

EDUCATION JOURNAL

Incorporating *Children's Services Weekly* and *Education*

Founded in 1996

Special issue No. 1

ISSN: 1364-4505

Wednesday 24 October, 2018

Reception Baseline Assessment

Over the last few years Margaret Clark has produced a series of original research articles challenging, with evidence, aspects of government education policy. Six of her articles on baseline assessment were published in the research section of *Education Journal* from 2015 to 2017. She reported on numerous researches questioning the validity and reliability of such assessment.

Yet it was announced on 11 April 2018 that: "The government plans to introduce a statutory reception baseline assessment in autumn 2020... Schools will administer this assessment soon after pupils enter reception. This assessment will last 20 minutes and teachers will record the results. We will use it as a baseline for measuring the progress primary schools make with their pupils".

The National Foundation for Educational Research (NFER) was announced as the preferred bidder and currently "trailing is taking place in a nationally representative selection of schools". On 4 July a report was published by the British Educational Research Association (BERA) entitled *A Baseline Without Basis: The validity and utility of the proposed reception baseline in England*. The report sets out the case against the government's proposal to use a baseline assessment of pupils in reception to hold schools in England to account for the progress their pupils have made by the end of key stage 2.

As Margaret Clark's six articles present further important evidence from the years 2015 to 2017 to which she had drawn the government's attention we decided to re-publish these in this special issue.

"The language of young children on entry to school as measured by baseline assessments. Why ignore the evidence from research?" Pages 2 to 6. From *Education Journal*, issue number 244, 5 October 2015.

"The views of teachers on the proposed baseline assessments on entry to reception class: their purpose and the dangers." Pages 7 to 9. From *Education Journal*, issue number 259, 16 February 2016.

"Baseline Assessment: what research is telling us. A report from an invitation research seminar held in Newman University on 24 February 2016" Pages 10 to 12. From *Education Journal*, issue number 261, 1 March 2016.

"Baseline assessments and young children's readiness on entry to primary school: where now?" Pages 13 to 16. From *Education Journal*, issue number 280, 27 September 2016.

"Synthetic Phonics and Baseline Assessment under the Searchlight in 2017: are they value for money in a time of cuts?" Pages 17 to 20. From *Education Journal*, issue number 299, 31 March 2017.

"Primary Assessment in England: Government consultation. Part II Baseline Assessment: where is the evidence?" Pages 21 to 23. From *Education Journal*, issue number 308, 20 June 2017.

The language of young children on entry to school as measured by baseline assessments. Why ignore the evidence from research?

Professor Margaret M Clark OBE

From 2016, the Department for Education will require all children in reception class in England to be tested on a baseline assessment, conducted in English, and within six weeks of starting school. The assessment must use one of the three commercial baseline assessments identified by DfE that remain as from June 2015, Early Excellence, CEM and NFER; these are currently being tried out in schools. CEM is a 15-20 minute programme; NFER focuses on maths, literacy and communication and language; Early Excellence is based on observation, rather than tests. A single score based on the baseline assessment chosen will be used to calculate how much progress the child has made by the end of primary school when compared with others with the same starting point, and to hold schools accountable. Many professionals have recorded their opposition to this policy.

I am currently undertaking research in the West Midlands on baseline assessment. We will interview head teachers and reception class teachers, analyze the results of this year's scores on baseline assessment measures and test selected children individually on the Preschool Language Assessment Instrument devised by Marion Blank. PLAI2 yields not only a score for the young children's understanding of questions of different levels of complexity, but also diagnostic information from their incorrect responses.

This article, the first of a series on baseline assessment, provides a brief outline of selected research findings, which although they took place in the 1970s and 1980s should provide insights for contemporary policy on the assessment of young children's language, in particular the effect of the context and the adult on young children's score. Yet such findings are being ignored. In *Understanding Research in Early Education* (Clark, 2005) there are more details and a critique of these researches and others. Here relevant findings from four researches, those by Tough, Wells, Tizard and Hughes and the present writer are summarized. Other aspects in assessment of young children by tests that should provoke disquiet are identified. In each study, not only the sample, but any groups excluded needs to be considered; this is indicated here. For example, none of the three studies by Tough, Wells or Tizard and Hughes had any children whose home language was other than English.

The Development of Meaning: Joan Tough

Tough's study is one of those referred to in *A Language for Life* (DES, 1975: 53-54). To quote: "there is a range of uses which children from 'educating' homes seem to have developed more extensively than children without these home advantages"; it is stressed that such abilities are important for learning in the school situation.

Tough's study was a longitudinal investigation of 64 children, samples of whose language in 'contrived' situations were tape recorded and analyzed at the ages of 3, 5+ and 7+ years. There were two groups, referred to as the 'advantaged' and 'disadvantaged' groups, within each of which was a group who attended nursery school or class and a group not attending. Excluded from the sample were: children whose IQ was below 105; from a family of more than six children; who showed evidence of known or suspected rejection or emotional stress; whose mother did not speak English as her first language; who were West Indian, because of problems with use of non-standard English; were shy, withdrawn or hostile to the observer; in the nursery if they were not happily settled in school; did not speak clearly enough for transcriptions to be made of what they said. Consider both the relevance of these findings, and the extent to which these omissions limit their generalizability.

At 5 years and 7 years of age there were striking differences in the groups in the complexity with which they expressed interpretations of a picture. The disadvantaged children tended to give short responses which treated the picture as a series of objects, and thus it is argued that, “the children were orientated to examine the situations differently” and so used “different strategies of language”. (Tough 1977: 103). Even in such a highly selected group of ‘disadvantaged’ children, response to such commonly used materials as pictures, unless supported and encouraged, may lead only to labelling of unrelated objects. Tough makes crucial points for understanding the language knowledge and use of young children (based on the evidence from the children at 7+ years of age). It should be noted that when retelling a story the disadvantaged children had a mean length of utterance almost twice as high as in other situations. They showed they were able to remember sequences and reproduce the story line with much of the detail. Note: the story was repeated to a companion in a meaningful context.

Tough stresses that all the children produced long, complex utterances at times. When retested at 7 years of age the disadvantaged group were inclined, for some types of items, to respond with ‘I don’t know’, or otherwise avoid answering the question. Thus tests may be an underestimate of their knowledge and understanding. However, frequently when pressed further the disadvantaged children ‘moved towards the answer given spontaneously by the children in the advantaged groups’ (Tough 1977: 170). Clearly in many situations in school, children are likely to be judged by the more limited responses and this may in turn lead to lowered expectations.

In her research there were children who were attending a preschool setting and even at three years of age there were already some statistically significant differences between the nursery and non-nursery disadvantaged groups. This could have important implications for baseline assessment in reception class where children who have had experience of preschool education, particularly in that school, might reveal to the teacher, even without probing, greater apparent competence in language situations in the classroom. The situation, and whether there were probes, influenced the children’s language differentially. However, there was considerable overlap between the groups. The experimental work of Donaldson was only beginning to appear when Tough’s researches were published. Donaldson was able to show that, even in experiments such as those on which Piagetian evidence was based, the precise context will influence the child’s apparent competence (see *Children’s Minds*, Donaldson 1978). Tough’s research reveals the danger of making generalizations about habitual use of language from speech samples in one limited situation.

Studies of Language at Home and at School: Gordon Wells

Wells in his critique of Joan Tough’s research, commented that what seems important for sustaining dialogue is, ‘the presence or absence of genuine reciprocity and collaboration’ The studies by Wells, and related studies by co-workers, were extensively funded between 1972 and 1984 (See Clark 2005 chapter 3). Wells’ study began as an investigation of language development and it was also hoped to describe the children’s conversational experience and to investigate the relationship between preschool experience and success in school. The sample excluded: children of multiple births; with known handicaps; whose parents were not native speakers of English; who were in institutions or full-time day care; likely to move soon; who had siblings already in the study.

The sample was 128 children from a larger randomly drawn sample in Bristol of children representative in sex, month of birth and family background. A sample of 32 children from the younger group was followed into school and over their first 2 years in primary school. A further assessment was made when the children were aged 10 years 3 months.

Wells found that although the ‘route’ of the children’s language development was similar the quality and quantity of the conversational experience was the best predictor of the child’s oral language at entry to primary school. He states that one of the most important features found in the homes of children whose success could have been predicted early was ‘the sharing of stories’. This he suggests, in relation to literacy development, is more important than any early introduction to features of print. He claims that “stories have a role in education that goes far beyond their contribution to the acquisition of literacy” (Wells 1986: 194). Tough had also found that retelling stories in a meaningful context was a rich source of language interaction in young children.

Language at Home and at Preschool: Barbara Tizard and Martin Hughes

This study was funded by SSRC in the later years of the 1970s. The evidence cited here is from articles in academic journals. *Young Children Learning* (Tizard and Hughes 1984, second edition 2002), does not give sufficient detail to allow a critical evaluation of the research (See Clark, 2005, chapter 3).

The research by Tizard and Hughes is based on recordings with a radio microphone of naturally occurring dialogue of girls with staff at preschool in the morning and at home with their mothers in the afternoon. The study was of 30 girls of about 4 years of age. Half the girls were working-class and half were middle-class. The other selection criteria were that the children should attend nursery school or class in the morning, and spend the afternoon at home with their mothers. The research involved recording, in the presence of an observer at home in the afternoons as well as recording in the nursery school. Children's talk with other children at school was recorded but not analyzed.

Excluded from the sample were: boys; children not attending part-time preschool education; children not at home with their mothers in the afternoon (thus any whose mothers worked full-time); children whose father was at home in the afternoon; children who came from families with more than three children; children in homes where the main language spoken was not English.

The children were to quote Tizard et al, probably typical of the majority of working-class children who attend half-day nursery school, and who are nevertheless seen by their teachers as in need of language enrichment (Tizard et al. 1980: 52).

Tizard and Hughes, in their much quoted book, seldom make reference to the fact that their sample was all girls, using the word children in the title *Young Children Learning*, and on almost every occasion in the book (Tizard and Hughes 1984). Some people who have read about the research only in that book have admitted to failing to appreciate that the sample was of girls only; yet there might have been rather different findings for young boys. Tizard and Hughes indicate how similar to the findings of Wells theirs are in showing extended and complex conversations in the homes. Taken together these studies gave new insights into the contribution of the home to the language development of young children. Both may undervalue the contribution of fathers to their children's language development.

No social class differences were found in the number and length of adult-child conversations; but there were fewer at school than at home. The majority of school conversations concerned play activity whereas at home a number were on a range of topics, including past and future events. Much the longest conversations both at home and at school concerned books that the adult was reading aloud or had just read aloud and/or when engaged in a joint activity. Children asked many questions at home, few at school, with no social class difference. The number was highest in relation to books and past and future events. An important observation was that a child who talked a lot or asked a lot of questions at home, or who tended to initiate conversation, was not necessarily likely to do so at school.

All the above points are taken from one of the articles by Tizard et al. (1980: 55–68). The findings of Tizard and Hughes raise the possibility that the level of language competence of 'working-class' children may well be underestimated in formal test situations. This may result in adults providing less-challenging dialogue with some children than they have potential to sustain, and may indeed show in more naturalistic settings.

The assumption in the early 1970s of deprivation of language and deficiencies in 'working class' homes and the incompetence of parents, are views still apparent in many current discussions on the role of parents in their children's language development.

Assessment of young children's language on entry to primary school: Clark, Barr and Dewhirst

The discussion will focus only on the part of the research concerned with assessment of the young children's language shortly after entry to school. For details of the other aspects of the study see *Understanding Research in Early Education* (Clark, 2005, chapters 8 and 9).

The research investigated similarities and differences in the classroom environments within which children spent their first year in primary school, the range of competence of the children and the extent to which this appeared to vary in different contexts and over time. The research was undertaken in five primary schools in The West Midlands with varied proportions of children from different ethnic backgrounds, for some of whom English was not their mother tongue. The children who were studied were the 247 who entered reception class during 1982/3. Most children entered reception class having attended

some form of preschool unit, though few attended playgroup or day nursery. We had observed some of the children and recorded their language in the preschool unit before the funded research commenced.

The children's language was assessed in a variety of settings, by test, observation and recording, with peers and with their teachers. All the children in the reception classes were tested individually in English, and children whose mother tongue was Punjabi were also tested in their mother tongue. The assessment used was the Preschool Language Assessment Instrument (PLAI) (Blank, 1985). We had used this test with the children with special needs and their controls in an earlier study. A revised version is now available PLAI2 which we plan to use in our new study.

PLAI was devised specifically for children about this age, at the point of entry to school, and consists of 15 questions at each of four levels of complexity, chosen to reflect the type of questions faced in the classroom. The levels involve: matching perception; selective analysis of perceptions; reordering perception; reasoning about perception. Most questions require only pointing or a few words for an adequate response, even for higher levels of questions. Valuable diagnostic information can be obtained not only from a study of the children's scores for each level of questions, but also from an analysis of their errors.

A number of the low-scoring children on the test were among the youngest in the sample. It is important to stress that in each of these schools there were some children entering reception class who had impressive abilities and who understood and responded to questions of high levels of complexity and perceptual distance. There were other children in each school who were able only to respond appropriately to simple questions tied closely to perception. Furthermore, even when assessed in English, there were within each ethnic group children who were able to answer appropriately questions on all four levels of difficulty. Likewise, there were children from each ethnic background with very limited understanding of anything beyond simple labelling questions. There were some young children who had impressive competence in both English and Punjabi. This we were able to show not only on test performance, but also in the dialogue between peers that we recorded.

Teachers' judgements of the children's language were also measured by questions paralleling the levels of complexity on PLAI. The reception class teachers were asked to make a judgement on which of their children they would expect to answer each question successfully. Teacher-child dialogue was also recorded using radio microphones. Some teachers were concerned when they realized the style they had adopted and its effects. Some discovered that had they adopted a different strategy and a more conversational style they could have developed a more interesting dialogue with a number of the children. The fact that children's language was assessed and recorded in a variety of settings, and over time, made it possible to appreciate the influence of a number of these variables, and not least how dangerous it is to assess a child's communicative competence from only one sample of language, or even one type of setting.

Group discussions between young children were recorded with and without an adult present. This aspect included 44 of the 215 children who had been assessed on PLAI, and as many children as possible for whom samples of language in other settings were available. The competence of these young children to engage in dialogue with peers when provided with sufficiently challenging and stimulating materials was revealed, children around five years of age (See Clark 2005 chapter 8).

Conclusions

The findings cited here make the proposal to use a single baseline assessment, limited contexts and a one score the basis for judging the competence of young children on entry to primary school a disturbing development. The following are important:

- The context in which any assessment is conducted.
- The adult conducting the assessment, their training and how familiar the child is with them.
- The child's home language and how competent they are in understanding and responding in the language of assessment, English being the language proposed for baseline assessments.
- The danger in using a single score as a measure of the child's competence.
- The problems in comparing the scores on different baseline assessments.
- The precise age of the child, as this will influence their attainment differentially.
- Whether or not the child has attended nursery class prior to entry to reception class as this will influence their score, since they may be more or less familiar in interacting with strange adults.

In our research on baseline assessments we plan in addition to interviewing teachers on their experience with baseline assessment, to analyze the scores on baseline assessment by sex, by age, by whether the children attended a nursery class in that school and whether or not their home language is English. Assessment of target children on PLA2 will add an additional dimension to the study.

- Anyone interested in further information contact margaret.clark@newman.ac.uk.

References

Blank, M. (1985) 'Classroom discourse: the neglected topic of the topic', in M.M. Clark (ed.) *Helping Communication in Early Education*, Educational Review Occasional.

Clark, M. M. (2005) *Understanding Research in Early Education: the relevance for the future of lessons from the past*. Second edition. Abingdon, Oxon: Routledge.

DES (1975) *A Language for Life*, The Bullock Report. London: HMSO.

Donaldson, M. (1978) *Children's Minds*, Glasgow: Fontana/Collins.

Tizard, B., Carmichael, H., Hughes, M. and Pinkerton, G. (1980) 'Four year olds talking to mothers and teachers', in L.A. Hersov and M. Berger (eds) *Language and Language Disorders in Childhood*, Oxford: Pergamon Press.

Tizard, B. and Hughes, M. (1984) *Young Children Learning: talking and thinking at home and at school*, London: Fontana (second edition with foreword by Judy Dunn was published in 2002, Oxford: Blackwell).

Tough, J. (1977) *The Development of Meaning: a study of children's use of language*, London: Allen Unwin.

Wells, G. (1986) *The Meaning Makers: children learning through language and using language to learn*, New Hampshire: Heinemann.

Originally published in *Education Journal*, issue number 244, 5 October 2015.

The views of teachers on the proposed baseline assessments on entry to reception class: their purpose and the dangers

By Professor Margaret M Clark OBE
Newman University

From 2016, the Department for Education will expect all children in England aged 4 to 6 years of age to be tested on a baseline assessment within their first few weeks in reception class. A single score based on this assessment carried out in English, uncorrected for age, is to be returned to the Department for Education. While this will not be mandatory most schools feel under pressure to conform as this will be used as a measure of accountability and children's progress at the end of Key Stage 2.

The assessment is expected to use one of three commercial measures approved by DfE (from an original choice of six). These are Early Excellence, observation based and two test based measures, NFER and CEM (Centre for Evaluation and Monitoring from the University of Durham). Most schools have used one of these in 2015; just 2,000 of England's 17,000 primary schools did not carry out the baseline, when it was introduced last September, according to an article in *Nursery World* online (12.2.16). Most schools selected Early Excellence due to its similarity to the EYFS and the promotion of the scheme as 'early years friendly', with some schools reporting that they were under pressure to do so from their local authority, or because other providers were removed, according to the new research report. According to the article in *Nursery World* about the new research report, "the union leaders are not anti the baseline provider but against baseline, whatever the provider, and wanted schools to use the EYFS Profile instead".

This is the third of a series of articles on baseline assessment in *Education Journal*, focusing on the current proposals by DfE. In an article in No. 244 in 2015 I questioned why the DfE ignored evidence from research on the dangers and unreliability in assessing the language of young children's on a single occasion, in a particular context, and by a strange adult. In a second article with the title *Baseline Assessment: who made the cuts and why?* (*Education Journal* No. 258, 2016:16-18) Louise Wormwell analysed the evidence of how the three remaining commercial measures came to be selected by DfE. Evidence is building showing that teachers are concerned at many aspects of these assessments, not least the plan to use a single figure based on one or other of these tests as a measure of schools' 'accountability'.

Researches on the views of teachers

There have now been two major studies published investigating the views of teachers on these plans. The first research report entitled: *Reception baseline research: views of teachers, school leaders, parents and carers* by NFER and funded by DfE was published in 2015 (Lynch, S. Bamforth, H. and Sims, D.). A further report funded by NUT and ATL was published on 12 February 2016 with considerable publicity, including news programmes on television and radio featuring interviews with a number of those involved and some parents on the issue of baseline assessment. However, only a few of the concerns that such an initiative should arouse were raised in the programmes. On the same day there were several reports online carrying details of the research, in *Nursery World*, *Huffington Post* and on the *BBC Educationonline* (<http://www.bbc.co.uk/news/education-35549611>). Key issues from the research evidence are clearly articulated in the articles. The union leaders express concern at the proposed imposition of new



Prof. Margaret M Clark OBE

commercial assessments and at the way the data will be used by the government. The article in BBC Education entitled, *Tests for four-year-olds 'unreliable and disruptive'*, was followed by 235 comments before this aspect was closed. Disappointingly few of those who commented appeared to have read, or at least appreciated the issues raised in the article. Most make a brief statement either pro the Government, against the Government or the unions, or state the need for assessment, appearing to believe that the research is against all assessment of young children's.

The NUT/ATL funded research into the views of teachers on baseline assessment

In this article the issues raised in this new report will be discussed. The report is entitled *They are children's... not robots, not machines*. The introduction of Reception Baseline Assessment (Bradbury, A and Roberts-Holmes, G.). Bradbury and Roberts-Holmes cite the earlier NFER report as finding that there was:

- Lack of understanding of the assessment.
- Existing on-entry assessments were used widely, often through observation; 'some evidence' of gaming results, in order to maximise progress measures.
- A desire to report results to parents orally. (Bradley and Roberts-Holmes: 8)

In their new research the questions they sought to answer were:

1. What impact does the reception Baseline Assessment have on the start of school for the children's concerned.
2. To what extent does the Baseline Assessment support or undermine existing transition and assessment strategies used in schools (especially the Early Years Foundation Stage Profile).
3. How do parents and teachers perceive and experience the assessment?
4. How can this assessment be adapted to the needs of different groups of pupils, such as children's with SEND and EAL, looked after children's, children's from BME groups and from disadvantaged backgrounds; Do teachers and school leaders believe it to be a valid assessment and how do they intend to use the resultant data? (Bradley and Roberts-Holmes: 10)

The research consisted of an online survey completed by 1,131 people, 50% of whom were Reception class teachers and 38% were EYFS or Phase Leaders, 7% Senior Leaders and the remainder support staff or other. This was followed by a case study in five primary schools across England. This involved 35 people interviewed including five headteachers, two assistant headteachers or EYFS coordinators, 13 Reception teachers and 15 parents. The researchers were unable to get access to an Academy school and all of the schools in the case study sample had chosen Early Excellence, reflecting the trend across the country. However, the other two assessments NFER and CEM were used in some of the schools in the larger online survey. The researchers admit that their sample of parents was limited not only in number but also in terms of gender and social diversity.

Executive Summary

The following are the most significant points made in the Executive Summary (see page 5 of the report):

1. Teachers and school leaders have serious doubts as to the accuracy of the assessment and its use in measuring progress, in relation to all three baseline providers.
2. Many teachers and school leaders doubt the use of measuring progress from Reception to Year 6.
3. The majority of schools already used informal on-entry assessments to plan teaching and to identify children's with particular needs; the Baseline Assessment is not seen as an improvement on these methods.
4. Most schools selected Early Excellence due to its similarity to the existing EYFS and the promotion of this scheme as 'early years friendly'.
5. Baseline was seen "as encouraging the practice of 'stopping teaching' and was not seen as helping teachers to know pupils better".
6. It has little use in the identification of additional needs.
7. There is a significant effect on teachers' workloads, and that was true with all three providers.
8. Few schools plan to provide information on Baseline Assessment scores to parents due to their concerns.

9. School leaders are uncomfortable with the use of private providers and the related marketing that they have received. (Bradley and Roberts-Holmes: 5)

Conclusions

The full report is available online on the webpages of both NUT (teachers.org.uk) and ATL (atl.org.uk). This report adds to the growing evidence of the concerns of teachers at this additional form of assessment, the purposes for which it will be used and the fact that it is carried out within the first few weeks of a child's entry to school. Teachers feel it is drawing them away from their important role of making sure that young children's have a stress-free entry into primary school and that it does not add to information they were already gathering on children's entry level. So far there has not been research evidence on the pattern of scores from these new baseline assessments. In a further article I will analyse scores from a sample of schools, based on this year's assessments, considering the results by sex, date of birth, language spoken by the children's and whether or not the child had attended a nursery class in that school. These are all important but ignored variables that could impact on how well a child would perform on such a one off assessment, thus on the reliability of predictions of progress based on such measures. A further article will report on a research seminar entitled Baseline Assessment: what research is telling us, which we are hosting at Newman University on 24 February. There a number of researchers will make presentations summarising their findings and the issues these raise will be discussed.

References

Bradbury, A. and Roberts-Holmes, G. (2016) *"They are children's... not robots, not machines" The introduction of Reception Baseline Assessment*. Available online at NUT (teachers.org.uk) and ATL (atl.org.uk).

Clark, M. (2015) *The language of young children's on entry to school as measured by baseline assessments: why ignore the evidence from research?* in *Education Journal* No. 244: 12-16.

Lynch, S., Bamforth, H. and Sims, D. (2015) *Research baseline research: views of teachers, school leaders, parents and carers*. Research Report. DfE online.

Wormwell, L. (2016) *Baseline Assessment: who made the final cuts and why?* in *Education Journal* No. 258: 16-18.

Originally published in *Education Journal*, issue number 259, 16 February 2016.

Baseline Assessment: what research is telling us

A report from an invitation research seminar held in Newman University on 24 February 2016

By Professor Margaret M Clark OBE
Newman University

This is the fourth in a series of articles on baseline assessment; the three previous articles were published in *Education Journal* issues 244, 258 and 259. The articles in issues 258 and 259 give further information on the papers at the seminar by Louise Wormwell and Guy Roberts-Holmes. From 2016 the Government has indicated that it expects all children in reception class in England to be assessed within a few weeks of starting school on one of three baseline assessments, and in English, the scores to be used for accountability and to assess children's progress by the end of Key Stage 2.

This policy has aroused concern among many professionals. On 24 February an invitation research seminar, *Baseline Assessment: what research is telling us*, which I convened jointly with Professor Terry Wrigley, Visiting Professor, Northumbria University, was held at Newman University. In this article I summarise the eight papers presented at the seminar and highlight the issues raised during the discussion. The summaries and discussion will be made available online shortly. A flavour of the range of research presented can be gained from the brief quotes from the summaries:

1. *Why it is better without baseline assessment. 4 core reasons.* Nancy Stewart, Principal Consultant, Early Learning Consultancy stated that many education experts and teaching unions are strongly opposed to the introduction of these standardised on-entry assessments on a number of well-evidenced grounds, including that:

- a)** Many children are already being wrongly labelled as achieving below typical standards, with harmful effects;
- b)** The assessments disrupt children's introduction into school;
- c)** The narrow focus on attainment in prescribed subject areas is harmful to children's learning and development in the early years;
- d)** The planned system will not provide a useful indicator of school quality.

See <http://www.betterwithoutbaseline.org.uk> for further information.

2. *Baseline assessments: who made the cuts and why?* Louise Wormwell, Newman University quoted The Standards and Testing Agency, May 2014 as claiming: "The purpose of the reception baseline is to support the accountability framework and assess school effectiveness by providing a score for each child at the start of reception....as the basis for an accountability measure of relative progress of a cohort of children through primary school." She also drew attention to a significant disturbing point in the guidance in 2014 under the sub-heading of 'Minimising bias', "A particular focus should be given to pupils with English as an additional language (EAL) to ensure that the progress measure is not unduly beneficial to schools with high proportions of EAL children." See *Education Journal* Issue 258 for her article with the same title.

3. *Baseline assessments: the delusions of 'predictive validity'.* Terry Wrigley, Visiting Professor, Northumbria University took as his focus "the claims and assumptions that baseline tests can form a reliable starting point for evaluating the value-added by the school, and the dangers posed to children by assumptions that they can accurately assess 'ability' or 'potential'. The presentation was based on a study of documents, interviews and data secured as the result of a Freedom of Information request. He stressed that his criticisms are not of the organisations per se, but of the whole policy "which is deeply flawed". He warned that many children are "particularly at risk of receiving teaching which is premised on the assumption that they have limited potential".

4. *Which abilities of 4 year olds predict later academic achievement? Developmental evidence and implications for early assessment.* David Whitebread, University of Cambridge argued that if the purpose of

"This policy has aroused concern among many professionals."

making baseline assessment of children's capabilities is "ultimately to improve the quality of their education" then we should bear in mind that "the BA models currently on offer are likely to be inaccurate, unreliable and potentially harmful, as they are in danger of negatively impacting on the crucial first few weeks of children's experience of their primary school". Whitebread, D., Pino-Pasternak, D. and Coltman, P. (2015) *Making learning visible: the role of language in the development of metacognition and self-regulation in young children*. In S. Robson and S. Quinn (eds) *The Routledge International Handbook of Young children's Understanding*. London: Routledge: 199-214.

5. *Reception Baseline Research: views of teachers, school leaders, parents and carers*. Sarah Lynch, National Foundation for Educational Research reported the findings of an independent research commissioned by DfE between October and December 2014. The aim of the research was to inform the implementation of the reception baseline and identify effective ways of communicating the results to parents/carers. The research was an online survey plus telephone interviews and focus groups of parents. The majority of respondents were positive to some extent about the introduction of baseline assessment. However, all the schools surveyed were already using some form of on-entry assessment and most schools were already using the outcomes of existing on-entry assessments as evidence for accountability. The report is available at <https://www.nfer.ac.uk/publications/DFER01home.cfm>.

6. *The Introduction of Baseline Assessment: views of head teachers, teachers and parents*. Guy Roberts-Holmes, UCL, IoE, London reported the main findings of the NUT/ATL sponsored research carried out in the autumn of 2015 involving a nationwide survey and five case studies in primary schools. They found that, "For many teachers, baseline assessment has had a negative impact on their working lives without benefiting the children they teach". Teachers and headteachers see all three baseline assessment providers as inaccurate, unreliable and lacking in validity and hence its ability to accurately measure the 'value added' by schools. The report was published online on 12 February 2016 on ATL's website atl.org.uk and NUT's website: www.teachers.org.uk/baseline.

7. *Reception teachers' and heads of schools' perspectives and experiences of an observation based assessment that focuses on child-initiated activity*, Sally Howe and Michelle Cottle, University of Roehampton. The research, funded by the Froebel Trust, between October 2014 and March 2015, studied the experiences of a group of teachers and head teachers as they piloted Early Excellence. "Some teachers talked about having to organise specific or more structured activities in order to complete the baseline assessment within the time constraints dictated by government policy". They felt that it may not be valid for children with special needs and children with English as an additional language and "raised questions about the validity of relying on assessments that focus mainly on literacy and numeracy on entry to school".

Brogaard Clausen, S., Guimaraes, S., Howe, S. and Cottle, M. (2015) *Assessment of young children on entry to school: informative, formative or performative?* in the *Journal for Cross-Disciplinary Subjects in Education*. Vol.6, Issue 1.

8. *Baseline Assessments: their value and validity in assessing young children on entry to school*, Margaret M. Clark, Newman University. This is a small ongoing study in five schools. Seven teachers involved in baseline assessment in 2015 were interviewed. What is unique is that we have details not only of the children's scores on the baseline assessment but also their sex, date of birth, whether they attended the school's nursery class and if so if that teacher assessed them, and if English is not their mother tongue, which language they do speak. In three schools with a total of 117 reception children assessed there are 52 children who speak at least one other language; 16 different languages in addition to English are spoken by children in these four reception classes. We plan to analyse the results for 2015, extend the study to

"For many teachers, baseline assessment has had a negative impact on their working lives without benefiting the children they teach". Teachers and headteachers see all three baseline assessment providers as inaccurate, unreliable and lacking in validity and hence its ability to accurately measure the 'value added' by schools."

include further schools and make an assessment of selected children on a diagnostic language test.

Key issues raised in the discussion

There were 60 participants at the seminar, staff from universities, teachers and advisers and representatives from the teachers' unions. No one challenged the strength of the research evidence against the proposed baseline assessment policy, stressing that this was not a criticism of specific providers but of the policy itself. The major focus in the discussion was on how to alert parents to the precise nature of this policy and its damaging effects on children by its deficit model of prediction and by labelling children; its effect on the practice in schools in the children's early weeks in reception class and the strain it was placing on teachers to implement a policy in which they did not believe and which they felt added nothing to the knowledge they had or could get from other sources. I noted two researches cited in *Education Journal* the previous day (Issue 260). Save the Children stated that "the research demonstrated that the most crucial determinant of success in Sats tests was how well children could communicate when they started school" (p. 6). A new paper by UNESCO's Global Education Monitoring Report revealed "that 40% do not access education in a language they understand" and, to quote from page 12: "A review of 40 countries' education plans found that less than half recognised the importance of teaching children in their home language, particularly in early grades."

In view of the data we are collecting in our ongoing research on the variety of languages spoken by young children in many reception classes in England, it would appear that DfE pays insufficient attention to researches such as these!

Footnote

On 25 February in <http://schoolsweek.co.uk> in an article "Baseline assessments could be scrapped over comparability concerns" it is claimed that following a study commissioned by DfE it has been found that "tests from the three chosen providers for baseline assessments of reception children cannot easily be compared – putting the policy's future in doubt". "Government officials would not confirm the content of the study".... To quote: "A mooted alternative to baseline test is the introduction of 'school readiness' checks – an option said to be preferred by No 10. Such checks assess children's ability to hold pencils or read basic words and have been widely debated in the US.

Should such an alternative be proposed not only is further research needed with a shift in focus, but reference back to existing research over many years showing the unreliability of such tests of so-called 'school readiness'.

First published in *Education Journal*, issue number 261, 1 March 2016.

Baseline assessments and young children's readiness on entry to primary school: where now?

Margaret M Clark OBE
Newman University

In 2015 and 2016 I published a series of articles in the *Education Journal* critiquing the Government's plans for baseline assessments of all children on entry to reception class in England due to become mandatory in 2016. In the first article I considered why the evidence from research was being ignored (Issue 244); the second article reported the views of teachers on the proposals (Issue 259); the third article summarised the papers from a research seminar at Newman University on 24 February 2016 (Issue 261). In Issue 258 Louise Wormwell described the procedure by which the Government had decided on the three baseline measures that would be acceptable.

With a team of colleagues at Newman University I am undertaking research in schools in the West Midlands on baseline assessment and we submitted a symposium on our research for the BERA conference on 15 September 2016. Shortly after our symposium was accepted, the Government made a dramatic change in its plans and for that reason we changed the focus of our research. In this and two following articles we will report on our presentation at BERA and in doing so trace the current situation in England with regard to assessment on entry to reception class and also raise our concerns for the future. In this first article I will outline the changes made so far; in the second article, Parminder Assi and Sue Reid will discuss the findings of a questionnaire we sent to reception class teachers to find out their views on readiness for entry to reception class. While the 100 teachers who responded were guaranteed anonymity we invited anyone willing to be interviewed to give us a contact email address. The questionnaires have been analysed and currently we are in process of completing the interviews. We are already able to highlight several issues that arose from these two sources. The third article by Chris Watts will look in more depth at the concept of EAL which is of particular importance in view of the government's intention that the baseline assessments be only in English.

Changing Government Plans

1. DfE announced its intention from 2016 to require all children in reception class in England to be tested on a baseline assessment, conducted in English, and within six weeks of starting school. The assessment was to use one of three commercial baseline assessments identified by DfE that remained in June 2015 from a larger number, namely Early Excellence, CEM and NFER. It was proposed that a single score from one of these measures should be used to calculate how much progress the child made by the end of primary school when compared with others with the same starting point, and, to hold schools accountable. Many professionals recorded their opposition to this policy. On 24 February 2016 at an invitation research seminar held at Newman University eight research papers were presented considering the available evidence on baseline assessments as proposed by the government (see www.newman.ac.uk/24feb).
2. On 3 March 2016 in a letter signed by Eric Lui at DfE to Christine Blower of the NUT it was announced that this policy would not be implemented in 2016 as it had become clear that the scores from these three different measures could not be compared; something that should have been clear from the onset. However, it is important to note that a commitment to some form of readiness measure and accountability is still government policy. This change was following the publication of research questioning the comparability of the three different reception baseline assessments used in the 2015 to 2016 academic year. (<http://gov.uk/government/news/reception-baseline-comparability-study-published>). I was surprised to find that schools could still use these as on-entry assessments in 2016 to 2017 and on 14 April (on Gov.UK) schools were not only encouraged to sign up with approved providers, but DfE would cover the basic cost of approved baselines.

3. With no explanation for its inclusion, now not only are the three previous providers listed, but a fourth has been added to the government website, GL Assessment (<http://www.gl-assessment.co.uk/products/baseline-reception-baseline-assessment>). On their webpage GL Assessment reports not only that they are a recognised provider, but that 'Baseline and now brand new Baseline Progress enables teachers to assess children on entry to school (Reception/P1/Y1) and measure their progress in literacy and mathematics at the end of the same year'. One must question why this addition of a previously rejected assessment, and whether this provider is also targeting Scotland, with its reference to P1. New plans are being outlined in Scotland for national testing of children in Primary 1 (there are no reception classes in Scotland). It is suggested that at P1 a short assessment will focus on a range of early reading, writing and numeracy skills. (see gov.scot/Resource/0048/00484452).

4. In an official notice to primary school heads by the Standards and Testing Agency on 9 August, entitled 'The early years foundation stage profile will remain statutory for the 2016 to 2017 academic year', yet another change has been announced. It is stated that: "to provide continuity and stability for schools, pupils and parents and avoid unnecessary change whilst we take time to review options for assessment in the reception year beyond 2016 to 2017", the early years foundation stage profile (EYFSP), a series of teacher assessments will be retained. Thus it seems unlikely that many schools will continue to use any of the four providers on the DfE website; however, funding appears still to be available for schools who choose to use these measures.

Our Research on Baseline Assessments

Following the Government's initial announcement in September 2015 we started a small research in five primary schools in the West Midlands, with a focus on the views of their reception class teachers on the DfE proposals, an analysis of the results of the 2015 baseline assessments and an assessment of selected individual children on a more appropriate diagnostic language assessment PLA12. We submitted our symposium proposal to BERA on this basis. We completed the first stages of this research with interviews of seven reception class teachers, and a detailed analysis of the results in three schools. I then with Terry Wrigley convened a research seminar on 24 February 2016 at Newman University. No one at the seminar challenged the strength of the research evidence against the proposed baseline assessment policy, stressing that this was not a criticism of specific providers but of the policy itself. The major focus in the discussion was on how to alert parents to the precise nature of this policy and its damaging effects on children by its deficit model of prediction and by labelling children; its effect on the practice in schools in the children's early weeks in reception class and the strain it was placing on teachers to implement a policy in which they did not believe and which they felt added nothing to the knowledge they had or could get from other sources. We reported our findings so far on the views of the reception teachers and evidence from our analysis from four reception classes in three primary schools.

We have details not only of the children's scores on the baseline assessment but also their sex, date of birth, whether they attended the school's nursery class, if so if that teacher assessed them, and if English is not their mother tongue, which language they do speak. In three schools with a total of 117 reception children assessed there are 52 children who speak at least one other language; 16 different languages in addition to English are spoken by children in these four reception classes. There is a year's difference between the oldest and youngest children. Already further children have entered these classes, for some of whom English is not their first language. All these confirmed our concerns at the proposed Government baseline assessments as a measure of progress and accountability.

Immediately following the seminar, the first of the Government's changes took place, and as a consequence we changed our priorities for the research, postponing further analysis of the test results and assessment of individual children on PLA12 as it was possible that these baseline measures would not be implemented. We have continued with our investigation into the issues surrounding the large number of children now in primary schools whose mother tongue is not English. That aspect was considered by Chris Watts. In a questionnaire for reception teachers we explored their views of the value and validity of assessment of young children on entry to primary school. The responses to this are anonymous but we invited teachers to indicate if they were willing to be interviewed. We now have completed this aspect, with 100 responses within two months. On the basis of our analysis of the responses to the questionnaire

we planned interviews with those teachers who expressed willingness to be interviewed.

A selection of relevant references

1. At the seminar on 24 February in *Why it is better without baseline assessment 4 core reasons*, Nancy Stewart, Principal Consultant, Early Learning Consultancy stated that many education experts and teaching unions are strongly opposed to the introduction of these standardised on-entry assessments on a number of well-evidenced grounds, including that:

- a) Many children are already being wrongly labelled as achieving below typical standards, with harmful effects.
- b) The assessments disrupt children's introduction into school.
- c) The narrow focus on attainment in prescribed subject areas is harmful to children's learning and development in the early years.
- d) The planned system will not provide a useful indicator of school quality.

2. In his presentation, *Which abilities of 4 year olds predict later academic achievement? Developmental evidence and implications for early assessment*, David Whitebread, University of Cambridge, argued that if the purpose of making baseline assessment of children's capabilities is "ultimately to improve the quality of their education" then we should bear in mind that "the BA models currently on offer are likely to be inaccurate, unreliable and potentially harmful, as they are in danger of negatively impacting on the crucial first few weeks of children's experience of their primary school".

3. In *The Introduction of baseline Assessment: views of head teachers, teachers and parents*, Guy Roberts-Holmes, UCL, IoE, London reported the main findings of the NUT/ATL sponsored research stating that: "For many teachers, baseline assessment has had a negative impact on their working lives without benefiting the children they teach."

Teachers and head teachers see all three baseline assessment providers as inaccurate, unreliable and lacking in validity and hence in ability to accurately measure the 'value added' by schools. Guy Roberts-Holmes and Alice Bradbury explore this further in an article in *Improving Schools* vol.19 (2), June 2016, entitled *The datafication of early years education: its impact on pedagogy*. (This journal is available online from the Sage website.)

4. In *Baseline assessments: the delusions of 'predictive validity'*, Terry Wrigley challenged the claims that these can accurately assess ability or potential of young children. These issues are considered in more depth in an article by Terry Wrigley and Louise Wormwell in *Improving Schools* vol. 19 (2), June 2016 entitled, *Infantile accountability: when big data meet small children*.

5. Shortly after our research seminar, the Government decided to scrap its plans for the baseline assessment it was to have introduced in 2016. My further article in *Primary First Summer Issue, Baseline Assessments: why ignore the evidence?* updates the information. This is now available online.

6. NB Research cited in the *Education Journal* immediately prior to our seminar is relevant (Issue 260). Save the Children stated that "the research demonstrated that the most crucial determinant of success in SATs tests was how well children could communicate when they started school" (p. 6). A new paper by UNESCO's Global Education Monitoring Report revealed "that 40% do not access education in a language they understand" and, to quote from page 12: "A review of 40 countries' education plans found that less than half recognised the importance of teaching children in their home language, particularly in early grades."

In view of the data we are collecting in our ongoing research on the variety of languages spoken by young children in many reception classes in England, it would appear that DfE pays insufficient attention to researches such as these!

7. Often attention has centred on the deficiencies of the child in the learning situation and the inadequacies of the parents. This view still permeates much of the discussion on baseline assessment and

school readiness in UK and the United States. On May 9 2016 in *Metro* (the free newspaper), the heading of a news item was 1 in 3 pupils “not ready for school”. “Almost a third of four-year-olds are not considered to be ready for the classroom, a report found. School readiness seems now to be claimed to be measurable by testing the young children’s competence on entry to primary school aged four to five years on very limited aspects of reading, writing and numeracy; no account is being taken of the many more important attributes shown by research to influence children’s readiness for learning. Parents may be encouraged to see these as priorities, and thus be tempted to purchase the many commercial programmes widely advertised to ensure their children are ‘school ready’.”

8. A research reported in *Improving Schools* vol. 19 (2), June 2016, provides important evidence for the debate on school readiness of young children around five years of age from Norway, at least for the formal testing and expectations placed on them currently. In his article, *School starters’ vision – an educational approach*, Gunvur Wilhelmsen questions whether the level of visual development and gender differences may influence how pupils react, behave and learn in school. He claims this ought to have an impact on our expectations with regard to young children’s school performance, stating that school starters have a less mature visual acuity at a close reading distance than at far distance.

9. The following new information from Peter Moss entitled *Is a pre-school PISA what we want for our young children?* has just been reported (<https://ioelondon.blog.wordpress.com/2016/08>). According to Peter Moss, an assessment of early learning outcomes among 5-year-olds IELS is being developed, and a call for tenders has been issued. The aim is for pilot testing to be undertaken in 3 to 6 countries in late 2017 and early 2018, and a group of 16 countries is involved. This includes UK, though Peter Moss isn’t sure if this means only England!

Assessing readiness for school or schools’ readiness for children

In April 2016, DfE stated that it was still committed to assessment in reception and that: “over the coming months we will be considering options for improving assessment arrangements in reception beyond 2016/17 and will engage stakeholders in that work”. (Press release on 7 April 2016: <https://www.gov.uk/government/publications/reception-baseline-comparability-study>.)

In spite of changes in policy and postponements in implementation over the past few months there is no evidence to show that DfE is not still committed to some form of readiness measure and accountability as early as shortly after young children’s entry to reception class.

Originally published in *Education Journal*, issue number 280, 27 September 2016.

Synthetic Phonics and Baseline Assessment under the Searchlight in 2017: are they value for money in a time of cuts?

By Margaret M Clark OBE

I Synthetic phonics as the method of teaching reading in England

Good literacy is the foundation for all achievement in education and critical for everyday life. We have strengthened the curriculum focus on developing pupils' reading and writing skills and placed renewed focus on the requirement for pupils to be taught to read using systematic phonics.

Nick Gibb, Minister of State (Department for Education) 13 February 2017. Written answer to questions by Jim Cunningham MP.

Background

Since 2013 in a series of articles in *Education Journal* and elsewhere I have challenged the evidence base for claims by government ministers in England that synthetic phonics should be the method of teaching all children to read and that there is value each year in a phonics check with a pass/fail score administered to all children in Year 1 and if they fail to reach a pass mark of 32 repeated in Year 2, and even less in repeating it again in Year 3 as has been considered.

In 2017 claims continue to be made by Nick Gibb both for synthetic phonics and for a positive effect of the phonics check on attainment. In a written answer to a question on 1 March 2017 he again claimed that "evidence shows that systematic phonics is the most effective approach to teaching young children to read". He also claimed that the phonics check now passed by 89% of children in Year 1 has caused an improvement in the level of literacy attainment. Both debatable claims. I have updated my evidence on this in Part IV of the revised edition of *Learning to be Literate: insights from research for policy and practice* (Clark, 2016a). I also challenged whether the phonics check had since 2012 achieved anything other than an improvement year on year in the percentage pass on that test, now a high-stakes test the results of which form an important element in Ofsted's inspection of schools and of courses of training for teachers.

My evidence came in part from the research by the National Foundation for Educational Research commissioned by DfE. My concerns about the effects of this policy on children's understanding of written language are set out briefly in *Flawed arguments for phonics in The Mismeasurement of Learning* (Clark 2016b downloadable from www.reclaimingschools.org). Freedom of Information Questions have enabled me to estimate the cost of the policy to DfE. In Issue 186 of *Education Journal* in January 2014 I reported on some of the expenditure on this policy and there is further information on expenditure by DfE up to 2015 in chapter 18 of Clark 2016a. I have continued to enquire as to the expenditure by DfE on synthetic phonics and will summarize this further information below.

The results of the phonics check for 2016, published in September, do show a gradual increase in the percentage pass on the phonics check over the years between 2012 and 2016, not surprising in view of the evidence from research as to how much time is being spent in preparing for what has become a high stakes test (SFR 42/2016, 29 September 2016). However, there are two disturbing aspects to which attention is not drawn by the Government. Table 5 in 2016 shows the difference in pass rate by month of birth with only 72% of the youngest children passing the check and 87% of the oldest children (only 68% of the youngest boys and 77% of the girls passed). I had to request this information in previous years, and did draw attention to its significance as being as important as the difference in percentage pass between girls and boys to which attention is frequently drawn. My further concern is that although the phonics check was expected to be a diagnostic assessment only pass/fail is reported with 32 a pass, and 31 a fail necessitating a retake of the check the following year, often with more practice with synthetic phonics. In

the first two years when the pass mark was known in advance NFER pointed out that there was a disturbing peak percentage pass at 32 as compared with 31 (recorded as a failure). Although the pass mark is now not revealed to teachers in advance it has remained at 32. Figure 1, the marks distribution, shows that over the years there has continued to be a peak at 32 as compared with 31, suggesting the possibility of a less than reliable test. Taken together with the lack of evidence that there is one best method of teaching reading, namely synthetic phonics, how justified is the continuation of this policy and the ongoing expenditure with cuts affecting other aspects of education? Have this policy and the high stakes test fulfilled the claims made for them?

Four initiatives related to synthetic phonics with their costs are summarised below together with approximate costs of the phonics check to DfE. I was informed that DfE does not hold any information on the cost to schools or colleges training teachers of the initiatives related to synthetic phonics. The cost of developing the phonics check has already been reported.

Initiatives funded with synthetic phonics as their focus

1. Eight schools each received £10,000 as in 2015-16 the DfE funded 8 partnerships * to enable schools to work together to improve the quality of phonics teaching (up to March 2016) £80,000

*On a press release on about Partnership schools the names of the eight schools were given

2. External advisers in 2013 (linked to phonics check). £11,750

3. Ten Ruth Miskin Road Shows. "After a competitive tender in March 2016" DfE funded Ruth Miskin Training to deliver ten roadshows during March 2016 "to promote effective phonics teaching and early reading". £30,690

4. NFER was commissioned to undertake a pilot study in 300 schools, pilot study to consider whether children who failed the phonics check in Year 2 should be retested in Year 3. £64,606

On 1 Feb 2017, following several requests, I received a copy of the report (of the administration of the pilot study) with the comment that "there are currently no plans to publish the report" and a brief resume of the release by DfE to appear the following day. There were no recommendations for policy either in the DfE short document or in the NFER report. It has been reported that NAHT are pleased to learn that the phonics check will not be repeated in Year 3. I have so far failed to establish any written source for this, in spite of extensive enquiries.

The four initiatives listed above cost £187,046

I have not been able so far to track any published reports on these four initiatives.

5. Match-funding for commercial synthetic phonics materials and/or training (matched by similar expenditure by schools) in 2011to 2013.

This was claimed by over 14,000 schools DfE expenditure only: £23.7 million
(This was reported again recently by Nick Gibb.)

Through my Freedom of Information enquiries, I did establish that approximately £22 million was spent on one or other of the approved commercial programmes and a further £1.3 million on training courses and found out the names of those publishers who received payment which accounted for part of this expenditure (see page 148 in Clark 2016).

Cost of printing, distribution, collation and analysis of the results of the phonics check

NB this is DfE expenditure and does not include any costs to schools, nor does it include the development of the check.

I have now obtained updated information on the expenditure by DfE specifically on the phonics check. This covers only printing, collating, distribution of the tests and the analysis of the results. Each year since 2012 to 2016 cost has been approximately £260,000 per year, with a similar expenditure anticipated in 2017. Approximately £1.2 million

II Baseline assessment on entry to primary school

Background

In 2014 The Standards and Testing Agency claimed: “The purpose of the reception baseline is to support the accountability framework and assess school effectiveness by providing a score for each child at the start of reception.... as the basis for an accountability measure of relative progress of a cohort of children through primary school.”

In answer to a question on 2 March 2017 on the international evidential basis on the validity and reliability of an accountability progress measure using an assessment at the start of school as a baseline .. Nick Gibb replied:

“There is evidence from The University of Durham that a baseline in reception can be used to measure progress by the end of primary school ... Baseline assessments are used in a number of countries for a wide range of purposes ...”

See *Education Journal* issue 297: 49 for the full reply and page 47 for his answer to a question on 1 March on expenditure so far on baseline assessment in 2016-17.

In a series of articles of mine and by colleagues in *Education Journal* since 2015 we have reported on research evidence that should hopefully influence government policy on baseline assessment, including well documented research on the unreliability of one off assessments of young children, the effect of the context on their apparent competence, and the poor predictive reliability of such measurements. In February 2016 at a research seminar at Newman University a series of research papers were presented. Summaries of all eight papers are available on line. No one challenged the strength of the argument against the proposed baseline assessment policy, stressing that it was not a criticism of specific providers but of the policy itself. Shortly after this it was announced, following a study commissioned by DfE, that baseline assessment would not be made a statutory requirement in 2016 because of compatibility concerns with regard to the scores on the tests from the three chosen providers. However, schools continue to be permitted to reclaim the basic costs for the use of the authorised providers, now four rather than the previous three, still in 2016, and no decision has yet been with regard to 2017. Judging by the Minister’s response to recent questions it appears that this may merely have been a postponement and that an announcement may be imminent.

In a new edition of *Understanding Research in Early Education: the relevance for the future of lessons from the past*, chapter 10 provides up to date information on the issues on baseline assessments as predictors of progress. (Clark in press expected June 2017). I have set out below the expenditure by DfE so far on baseline assessment, mainly based on Freedom of Information questions.

“In 2017 claims continue to be made by Nick Gibb both for synthetic phonics and for a positive effect of the phonics check on attainment. In a written answer to a question on 1 March 2017 he again claimed that “evidence shows that systematic phonics is the most effective approach to teaching young children to read”. He also claimed that the phonics check now passed by 89% of children in Year 1 has caused an improvement in the level of literacy attainment.”

Expenditure in Baseline Assessment by DfE since 2015

Amount refunded to schools for 2015-16

CEM £297,000; EE £2,391,000; NFER £475,000 (2015-16) Total £3,163,000
(GL only in 2016-17 after it was added as an authorised provider £4,477. This is included in the figures from a written question on 17. February 2017 answered on 1 March by Nick Gibb. It is also stated in his answer that “further details on whether the optional baseline will be available for schools to use in the 2017/18 academic year will be provided in due course”.

Amount refunded to schools in 2016-17 expenditure to date: £745,000 and a further estimated £125,000

Approximately £870,000

Research Related costs (The results of some of these researches have been published)

NFER	£49,756
CEM	£73,667
Early Years review group	£12,131
Scottish Quality Authority to “examine the comparability from the different providers”	£192,877
Tribal Education to develop a trial of several approaches to reception baseline	£195,965

Total	£524,396

Written answer to question as to how many schools have used the various baseline measures in 2016-17 question on 2 March 2017 answer by Nick Gibb on 8 March.

Number of schools still using these commercial approved assessments in 2016-17

Centre for Evaluation and Monitoring Durham University:	913 schools
Early Excellence:	2,038 schools
GL Assessment:	33 schools
National Foundation for Educational Research:	917 schools

Concluding Comments

Following a speech by the education minister Nick Gibb on 10 September 2016 on the importance of education research I published an article in *Education Journal*, issue 281: 14-16, *An appeal for a research literate teaching profession: what could be the implications for policy?* In a further article in issue 186: 12-16 I asked: “Whose knowledge counts in Government literacy policies and at what cost?” As I stated in that article I am not asking readers to accept my analysis of research. I am pleading that teachers both in training and later are treated as professionals, encouraged to question and evaluate statements whatever their origins, whether from research or from government policy. In this latest article, I not only provide sources indicating where to find evidence to evaluate these two government policies but also information of some of the expenditure by DfE. Unfortunately, there is no way of estimating the cost to schools’ budgets of these and other policies and only in the longer term are we likely to be able to assess their effects on the education of the young children.

Originally published in *Education Journal*, issue number 299, 31 March 2017.

Primary Assessment in England: Government consultation Part II Baseline Assessment: where is the evidence?

By Margaret M. Clark OBE

The DfE has issued a consultation document on Primary Assessment, launched in March. The consultation has continued over the period of the General Election, with the closing date for responses 22 June. Over recent years I have made a careful assessment of the evidence for two policies relevant to the consultation, synthetic phonics and baseline assessment, with articles in *Education Journal* and elsewhere, updated in my two recent books (2016a for literacy and 2017b for baseline assessment).

I decided to summarise my evidence on synthetic phonics and baseline assessment in two articles prior to submitting evidence to the consultation. In these articles, I can only refer briefly to the most important aspects, but will give sources enabling readers to draw on this evidence for their submissions. In the previous article my focus was on the phonics screening check and in this second article, I will consider the justification for baseline assessment as proposed by the government in the light of published evidence. Concern has been expressed at the unreliability of such measures, particularly to measure children's progress and make teachers accountable. Commercial interests appear to have a powerful influence in determining government policy in England, as in some other countries, and on the materials recommended by government. These not only account for expenditure by DfE, where some of it can be tracked by Freedom of Information questions, but also eat into school budgets.

Baseline assessment

Freedom of Information questions have enabled me to make some estimate of the large amount of money spent by government, in particular on commercial materials, not only on synthetic phonics but also already on baseline assessment (see Clark 2017a in *Education Journal* 299: 16-19). It had been DfE's intention to require all schools to assess children on one of three measures EE, CEM or NFER within six weeks of the children starting in reception class, commencing in 2016. A single score was to have been returned to DfE, assessed in English, whether or not that was the child's first language. However, the Government announced it would reverse its decision and not implement baseline assessment for children entering reception class in 2016, having discovered that the scores from the three selected tests could not reliably be compared! The decision was made for 2016 to continue to use the Foundation Stage Profile while further consultation took place.

In spite of this decision, DfE has in 2016 continued to refund schools the basic costs should schools choose to use any of these three tests, EE, CEM or NFER. To my surprise not only has that cost been incurred but a fourth assessment GL has been added for which schools can reclaim costs. Using Freedom of Information questions, not only did I seek an explanation for this addition, I was also able to gain an estimate of how much money has so far been spent by DfE on baseline assessment, this is even before baseline assessment has become a statutory requirement and while the Foundation Stage Profile is still in place (see Clark 2017a *Education Journal* 299: 16-19).

I have published a series of articles critiquing the developments in *Education Journal* (244: 11-16, 259: 14-16, 261: 10-12, 280: 14-17, and *Primary First* 16: 8-13). These publications form the basis for chapter 10 in my new book, published in June 2017 (Clark 2017b). In February 2016, we held an invitation research seminar on baseline assessment at Newman University with papers from leading researchers who had recently published on the topic. An outline of these papers is available online at www.newman.ac.uk/24feb and these were reported in *Education Journal* Issue 261: 10-12. The researchers expressed their opposition to the imposition of any form of baseline assessment and challenged the

reliability of such measures to monitor children's progress even over a relatively limited period. Several of the presenters reported the concern of teachers at the proposed timing of the assessments, the fact that they were to be conducted in English and that DfE wished merely for a single score to be reported and that no account was being taken of the children's age. They were worried that at a time when teachers would have been settling children into reception class their focus would instead be on a high stakes test. The criticism was not confined to any one of the proposed measures.

The following are two of the research presentations available on line: *Baseline assessment: the delusions of 'predictive validity'* (T. Wrigley) and *Which abilities of 4 year olds predict later academic achievement? Developmental evidence and implications for early assessment* (D. Whitebread). My paper, *Baseline assessments: their value and validity in assessing young children on entry to school*, was a report on our ongoing research in the West Midlands. I expressed concern at the idea of assessing young children immediately on entry to school, and in English. I drew attention to evidence from three primary schools in the West Midlands where 16 different languages were represented in four reception classes in addition to English. In these classes where 117 children were tested in 2015, for 52 English was not their first language. A further matter for concern was that, as in other schools, there was a year's difference in age between the oldest and youngest children and no allowance was being made for this (as is also true with the phonics screening test).

Subsequently a powerful article has appeared in the Sage journal *Improving Schools* in 2016 19: 2 by Terry Wrigley and Louise Wormwell, *Infantile accountability: when big data meet small children*. This can be downloaded from the Sage website and gives a carefully argued critique against the type of accountability measures the government is still proposing. It argues against "tacit assumptions of linear progress underpinning large-scale data-based accountability processes", and to quote from the abstract: "This article examines a government attempt to impose testing on four-year-olds as a baseline against which to hold primary schools accountable for children's subsequent progress. It examines the various forms of baseline testing in this experiment and analyses the misleading claims made for the "predictive validity" of baseline scores.

As I commented in my previous article there are no questions in the consultation document on the phonics check. However, among the twenty questions raised are some that relate to baseline assessment. I have already seen the draft response from a group of over 20 organisations to be submitted under the title *Better Without Baseline* and in their response it is clear they oppose any introduction of baseline assessment in reception class and they appeal for DfE to abandon this policy which has been tried previously and failed. Two questions in particular address the question of baseline assessment, namely questions 5 and 6.

Question 5. "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."

Question 6 also relates to baseline assessment and asks if a reception baseline were to be introduced when should it take place. One could only respond to this question positively were one to accept the premise that baseline assessment of young children is a valid predictor of progress. Based on the references cited here I would argue that this policy should not go ahead.

Footnote:

In an article in *Education Journal* in 2015 244: 12-16 entitled *The language of young children on entry to school as measured by baseline assessments. Why ignore the evidence from research?* I drew attention to the volume of research as early as 1970s and 1980s which showed just how unreliable are one off assessments of the language of young children. The score will be affected by the context, the particular adult, whether familiar or strange, administering the test and the child's perception of the purpose of the assessment. There may even be differences depending on the extent to which the young child's home experiences have prepared them for this more formal questioning where the child is expected on occasion to respond, playing the game of answering questions where he or she may be asked questions to which the adult already knows the answers. Some young children, but not all, know how to play such games on entry

to school. Clearly the subtitle of my new book remains apt, *The relevance for the future of lessons from the past!*

References

Clark, M.M. (2016a) *Learning to be Literate: insights from research for policy and practice*. Revised edition Routledge.

Clark, M. M. (2017a) *Synthetic Phonics and Baseline Assessment under the Searchlight in 2017: are they value for money in a time of cuts?* In *Education Journal* 299: 16-19.

A series of articles in *Education Journal* and *Primary First* present further evidence for the views expressed in this article, These articles were updated to form a new chapter 10 in the third edition of my book referenced below.

Clark, M.M (2017b) *Understanding Research in Early Education: the relevance for the future of lessons from the past*. Third Edition Routledge.

Wrigley, Terry and Wormwell, Louise. (2016) *Infantile accountability: when big data meet small children*. In *Improving Schools*. 19: 2. Available online on The Sage website.

Originally published in *Education Journal*, issue number 308, 20 June 2017.

EDUCATION JOURNAL

Editor	Demitri Coryton FRSA
Associate Editor	Chris Waterman FRSA
Parliamentary Editor	Tracy Coryton
Reference and Research Editor	Arabella Hargreaves
Writers/contributors	Professor Margaret M Clark OBE

Education Journal
Special issue No. 1
Wednesday 24 October 2018

© The Education Publishing Company Ltd. 2018.

Published by
The Education Publishing Company Ltd., Devon.

Education Journal

Email: info@educationpublishing.com

Web: www.educationpublishing.com

The Education Publishing Company Ltd.

15A East Street, Oakhampton, Devon, EX20 1AS.

Email: info@educationpublishing.com

The Education Publishing Company of Australia

PO Box 390, Sandy Bay, Tasmania 7006, Australia.

Email: epca@educationpublishing.com

The Education Publishing Company of New Zealand Ltd,

PO Box 109,481, Newmarket, Auckland, New Zealand.

Email: epcnz@educationpublishing.com

The Education Publishing Company (Europe) Lda.

Largo Professor Jose Bernardo Cotrim No. 35,
2240-606 Carril, Dornes, Ferreira do Zezere, Portugal.

Email: epceurope@educationpublishing.com

This is the first of a series of special issues of *Education Journal* that will bring together, in one easily accessible space, a series of articles that were originally published in *Education Journal*, and sometimes other of our publications, on a particular subject.

This first special issue is on a review of the evidence of Reception Baseline Assessment becoming statutory in England. It reproduces a series of articles by Professor Margaret Clark first published in *Education Journal* between 2015 and 2017.

Subscription Rates

UK electronic: £100 + VAT = **£120**

International electronic: £100 (No VAT) **£100**

Multiple user licence, electronic £250 + VAT = **£300**

Education Journal is published weekly except during academic holidays.

ISBN: 977-1-36-445000-8

ISSN: 1364-4505

Order details

To order *Education Journal* please send your details to us by email, telephone or post. Payment can be made by BACS, cheque or credit card.

We need to know:

Your name and email address.

Your postal address and the name of your organisation if relevant.

How you wish to pay and whether you want an invoice.

Your purchase order number if you have one.

If paying by credit card please let us have the name on the card, the type of card, the address to which the card statement is sent if different from the address given above, the long number, the expiry date and the security code (the three digit number on the back).

Cheques should be made payable to the Education Publishing Company Ltd and sent to the Subscription Department, The Education Publishing Company Ltd., 15A East Street, Oakhampton, Devon, EX20 1AS.

THE
EDUCATION PUBLISHING
COMPANY LIMITED