. Unless the writing assessments can be made substantially more effective in differentiating between pupils, they are not useful for accountability purposes. However, novel assessment approaches that can afford teachers greater flexibility in judging whilst also ensuring greater reliability – comparative judgement for instance – may be worth piloting and evaluating to see if they alter the position by providing better information.

In the meantime, making these teacher assessments non-statutory would not be to signify that writing is unimportant, but to recognise that a good assessment of pupil performance for accountability purposes does not need to measure everything a pupil can do. This may, however, skew teaching focus again highlighting the need to keep stakes low.

An important exemption would be needed to retain statutory status for the current teacher assessments in the case of children who cannot access the tests at key stage 2, or the baseline assessments at ages five to seven even when designed to span ages three to seven for most children. This would be needed to ensure some accountability is possible for subsets of children with special educational needs or disabilities.

Further, the phonics screening check, the spelling, punctuation and grammar test, and the proposed multiplication check are not needed for accountability – they simply signal to schools the government's current priorities, instead of serving as rigorous accountability assessments. In addition, these tests have the potential to prescribe particular methods of teaching and skew

teaching time; they also do not fully demonstrate pupil ability. We believe they should be non-statutory, available through a test bank for teachers to use as they see fit.

Ensuring equality

The use of assessments in English only will disproportionately affect the results of children who speak a different language at home. Only 19.5 per cent of Indians are considered native English speakers, and likewise a high percentage of other ethnic minorities, apart from Black Caribbean, also grow up with a different home language.² Having an age-standardised assessment taken up to three times will help to moderate this effect as the impact of a different first language declines as children become older.

The use of elicited response assessments can also have negative impacts on some cultural subgroups, due to the different cultural variations in the ways that adults and children communicate. Certain children may not show their full verbal ability in assessments based on the elicited response model if they come from cultures which prioritise learning through observing rather than responding, or if children were not raised as direct conversational partners with adults (although this is common practice in white middle-class families). Thus, we must be wary not to associate developmental status with norms of dominant middle-class culture as they can misjudge a child's actual functional abilities based on their ethnic or social class. These can include both test design that would favour white middle-class children, such as verbal question and answer dialogues as shown above, and content developed from majority experiences and values.³

Thus, the interaction of a particular test style with a child's family background, such as a different home language or approach to child rearing, can negatively impact British ethnic minorities, and has traditionally done so. Ethnic minority children (Black Caribbean and non-Caribbean, Indian, Pakistani, Bangladeshi and Chinese) between the ages of three and five significantly underperform in early cognitive tests compared to White British-born pupils although they make greater progress and often subsequently outperform them in educational achievement.⁴

Young children with disabilities are also often negatively impacted by standardised assessment and will require flexibility in the choice of assessment methods, potential for modification of the instruments and a multidimensional, team-based assessment approach.

All of these issues further highlight the potential benefits of using standardised assessments that can be implemented by trained professionals who understand cultural norms and how they might affect assessments. While there are drawbacks to young children being tested by strangers they might not feel comfortable with, professionals would not only have the training to recognise and counteract this (for example, by short bonding games), but also be better prepared to implement different assessment methods, such as clinical interviews and multicultural assessments, when needed.

² Dustmann, Machin and Schonberg, 2010, 'Ethnicity and educational achievement in compulsory schooling', *The Economic Journal*, 120(August), p. F272-F297.

³ Reynolds, C.R. and Suzuki, L.A., 2013, 'Bias in Psychological Assessment: An empirical review and recommendations'.

⁴ Dustmann, Machin and Schonberg, 2010, 'Ethnicity and educational achievement in compulsory schooling', *The Economic Journal*, 120(August), p. F272-F297.

Trained professionals can also provide the further benefit of being less biased and unaffected by perverse incentives linked to school accountability outcomes.

Limitations

While we realise that there is no perfect way of eliminating cultural or other biases, the assessments should aim to minimise them as far as possible.

Delivery of all assessments by independent trained professionals is not currently plausible on current local authority staffing models. The proposed assessments could be implemented by teachers, possibly with some training required depending on the suite of assessments selected. However, to gain further benefits in terms of minimised bias, better diagnostic screening to enable early intervention for children with SEND and further reduced burdens on teachers, the ideal model would be one whereby other trained professionals deliver the assessments independently and is one worth considering. Teachers could then concentrate on formative assessment to support teaching and learning, separate from the accountability system.

Additionally, averaging across three baseline assessments where the results indicate that this is advisable might mean not all progress, particularly significant gains made over reception to year 2, can be reflected for every child. However, it is our view that this drawback is outweighed by the advantage gained from establishing a more reliable baseline for students.

Lastly, as acknowledged above, this system of assessments would not be perfectly comparable across all school types. However, it will provide better information on how infant and junior schools compare with other school systems. This will improve the reliability of information for infant schools and the fairness of junior school performance measurement. We recognise that middle schools will remain an unsolved problem in accountability; there is no solution that involves for baseline assessments for nine year olds that are comparable to those for younger children and so middle schools will need a separate system if they are to be as accountable as other schools.

Key Recommendations

To summarise, we recommend a holistic approach to primary assessments which would include:

1. Implementing a single suite of validated baseline assessments, age-standardised and completed annually during reception year, year 1 and year 2.
2. Making all teacher assessments in key stages 1 and 2 non-statutory, except for children who cannot take the tests.
3. Making the phonics screening check, the spelling, punctuation and grammar test, and the proposed multiplication check non-statutory, but available through a national test bank.
4. Allowing flexibility in assessment methods and content to meet the needs of ethnic minorities and students with SEND, along with the potential to be implemented by trained professionals.

	Current System		Proposed System	
	Statutory	Non-statutory	Statutory	Non-statutory
Reception Year	Early Years Foundation Stage Profile (EYFSP)		Baseline Assessment 1	EYFSP
Year 1	Phonics screening check		Baseline Assessment 2	Phonics screening check
Year 2	Teacher assessments in maths, science, English reading and English writing	English spelling, punctuation and grammar, test	Baseline Assessment 3	English spelling, punctuation and grammar test Teacher assessments in maths, science, English reading and English writing
Year 6	Tests in maths, English reading and English spelling, punctuation and grammar Teacher assessments in English reading, English writing, maths and science		Tests in maths and English reading	Teacher assessments in English reading, English writing, maths, and science English spelling, punctuation and grammar test