

Occasional paper, March 2017

Establishing the Effects of Quality in Early Childhood: Comparing evidence from England

Professor Pam Sammons, Professor Kathy Sylva, Dr James Hall, Professor Iram Siraj, Professor Edward Melhuish, Brenda Taggart, and Sandra Mathers

In February 2017 Blanden, Hansen and McNally published a document that sought to investigate the effects of quality within early education and care settings in England (Quality in Early Years Settings and Children's School Achievement, CEP Discussion Paper 1468, The London School of Economics). Within this document was an argument that contradicts over 30 years of research showing that the quality of early childhood settings has both significant and lasting positive effects on children's development and educational progress (e.g. Anders et al., 2011; 2015; Phillips and Lowenstein, 2011; Vandell et al., 2010; Zaslow et al., 2010; Hall et al., 2013; Mathers et al., 2014; Melhuish et al., 2008; Sammons et al., 2008; Siraj-Blatchford et al., 2008 Sylva et al., 2011; Taggart et al., 2006). Blanden and colleagues conclude that quality (as measured by the presence of a graduate and by Ofsted inspection grades) has only very weak associations with children's outcomes.

The controversial findings of this study have since been widely covered in the media - and by those with an interest in keeping staff costs low in the sector. Like many, we were surprised by the authors' counter intuitive conclusions. We have now scrutinised the LSE paper and have identified a number of serious limitations in the study's conceptualisation, research design and research methods. We argue that these limitations are so important that the study's conclusions and apparent policy implications should be treated with great caution. However, we also welcome the

opportunity to reflect on the need for appropriate research designs and methods for studying the effects of early education; and in this critique we draw attention to the ways that variation in research methods can lead to important differences in conclusions. We suggest that research designs need to be appropriately constructed to identify and separate the effects of pre-school and primary schools, with careful measurement at pre-test and later follow ups at appropriate ages, in order to have confidence in the results.

Research methods in the CEP study

The paper combines data from The National Pupil Database (NPD) and the Early Years' Census (EYC), and analyses the information they provide for about 1.8 million children resident in England who were born between September 2003 and August 2006. The study of 1.8 million children is the paper's unique selling point, but it comes at the price of very weak, and in our judgment flawed, measurement of the quality of provision within the early years, and of child outcomes. Those engaged with early years education know that there are many measures of 'quality', and that process quality, i.e. children's daily experiences, is key for child development and is particularly relevant in supporting disadvantaged children. Furthermore, measures of quality are best established through detailed observation. Likewise, measures of child development are most robust when completed using trustworthy standardized assessments, by individuals trained to use these assessments consistently and accurately, rather than relying only on more subjective teacher judgments that are not intended to show fine distinctions in children's learning or socio-emotional development .

The LSE paper uses official government data to link the development of the children, at ages 5 and 7, with the quality of the early years settings they attended during their entitlement provision. However, the measures of 'quality' used are weak, and cannot be relied upon for the purpose of demonstrating links between quality in early years settings and child development. Quality is assessed only by children's access to a graduate and by global Ofsted inspection grades. Previous research (Mathers et al., 2012) has shown that Ofsted inspection grades are only very weakly correlated with robust observational quality measures. Similarly, although staff qualifications have been shown to relate to quality and to child outcomes (Fukkink & Lont, 2007), this is a very crude structural measure which takes little account of the myriad other

influences on children's actual experience within their pre-school setting. Even as a measure of staff qualifications, the simple presence or absence of a graduate is very crude. The study uses information in the PVI Census (PVI) to estimate the ratio of children to the teacher-qualified staff who taught them, assuming all children in the maintained sector have access to a qualified teacher. This does not take into account the many ways in which graduates may be deployed and work with other staff in pre-school settings; or the qualifications of the many other staff working within the setting.

The measures of child outcomes are also weaker than those used in much previous research. The main outcome measure is the child's total score on the Early Years' Foundation Stage Profile (EYFSP), completed by their teacher at the *end of* Reception Year. In the period relevant to this study, the EYFSP used a 9-point rating system to measure 6 areas of learning and development: 1. social development; 2. literacy; 3. numeracy; 4. knowledge and understanding of the world; 5. physical development; 6. creative development. The LSE paper uses the 'broadbrush' total EYFS score rather than the more detailed individual scores for the six areas.

Child outcomes at age 7 years are also based on teacher assessments: a single Key Stage 1 (KS1) points score that conflates children's attainment in different domains including literacy and maths. This is a surprise as educational effectiveness research consistently shows school effects tend to be larger for maths and science rather than literacy. Teacher assessment is recognised as less reliable than standardised externally administered assessments, and is therefore less useful for a research study that seeks to tease out continuing pre-school influences and separate them from school effects. Each assessment is subject to unique measurement error, and the single-person rating system means that scores are subject to the unconscious biases of each teacher; an effect particularly noted to risk the accuracy of teacher-ratings for disadvantaged children (e.g. Burgess & Greaves, 2013), sometimes termed 'halo effects'. The use of teacher assessment data therefore considerably weakens the claims made by the authors. Further, use of EYFSP data at age 5 means that the LSE study's first measure of pre-school effects is taken at the end of the Reception year, and therefore includes the effects of that Reception year as well as the effects of the pre-school settings children attended prior to school.

In addition to the weakness of their quality and child outcome measures, the LSE study faces difficulties with alignment of these measures over time. Of particular concern is the use of the EYFS measures at the *end* of reception (when most young children have spent many months in full time reception classes). This is a limitation of reliance on the NPD. Previous studies of pre-schools have sought to collect outcomes measures at the end of pre-school or start of reception to avoid confusion with likely school effects, an important issue that we address later in this piece. Children and their settings are matched to the Ofsted inspection judgment closest to the time they were enrolled – even though the inspection might have taken place up to six years before or one year after the child’s attendance. This time difference is particularly important given the high staff turnover in early years settings. The waters are further muddied by changes in the Ofsted inspection regime over time. In 2005-2008, PVI settings were inspected on different criteria from the maintained sector, but in 2008-2011 all providers experienced the same inspection criteria and a single overall judgment. It must be recognised that Ofsted judgments are not intended to provide detailed measures of quality and this is an important problem for their use in the LSE study because settings that received only a satisfactory or inadequate judgment would have been under much pressure to improve. Thus the measure of poor quality (Inadequate rating) is likely to be a poor guide to the quality of children’s experiences since settings would have been responding to the judgment and taking steps to try to improve.

Summarising the study’s major limitations

(i) **It has a very weak set of outcome measures.** Although there are some government attempts to ensure that teacher EYFSP ratings are consistent, there are obvious and inevitable problems of consistency with assessments by thousands of different teachers. Other research mitigates this by using rigorous psychometric assessments conducted by a small group of researchers trained to use stringent assessments (such as the British Ability Scales vocabulary or early number tests). Furthermore, because EYFSP is conducted at the *end of* reception, it measures *both* the effect of a child’s ‘early education’ as well as the effect of their Reception Year. This means that it cannot isolate and identify the effect of a child’s time in their pre-school setting. Other research designed to study pre-school effects avoids

confounding school and pre-school influences by collecting outcome measures at the end of pre-school, or within a few weeks of entry to reception.

There are also many problems with the use of one summed KS1 score for the later study of teacher assessed outcomes at age 7, as discussed earlier.

- (ii) **It relies on poorly timed measures and has no baseline.** First, the paper has no measures of development carried out when children entered early years provision at the start of the 3 year old offer. This means that the 'value added' contribution made by preschool to children's development from age 3 to the point of primary school entry (at the start of reception when children are typically rising 5 years) cannot be calculated. Other studies avoid this problem by taking several measures of child development when they enter pre-school provision and then conducting follow ups. Measures commonly assess receptive and productive language, non-verbal skills, and social-emotional development in order to establish developmental progress (change) over the time in pre-school. Second, the Ofsted measures of quality were not collected when the children were in their settings: some were up to six years adrift and nearly half over two years out. Given the time lapse and accountability role it cannot be assumed that low Ofsted ratings of quality can accurately reflect the quality experienced by children in their settings despite the national sample studied.
- (iii) **It relies upon unsuitable measures of 'quality' within the early years.** Ofsted measures can only provide crude judgments. Other research designed to investigate quality in depth uses finely differentiated measures such as the Early Childhood Environment Rating scales (ECERS) taken in the setting when the children who participate in the statistical analysis are *actually* present. Furthermore, Ofsted judgments have a low correlation with research-validated measures such as the systematic observation rating scales: ECERS-R, ECERS-E and SSTEW. Such measures are better suited to capture 'process' quality being based on well tested observations of children, staff and settings. The Ofsted judgment is based on a wide range of criteria that go beyond the observed 'learning environment' of the children and the timing of inspections fits poorly with the time children were in their pre-schools.
- (iv) **It does not take account of the many important differences between families that affect child development.** For decades, research has shown that

parental education and social class have the largest influence on children's development. The LSE paper accounts for the children's social background using only two coarse measures: free school meals and a post-code measure of neighbourhood disadvantage. Other research investigating the effects of quality has collected much more detailed measures of parents' education level, the size of families, parents' salary, and, especially important, the early years home learning environment. These measures permit us to safely compare the effects of early years provision, despite sometimes large differences between families and children. The lack of such measures in the LSE study means that we cannot read the results with the same level of confidence as those from other studies with better controls for family factors. In addition, the absence of control for baseline cognitive and social-emotional development at the start of the three year old offer is of great concern since it is known that there are large differences in development that are well established before the age of three. The LSE paper therefore cannot see how far pre-schools may shape children's progress during their time in pre-school.

(v) **The statistical methods used in the LSE paper do not adequately account for the fact that children are 'clustered' within pre-school settings and schools.** It is well established from educational effectiveness research that studies seeking to identify the effects of schools and preschools to take proper account of the clustering in the data (children nested by pre-schools and then later by primary schools) using 'multilevel modelling' techniques. Failure to control for such clustering leads to biased estimates of the effects of any predictors in the models used. The fact that the LSE paper does not use such techniques weakens the conclusions which can be drawn.

Conclusions

Blanden and colleagues claim their research represents the first comprehensive assessment of the link between children's outcomes and the characteristics of their pre-school experience. However, in contrast to earlier studies, the research was not carefully designed as a study of children's pre-school experiences and their subsequent development; rather it is a secondary data analysis that has relied on unsuitable measures from the national data sets it has linked. Given the serious limitations we have outlined in this critique we think this claim cannot be substantiated. Moreover, in view of these serious methodological problems, we are

not surprised that the paper suggests 'quality' has only a small effect. The study does not have the necessary design and careful measures needed to identify pre-school effects reliably; its measures of quality are coarse and lack external validity (because the timing of inspection ratings fits poorly with the period children attended pre-schools and the intended accountability consequences of inspection judgments will have influenced settings' practices).

A profound understanding of the effects of quality in early childhood requires precise/differentiated outcome measures at entry to primary school, strong control for family background and good baseline measures of child development at the start of the three year old offer/entry to pre-school in both cognitive and socio-emotional domains. In addition, careful observations of the daily experiences of children during the period of their entitlements and the role qualified teachers play in supporting these are needed to understand the complexity of early childhood quality in pre-schools.

The authors freely acknowledge that they have identified 'substantial differences' in preschool settings in their apparent 'effect' on children's development (despite the documented flaws in their research design and measures we have outlined above). This tends to support earlier international as well as English studies that have suggested pre-school does indeed matter, especially if of high quality. For example, they claim that there was a very large gap between the developmental outcomes of children in settings in the top and bottom quartiles. They conclude that the predictors they used to explain quality effects (teacher presence, Ofsted judgments) may not be the best ones to account for why some settings seem to boost children's development while others do not. We conclude that the serious limitations of the data available and the research design mean that the study by Blanden and colleagues cannot give an accurate understanding of the effects of quality in early years settings on child development from pre-school entry or start of the three year old offer up to entry to Reception, and then finally up to age 7. Given the flaws, there is likely to be an underestimation of the effects of both 'quality' and of teacher staffing. For these reasons we believe it would be dangerous for the LSE study to be used in drawing policy conclusions about the importance of pre-school quality and the role of qualified teachers in supporting young children's development

To conclude we believe it is important for researchers and policy makers to draw on current understanding of what is meant by early childhood quality, including both structural and process quality and support more studies that explore the links between these. This has been well discussed by Anders (2015) in her comprehensive review and comparisons of pedagogical practices in ECE in England and in six other countries. 'The quality of preschool learning is seen as a multidimensional concept covering structural characteristics, teachers' beliefs and orientations, and process quality Structural quality refers to aspects such as class size, teacher-child-ratio, formal staff qualification levels, provided materials and size of the setting. Process quality refers to the nature of the pedagogical interactions between preschool teachers and children, the interactions among children and the interaction of children with space and materials. Recent approaches also highlight the quality of interactions between staff and parents Conceptualizations of preschool quality cover global aspects (such as warm climate or child-appropriate behaviour), as well as domain-specific stimulation in learning areas such as literacy, emerging mathematics and science. It is hypothesized that process quality has direct effects on children's learning and development, while structural and orientation quality have indirect effects through their influences on process quality (Pianta et al., 2005). (Anders, 2015:7-8)

The LSE study has not built on past pre-school research designs and we do not think its claim to be comprehensive is justified. We urge policy makers and practitioners to take account of the serious limitations we have identified here and to avoid taking at face value the erroneous conclusions that staff qualifications and quality of pre-school experiences do not matter because they make little differences to young children's development.

Note: An earlier and shorter version of this Commentary was published by Nursery World on 6 March 2017

<http://www.nurseryworld.co.uk/nursery-world/opinion/1160509/do-graduates-and-ratings-really-make-no-difference>

**Professor Pam Sammons, Professor Kathy Sylva, Dr James Hall,
Professor Iram Siraj, Professor Edward Melhuish, Brenda Taggart, and
Sandra Mathers**

University of Oxford Department of Education, University College London Institute of Education, and University of Exeter Graduate School of Education.

References

- Anders, Y. (2015) Literature Review on Pedagogy Published by OECD EDU/EDPC/ECEC (2015)7
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC\(2015\)7&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2015)7&docLanguage=En)
- Anders, Y., Sammons, P., Taggart, B., Sylva, K., Melhuish, E., Siraj-Blatchford, I. (2011). The influence of child, family, home factors and pre-school education on the identification of special educational needs at age 10, *British Educational Research Journal*, 37(3), 421-441. First published on: 17 May 2010 (iFirst)
- Blanden, J., Hansen, K., McNally, S. (2017). Quality in Early Years Settings and Children's School Achievement, CEP Discussion Paper 1468, *The London School of Economics*
- Burgess, S., Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities, *Journal of Labor Economics*, 31(3) 535-576
- Hall, J., Sylva, K., Sammons, P., Melhuish, E., Siraj-Blatchford, I., Taggart, B. (2013). Can pre-school protect young children's cognitive and social development? Variation by center quality and duration of attendance. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 24(2), 155-176
- Fukkink, R., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22, 294-311.
- Mathers, S., Singler, R. & Karemaker, A. (2012) *Improving Quality in the Early Years: a comparison of perspectives and measures*. London Oxford: University of Oxford and Daycare Trust
- Mathers, S., Eisenstadt, N., Sylva, K., Soukakou, E., Ereky-Stevens, K. (2014). Sound Foundations: A review of the research evidence on quality of early childhood education and care for children under three– Implications for policy and practice. *UK: The Sutton Trust and the Education Endowment Foundation*.
- Melhuish, E.C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M.B., & Malin, A. (2008). Preschool influences on mathematics achievement. *Science*, 321(5893), 1161–1162
- Phillips, D. A., & Lowenstein, A. E. (2011). Early care, education, and child development. *Annual Review of Psychology*, 62, 483-500. doi: 10.1146/annurev.psych.031809.130707
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D. & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9(3), 144-159.
- Sammons, P., Anders, Y., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Barreau, S. (2008). Children's cognitive attainment and progress in English primary schools during Key Stage 2: investigating the potential continuing influences of pre-school education. *Zeitschrift für Erziehungswissenschaft Special Edition*, 10, 179–198
- Siraj-Blatchford, I., Taggart, B., Sylva, K., Sammons, P., & Melhuish, E. (2008), Towards the transformation of practice in early childhood education: the effective provision of pre-

- school education (EPPE) project (2008). *Cambridge Journal of Education*, Vol. 38, (1), March, pp. 23-36.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: low quality has little benefit. *Journal of Early Childhood Research* 9(2), 109–124
- Taggart, B., Sammons, P., Smees, R., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Elliot, K., Lunt, I. (2006). Early identification of special needs and the definition of 'at risk': the Early Years Transition and Special Education Needs (EYTSSEN) Project. *British Journal of Special Education*, 33, 40–45
- Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, 81(3), 737-756.
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). Quality dosage, thresholds, and features in early childhood settings: A review of the literature. OPRE 2011-5. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.