Department for Education Primary Assessment in England Government Consultation: NFER Response, 21st June 2017

Contact: Catherine Kirkup, c.kirkup@nfer.ac.uk

Overview of our response

1. **We welcome the Department for Education's decision to hold a consultation on primary assessment in England.** NFER has been developing high-quality, robust, reliable and informative assessments for 70 years, from classroom assessments to large-scale national and international assessments that support system-wide improvement. We have an extensive track record in undertaking research on assessment related issues and believe that the use of reliable assessments is an important part of effective teaching and learning. Our range of assessments and related services include NFER's own suite of robust primary phase tests for schools. These are widely known and recognised for providing teachers and schools with targeted information on their learners' performance and for providing high-quality feedback to learners and schools.

2. Our key recommendations are:

Preparing children to succeed at school

a) In order to provide evidence as to the most appropriate early learning goals to be assessed at the end of the reception year, we recommend that the profile should be reviewed by a panel of early years curriculum experts.

b) The panel should be commissioned to map detailed learning progressions and the key developmental milestones between the early years foundation stage, and the requirements of the 2014 KS1 (Key Stage 1) national curriculum. Based on their findings, the panel would recommend improvements to the EYFSP to strengthen continuity between the two phases.

The best starting point for measuring progress in primary school

c) The best starting point for measuring progress is a baseline assessment in reception. In order for it to be reliable, valid and manageable, the baseline should be a standardised assessment with all the measurement properties detailed in full in our response to question 5. The content of the baseline should be based on known predictors of later attainment as evidenced in the research literature.

d) If a baseline for measuring progress is introduced into reception, KS1 teacher assessment data should continue to be used as the baseline for measuring progress in the interim years (with all stakeholders being made aware of the limitations inherent in the data as detailed in our responses to questions 7 and 8).

e) Once a reception baseline for measuring progress is established, assessments at the end of KS1 should be made non-statutory. Annual monitoring of standards should occur at the end of KS1 via nationally representative sampling and using externally-marked KS1 tests. Implications for teaching and learning could be extracted from an analysis of the data and fed back to schools.

f) Where pupils have spent time in more than one school (whether this is an infant and a junior school or two all-through primaries), schools should be jointly held to account for the progress those pupils have made. Alternatively, new end of KS1 sample measures of national standards could include all infant schools to enable a measures of progress in infant and junior schools to be created, and compared to national progress in all-through primaries during each key stage.

A proportionate assessment system

g) Teacher assessment in reading and mathematics at KS2 (Key Stage 2) should be made non-statutory.

h) If the KS1 English grammar, punctuation and spelling test remains non-statutory beyond 2016/17, test papers should still be made available for optional teacher use.

i) The multiplication tables check should be administered during year 5.

Improving end-of-key stage statutory teacher assessment

j) The system of writing assessment should offer greater flexibility for teachers to use their professional judgement (e.g a best-fit model).

k) A more robust and credible teacher assessment framework needs to be developed to support teachers' judgement-making in writing and teachers should be fully supported in developing their skills in the assessment of writing (e.g. with guidance materials and professional development).

l) A robust evaluation of the use of comparative judgement approaches should be carried out, considering not only the technical and statistical evidence but also the implications of the use of such an approach for the assessment of writing.

m) Consideration should be given as to how to develop and support teachers' analytical marking of writing in the classroom in order to provide formative feedback that will enable pupils to develop their writing skills.

n) We recommend that the use of a comparative judgement approach should be evaluated as a tool to support the moderation of teacher assessment judgments.

3. Preparing children to succeed at school

Q1. The EYFSP measures a child's development against the ELGs set out in the EYFS statutory framework. Should the profile be improved to better assess a child's knowledge, skill, understanding and level of development at the end of the early years? If so, please describe which elements could be added, removed or modified.

Recommendations

- **In order to provide evidence as to the most appropriate early learning goals to be assessed at the end of the reception year, we recommend that the profile should be reviewed by a panel of early years curriculum experts.**

- **The panel should be commissioned to map detailed learning progressions and the key developmental milestones between the early years foundation stage, and the requirements of the 2014 KS1 national curriculum.**

- **Based on their findings, the panel would recommend improvements to the EYFSP to strengthen continuity between the two phases.**

The purpose of the EYFSP is to summarise a child's level of development at the end of the foundation stage. Whilst still fulfilling this purpose, we believe that improvements to the EYFSP could be made to provide clearer links between achievements at the end of the early years foundation stage and progress towards the KS1 national curriculum. Such improvements would provide a clearer picture for parents and carers about the developmental steps between the early years foundation stage and the end of KS1 and clearer links for teachers between the two curricula.

In the Rose review of the primary national curriculum published in 2009, it was recommended that there was a stronger focus on curriculum progression throughout the primary phase, including strengthening 'the continuity and progress in learning between the EYFS and KS1' (Rose, 2009). The 2011 Tickell review made similar recommendations and following this changes were made to the EYFSP, with the intention of aligning the level of 'exceeding' with the year 1 curriculum (Tickell, 2011). However, those changes were based on the pre-2014 national curriculum and therefore the alignments made then need to be reconsidered and updated.

References

Rose, J. (2009). *Independent Review of the Primary Curriculum: Final Report* [online]. Available: http://www.educationengland.org.uk/documents/pdfs/2009-IRPC-final-report.pdf [25 May, 2017].

Tickell, C. (2011). *The Early Years: Foundations for Life, Health and Learning. An Independent Report on the Early Years foundation Stage to Her Majesty's Government* [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/180919/DFE-00177-2011.pdf [25 May, 2017].

Q2. The EYFSP currently provides an assessment as to whether a child is 'emerging, expecting or exceeding' the level of development in each ELG. Is this categorisation the right approach? Is it the right approach for children with SEND?

No response to this question.

Q3. What steps could we take to reduce the workload and time burden on those involved in administering the EYFSP?

No response to this question.

Q4. How could we improve the consistency and effectiveness of the EYFSP moderation process whilst reducing burdens?

No response to this question.

4. The best starting point for measuring progress in primary school

Q5. Any form of progress measure requires a starting point. Do you agree that it is best to move to a baseline assessment in reception to cover the time a child is in primary school (reception to key stage 2)? If you agree, then please tell us what you think the key characteristics of a baseline assessment in reception should be. If you do not agree, then please explain why.

Recommendations
- **We agree that the best starting point for measuring progress should be a baseline assessment in reception.**
- **In order to be reliable, valid and manageable, the baseline should be a standardised assessment with all the measurement properties detailed in our response below.**
- **The content of the baseline should be based on known predictors of later attainment as evidenced in the research literature.**

Timing

The case for placing the starting point for the progress measure in reception is extremely strong. Placing the baseline as near as possible to the point of school entry would give schools credit for the progress children make across the whole of the primary phase and, as advocated by the then Chief Inspector of Schools Michael Wilshaw in 2013, it would acknowledge the valuable contribution made by schools in reception and in KS1 (Wilshaw, 2013).

In May 2017, we asked a representative sample of primary senior leaders and classroom teachers (via our Teacher Voice omnibus survey) to answer some questions related to the primary assessment consultation[1]. In one of the questions we asked whether teachers

---

[1] The survey sample was representative, in terms of our deprivation indicator (FSM), of all schools with a KS1 or KS2 cohort (excluding special schools, independent schools and alternative provision schools). Of the 653 respondents approximately one third were senior leaders and two thirds were classroom teachers. Respondents were presented with questions relevant to their school, i.e. they

agreed that measuring children's progress is a better way to assess the contribution schools make, rather than relying on absolute measures of attainment. Almost 90 per cent of respondents answered 'yes' to this question.

Placing the baseline in reception would maximise the credit given to schools for the value that they add. The introduction of a reception baseline has been advocated by an independent review group, commissioned by the NAHT (Assessment Review Group, 2016). We support this greater emphasis on progress throughout the whole school as a fairer means of assessing the effectiveness of a school.

An alternative option, not directly addressed in the consultation, would be to place the starting point for measuring progress at the beginning of year 1. This would align the starting point with the beginning of statutory school attendance and the national curriculum. If the EYFSP is to remain as a statutory assessment in reception, placing the baseline at the beginning of year 1 would avoid two statutory assessments in one year group and any workload issues that might arise as a result. A year 1 baseline would also avoid any difficulties that would arise in comparing cohorts from schools with different reception intake policies (e.g. staggered entry).

Our recent survey asked respondents about their preferred starting point from which to measure progress, setting out three options: early in the reception year, start of Y1 and end of KS1. Overall, 49 per cent of respondents chose the start of year 1, with a further 37 per cent choosing reception. However, the preferences of senior leaders and classroom teachers were somewhat different. Amongst the senior leaders 46 per cent indicated a preference for early in the reception year, with 40 per cent opting for the start of year 1. The equivalent figures amongst classroom teachers were 32 per cent and 54 per cent respectively (NFER, 2017, unpublished).

However, a year 1 baseline would not credit schools for the progress children make in reception.

As noted above, we believe the preferred option and the most appropriate starting point for a progress measure would be in reception. We would therefore recommend that, if this were adopted, the EYFSP should become non-statutory. This would reduce the workload burden on schools, particularly in reception. The EYFSP could be retained for optional school use to support formative assessment and to inform reports to parents.

Key characteristics: properties

Assessments need to be matched to the purpose for which they are going to be used. Wherever the starting point for a progress measure is situated, if it is to be used for school accountability, the baseline assessment needs to be reliable, valid and manageable. From a measurement perspective, the baseline assessment should have all of the following characteristics:

- The assessment should be a valid age-appropriate assessment of what children know and can do at the start of reception, designed by curriculum and assessment specialists.

---

only saw questions relating to KS1 if their school includes KS1 pupils. The number of respondents with pupils in KS1 was 556 and the number of respondents with pupils in KS2 was 590.

- In order to be administered fairly and consistently by teachers / practitioners, within and between schools, the assessment should have standardised administration instructions and standardised assessment criteria. All children should be assessed against the same tasks in the same way, giving all children an equal opportunity and reducing the impact of contextual differences inherent in observational assessments made during classroom activities.

- Given the age of the children, the assessment would need to be mediated by a teacher / reception practitioner. Providing prescriptive, standardised, objective yes/no criteria would minimise the potential for teacher bias.

- As well as being easy to administer, it should be extremely easy to record the yes/no judgements. Online recording would reduce the burden on reception practitioners.

- In order for it to produce an accurate cohort baseline, the assessment should have an adequate score range to cover the whole ability range, some parts of the assessment should be accessible to almost all pupils and there should be no ceiling effect.

- As the purpose of the baseline is to form the starting point for the measurement of progress in English and mathematics, it should focus on predictors of later attainment evidenced in the research literature. An appropriate baseline should correlate strongly with later attainment measures. See below for a discussion of appropriate content. As the EYFSP can still be used to gain a detailed picture of children's development, skills and understanding in all other areas of the Early Years curriculum, it would not be necessary to measure all areas of learning within the baseline.

- The baseline assessment should be accurate and reliable, demonstrating high internal reliability and test-re-test values.

Our own experience in developing the NFER Reception Baseline Assessment demonstrates that it is possible to develop an age-appropriate, manageable assessment that is sufficiently granular to differentiate between children and provides accurate and consistent results (GB. Parliament. HoC. Education Committee, 2017).

We do not believe it would be appropriate to modify the use of the EYFSP so that it could be used as a baseline for measuring progress. Effective assessments are developed to match the purpose for which they are being used. The EYFSP was developed not as a baseline but to summarise attainment at the end of the foundation stage and share information with parents and carers about how their children are progressing across all areas of the early years curriculum.

The assessment approach of the EYFSP is not appropriate for a baseline progress measure; criteria in an observational assessment are often open to subjective interpretation and therefore may lead to teacher bias and/or deflating of results (Campbell, 2013). Also because it is based on observing children during normal classroom activities, this inevitably results in children being assessed in different contexts, which may provide an unfair advantage for some pupils. Finally, the score distribution of the current EYFSP (see below) provides insufficient differentiation between children's starting points (and therefore insufficient differentiation between different cohorts).

Fig 1 Total points score distribution in the EYFSP for all pupils

England, 2016
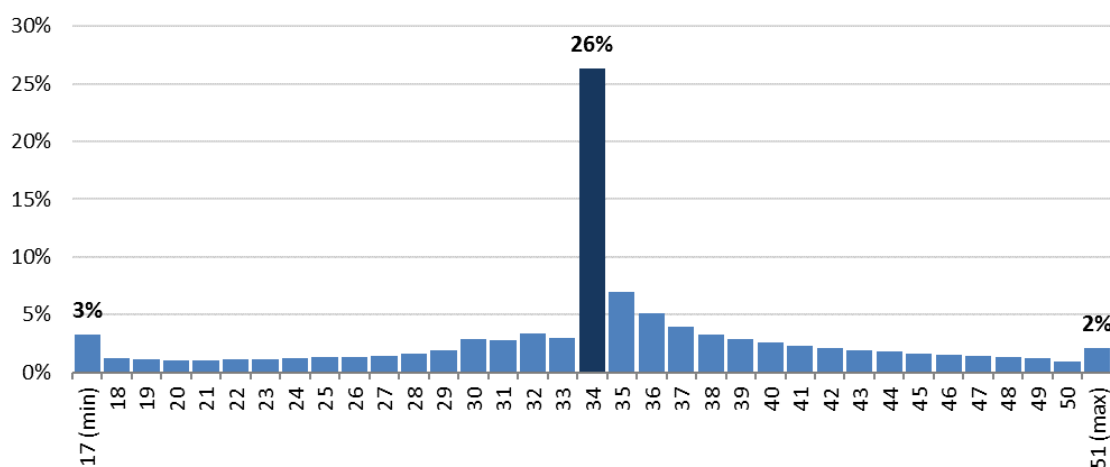
Source: DfE 2016

Key characteristics: content

In terms of appropriate content for a baseline assessment, as noted, consideration should be given to predictors of later attainment, particularly in reading and mathematics. Examples from the research literature provide evidence of positive relationships between some aspects of development at school entry and later attainment. One important consideration in deciding on the content of the baseline is the strength of the relationships between particular aspects of development and later attainment. For example, some aspects of personal, social and emotional development may be important in identifying children's developmental needs but may be only weakly associated (correlated) with later attainment. Even where research has shown that there is a statistically significant association when measured independently, this correlation may become insignificant when other more powerful relationships are taken into account (Duncan *et al.*, 2007).

In the case of baseline, it makes intuitive sense that the best predictors of later attainment in English and mathematics will be early skills in literacy and numeracy. This is supported by evidence from the research literature as shown below.

Language and literacy

Research has shown that language development is central to a child's ability to access the curriculum and develop literacy skills (Bowman *et al., 2000*).

Evidence shows that both receptive and expressive oral language skills are strongly related to literacy development (Cooper *et al.*, 2002). Children whose oral language is compromised are at risk of academic failure (Kieffer, 2008) and it is therefore important to measure early language and communication skills on school entry.

The Simple View of Reading provides a useful and well established framework for the understanding of reading development (Hoover and Gough, 1990). This model proposes that there are two sets of skills which contribute to reading: word recognition abilities and language comprehension abilities. Research has shown that reading accuracy is predicted

by single word reading ability, phonological awareness (Oakhill *et al.*, 2003) and additionally by letter knowledge (McGill-Franzen, 2010; Muter *et al.,* 2004). Furthermore, there is strong evidence to show that phoneme manipulation skills, rather than onset-rime awareness, predict later literacy skills (Savage and Carless, 2008).

Reading comprehension is predicted from the ability to draw inferences, the understanding of story structure, comprehension monitoring ability (Oakhill *et al.*, 2003) and also from vocabulary knowledge and grammatical skills (Muter *et al.*, 2004).

Numeracy

Competence in early mathematics is crucial for later school success. For example, early mathematics ability (at 54 months) has been found to predict mathematics attainment up to age 15, even after accounting for early reading, cognitive skills, and family and child characteristics (Watts *et al.,* 2014). The relationship between early number competence and later mathematical achievement has been well established (Aunio and Niemivirta, 2010; Jordan *et al.*, 2009) and there is clear evidence that numeral identification is related to the acquisition of numeracy skills (Wright *et al.,* 2006).

In addition to number recognition, counting skills, such as number- word sequence skills and enumeration skills, an early understanding of relations in shape, order or quantity and general number knowledge have been found to be good predictors of later performance in mathematics (Aubrey *et al.,* 2006; Aubrey and Godfrey, 2003). There is also increasing evidence that an awareness of mathematical pattern and structure is crucial to mathematical competence in young children (Mulligan and Mitchelmore, 2009). More recently, a longitudinal study (Nunes *et al.*, 2011) identified that both mathematical reasoning and arithmetic make independent contributions to the prediction of mathematical achievement, with mathematical reasoning being the strongest predictor. This includes children's abilities to understand and make relational statements, compare, classify and understand one-to-one correspondence and seriation.

There is also a substantial body of evidence that shows a strong relationship between numeracy skills and early literacy skills (Welsh *et al.*, 2010). Research shows that early knowledge of numbers and mathematical concepts are not only strong predictors of later achievement in mathematics, but also of word identification and reading (Duncan *et al.*, 2007; Scanlon and Vellutino, 1996). Studies report that numeral and letter identification are correlated at an early age and that both are equally predictive of word identification (Scanlon and Vellutino, 1996). Underpinning this is the ability to understand and manipulate symbol systems and the fact that numbers and letters share similar perceptual qualities (Cook, 1996). In a study carried out by CEM (Tymms *et al.*, 2012), early mathematical skills, particularly the ability to identify numbers and do simple sums, have been found to be the best indicators of later achievement in mathematics, reading and science.

Attention / self-regulation skills

Teachers tend to conceptualise school readiness in terms of children's social and emotional skills rather than their cognitive abilities. Recent research has focussed on the concept of self-regulation and the ways in which it is closely interrelated with cognitive abilities (Blair and Raver, 2015), possibly mediated through a 'positive adaptation to school' (Blair and Diamond, 2008). It has further been claimed - in a review of the literature - that self-

regulatory and metacognitive activities are not only important to later academic achievement but are also teachable skills (Whitebread and Basilio, 2012).

Self-regulation, in particular the regulation of attention (e.g. following instructions, resisting distractions, etc.) may contribute to later academic attainment; although the strength of the association appears to be much weaker than the more direct links detailed above for early literacy and numeracy skills. In a meta-analysis of six international studies (Duncan *et al.*, 2007), early skills in reading and mathematics were most closely associated with later academic achievement. However, the ability to sustain attention was found to be modestly, but consistently, associated with achievement outcomes even after controlling for cognitive ability and background factors. Other aspects of social and emotional development, including social skills, did not predict later academic performance.

References

Assessment Review Group (2017). *Redressing the Balance* London: NAHT [online]. Available: https://www.nahtedge.org.uk/Portals/4/Content/Assessment%20Review%20Group%20report%20January%202017.pdf [25 May, 2017].

Aubrey, C. and Godfrey, R. (2003). 'The development of children's early numeracy through key stage 1', *British Educational Research Journal,* **29,** 6, 821–840 [online]. Available: http://onlinelibrary.wiley.com/doi/10.1080/0141192032000137321/abstract?systemMessage=WOL+Usage+report+download+page+will+be+unavailable+on+Friday+27th+January+2017+at+23%3A00+GMT%2F+18%3A00+EST%2F+07%3A00+SGT+%28Saturday+28th+Jan+for+SGT%29++for+up+to+2+hours+due+to+essential+server+maintenance.+Apologies+for+the+inconvenience [25 May, 2017].

Aubrey, C., Dahl, R. and Godfrey, S. (2006). 'Early mathematics development and later achievement: further evidence', *Mathematics Education Research Journal,* **18,** 1, 27–46 [online]. Available: http://link.springer.com/article/10.1007/BF03217428 [25 May, 2017].

Aunio, P. and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Difference*, **20**, 427-435 [online]. Available: http://www.helsinki.fi/~niemivir/AunioNiemivirtaL%26ID2010.pdf [25 May, 2017].

Blair, C and Diamond, A. (2008). 'Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure', *Development and Psychopathology,* **20,** *3*, 899–911 [online]. Available: https://www.cambridge.org/core/journals/development-and-psychopathology/article/div-classtitlebiological-processes-in-prevention-and-intervention-the-promotion-of-self-regulation-as-a-means-of-preventing-school-failurediv/0ABF16C72FF44FF2CD6D15E6433962A0 [25 May, 2017].

Blair, C. and Raver, C. C. (2015). 'School readiness and self-regulation: A developmental psychobiological approach', *Annual Review of Psychology,* **66,** 711–731 [online]. Available: http://www.annualreviews.org/doi/abs/10.1146/annurev-psych-010814-015221?journalCode=psych [25 May, 2017].

Bowman, B.T., Donovan, M.S. and Burns, S.M. (2000). 'Eager to Learn: Educating our Preschoolers'. *National Academy Press* [online]. Available: https://www.nap.edu/read/9745/chapter/1 [25 May, 2017].

Campbell, T. (2013). Stereotyped *at Seven? Biases in Teacher Judgements of Pupils' Ability and Attainment.* (CLS Working Paper 2013/8). London: Centre for Longitudinal Studies [online]. Available: http://www.cls.ioe.ac.uk/shared/get-file.ashx?itemtype=document&id=1715%20 [25 May, 2017].

Cook, D. 1996. 'Mathematical sense making and role play in the nursery'. *Early Child Development and Care* **121,** 55–66 [online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0300443961210106 [25 May, 2017].

Cooper, D.H., Roth, F.P., Speece, D.L. and Schatschneider,C. (2002). 'The contribution of oral language skills to the development of phonological awareness', *Applied Psycholinguistics, 23,* 3, 399–416. (DOI: 10.1037//0033-295x.108.1.204).

Department for Education (2016). *Early Years Foundation Stage Profile Results in England, 2016* (SFR 50/2016). London: DFE [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/561224/SFR50_2016_Text.pdf [25 May, 2017].

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007). 'School readiness and later achievement', *Developmental Psychology, 43,* 6, 1428–1446 [online]
Available: http://psycnet.apa.org/journals/dev/43/6/1428/ [25 May, 2017].

Great Britain. Parliament. House of Commons. Education Committee (2017). *Supplementary Written Evidence Submitted by the National Foundation for Educational Research* [online]. Available: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/education-committee/primary-assessment/written/46274.html [7 June, 2017].

Hoover, W.A. and Gough, P.B. (1990). 'The simple view of reading', *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160 [online]. Available: https://link.springer.com/article/10.1007/BF00401799 [25 May, 2017].

Jordan, N.C., Kaplan, D., Ramineni, C. and Locuniak, M. N. (2009). 'Early math matters: kindergarten number competence and later mathematics outcomes'. *Developmental Psychology*, **45,** 3, 850–867 [online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782699/ [25 May, 2017].

Kieffer, M.J. (2008). 'Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States', *Journal of Educational Psychology,* **100,** 4, 851 [online]. Available: https://www.researchgate.net/publication/232513126_Catching_Up_or_Falling_Behind_Initial_English_Proficiency_Concentrated_Poverty_and_the_Reading_Growth_of_Language_Minority_Learners_in_the_United_States [25 May, 2017].

McGill-Franzen, A. (2010). 'The national early literacy panel report: summary, commentary, and reflections on policies and practices to improve children's early literacy', **39,** 4, 275–278

Public

[online]. Available: http://journals.sagepub.com/doi/pdf/10.3102/0013189X10370619 [25 May, 2017].

Mulligan, J. and Mitchelmore, M. (2009). 'Awareness of pattern and structure in early mathematical development', *Mathematics Education Research Journal,* **21**, 2, 33–49 [online].
Available: https://pdfs.semanticscholar.org/eed3/da2272674128071352c79ed3e2e2aab9b09a.pdf [25 May, 2017].

Muter, V., Hulme, C., Snowling, M.J. and Stevenson, J. (2004). 'Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study', *Developmental Psychology*, **40**, 5, 665–681
[online]. Available: http://www.itari.in/categories/ability_to_learn/phonemes_rimes_vocabulary_and_grammatical_skills_as_foundations_of.pdf [25 May, 2017].

Nunes, T., Bryant, P., Evans, D. and Barros, R. (2011). *An Intervention to Improve Children's Performance in Mathematics in KS1.* London: DfE [online].
Available: http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2015/01/Nunes_Numeracy2014.pdf [25 May, 2017].

Oakhill, J.V., Cain, K. and Bryant, P.E. (2003). 'The dissociation of word reading and text comprehension: Evidence from component skills' *Language and Cognitive Processes*, **18**, 4, 443–468 [online]. Available:
https://pdfs.semanticscholar.org/5433/dc562dbdc20d8be52dc3578124038b4bd1b3.pdf [25 May, 2017].

Savage, R. and Carless, S. (2008). 'The impact of early reading interventions delivered by classroom assistants on attainment at the end of year 2 '*British Educational Research Journal*, **34**, 3, 363–385 [online]. Available:
http://onlinelibrary.wiley.com/doi/10.1080/01411920701609315/abstract [25 May, 2017].

Scanlon D.M. and Vellutino F.R. (1996). 'Prerequisite skills, early instruction, and success in first grade reading: selected results from a longitudinal study', *Mental Retardation and Developmental Disabilities Research Reviews,* **2**, 1, 54–63 [online].
Available: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1098-2779(1996)2:1%3C54::AID-MRDD9%3E3.0.CO;2-X/full [25, May, 2017].

Tymms, P., Merrell, C., Henderson, B. and Jones, P. (2012). 'Learning difficulties in the primary school years: predictability from on-entry baseline assessment'. *Online Educational Research Journal* [online]. Available:
https://www.researchgate.net/publication/255180648_Learning_Difficulties_in_the_Primary_School_Years_Predictability_from_On-entry_Baseline_Assessment [25 May, 2017].

Watts, TW., Duncan, G.J., Siegler, R.S. and Davis-Kean, P. E. (2014). 'What is past is prologue: relations between early mathematics knowledge and high school achievement.' *Educational Researcher,* **43,** 7, 352–360 [online]. Available:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4719158/ [25 May, 2017].

Welsh, J.A., Nix, R.L., Blair, C., Bierman, K.L. and Nelson, K.E. 2010. 'The development of cognitive skills and gains in academic school readiness for children from low-income families'. *Journal of Educational Psychology,* **102,** 1, 43–53 [online]. Available:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856933/ [25 May, 2017].

Whitebread, D. and Basilio, M. (2012) 'The emergence and early development of self-regulation in young children', *Profesorado: Revista de Curriculum Y Formación de Profesorado*, **16**, 1 [online]. Available: http://www.ugr.es/~recfpro/rev161ART2en.pdf [25 May, 2017].

Wilshaw, M. (2013). 'Unseen children.' Speech at Church House, Westminster, London, 20 June 2013 [online]. Available: https://www.gov.uk/government/speeches/unseen-children

Wright, R.J., Martland, J. and Stafford, A.K. (2006). *Early Numeracy. Assessment for Teaching and Intervention. Second Edition.* London: Sage Publications Ltd.

Q6. If we were to introduce a reception baseline, at what point in the reception year do you think it should be administered? In particular, we are interested in the impact on schools, pupils and teaching of administering the assessment at different times.

The arguments in favour of the first or second half terms are finely balanced. If a baseline assessment is to be introduced in reception, placing the assessment at the beginning of second half term would give all children time to settle into school. However, it would not credit schools for the progress children make in the first half term.

Our experience with the NFER Reception Baseline Assessment has shown that many schools wanted to carry out the assessment as soon as possible. In 2015, when schools using the NFER Reception Baseline Assessment were asked to assess children within the first half term, over half of those schools chose to carry out the assessments within the first four weeks.

Irrespective of whether the baseline is to be administered during the first or second half term of reception, teachers should be given discretion within the specified assessment window to choose an appropriate time to assess each child, taking into account their age and the extent to which they have settled into school.

Some schools have more than one intake and children may join the school at different points in the reception year. From a measurement perspective, the impact of different intake policies on the average age of the cohort at the time of the assessment may need to be considered if the baseline is not age-standardised. Finally, care would need to be taken in the transfer of data to avoid children being assessed more than once should they move schools within the reception year.

Q7. Our view is that it would be difficult to change key stage 1 assessment in order that it could be used as the baseline for progress in the long term. If you disagree, what could be done to improve the key stage 1 assessments so that they would be sufficiently detailed, and trusted as a fair and robust baseline?

> **Recommendation**
>
> - **We agree with the recommendation that instead of changing KS1 assessments, the baseline for measuring progress should be moved from the end of KS1 into reception.**

Using KS1 assessment data as a baseline would not credit schools for the contribution they make to children's development in the first three years of school. We would therefore recommend placing the starting point for the measurement of progress in reception, as stated in our response to Q5.

The use of KS1 data as the baseline for measuring progress is problematic due to its reliance on teacher assessment data. There is already evidence that teacher assessment at KS1 can be biased (Campbell, 2013) or unreliable (Johnson, 2013) or both (Harlen, 2007). Introducing the progress threshold at KS2 has raised the stakes of the KS1 assessments, creating incentives for teachers to game the system, deflating results to demonstrate greater progress at the end of KS2.

The introduction of the current teacher assessment frameworks has reduced the number of teacher assessment categories into which pupils can be placed (compared to the previous system of levels) reducing the differentiation available between pupils and making it more difficult to use this data to measure progress effectively.

A potential improvement to the current system as outlined in the consultation document i.e. increasing the number of teacher assessment categories, could increase the workload burden without necessarily improving the reliability and robustness of the KS1 data. There would need to be a significant and costly level of moderation to ensure consistency across schools in the use of the teacher assessment categories and to ensure schools were not gaming the system.

An alternative approach could be generated if the KS1 tests were externally marked in the same way as the KS2 tests and the data used to provide the baseline. Such an approach would provide sufficient score differentiation between pupils and eliminate teacher bias, potential gaming and the need for moderation of teacher assessment judgements. However, as noted in the consultation document, this would raise the stakes of the tests for both pupils and schools with potential adverse impacts on teaching and learning. Using test data also has the disadvantage of omitting some areas of the curriculum that cannot be assessed by means of written tests (e.g. speaking and listening; reading aloud).

Making changes to KS1 assessments would not remove the key disadvantage of using a KS1 baseline; that it is too late to be used as an effective starting point for measuring pupil progress across the primary phase. And therefore our recommendation is that the baseline should be in introduced into reception.

References

Campbell, T. (2013). *Stereotyped at Seven? Biases in Teacher Judgements of Pupils' Ability and Attainment.* (CLS Working Paper 2013/8). London: Centre for Longitudinal Studies [online]. Available: http://www.cls.ioe.ac.uk/shared/get-file.ashx?itemtype=document&id=1715%20 [25 May, 2017].

Harlen, W. (2007). 'Teachers' summative practices and assessment for learning – tensions and synergies', *The Curriculum Journal*, **16,** 2, 207–223 [online]. Available: http://www.tandfonline.com/doi/abs/10.1080/09585170500136093 [7 June, 2017].

Johnson, S. (2013). 'On the reliability of high-stakes teacher assessment', *Research Papers in Education,* **28,** 1, 91–105 [online]. Available: http://www.tandfonline.com/doi/abs/10.1080/02671522.2012.754229 [7 June, 2017].

Q8. If we were to introduce a new reception baseline measure, do you agree that we should continue to use key stage 1 teacher assessment data as the baseline for measuring progress in the interim years before a new measure was in place? If you disagree, what do you think we should use as the baseline instead?

Recommendation

- **We agree with the recommendation that if a baseline for measuring progress is introduced into reception, KS1 teacher assessment data should continue to be used as the baseline for measuring progress in the interim years.**

If a new baseline assessment were introduced in reception, the least disruptive means of measuring progress in the interim years would be to continue to use the current KS1 teacher assessment data as the baseline. However, as noted in our response to Q7 there are concerns about the consistency and accuracy of teacher assessment judgements, the limited range of categories from which to generate a progress measure and the perverse incentives to deflate results.

The alternative outlined in the consultation document would be to collect KS1 test data to use as the baseline for measuring progress in the interim years. The requirement for schools to report teacher assessment judgements in reading and mathematics could then be removed to reduce teacher workload. The use of KS1 test data would provide more differentiated baseline scores and test results could be moderated more effectively and efficiently than moderating teacher assessment judgements if required (e.g. by externally marking sampled scripts). However, as noted in our response to Q7, this would raise the stakes of the tests for both pupils and schools with potential adverse impacts on teaching and learning. The benefits of using test data may not outweigh the disruption that further changes to the system would entail, particularly as this would only be an interim measure.

Although there is as yet very limited evidence, a small study (900 pupils) looking at the relationships between teacher assessment and test scores at KS1 found that a wide range of scores was associated with 'working at the expected standard' (scaled score range of 85-

115 in reading and 92-115 in mathematics). Although this finding casts doubt on the assertion that teachers use test scores to inform their teacher assessment, there was some evidence that the 100 scaled score, which represents the expected standard, was exerting some influence on such decisions (Pembroke, 2016).

Given that making changes to the KS1 assessments would be both costly and disruptive, in the interim years, until a more robust baseline is established, it might be simpler to continue with the current system, with both schools, Ofsted and all other stakeholders being aware of the limitations inherent in the data.

References

Pembroke, J. (2016). 'Analysis of the relationships between teacher assessment and test scores at KS1', *School Data Updates,* 30 June [online]. Available: http://sigplus.blogspot.co.uk/2016/06/analysis-of-relationships-between.html [7 June, 2017].

Q9. If a baseline assessment is introduced in reception, in the longer term, would you favour removing the statutory requirement for all-through primary schools to administer assessments at the end of key stage 1?

Recommendations

- **We agree with the recommendation that if a baseline for measuring progress is introduced into reception, assessments at the end of KS1 should be made non-statutory, once the baseline is established.**

- **We recommend annual monitoring of standards at the end of KS1 via nationally representative sampling and using externally-marked KS1 tests. Implications for teaching and learning could be extracted from an analysis of the data and fed back to schools.**

We agree with the recommendation that if a baseline assessment is introduced in reception, KS1 assessments (both teacher assessment judgements and national curriculum tests) in all-through primary schools should eventually be made non-statutory. This would have a positive impact on teacher workload.

When asked how much impact each of the potential changes to primary assessment arrangements proposed in the consultation document would have in reducing workload in their school, over 90 per cent of primary classroom teachers and senior leaders indicated that the proposal to remove the obligation for schools to assess pupils against statutory teacher assessment frameworks at the end of KS1 would result in 'some reduction' or a 'significant reduction' in workload. Similarly, over 90 per cent of respondents indicated that the proposal to make the end-of-KS1 national curriculum tests non-statutory would result in 'some reduction' or a 'significant reduction' in workload. In each case, over half of these respondents indicated it would result in a 'significant reduction' in workload. In addition, 52 per cent of teachers said they would still want to use optional test materials to assess pupil attainment and progress if KS1 assessments become non-statutory; amongst senior leaders the percentage was slightly higher, at 59 per cent (NFER, 2017, unpublished survey).

Public

This suggests that the majority of schools would still want to benchmark their performance at the end of KS1 against the national picture. The government, parents and other stakeholders would also want to know that standards of attainment in English and mathematics were being maintained.

Based on NFER's long experience of developing both statutory and non-statutory assessments our recommendations would be as follows:

- Annual monitoring of national standards at the end of KS1 could be achieved via representative sampling using KS1 tests (similar to the system currently used for KS2 science sampling). To reduce the burden on schools and to ensure accuracy we would recommend that these tests are externally marked.

- Unseen secure tests used for national monitoring could be re-used in subsequent years as optional tests for teacher use, allowing them to benchmark their performance against a nationally representative sample.

National sampling using KS1 tests would also provide an opportunity to code and analyse pupil responses in detail, reporting patterns of performance and common errors and misconceptions (made by different groups of pupils). This could not only assist in the monitoring of national standards but also provide very useful feedback for schools to inform teaching and learning in KS2.

Q10. If we were to introduce a reception baseline to enable the creation of reception to key stage 2 progress measures for all-through primaries, what would be the most effective accountability arrangements for infant, middle and junior schools' progress measures?

Recommendation

- **Where pupils have spent time in more than one school (whether this is an infant and a junior school or two all-through primaries), schools should be jointly held to account for the progress those pupils have made. Alternatively, new end of KS1 sample measures of national standards could include all infant schools to enable measures of progress in infant and junior schools to be created, and compared to national progress in all-through primaries during each key stage.**

All schools will experience some level of pupil mobility during the primary phase, with pupils moving between schools for many reasons. Although infant, middle and junior schools are a special case, all schools will have some children in their schools who have not been in their school for seven years.

Judging infant, middle and junior schools on a different basis to all-through primaries, using different baseline measures or different outcome measures, could be costly, confusing for parents and burdensome for schools.

Where pupils have spent time in more than one school (whether this is an infant and a junior school or two all-through primaries), consideration should be given as to how schools can be jointly held to account for the progress those pupils have made.

During this period of transition, until the new reception baseline to end KS2 progress can be measured, it would be beneficial to carry out robust research into whether it is possible to measure the absolute progress on average that pupils make in each phase. Such research could potentially inform decisions as to how responsibility for progress should be apportioned between two or more schools.

An alternative option that could be considered would be to include all infant school pupils in an annual sample used for monitoring standards at KS1 (see our response to Q9). The infant school data would be analysed separately and could then form the outcome measure for infant schools and the baseline measure for junior schools. This could be compared to national progress measures during each key stage based on the wider sample of all-through primary schools.

5. A proportionate assessment system

Q11. Do you think that the department should remove the statutory obligation to carry out teacher assessment in English reading and mathematics at key stage 2, when only test data is used in performance measures?

Recommendation

- **We support the recommendation to make teacher assessment in reading and mathematics at KS2 non-statutory.**

Balanced against other considerations, and with the aim of reducing the overall burden of statutory assessment, we support the recommendation to make teacher assessment in reading and mathematics at KS2 non-statutory.

When asked how much impact each of the potential changes to primary assessment arrangements proposed in the consultation document would have in reducing workload in their school, over 50 per cent of primary classroom teachers and senior leaders indicated that the proposal to remove the statutory obligation to carry out teacher assessment in English reading and mathematics at end of KS2 would result in a 'significant reduction' in workload, and a further 38 per cent indicated it would result in 'some reduction' in workload (NFER, 2017, unpublished survey).

Although there may be some disquiet that removing this statutory obligation is devaluing teachers' professional judgments, it is important to note that teacher assessment would still be incorporated into the statutory assessment system in the reporting of assessments in English writing and science.

Although teachers would still be required to report pupils' test performance in reading and mathematics in a broader context when reporting to parents, removing the statutory

obligation to use the teacher assessment frameworks for this purpose would give schools more autonomy in how they do this.

In the subject areas where teacher assessment is retained as part of statutory assessment, it is important that training and support is provided to build common understanding of standards locally / regionally / nationally. Evidence suggests teachers find it difficult to be objective in their teacher assessments and may explain why some teachers welcome nationally standardised tests in other year groups to support their assessment judgements. In a report published in 2013, concerning teacher assessment in Wales, ACER found that confidence in the accuracy and reliability of teacher assessment judgments, including the moderation processes, was extremely low within the teaching profession and the wider educational community (ACER, 2013). One of the recommendations of the report authors was further training in making best-fit judgements, internal standardisation and internal and external moderation. A shared understanding of standards was believed to be essential for the development of more effective teacher assessment practice.

As the use of scaled scores for the national curriculum tests was introduced in 2016, we believe there is also a need to offer further guidance and professional development in the reporting and interpretation of KS2 test outcomes.

References

Australian Council for Educational Research (ACER) (2013). *An Investigation into Key Stages 2 and 3 Teacher Assessment in Wales.* Cardiff, DfES [online]. Available: http://gov.wales/docs/caecd/research/130718-investigation-key-stages-2-3-teacher-assessment.pdf [25 May, 2017].

Q12. Do you agree that the key stage 1 English grammar, punctuation and spelling test should remain non-statutory beyond the 2016 to 2017 academic year, with test papers available for teachers to use as they see fit?

Recommendation

- **We agree that the KS1 English grammar, punctuation and spelling test should remain non-statutory. If this occurs we recommend that test papers should still be made available for optional teacher use.**

In NFER's recent survey, 58 per cent of teachers reported that they had used the non-statutory 2016 KS1 English grammar, punctuation and spelling test to support their teacher assessment judgements of writing. Of these, 65 per cent had found them either 'useful' or 'very useful' (NFER, 2017, unpublished survey).

We think it would be beneficial to carry out further research to find out how the non-statutory KS1 English grammar, punctuation and spelling tests are being used and for what purpose(s). It would be interesting to discover the extent to which teachers are confident in teaching and assessing this area of the curriculum and whether usage of the tests relates to their level of confidence.

Q13. At what point in key stage 2 do you think the multiplication tables check should be administered? Please explain the basis for your views.

a) At the end of year 4

b) During year 5

c) During year 6

> **Recommendation**
>
> - **Our recommendation is that the multiplication tables check should be administered during year 5.**

Our recommendation is that the multiplication tables check should be administered during year 5. By the end of year 4, pupils should have learnt all the multiplication tables but they may not have had time to fully consolidate that learning. Placing the check in year 4 may cause some schools to introduce practice tests in year 4, rather than maximising the time available for teaching and learning of the tables. Placing the multiplication tables check in year 5 allows time for knowledge to embed before the check is administered. It also allows for a follow up year of further support where necessary before pupils leave primary education.

We would not recommend administration of the multiplication tables check in year 6. Adding to the assessments carried out in year 6 would add to teacher workload as well as having a potential adverse impact on year 6 pupils. Although research has shown considerable differences in the ways pupils experience the KS2 tests (Putwain *et al,* 2012), any increase in the number of assessments in year 6 would be likely to increase concerns about adverse impacts on pupils' well-being.

When asked at what point the multiplication tables check should be administered, 17 per cent of primary classroom teachers and senior leaders opted for year 5, compared to 10 per cent for year 4 and 10 per cent for year 6. However, 30 per cent of respondents felt that none of these points were appropriate (NFER, 2017, unpublished survey).

References

Putwain, D.W., Connors, L., Woods, K. and Nicholson, L.J. (2012). 'Stress and anxiety surrounding forthcoming standard assessment tests in English schoolchildren'. *Pastoral Care in Education,* **30,** 4, 289–302 [online]. Available: https://repository.edgehill.ac.uk/4777/ [25 May, 2017].

Q14. How can we ensure that the multiplication tables check is implemented in a way that balances burdens on schools with benefit to pupils?

Firstly, as noted in response to Q13, introducing the multiplication tables check in year 5 will avoid adding to the assessment burden in year 6.

An online test model with an effective user-friendly design will provide a multiplication tables check of five minutes or less, with an instant score and no need for marking. Provided any

input or familiarity issues can be resolved, this would maximise benefit to pupils with minimal burden on schools.

Q15. Are there additional ways, in the context of the proposed statutory assessments, that the administration of statutory assessments in primary schools could be improved to reduce burdens?

In the longer term, consideration should be given to wider use of e-assessment, particularly at the end of KS2. This would need to be subject to resolving any validity and reliability issues and any logistical and technical difficulties.

6. Improving end-of-key stage statutory teacher assessment

Q16. Do you agree that the statutory assessment of writing should afford teachers greater flexibility in determining a pupil's overall standard of attainment than is currently the case? Please give reasons for your answer.

Recommendations

- **We support the recommendation that the system of writing assessment should move to a best fit model, offering greater flexibility for teachers to use their professional judgement.**

- **A more robust and credible teacher assessment framework needs to be developed to support teachers' judgement-making in writing.**

- **Teachers should be fully supported in developing their skills in the assessment of writing (e.g. with guidance materials and continuing professional development training) so that teachers are confident in making judgements about aspects of writing composition and understand how to carry out effective internal moderation.**

The challenge in the assessment of writing is to devise a process that is manageable for teachers, fair for pupils and one that produces accurate and reliable data. The assessment process also needs to achieve a balance between the creative and technical aspects of writing.

We support the idea of the system of writing assessment offering greater flexibility for teachers to use their professional judgement but the system of assessment needs to provide a more robust and credible framework to support teachers' judgement-making in writing. The current approach to statutory assessment of writing is not the most purposeful or manageable way of determining a pupil's overall standard of attainment in writing for the following reasons:

- The *Interim teacher assessment framework for the end of KS2 – writing* (hereafter 'Interim Framework') states that teachers '*need to have evidence that a pupil demonstrates attainment of **all** of the statements within that standard **and all** the*

statements in the preceding standard(s)' (i.e. what has become known as a 'secure fit') (STA, 2016a). This means that teachers are currently unable to recognise or credit some positive elements of pupil performance in writing. For example, a piece of writing that shows evidence of seven or eight of the nine criteria for 'working at the expected standard' will be relegated to 'working towards the expected standard'.

- The criteria in the Interim Framework emphasise the technical elements of writing (grammatical structure, punctuation and spelling) above elements of composition and effect. This means that many of the creative and compositional elements of writing are undervalued by the framework. This emphasis seems contradictory, given that the Bew Review (Bew, 2011) advocated an approach to writing assessment that would enable pupils to demonstrate what they can do across a range of genres and that 'teachers will be encouraged to approach writing composition in a richer and broader way'. There is a danger that the rigid focus on a set of bullet points can lead to a 'box ticking' exercise. In addition, the compositional criteria seem to prioritize story writing over other text types, which is unexplained and undesirable.

- The Interim Framework does not offer a suitable model for the assessment of more able pupils (only three additional criteria) and so does not accommodate fully the needs of high ability writers at the end of KS2.

- Whereas the previously widely used Assessing Pupils' Progress (APP) guidance was arguably too detailed (QCA, 2010) – the current Interim Framework may not give sufficient detail for confident judgements to be made. This can lead to concerns from teachers about how to apply it correctly – and concerns about over-claiming. Although exemplar materials have been provided, these are very time consuming to read in detail (STA, 2016b). It might be more effective if some of the exemplification was embedded in the framework itself.

In NFER's recent survey, 62 per cent of primary classroom teachers and senior leaders indicated that the proposal to give teachers greater flexibility in determining a pupil's overall standard of attainment in writing (by moving from a 'secure fit' model to a 'best fit' approach) would result in a significant reduction in workload, and a further 31 per cent indicated it would result in some reduction in workload. When asked about resources to support the assessment of writing, 67 per cent reported that they would welcome a 'large bank of online exemplar materials' and 64 per cent said 'exemplification embedded within the frameworks' would be beneficial (NFER, 2017, unpublished survey).

References

Bew, L. (2011). *Independent Review of Key Stage 2 Testing, Assessment and Accountabililty. Final Report.* [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf [26 May, 2017].

Qualifications and Curriculum Development Agency (QCA). (2010). *Assessing Pupils' Progress (APP): Assessment Guidelines* [online]. Available: http://webarchive.nationalarchives.gov.uk/20110809101133/nsonline.org.uk/node/20683 [26 May, 2017].

Standards and Testing Agency (2016a). *2017 National Curriculum Assessments Key Stage 2. Interim Teacher Assessment Frameworks at the End of Key Stage 2* [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/538415/2017_interim_teacher_assessment_frameworks_at_the_end_of_key_stage_2_150716_PDFa.pdf [26 May, 2017].

Standards and Testing Agency (2016b). *2016 Teacher Assessment Exemplification: KS2 English Writing.* (Guidance) [online]. Available: https://www.gov.uk/government/publications/2016-teacher-assessment-exemplification-ks2-english-writing [26 May, 2017].

Q17. Please give details of any robust alternative approaches to the assessment of English writing, which the Department for Education should explore.

Recommendations

- **As noted in our response to Q16, a more robust and credible teacher assessment framework needs to be developed to support teachers' judgement-making in writing.**

- **We would recommend that a robust evaluation of the use of comparative judgement approaches is carried out, considering not only the technical and statistical evidence but also the implications of the use of such an approach as an ongoing tool for the assessment of writing.**

- **Consideration should be given as to how to develop and support teachers' analytical marking of writing in the classroom in order to provide formative feedback that will enable pupils to develop their writing skills.**

We recognise the need to ensure that 'the assessment of writing for national statutory purposes is as robust and useful to schools as possible.

Although external marking of statutory writing tasks was abandoned following the Bew review (Bew, 2011), some positive aspects of the marking of those writing tasks have been lost. Separating the assessment of writing into separate strands or categories, with banded mark schemes showing a clear progression in the criteria of each strand, recognised that pupils' skills in different areas of compositional writing may progress at different rates and facilitated best-fit judgements in different strands. This method of assessing writing could be utilised within a teacher assessment model.

An alternative approach to marking writing for statutory assessment purposes would be to use a comparative judgements system. In such an approach, rather than marking the writing, teachers compare the work of two pupils and decide which piece is better. Using multiple comparisons, pieces are ranked demonstrating the relative quality of pupils' writing.

Research into comparative judgement approaches has shown that such processes can be highly reliable, with reliability statistics that are often higher than marking approaches (Pollitt, 2012; Steedle and Ferrara, 2016). In one of the studies reported by Pollitt (2012) focussing on primary writing, comparative judgements were favoured by participating teachers as being quick, fairer to pupils and less onerous than using a complex mark scheme. However, in another study (science marking) comparative judgments were found to be more time consuming that traditional marking (McMahon and Jones, 2015).

Another key benefit of this type of approach is that teachers participating in online comparative judgement exercises compare not only the writing of their own pupils but the writing of pupils from many different schools. In this way they may see a greater range of writing ability than they might normally see within their school. This enables them to evaluate the overall standard of the writing of their pupils more accurately.

However, comparative judgement approaches to the assessment of writing facilitate judgements about the overall standard of a piece of writing but do not involve teachers in making judgements about different aspects of writing; writing is judged holistically rather than analytically (Pollitt, 2012). It is possible that although such approaches may be useful in deciding if a pupil has reached the expected KS2 standard, they may not support teachers in their ongoing assessment of writing in the classroom i.e. making judgements about specific aspects of a piece of writing and providing formative feedback to individual pupils. It has been suggested by Alistair Pollitt that comparative judgements might be used to judge individual components of a task but whether this could be implemented in a manageable and cost-effective way would need to be evaluated.

NFER's recent survey found that 22 per cent of the sample of primary classroom teachers and senior leaders had used a comparative judgement approach to assessing writing. Of these, over 90 per cent had found the approach 'useful' or 'very useful' for determining if a pupil had reached the expected standard in writing. 78 per cent had found a comparative judgement approach 'useful' or 'very useful' for providing formative feedback to pupils and 89 per cent had found it 'useful' or 'very useful' for developing their skills in the assessment of writing (NFER, 2017, unpublished survey). However, it should be noted that these ratings were based on a relatively small sub-set of the sample (140 respondents).

We would recommend that a robust evaluation of the use of comparative judgement approaches is carried out. In our view, it is crucial to look not only at the technical and statistical evidence but also consider the implications for teaching and learning of the use of a comparative judgement approach as an ongoing tool for the assessment of writing.

If a comparative judgement approach were to be adopted for the statutory assessment of writing, consideration would need to be given as to how many pieces of writing would need to be submitted and compared in order to reach a valid judgement of a pupil's writing ability. Although different forms of writing can be compared (Jones *et al.*, 2014) with high levels of reliability, each submitted task would only provide evidence of performance on one piece of writing. As one of the recommended reasons for a move away from statutory assessment writing tasks was to allow teachers to judge writing across a range of genres (Bew, 2011), limiting the number of tasks within a comparative judgment approach could disadvantage some pupils.

Consideration would also need to be given as to how to develop and support teachers' analytical marking of writing in the classroom in order to provide formative feedback that will enable pupils to develop their writing skills.

References

Bew, L. (2011). *Independent Review of Key Stage 2 Testing, Assessment and Accountabililty. Final Report.* [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf [26 May, 2017].

Jones, I., Swan, M., and Pollitt, A. (2014). 'Assessing mathematical problem solving using comparative judgement'. *International Journal of Science and Mathematics Education*, **13,** 1, 151–177 [online]. Available: https://link.springer.com/article/10.1007/s10763-013-9497-6 [26 May, 2017].

McMahon, S. and Jones, I. (2014). 'A comparative judgement approach to teacher assessment.' *Assessment in Education: Principles, Policy & Practice*. **22,** 3, 368–89. [online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0969594X.2014.978839 [26 May, 2017].

Pollitt, A. (2012). 'The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice.***19,** 3. [online]. Available: http://www.tandfonline.com/doi/full/10.1080/0969594X.2012.665354?scroll=top&needAccess=true [26 May, 2017].

Steedle J T. and Ferrara S. (2016). 'Evaluating comparative judgment as an approach to essay scoring.' *Applied Measurement in Education*. **29**, 3. [online]. Available: https://eric.ed.gov/?id=EJ1101445 [26 May, 2017].

Q18. Please give details of any effective models of moderation or standardisation of teacher assessment that the Department for Education should explore.

Recommendation

- **We recommend that the use of a comparative judgement approach should be evaluated as a tool to support the moderation of teacher assessment judgments.**

Models of standardisation and moderation need to avoid the problems of 'localism' i.e. if the group of schools is too small, there is the danger that local standards proliferate and do not have a meaningful relationship with an 'overall' national standard of judgement. In the future it may be possible for groups to meet and moderate judgements online to avoid geographical clusters / standards.

Good models of moderation / standardisation would allow teachers to be involved as a way of supporting professional development and strengthening inter-school networks of assessment expertise. The aim would be to support consistency in terms of how the

judgements relate to an overall standard and progression against the national curriculum for writing.

Comparative judgement approaches may be helpful in moderating teacher assessment judgments rather than replacing the current system of teacher assessment in its entirety. Teacher assessment judgments would be based, as now, on the assessment of a pupils' work across a number of different genres. Samples of writing from different schools could be submitted by participating teachers, using online comparative judgment software, allowing them to see the relative performance of their pupils and validate / moderate their teacher assessment judgments. The inclusion of writing from previous cohorts could be used to monitor standards over time (McMahon and Jones, 2015) both locally and nationally. As part of such an annual moderation exercise, a nationally representative sample of pupils' writing could be collected. From this exercise, pieces of writing exemplifying the expected standard could be disseminated regularly to schools, which in turn would support further professional development. Consideration should be given as to whether an online comparative judgement approach would be more cost-effective and efficient than training moderators and administrating face-to-face local/regional moderation.

References

McMahon, S. and Jones, I. (2014). 'A comparative judgement approach to teacher assessment.' *Assessment in Education: Principles, Policy & Practice*. **22,** 3, 368–89. [online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0969594X.2014.978839 [26 May, 2017].

Equalities

Q19. Do you think that any of our proposals could have a disproportionate impact, positive or negative, on specific students, in particular those with 'relevant protected characteristics' (including disability, gender, race and religion or belief)? Please provide evidence to support your response.

No response to this question

Q20. How could any adverse impact be reduced and are there any ways we could better advance equality of opportunity? Please provide evidence to support your response.

No response to this question

Date submitted: 21/06/2017

Public