This is the accepted version of Campbell, T. (2021), *Relative age and the Early Years Foundation Stage Profile: How do birth month and peer group age composition determine attribution of a "Good Level of Development" – and what does this tell us about how "good" the Early Years Foundation Stage Profile is?* British Education Research Journal: https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1002/berj.3771

# Relative age and the Early Years Foundation Stage Profile: How do birth month and peer group age composition determine attribution of a 'Good Level of Development' – and what does this tell us about how 'good' the Early Years Foundation Stage Profile is?

Tammy Campbell

Centre for Analysis of Social Exclusion

London School of Economics

t.campbell1@lse.ac.uk

This work was supported by a British Academy Postdoctoral Fellowship, under Grant [PF2 \180019].

## Abstract

The Early Years Foundation Stage Profile (EYFSP) is a statutory, summative teacher assessment against nationally prescribed criteria. It is completed for all children in state education in England in summer of the reception year, when September-born children are nearly six, and summer-born children are turning five. This research uses quantitative descriptive and regression analyses of de-identified National Pupil Database records spanning 2008-2018 (N=6 million+). It considers how children's own age and the mean age of their school year-group peers interplay in determining attribution of the EYFSP 'Good Level of Development' ('GLD'). Birth month is starkly predictive of 'GLD' attribution: on average, across years, August-borns are 30 percentage points less likely to be deemed 'Good' than September-borns. Moreover, an older year-group predicts lowered chances of 'GLD' attribution, for children born across all seasons. In 2018, for example, a summer-born in a much older year-group had an estimated 58% chance of being attributed a 'GLD;' a comparable summer-born in a young year-group an estimated 65% chance; an autumn-born in an old year-group an estimated 79% chance; and an autumn-born in a young year-group, an estimated 83% chance. Results build upon previous research indicating teachers' judgements are relative, and also complement studies describing 'moderation' requirements forcing teachers to shape EYFSP scores into prescribed patterns within schools. The latest EYFSP revision continues to ignore the dominance of age in determining 'GLD' attribution and the effects of context and relative judgement. Therefore the EYFSP cannot be an entirely 'reliable, valid and accurate' measure of 'child development' as intended by the Department for Education.


## Key Insights

What is the main issue that the paper addresses?

- Whether the EYFSP functions as intended: as a 'reliable, valid and accurate' measure of 'child development,' that can / should be used at the level of the individual child to inform parents and teachers, and at the aggregate level to form a national dataset to be used in policy-making.

What are the main insights that the paper provides?

- EYFSP 'GLD' assessments are systematically predicted by children's birth month (thus to a large degree measuring age rather than individual 'development'). Additionally, across birth months, all children in year-groups averaging more older

children are less likely to be attributed a 'GLD.' Findings challenge the EYFSP's 'validity,' 'accuracy,' and uses.

**Keywords:** Early Years Foundation Stage Profile; relative age effects; peer effects; assessment bias

## The Early Years Foundation Stage Profile: history, aims, and use

In 2000, a national curriculum for three- to five-year olds – the Foundation Stage – was introduced, and in 2003, the Foundation Stage Profile followed: the 'first statutory assessment regime for early childhood education in England' (Bradbury, 2012). The Foundation Stage was revised in 2008 to incorporate the years from birth to five (becoming the Early Years Foundation Stage [EYFS]), and again in 2012, with associated amendments to the Early Years Foundation Stage Profile (EYFSP) assessment (Moss 2014). Currently, the curriculum and assessment are being revised once more, with 'early adopter' schools trialling a new framework from September 2020 (Department for Education [DfE], online; 2020a; 2020b).

The EYFSP is a teacher assessment, submitted at the end of the first year of primary school. The vast majority of children in England are educated in age-grouped cohorts that correspond to the structure of the academic year, which runs from September to August: so, at the point that the EYFSP is competed, September-born children are nearly six, while summer-born children are turning five. Teachers are told to judge each child against the 'national standards' prescribed in the EYFS framework, and to denote whether criteria for these 'standards' are met. Each child is allocated scores on subscales for different 'areas of development,' a summed total score, and an attribution of having 'achieved a good level of development' ('GLD') – or not. (DfE, 2017.)

The component scales and scores that are combined to construct the prescribed 'GLD' have altered over the years with revisions to the EYFSP: currently, children marked with a score of 2 or more (the 'expected level'; scale: 1-3) across criteria covering the areas of communication and language; physical development; personal, social, and emotional development; literacy; and mathematics, are denoted thus (Department for Education, 2017). Throughout these iterations, the 'GLD' has continued as a key target and metric, with teachers describing how, 'It's all about who's going to achieve the GLD…the ones who may make it…' (Roberts Holmes and Bradbury 2016, p126). It has consistently been reported at the aggregate level by the Department for Education (DfE), as shown in Figure 1, again supporting the primacy of the measure. The proportion of children denoted nationally as reaching a 'GLD' has risen sharply from 51.7% in 2013 to 71.8% in 2019, but the average total point score, summed from the component scales of the EYFSP, has increased very little, from 32.8 to 34.6 (max=51). This suggests a shuffling of sub-scale scores into the configuration required to meet the criteria for the crucial 'GLD,' once more indicating that the 'GLD' is the focus in scoring.

England, 2013 to 2019

| | % achieving good level of development | % achieving at least expected in all 17 early learning goals | Average total point score (out of 51) |
|---|---|---|---|
| 2013 | 51.7 | 48.9 | 32.8 |
| 2014 | 60.4 | 58.0 | 33.8 |
| 2015 | 66.3 | 64.1 | 34.3 |
| 2016 | 69.3 | 67.3 | 34.5 |
| 2017 | 70.7 | 69.0 | 34.5 |
| 2018 | 71.5 | 70.2 | 34.6 |
| 2019 | 71.8 | 70.7 | 34.6 |

**Figure 1: Department for Education national statistics for EYFSP scores (Department for Education, 2019, p5)**

The research reported in this paper analyses patterns in attributions of a 'GLD' using national EYFSP data spanning 2008 to 2018. Figure 2 is taken from the DfE's guidance for the last of these years, and represents the Department's conceptual framework regarding: 'How an EYFSP is completed' (Department for Education, 2017). It illustrates that the manifest skills of each child are filtered through both teacher judgements and through 'moderation' processes.

The DfE states that the aims of the EYFSP are manifold, though primarily, 'The main purpose of the EYFS profile is to provide a reliable, valid and accurate assessment of individual children at the end of the EYFS.' The EYFSP should also: 'inform parents about their child's development,' and provide 'an accurate national data set relating to levels of child development.' This aggregate data is used for policy-making 'both nationally and locally' (DfE, 2017, p10), and national and area-level figures on proportions of children designated as having reached a 'GLD' are reported annually in the Department's statistical releases. These releases are designated National Statistics, 'assessed as fully compliant' with the Code of Practice for Statistics, on the basis that, 'they meet the highest standards of trustworthiness, quality and value' (Office for Statistics Regulation, online).

In 2010, the DfE published a report analysing predictors of EYFSP assessment results, noting 'children born later in the school year' (i.e. summer-borns) as tending to receive lower scores (DfE, 2010a, p30). The Department also acknowledges this disparity in its most recent consultation (DfE, 2020c, p27). However, the EYFSP does not incorporate age adjustment, or differentiation of criteria according to children's age, despite the year's difference between the youngest and oldest reception children.

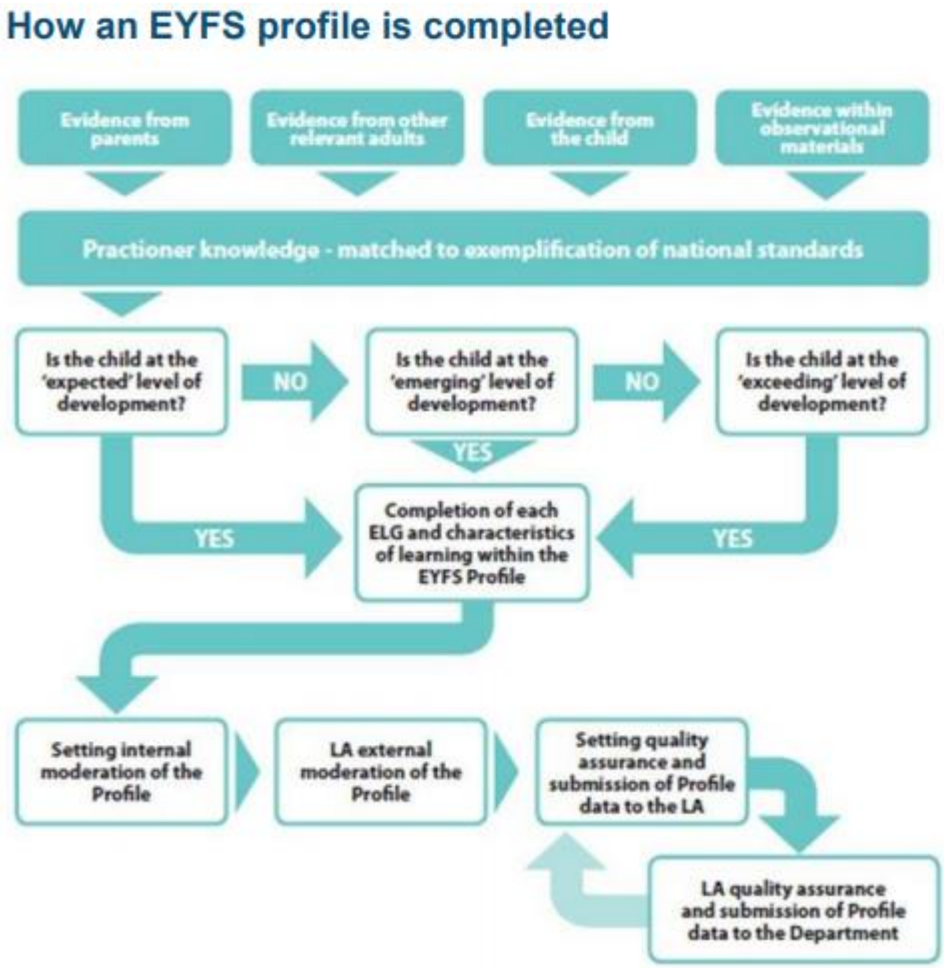


**Figure 2: Department for Education guidance / conceptual framework on completion of the EYFSP (Department for Education, 2017, p11)**

## Month of birth effects and children's ages

### *Individual birth month / age*

There is a longstanding and comprehensive body of English and international evidence that demonstrates children who are relatively younger within year-group are scored perennially lower on assessments throughout their school career (Sharp et al, 2009), and that, at the aggregate level, this is predominantly because younger children are less developed, and that their skills have not yet progressed to the same stage as those of their older peers (Crawford et al, 2013a; 2014). Evidence, including a report published by the DfE itself, also shows consistently that these month of birth effects are most pronounced when children are younger, and the gap in age within a cohort is proportionally largest (Crawford et al, 2007; DfE, 2010b; Sharp et al, 2009).

When assessments of young children are age-standardised, when children are measured at the same absolute age, and when children's skills are examined outside of the school system, structural month of birth patternings and discontinuities in line with the shape of the school year fall away: fundamentally there is a general, incremental increase in skills and therefore scores as children grow older (Crawford et al, 2013a; 2014). The skills of four-year-olds are on average less mature and developed than those of five-year-olds.

### *Peer birth month / age*

Thus the relationship between a child's own birth month and their chances of being awarded a 'GLD' is straightforward: younger children have developed for less time. But is it only a child's own age and birth month that determine their chances, or do the ages of their peers matter too? Might the age constitution of a year-group – whether the reception year is older on average, and contains more children born towards the autumn end of the year, or younger on average, with more children born later, towards the summer end– make a difference?

The previous evidence supports the likelihood that peer age composition will matter, but it falls into two camps: that indicating more older peers can increase children's chances of being attributed a 'GLD,' and that indicating the opposite. Evidence informing these two conflicting possibilities is examined in turn below.

### *Possibility one: More older peers can be beneficial for children's development, so an older year-group will boost chances of 'GLD' attribution*

The first body of evidence comes from quantitative (quasi-)experimental studies using standardised tests and validated research measures to explore relationships between peer group age and young children's development. This literature suggests that being situated with more older and fewer younger peers is beneficial – and should therefore elevate the likelihood of reaching a 'GLD.'

As noted by Cascio & Shanzenbach (2007; p1), this is congruent with 'the broader peer effects literature,' which documents 'positive spillovers from higher scoring peers.' Given that older children, by virtue of their age, are more skilled and developed, more older peers should facilitate development: scaffolding learning, and boosting progress. Correspondingly, studying five- and six-year old Kindergarten children in the US, Cascio & Shanzenbach find that children in classes where the mean age of their classmates is higher have elevated scores in standardised tests taken at the end of the Kindergarten year.

Researching slightly younger three- to four-year-old US children, but also using individual-level standardised tests as outcome measures, Ansari et al (2016, p53) find correspondingly that 'four-year-olds displayed fewer gains in academic skills […] when they were enrolled in classrooms with more three-year-olds.' With cognitive tests administered as part of their study, Moller et al (2008) also analysed how scores were associated with whether children were with more younger peers, and like Ansari et al found a negative relationship. Guo et al (2014) again examined US pre-schoolers, using scores on an established test of vocabulary. They found that the presence of more older peers was associated with higher scores for younger children, and highlight, 'the importance of children's social interactions with more knowledgeable conversational partners in promoting their vocabulary development' (p1016).

Borbely et al (2021) use national data and exploit class size rules in Scotland that result in 'as-if random' allocation meaning that some five-year-old P1 (reception) children are taught in composite classes with six-year old P2 peers, and others with younger four/five-year-old P1 classmates. Examining assessments at the end of P1 informed by online standardised tests (https://standardisedassessment.gov.scot/), they find that children who spend their first year of school with older peers score at a higher level, with 'larger and more pronounced peer effects for literacy than for numeracy' (p29). Employing a similar research design within a comparable institutional context, where some children are assigned to composite classes, Barbetta et al (2019) find for seven-year-old Italian second-graders 'a positive multigrade effect for younger children in the class and some possible negative multigrade effects for older children' (p4), in terms of maths and language standardised test scores.

Therefore, these international studies of young children in pre-school and the early primary years, who span the ages of reception pupils in the UK, suggest that being with a relatively older peer group is associated positively with development. Therefore, assuming the EYFSP to be a 'reliable, valid, and accurate' instrument at the individual level, as the DfE intends, more older peers should increase the chances of a child reaching and being attributed a 'GLD.'

### *Possibility two: More older peers will depress children's likelihood of being awarded a 'GLD'*

However, an alternative body of evidence seems to predict, on the contrary, that more older peers may depress children's chances of 'GLD' attribution. This evidence comes from work considering outcomes represented by in-school teacher judgements and assessments – like the EYFSP – rather than by standardised research instruments.

Elder & Lubotsky (2009) explicitly contrast relationships between the peer age composition experienced by US Kindergarten children and their outcomes as measured, on the one hand, by reading and maths cognitive tests, and on the other, by diagnoses with 'learning disabilities' such as ADHD, and grade retention. These latter measures result in part from in-school teacher judgements and decisions, rather than from standardised assessment. They report positive effects of older peers on reading and maths test scores, but negative effects on diagnoses and retention. Having more older Kindergarten peers leads to an increased likelihood of attribution of ADHD and of being held back a grade.

These findings echo studies using alternative measures to consider the impact of average peer skill-level on in-school teacher assessments. Meissel et al (2017) examine a sample of eight-to-13 year-olds in New Zealand, where provision of 'overall teacher judgements' of each child is a national requirement. Very similarly to the EYFSP in England, 'judgements [are] commonly referred to as national standards, and are intended to reflect a student's achievement in relation to the standard expected of students at the same year level nationally' (p49). Meissel et al exploit information for each child on performance in concurrent standardised tests, and compare individuals in classrooms with different average test scores. They find that, overall, equivalently-scoring children whose peers score higher on the standardised tests tend to receive lower teacher judgements. Because absolute age is related directly and incrementally to children's skills, this study suggests that children in schools with more older (and therefore more skilled) peers may have lower chances of being denoted by their teacher to have reached a 'GLD' at the EYFSP.

Calsamiglia & Loviglio (2019) perform a somewhat similar study of children in Catalonia, using administrative data to compare the teacher-assessed grades of children in classes with varying class-average performance on externally-marked tests. Like Meissel et al, they find that equivalently-scoring children in classes with higher average test scores receive lower assessments by their teacher. Calsamiglia & Loviglio conclude that, 'teachers grade on a curve, leading to negative peer effects,' and argue that: 'This puts forth a source of distortion that may arise in any system that uses internal grades to compare students across schools' (p1). Lastly, Crawford et al (2013b) examine national Key Stage 1 teacher assessments in English schools, and find that, controlling for a child's own age, being ranked by age in a relatively lower position within school year-group (and therefore having more older year-group peers) is negatively related to scores.

Moreover, during EYFSP scoring, children's actual skills are filtered not only through teacher (relative) judgements but also through 'moderation' processes. As shown in Figure 1, the DfE requires an iterative process of 'moderation' involving the school, the local authority, and the Department itself. Though its guidance states that the scores attributed to each child are required to be based on 'practitioner knowledge,' which 'isn't often formally recorded,' (DfE 2017, p14), there is also a requirement that this knowledge and resultant attributions to the child must 'be subject to moderation to ensure national consistency and accuracy' (p17).

Given that there is no concrete, stable, existent entity to be moderated, it is unclear how this process is reliably to be enacted. Bradbury (2011; 2012) has researched the implementation of EYFSP 'moderation' using in-depth qualitative case studies, and describes explicitly how the process forces teachers to alter scores to 'fit into a bell curve' (p663; 2011) – echoing and providing context in England for Calsamiglia & Loviglio's conclusion that 'teachers grade on a curve.' If Bradbury's teachers did not adhere to a prescribed, pre-determined patterning of results, according to the characteristics of their pupils, they were 'told off' (p663, 2011), and required to complete the scoring process once more. This resulted in tactical compliance and deliberate, conscious amendment of scores in order to fit the distribution expected for their school.

Bradbury & Robert-Holmes (2017) also describe how patterns of scores on the EYFSP inform both Ofsted judgements and decisions to subject a school to inspection, resulting in 'pressure to deflate early years assessments in order to provide a lower baseline for measures of progress' (p 947), which is necessary to fulfil Ofsted's required 'narrative of progress.' Within this structure and according to the process of 'moderation' into a 'bell curve' at the school level, then mechanically, if an individual child is in a year-group that averages more older children (who will be more likely to be attributed a 'GLD' due to their higher skills and maturity) then that individual child will have lower chances of attribution, as they will be 'pushed' towards the lower part of the curve, or across the threshold for the 'quota' of 'GLD's.

The evidence on in-school assessment processes therefore predicts that being in a year-group with more older children may in fact lessen an individual child's chances of being attributed a 'GLD.' Additionally, qualitative and theoretical research synthesising and drawing from the international evidence on month of birth effects suggests that peer age-group composition may have different effects depending on a child's own age, within an environment that is not necessarily equitably accessible and facilitating for relatively older and younger children. Sharp et al (2009) posit situations where, 'younger children cannot access a curriculum aimed at older children…experience failure or stress, or compare themselves unfavourably to older classmates' (p2). If a summer-born child is in an older year-group, these processes may therefore amplify and extend the effects of their absolute age. The chances of a summer-born child proceeding to a 'GLD' might be depressed if they are situated with more autumn-born peers, to whom the expectations of the EYFS may be better suited. Conversely, an autumn-born child with more autumn-born peers might thrive, if the environment is better tailored to their age and stage.

### *Predictions and uncertainties*

Previous research thus indicates unambiguously that birth month (i.e. absolute age) will strongly be related to children's likelihood of being attributed a 'GLD,' simply because children's skills develop as they age, and because September-born reception children have lived and grown for a year longer than August-borns who are not yet five. That this will be the case in the current study is a certain prediction.

But the bodies of evidence informing the possible direction of effect of having relatively more older or more younger reception year-group peers produce predictions seemingly at odds. On the one hand, *if* the EYFSP directly measures the skills of an individual child, as intended by the DfE (and assuming an equitably accessible educational environment), then having more older peers is likely to be positively associated with being recorded as reaching a 'GLD' – because older peers facilitate and scaffold key areas of development, such as language, that the EYFSP is intended to measure.

However, conversely, research showing in-school teacher judgements and assessments to be relative and contextualised suggests that more older peers may decrease the probability of 'GLD' attribution for an individual. Denotation may be dependent on context, and the same child might have different chances of 'GLD' ascription depending on the distribution of ages in the reception year-group in the school they attend. Alongside this, studies on the nature and practice of 'moderation' of EYFSP teacher judgements indicate that the process may compound the impact of relative judgements, as it requires scores within school to fit an expected distribution. Year-group age composition will to some extent determine place in this distribution, and whether an individual child is attributed a 'GLD.' Finally, it is also possible that the effects of peer age will depend on whether a child themselves is older or younger within the cohort: an autumn-, spring-, or summer-born.

**The current study**

The current study firstly comprehensively updates the evidence on month of birth and the EYSFP, by mapping patterns by birth month in attribution of a 'GLD' across the decade from 2008 to 2018, through revision of the assessment. Secondly, it presents novel evidence of the relationship between the age constitution of a child's school year-group peers and that individual's likelihood of being awarded a 'GLD.' Thirdly, each child's own age is considered in combination with the mean age of their schoolmates, to allow for differential effects of peer age according to whether a child themselves is autumn-born, spring-born, or summer-born.

**Data and analytical strategy**

The National Pupil Database (NPD), which includes records for all children in statutory state-funded education in England, is used for this research. Permission for use was given after application to the DfE, and remote access was provided through the ONS's Secure Research Service. De-identified Foundation Stage Profile and Spring School Census records for children with recorded EYFSP scores in the summers of 2008 to 2018 are considered, to compare patterns over years. Information on scores, month and year of birth, and school attended is drawn from the Foundation Stage Profile dataset; information on child characteristics (whether recorded as eligible for free school meals [FSM], recorded ethnicity, home language [EAL], and gender) is matched from the Spring School Census.

The eleven main samples comprise all children who are in a school reception year-group in a mainstream school, in the summer of the given year, whose birthday falls into the range for the cohort (September of the previous calendar year – August of the EYFSP calendar year), who are in schools with 12 or more in-cohort reception children (including themselves), and where no out-of-cohort children are recorded as also being present and receiving an EYFSP score that summer. Thus the sample is limited to schools where only in-cohort, 'normal'-year-group children with the 12-month range of births from September (oldest) to August (youngest) are present. A minimal number of schools with infeasible age distributions that seem likely to be due to recording error (e.g. 30 February-borns) are removed from each sample. All retained schools then have a distribution of pupils across the year: there are none with only autumn (September-December), spring (January-April), or summer (May-August)-borns.

In quantifying (mean) age for analyses, month-within-year is the focus. September is coded as 12, October 11, and so on, through to August, as 1 - because September-borns are the oldest and August-borns the youngest within each school year-group, and all samples here fall within the 12-month, in-cohort range. This coding approach is taken for two reasons: firstly because, though the EYFSP assessment is reported in June of the Reception year, it is informed by teachers' judgements throughout the year (Department for Education, 2017) – so using children's absolute age at June would suggest false precision. Secondly, using the in-cohort range of months of 12 (oldest) through to 1 (youngest) emphasises the relative nature of age-within-cohort, and that it is not just absolute age that may be important. For information, however, note that the range of absolute ages of sample children in each reception year, on June 1st, is four-and-nine-months (a child born in late August) through to five-and-nine-months (a child born in early September).

As can be seen in Table 1, column i), the average age, coded as above, of all children in the main samples ranges therefore from 6.44 to 6.57. This around the mid-point of the academic year and the range of ages coded 12-1. The age constitution of children's peers is calculated at the school-level by taking the mean age of all those in the cohort year-group, minus the child themselves, and the average mean age of peers ranges in the main samples also from 6.44 to 6.57. Columns h) and i) in Table 1 show that the average age of children in each year's EYFSP cohort is slowly increasing, though it is not clear whether this is due to growing numbers of summer-borns deferring reception entry, or a shift in birth timings. This is interesting in itself, but not the focus here, and it will be investigated in a separate paper, along with other changing patterns.

To identify the independent influence of peer age composition on children's chances of being attributed a 'GLD,' it is important to account for confounding factors that may influence both composition and 'GLD' attribution. Table 2 shows the mean age of peers calculated for children born in each month, separately, and indicates a small amount of clustering where older children are generally very slightly more likely to be with older peers. This probably results from structural factors in the pre-school years and the transition from pre-school to primary school; see Campbell et al (2019) for discussion. However, this clustering is extremely minimal, so unlikely to cause confounding or reflect confounding factors. Crucially, Table 3 shows that, every year, there are children born in each birth month across the full range of peer constitutions. Figure 3 illustrates this by showing the distribution of mean peer ages experienced by children in the main samples in reception in 2008, 2013, and 2018. Within each cohort, there is a largely symmetrical distribution of mean peer age across children born within each month, with clustering around the middle of the school year – suggesting that variation 'naturally' arises mainly by chance. This spread allows investigation of absolute age and peer age constitution as independent factors that may each have an influence on attribution of a 'GLD.'

**Table 1: Numbers of children and schools in each year's main sample; average ages of children and school peers, and standard deviations around average peer age**

| a) EYFSP year | b) Birth cohort | c) Total N with EYFSP 'GLD' info in cohort | d) N in main sample | e) N schools in main sample | f) N pupils in schools, main sample: range | g) N pupils in schools, main sample: mean (median) | h) Average age of children with EYFSP 'GLD' info in cohort | i) Average age of cohort – main sample | j) Average mean age of peers – main sample | k) Average SD of mean age of peers – main sample |
|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 2002-2003 | 555,704 | 492,232 | 13,061 | 12-173 | 47.9 (45) | 6.39 | 6.44 | 6.44 | 0.60 |
| 2009 | 2003-2004 | 572,418 | 519,469 | 13,379 | 12-183 | 49.3 (47) | 6.42 | 6.48 | 6.48 | 0.58 |
| 2010 | 2004-2005 | 585,458 | 545,195 | 13,684 | 12-185 | 50.7 (50) | 6.43 | 6.48 | 6.48 | 0.58 |
| 2011 | 2005-2006 | 597,529 | 560,776 | 13,796 | 12-186 | 51.8 (52) | 6.40 | 6.45 | 6.45 | 0.57 |
| 2012 | 2006-2007 | 619,471 | 582,141 | 13,864 | 12-213 | 53.8 (55) | 6.41 | 6.46 | 6.46 | 0.56 |
| 2013 | 2007-2008 | 644,398 | 603,581 | 13,940 | 12-228 | 55.7 (58) | 6.43 | 6.48 | 6.48 | 0.55 |
| 2014 | 2008-2009 | 642,110 | 602,999 | 13,984 | 12-240 | 55.6 (57) | 6.45 | 6.49 | 6.49 | 0.55 |
| 2015 | 2009-2010 | 655,440 | 615,096 | 14,010 | 12-241 | 56.5 (58) | 6.47 | 6.51 | 6.51 | 0.55 |
| 2016 | 2010-2011 | 669,280 | 615,497 | 13,887 | 12-241 | 57.1 (58) | 6.49 | 6.53 | 6.53 | 0.55 |
| 2017 | 2011-2012 | 669,446 | 589,310 | 13,433 | 12-239 | 56.6 (58) | 6.48 | 6.52 | 6.52 | 0.54 |
| 2018 | 2012-2013 | 650,632 | 523,846 | 12,370 | 12-241 | 54.3 (56) | 6.53 | 6.57 | 6.57 | 0.57 |

Note: September=12, October=11,…July=2, August=1.

**Table 2: Average mean peer age experienced by children born in each month: main samples**

| | Aug (1) | Jul (2) | Jun (3) | May (4) | Apr (5) | Mar (6) | Feb (7) | Jan (8) | Dec (9) | Nov (10) | Oct (11) | Sep (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 6.43 | 6.43 | 6.44 | 6.44 | 6.44 | 6.44 | 6.45 | 6.45 | 6.45 | 6.46 | 6.45 | 6.46 |
| 2009 | 6.47 | 6.48 | 6.47 | 6.48 | 6.48 | 6.48 | 6.49 | 6.48 | 6.49 | 6.48 | 6.49 | 6.49 |
| 2010 | 6.47 | 6.47 | 6.47 | 6.47 | 6.48 | 6.48 | 6.48 | 6.48 | 6.49 | 6.49 | 6.48 | 6.48 |
| 2011 | 6.44 | 6.45 | 6.44 | 6.45 | 6.45 | 6.45 | 6.44 | 6.45 | 6.46 | 6.45 | 6.46 | 6.45 |
| 2012 | 6.46 | 6.45 | 6.46 | 6.46 | 6.46 | 6.46 | 6.47 | 6.46 | 6.47 | 6.47 | 6.47 | 6.47 |
| 2013 | 6.47 | 6.47 | 6.47 | 6.47 | 6.47 | 6.48 | 6.48 | 6.48 | 6.48 | 6.48 | 6.48 | 6.49 |
| 2014 | 6.49 | 6.49 | 6.49 | 6.49 | 6.49 | 6.49 | 6.49 | 6.50 | 6.50 | 6.50 | 6.50 | 6.50 |
| 2015 | 6.50 | 6.50 | 6.50 | 6.51 | 6.50 | 6.51 | 6.51 | 6.51 | 6.51 | 6.52 | 6.52 | 6.52 |
| 2016 | 6.53 | 6.53 | 6.53 | 6.53 | 6.53 | 6.53 | 6.53 | 6.53 | 6.54 | 6.54 | 6.54 | 6.54 |
| 2017 | 6.51 | 6.51 | 6.51 | 6.51 | 6.51 | 6.52 | 6.52 | 6.52 | 6.51 | 6.52 | 6.52 | 6.51 |
| 2018 | 6.56 | 6.56 | 6.56 | 6.57 | 6.56 | 6.56 | 6.57 | 6.57 | 6.58 | 6.58 | 6.58 | 6.58 |

Note: September=12, October=11,…July=2, August=1.

**Table 3: Standard deviation around average mean peer age experienced by children born in each month: main samples**

| | Aug (1) | Jul (2) | Jun (3) | May (4) | Apr (5) | Mar (6) | Feb (7) | Jan (8) | Dec (9) | Nov (10) | Oct (11) | Sep (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 0.60 | 0.59 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.61 |
| 2009 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.58 | 0.59 | 0.59 | 0.58 |
| 2010 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.58 | 0.59 | 0.58 | 0.59 | 0.58 | 0.59 |
| 2011 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 2012 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.57 | 0.57 |
| 2013 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.56 | 0.55 | 0.55 | 0.56 |
| 2014 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 |
| 2015 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 |
| 2016 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.54 | 0.54 | 0.54 | 0.54 | 0.55 |
| 2017 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 |
| 2018 | 0.56 | 0.57 | 0.57 | 0.56 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |

Note: September=12, October=11,…July=2, August=1.

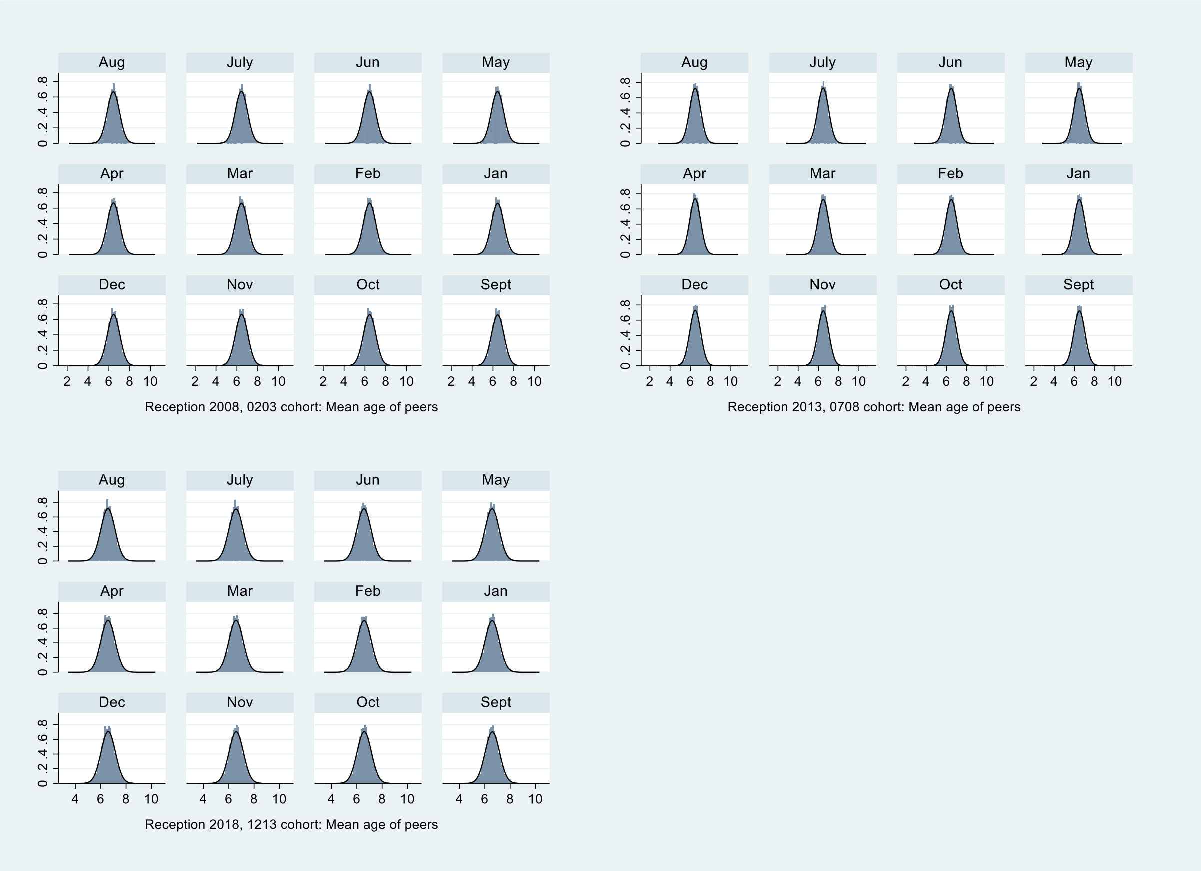


**Figure 3: Distribution of mean peer ages experienced by children born in each month: main samples, Reception 2008, 2013, 2018.** Note: September=12, October=11,…July=2, August=1

That the age constitution of a reception year-group should result largely from 'natural' variation makes sense: firstly because mothers have limited control over the month in which their child is born. Additionally, though parents can have some volition over birth planning, it is more difficult to see effective mechanisms within the English schools admissions system by which they can deliberately control to any meaningful extent the age constitution of their child's peers. Some threat to the independence of other factors from peer composition does exist however if enough mothers manage to time births *and* if other characteristics of these families correspond both to achieved month of birth and to school selection decisions. Of course, this is also a threat to concluding that birth month variation in attribution of a 'GLD' arises solely from children's maturity and absolute age, rather than from related characteristics. If, for example, parents who aim for and attain autumn birthdays also apply for certain schools, and are similar in other ways, this would lead these schools to have relatively older peer groups that are also selected on covarying factors that may influence children's scores, those of their peers, and interrelated patterns in attribution of the 'GLD.'

Additional analyses (t-tests and ANOVAs; not shown here but available on request) suggest generally minor bi-variate correspondences between the background factors available in the data – recorded ethnic group, whether languages other than English are spoken at home (EAL), whether the child receives free school meals (FSM) – and month of birth. Gender is never significantly related. When there are statistically significant relationships, they are usually substantively small, and they are inconsistent across cohorts, suggesting that some may have arisen by chance in these multiple large samples. However, in order to guard against any possibility that these key characteristics, which are all strongly related to EYFSP scores (see, for example, DfE [2010]), may account for any of the patterns by birth month or peer age constitution, results here firstly present simple descriptive findings, then use logistic regressions to control for the potential influence of the three factors.

For each year-group, six separate cumulative logistic regression models are estimated (Table 4). The outcome is being attributed a 'GLD' (0/1).The key predictors are: child's birth term (autumn: September-December; spring: January-April; summer: May-August), entered alone in specification one, and mean peer age (entered alone in specification two). These two predictors are then entered together in specification three, in order to ascertain whether each is related to attribution of a 'GLD' controlling for the other. They are then interacted in specification four, to explore whether one moderates the other. Model five then adds controls for child's own ethnicity, child's own EAL status, whether the child receives FSM in reception. Finally, model six adds controls for number of children in the school reception year-group, proportion White British children in the year-group (proportions of alternative ethnic groups are also checked in additional models, available on request), proportion EAL children, and proportion FSM children. Standard errors are clustered at the school level in all specifications.

**Table 4: Variables included in each logistic regression model specification (outcome is always attribution or not of an EYFSP 'Good Level of Development')**

| | **Specification** | | | | | |
|---|---|---|---|---|---|---|
| **Predictors** | 1 | 2 | 3 | 4 | 5 | 6 |
| Term of birth (Autumn / Spring / Summer) | x | | x | x | x | x |
| Mean age of peers | | x | x | x | x | x |
| Term of birth*mean age of peers | | | | x | x | x |
| Child-level factors: ethnicity, English as an additional language (EAL), Free School Meals (FSM) | | | | | x | x |
| School-level factors: N children in school, proportion White British children, proportion children with EAL, proportion children recorded FSM | | | | | | x |

## Results

Figure 4 shows the proportion of children born in each month within the main samples who are attributed a EYFSP 'GLD' each year. As expected, in line with previous research, the linear incremental birth month gradient is stark. Though there is inflation overall of the proportion of children denoted as reaching a 'GLD' from 2008 to 2012, and then again from 2013, after introduction of the new version of the EYFSP, the August-September gap remains consistently pronounced.

In 2008, 32% of August-borns were attributed a GLD, compared to 65% of September-borns; in 2013, 35% of August-borns compared to 67% of September-borns; and in 2018, 57% of August-borns, compared to 84% of September-borns. This suggests that to a significant extent the 'GLD' measure simply reports children's age, rather than whether their individual development follows any trajectory 'expected' by the EYFSP.


**Figure 4: Proportion of main sample children each year attributed a 'Good Level of Development': by birth month**

These birth month differences are not unexpected, given previous research: they were predicted with certainty. But they are worth emphasising, not least because they are not reported in the text of the DfE's National Statistical Releases, despite being greater than gaps according to other key characteristics documented therein – including those by gender, which continues to be described every year (Department for Education, 2019). In the main samples here, in 2008, 41% of boys were denoted as reaching a 'GLD' compared to 58% of girls; in 2013, 44% vs 60%; and in 2018, 66% vs 79%: gaps far smaller than those by birth month.

Figure 5 goes on to show the correspondence between the mean age of reception peers experienced by children and attribution to the individual child of a 'GLD.' Mean peer age is split categorically by standard deviation (SD) within each respective year's main sample. This is for consistency, because the absolute values for mean peer age vary slightly by year (Table 2). Children falling at the bottom of the distribution (with peers who are much younger than the average within the whole cohort), more than one SD below, are included in the first category; those falling from one SD below to the average (who thus have peers somewhat younger than the average) the next; those with mean peer ages falling from the average to one SD above (who have somewhat older peers) the next, and those with peers whose mean age falls more than one SD above the average, and whose peers are therefore oldest, the last. While differences here are less pronounced than those due to a child's own age, they are consistent, and again, they are linear: each year, children situated in schools averaging relatively more older peers are less likely to be recorded as attaining a 'GLD.' In 2008, 52% of those with the youngest peers were attributed a 'GLD,' compared to 48% of those with the oldest peers; in 2013, 54% vs 51%, and in 2018, 73% vs 71%.


**Figure 5: Proportion of children with more younger(left bars) vs more older (right bars) peers attributed a 'Good Level of Development' Note: The distribution is split by standard deviation within each year: the far left bar represents children with the youngest peers, (min/-1); the middle-left bar those with somewhat younger peers (-1/0); the middle right bar those with somewhat older peers (0/1); and the far right bar those with the oldest peers (1/max).**

But do these linear relationships interact with one another? Does the association between peer age constitution and denotation of a 'GLD' depend on a child's own age? Previous research suggests that more older peers may result in a beneficial environment for older children, while inversely being detrimental to younger children, or vice versa – so it is important to consider the interaction.  It is also important to control for other observable factors that may account for patterns by age or peer mean age.

Firstly, Table 5 therefore shows log odds from six logistic regression specifications for each year's sample, conveying the direction of relationships between the predictors detailed in Table 4 and odds of attribution of a 'GLD.' It also denotes the precision of these estimates (proxied by statistical significance). Specification 1 simply shows the relationship of a child's own birth term to their chances of being attributed a 'GLD' – as above, each year, spring-borns are less likely and summer-borns much less likely than autumn-borns. Specification 2 shows the same for mean age of peers, by itself – again, having more older peers is related to lower odds of attribution of a 'GLD.' When both a child's own birth term and the mean age of their peers are included simultaneously, at Specification 3, both remain related to odds of a 'GLD.'

In Specification 4, term of birth is interacted with mean peer age. This examines whether the impact of having more older year-group peers varies according to whether a child themself is summer, spring, or autumn-born. In several years there is an interaction where the relationship between having more older peers and lowered chances of a 'GLD' is more pronounced for summer-borns than for autumn-borns. However, this not always evident, and is statistically significant at the chosen 5% level only within the 2008 and 2017 cohorts, so it is not a precise nor consistent finding. What instead is consistently indicated is a substantial, significant relationship in the same direction across autumn-borns, spring-borns, and summer-borns every year. It is not the case that an older year-group is negatively related to chances of 'GLD' attribution only for younger children: more older peers predict lessened chances of a 'GLD' regardless of a child's own age.

Specification 5 controls for observable child-level factors that may feasibly confound relationships (FSM, EAL, and ethnicity). Overall, neither the estimates of the difference between summer-borns, spring-borns, and autumn-borns nor the relationships between mean peer age and the 'GLD' are attenuated by these factors. Specification 6 adds controls at the school-level: proportion FSM children, proportion EAL children, proportion White British children, number of children in the year-group. Again, overall, these factors do not attenuate associations by birth term or average peer age. Together, these specifications indicate both a child's own age and the age of their peers as being substantial independent predictors of attribution of a 'GLD.'

**Table 5: Relationships between birth term, mean age of peers, and odds of attribution of a 'Good Level of Development'**

| | Spec 1 | Spec 2 | Spec 3 | Spec 4 | Spec 5 | Spec 6 |
|---|---|---|---|---|---|---|
| EYFSP 2008 (N= 492,232) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.49*** | | -0.49*** | -0.47*** | -0.49*** | -0.49*** |
| Summer | -0.99*** | | -1.00*** | -0.69*** | -0.75*** | -0.78*** |
| Mean peer age (range: 2.2-10.5) | | -0.09*** | -0.10*** | -0.08*** | -0.07*** | -0.06*** |
| Spring*mean age peers | | | | -0.00 | -0.00 | -0.00 |
| Summer*mean age peers | | | | -0.05*** | -0.05*** | -0.04*** |
| Intercept | 0.46*** | 0.51*** | 1.09*** | 0.99*** | 1.18*** | 1.75*** |
| EYFSP 2009 (N=519,467) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.48*** | | -0.48*** | -0.37*** | -0.41*** | -0.42*** |
| Summer | -1.02*** | | -1.03*** | -0.91*** | -0.97*** | -0.97*** |
| Mean peer age(range: 2.7-10) | | -0.05*** | -0.06*** | -0.04*** | -0.05*** | -0.05*** |
| Spring*mean age peers | | | | -0.02 | -0.02 | -0.01 |
| Summer*mean age peers | | | | -0.02 | -0.02 | -0.02 |
| Intercept | 0.58*** | 0.38*** | 0.94*** | 0.87*** | 1.14*** | 1.71*** |
| EYFSP 2010 (N=545,195) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.47 | | -0.47*** | -0.47*** | -0.47*** | -0.47*** |
| Summer | -0.98*** | | -0.98*** | -0.84*** | -0.86*** | -0.87*** |
| Mean peer age(range: 2.3-9.9) | | -0.07*** | -0.09*** | -0.08*** | -0.07*** | -0.07*** |
| Spring*mean age peers | | | | -0.00 | -0.00 | -0.00 |
| Summer*mean age peers | | | | -0.02 | -0.03* | -0.03 |
| Intercept | 0.72*** | 0.71*** | 1.28*** | 1.23*** | 1.44*** | 2.01*** |
| EYFSP 2011 (N=560,776) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.46*** | | -0.46*** | -0.47*** | -0.50*** | -0.51*** |
| Summer | -0.96*** | | -0.96*** | -0.90*** | -0.95*** | -0.96*** |
| Mean peer age(range: 3.1-9.9) | | -0.07*** | -0.08*** | -0.08*** | -0.08*** | -0.07*** |
| Spring*mean age peers | | | | 0.00 | 0.00 | 0.00 |
| Summer*mean age peers | | | | -0.01 | -0.01 | -0.01 |
| Intercept | 0.85*** | 0.82*** | 1.37*** | 1.34*** | 1.58*** | 2.27*** |
| EYFSP 2012 (N=581,779) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.43*** | | -0.43*** | -0.42*** | -0.42*** | -0.42*** |
| Summer | -.090*** | | -0.90*** | -0.89*** | -0.92*** | -0.93*** |
| Mean peer age(range: 2.3-9.9) | | -0.06*** | -0.07*** | -0.07*** | -0.07*** | -0.07*** |
| Spring*mean age peers | | | | 0.00 | 0.00 | -0.00 |
| Summer*mean age peers | | | | -0.00 | -0.00 | -0.00 |
| Intercept | 1.03*** | 0.96*** | 1.47*** | 1.47*** | 1.76*** | 2.49*** |
| EYFSP 2013 (N=603,581) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.45*** | | -0.45*** | -0.32*** | -0.33*** | -0.33*** |
| Summer | -0.95*** | | -0.95*** | -0.80*** | -0.87*** | -0.88*** |
| Mean peer age(range: 2.8-10.7) | | -0.07*** | -0.08*** | -0.06*** | -0.07*** | -0.07*** |
| Spring*mean age peers | | | | -0.02* | -0.02* | -0.02 |
| Summer*mean age peers | | | | -0.02* | -0.02 | -0.02 |
| Intercept | 0.55*** | 0.52*** | 1.05*** | 0.95*** | 1.25*** | 1.85*** |
| EYFSP 2014 (N=602,999) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.47*** | | -0.47*** | -0.39*** | -0.41*** | -0.40*** |
| Summer | -0.98*** | | -0.98*** | -0.85*** | -0.90*** | -0.91*** |
| Mean peer age(range: 2.6-10.1) | | -0.08*** | -0.09*** | -0.07*** | -0.08*** | -0.09*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| Spring*mean age peers | | | | -0.01 | -0.01 | -0.01 |
| Summer*mean age peers | | | | -0.02 | -0.02 | -0.02 |
| Intercept | 0.93*** | 0.94*** | 1.49*** | 1.42*** | 1.73*** | 2.43*** |
| EYFSP 2015 (N=615,096) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.48*** | | -0.48*** | -0.42*** | -0.45*** | -0.46*** |
| Summer | -0.98*** | | -0.98*** | -0.81*** | -0.85*** | -0.87*** |
| Mean peer age(range: 2.5-9.8) | | -0.07*** | -0.08*** | -0.06*** | -0.07*** | -0.08*** |
| Spring*mean age peers | | | | -0.01 | -0.01 | -0.01 |
| Summer*mean age peers | | | | -0.03** | -0.02* | -0.02 |
| Intercept | 1.20*** | 1.13*** | 1.71*** | 1.62*** | 1.92*** | 2.58*** |
| EYFSP 2016 (N=615,497) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.49*** | | -0.49*** | -0.56*** | -0.58*** | -0.57*** |
| Summer | -0.99*** | | -0.99*** | -0.99*** | -1.05*** | -1.04*** |
| Mean peer age(range: 2.5-10) | | -0.08*** | -0.09*** | -0.09*** | -0.10*** | -0.10*** |
| Spring*mean age peers | | | | 0.01 | 0.01 | 0.01 |
| Summer*mean age peers | | | | 0.00 | 0.00 | 0.00 |
| Intercept | 1.35*** | 1.34*** | 1.92*** | 1.95*** | 2.24*** | 2.74*** |
| EYFSP 2017 (N=589,310) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -0.50*** | | -0.50*** | -0.43*** | -0.44*** | -0.44*** |
| Summer | -0.99*** | | -0.99*** | -0.77*** | -0.83*** | -0.83*** |
| Mean peer age(range: 3.3-10.5) | | -0.07*** | -0.07*** | -0.06*** | -0.08*** | -0.08*** |
| Spring*mean age peers | | | | -0.01 | -0.01 | -0.01 |
| Summer*mean age peers | | | | -0.03** | -0.03** | -0.03** |
| Intercept | 1.43*** | 1.34*** | 1.91*** | 1.80*** | 2.17*** | 2.69*** |
| EYFSP 2018 (N=523,846) | | | | | | |
| Autumn (ref) | - | - | - | - | - | - |
| Spring | -.050*** | | -0.50*** | -0.37*** | -0.39*** | -0.39*** |
| Summer | -.100*** | | -1.00*** | -0.84*** | -0.89*** | -0.90*** |
| Mean peer age(range: 3.4-10.3) | | -0.06*** | -0.07*** | -0.06*** | -0.08*** | -0.08*** |
| Spring*mean age peers | | | | -0.02 | -0.02 | -0.02 |
| Summer*mean age peers | | | | -0.02 | -0.02 | -0.02 |
| Intercept | 1.46*** | 1.34*** | 1.94*** | 1.84*** | 2.20*** | 2.70*** |

Table presents log odds. ***p<.0001, **p<.01, *p<.05. See Table 4 for details of the variables included at each specification. All specifications cluster standard errors at the school level. Note: September=12, October=11,…July=2, August=1.

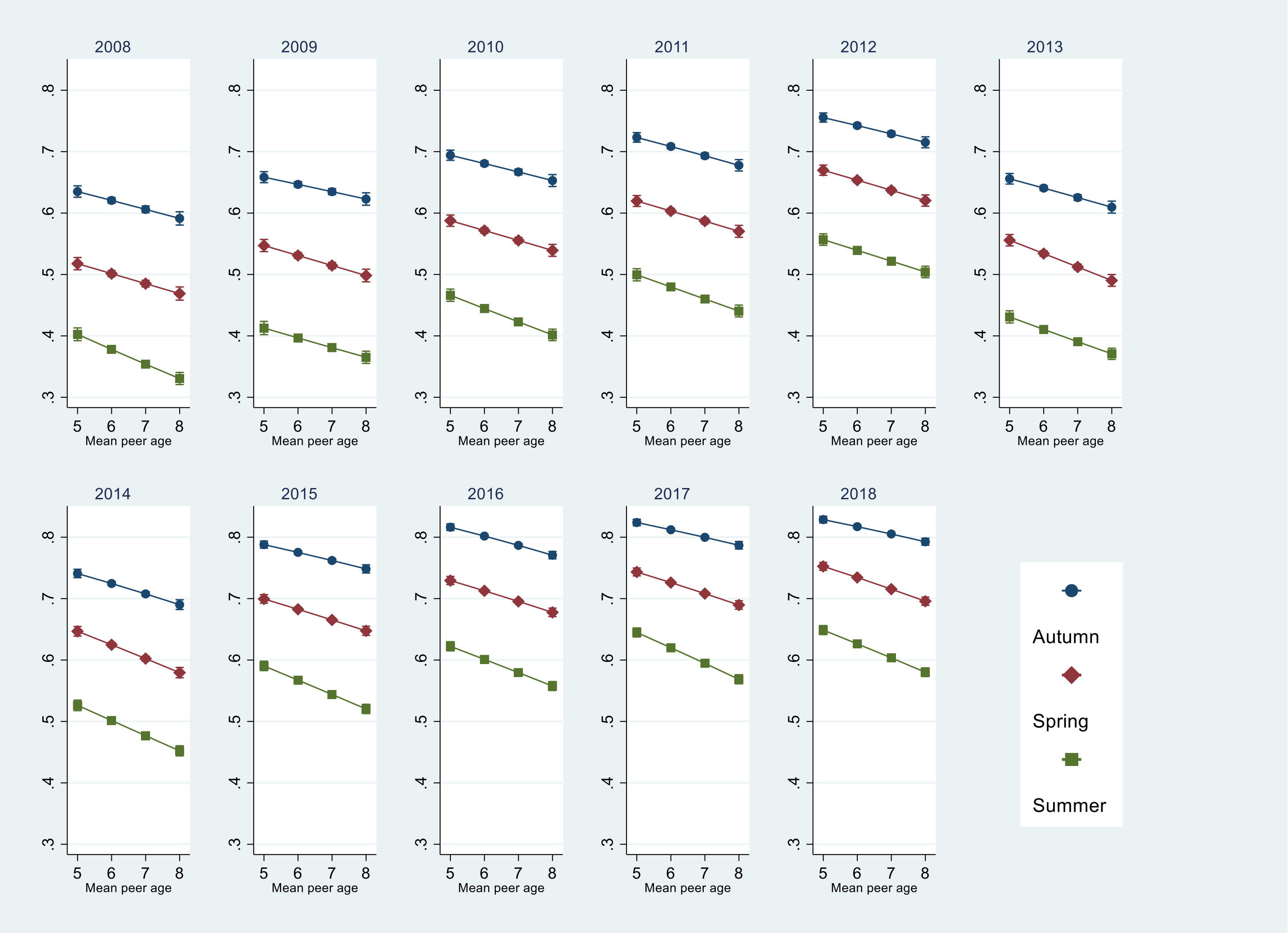


**Figure 6: Predicted probabilities of children born in each term and with each level of mean peer age being attributed a 'Good Level of Development': main samples, Specification 6**. Interpret with reference to Table 4 which also includes Ns. Whiskers are 95% confidence intervals around each estimate. Source: National Pupil Database Spring Schools Census 2008-2018; Early Years Foundation Stage Profile Results 2008-2018.
Note: September=12, October=11,…July=2, August=1.

In order to illustrate these relationships more intuitively and to provide comparisons between years, Figure 6 presents predicted probabilities from specification 6 for each year's cohort, calculated at different points within the distribution of mean peer ages. Table 1, column j) shows that the average mean peer age, as coded within cohort year from 12 (September-borns) to 1 (August-borns) across years is 6.5, and column k) showed that the average standard deviation from the mean across years is 0.56. Therefore these marginal predicted probabilities are estimated at the bottom of the distribution experienced by children (5), towards but below the middle (6), above the middle (7), and towards the top end (8). Figure 6 shows for example that, in the 2018 cohort, a summer-born situated in a much older ('8') year-group has an estimated 58% chance of being attributed a 'GLD,' while a comparable summer-born in a younger ('5') year-group at the other end of the distribution has an estimated 65% chance. An autumn-born situated in an older ('8') year-group has an estimated 79% chance, while an autumn-born in an younger ('5') cohort has an estimated 83% chance. Once more this demonstrates that both a child's own absolute age and the age of their peers contribute to their chances of being denoted as reaching a 'GLD:' considering one without the other would therefore underestimate relative age-related inequalities.

At this point it is useful to revisit the possibilities suggested by the previous research on peer age constitution. Firstly, studies using standardised instruments with pre-school and primary school pupils indicated positive effects of more older peers – effects not born out in findings so far on 'GLD' attribution. Results are instead congruent with evidence on relative judgement and assessment bias in in-school grading, and/or construction by 'moderation' and requirements of the wider system into a forced distribution, or curve.

***Additional analyses and sensitivity checks***

The main samples include schools where the mean number of reception children ranges from 47.9 (median 45) to 57.1 (median 58) and where the maximum number is 241 (Table 1, columns f) and g). It is therefore possible that classroom effects are being distorted or obscured by the inclusion of children in larger schools, who are not in the same base class group nor necessarily in sustained direct contact with one another. The regressions for specification 6 are therefore repeated for children in small schools, with 12-30 children, where proximity and shared environment can be assumed. Annex A, Figure A1 shows the distributions of mean peer age by birth month for these samples are similarly symmetrical (and wider) than those in the main samples.

Overall, across years, Figure 7 shows that there is no notable change in the pattern of relationships between term of birth, mean peer age, and children's chances of being awarded a 'GLD' when considering only these smaller schools. Associations are not different, nor more (or less) pronounced in these samples where children are more certain to interact with one another.

Further analyses, illustrated in Figure 8, use a third, less skewed set of samples (see Annex A, Figure A2 for distributions), where more extremely young (with 50%+ summer-borns) and extremely old (50%+ autumn-borns) year-groups are removed, and where children are less likely to be in direct contact, as the remaining included schools are bigger. Again, overall, patterns of results are all very similar. This supports the theory that moderation into a forced distribution or quota of children attaining a 'GLD' per school is contributing to the patterns by peer age constitution. Regardless of school size, of whether a year-group is more or less dominated by autumn-borns or summer-borns, and of whether children are more or less likely to be in proximity, the pattern of likelihood of an individual child being attributed a 'GLD' is similarly shaped by the average age of their school year cohort.

All analyses are replicated using mean year-group age constitution, including the child themselves (rather than mean peer age without the child), and all are equivalent. They are also checked and replicated having excluded those extremely rare (for example, in the 2018 cohort, N=18) schools which contain no children born in the autumn, or spring, or summer. Alternative specifications using linear probability models and substituting local authority fixed effects for clustering at the school-level are also estimated, thus comparing children within the same admissions area in order to address the possibility that bias or confounding may exist at this level: again, patterns are substantively similar. Lastly, as a final check against confounding through selection into schools, all cohorts are pooled across years, and only children in institutions with an identifier that is present at all points from 2008 to 2018 are included. This retains a stable longitudinal sample of schools that have not changed governance (and therefore identifier), for example as a result of academisation. School fixed effects are used again in linear probability models to exploit variation from year to year in the age constitution of children attending reception, while controlling for other unmeasured school factors that are common across years. Once more, results are consistent with the main analyses.

Results are robust across all these alternative specifications, as well as for all years – despite some changes over time to underlying patterns, including average pupil and mean peer age, overall proportions of children denoted as reaching a 'GLD,' amendment of the EYFSP measures, and associations of other factors to birth month and peer age constitution. Given also that this research uses the NPD, which is a census, there can therefore be a high level of confidence in the stability of findings

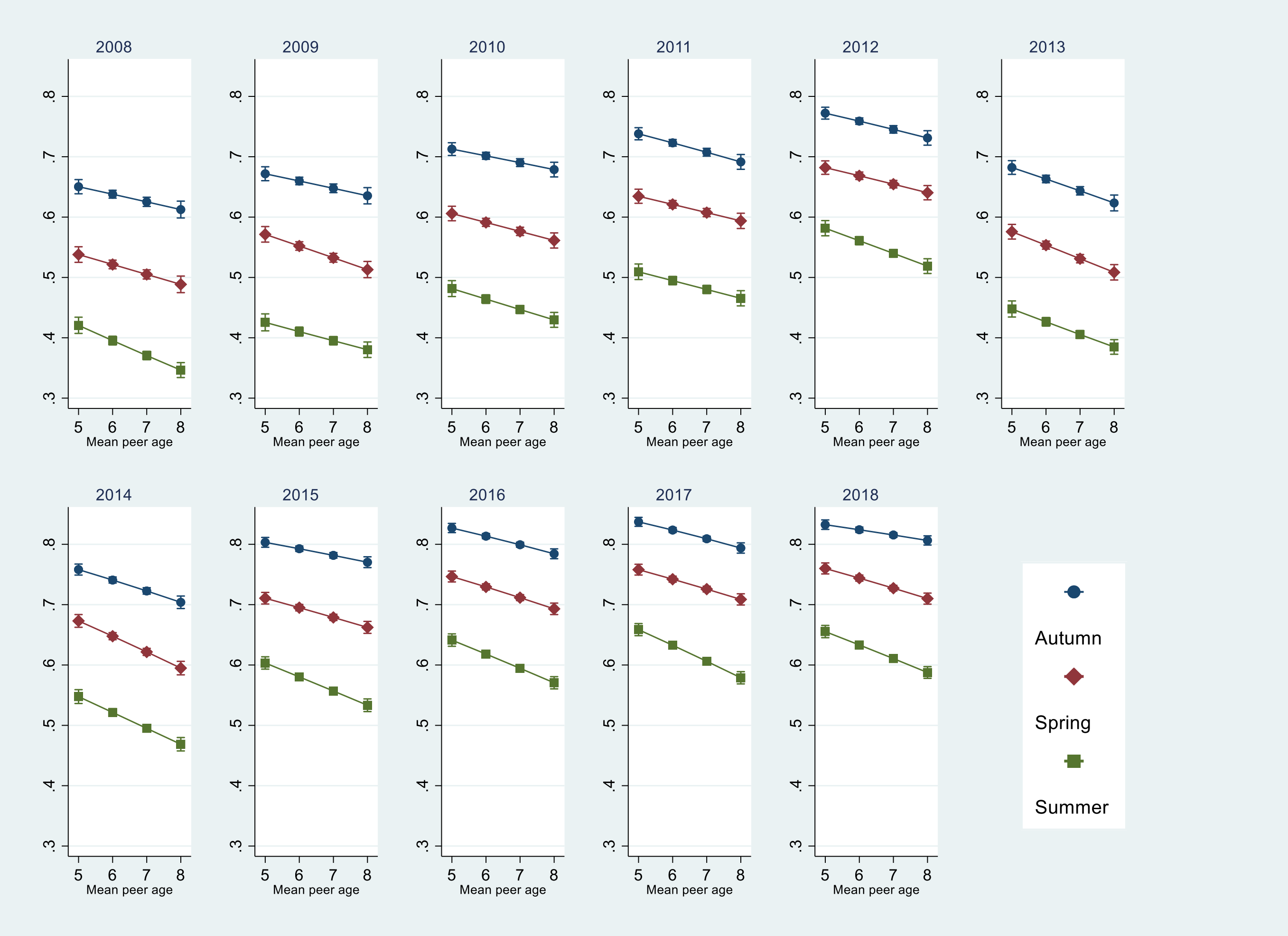


**Figure 7: Predicted probabilities of children born in each term and with each level of mean peer age being attributed a 'Good Level of Development': small schools samples, Specification 6.** Whiskers are 95% confidence intervals around each estimate. Source: National Pupil Database Spring Schools Census 2008-2018; Early Years Foundation Stage Profile Results 2008-2018. Controlled for child's recorded ethnic group, EAL, and FSM; N children in school, proportion White British children in school, proportion EAL children in school, proportion FSM children in school. All specifications cluster standard errors at the school level. 2008 N=159,071; 2009 N=160,886; 2010 N=163,329; 2011 N=161,502; 2012 N=158,173; 2013 N=152,370; 2014 N=153,772; 2015 N=150,776; 2016 N= 147,990; 2017 N=145,457; 2018 N=138,933. Note: September=12, October=11,…July=2, August=1.

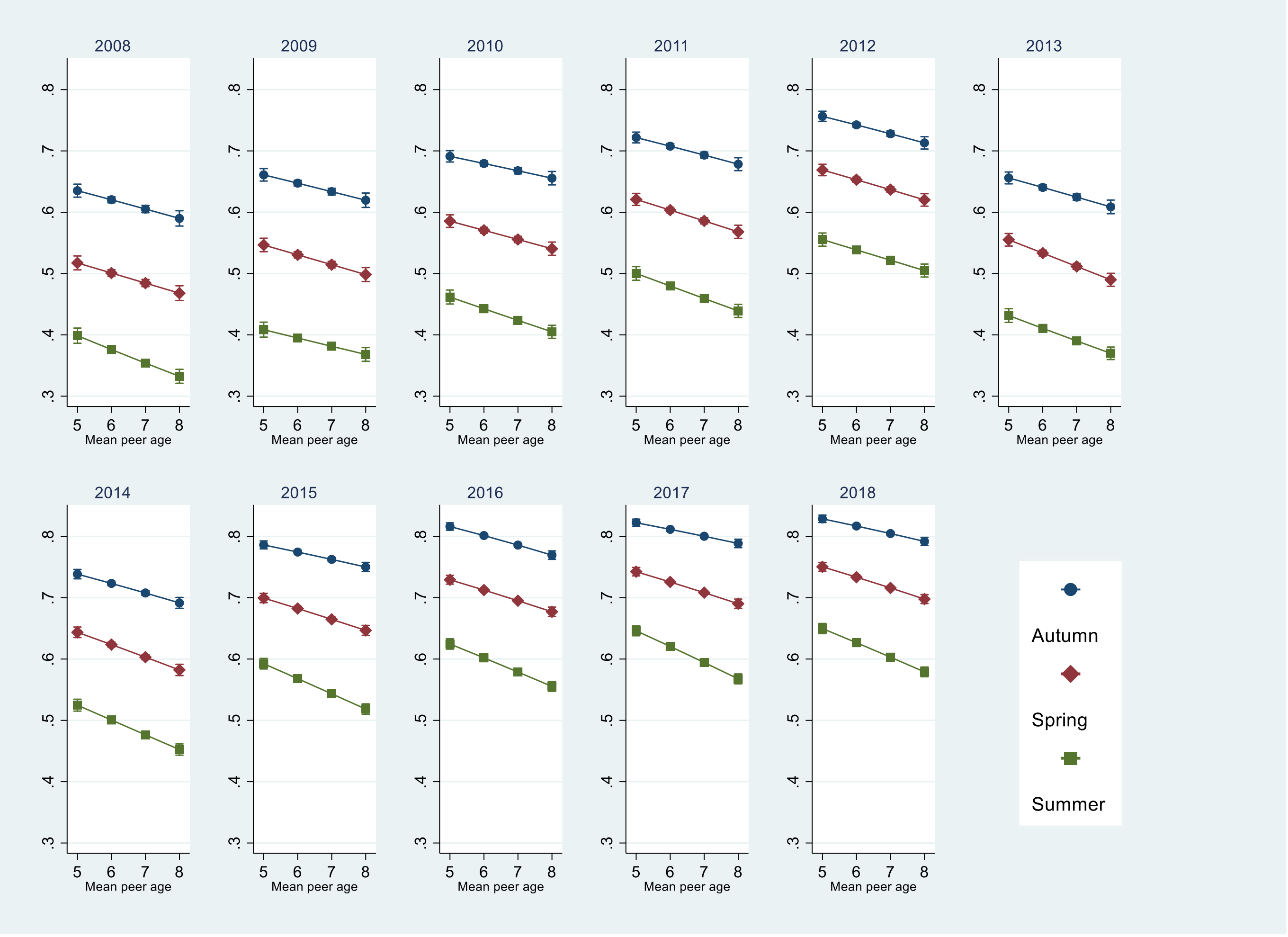


**Figure 8: Predicted probabilities of children born in each term and with each level of mean peer age being attributed a 'Good Level of Development': less skewed sample, Specification 6.** Whiskers are 95% confidence intervals around each estimate. Source: National Pupil Database Spring Schools Census 2008-2018; Early Years Foundation Stage Profile Results 2008-2018. Controlled for child's recorded ethnic group, EAL, and FSM; N children in school, proportion White British children in school, proportion EAL children in school, proportion FSM children in school. All specifications cluster standard errors at the school level. 2008 N=470,391; 2009 N=499,720; 2010 N=524,696; 2011 N=541,003; 2012 N=561,518; 2013 N=585,097; 2014 N=585,173; 2015 N=595,875; 2016 N=596,610; 2017 N=574,237; 2018 N=505,851.Note: September=12, October=11,…July=2, August=1.

## Discussion

This study finds that both a child's own age and the average age of their peers determine their chances of being attributed an EYFSP 'Good Level of Development.' Firstly, it mapped by birth month national patterns for each year from 2008 to 2018, showing consistent linear inequalities according to children's age. Every year, despite revisions to the EYFSP, August-born children are far less likely than September-born children to be attributed a 'GLD': on average, 30 percentage points less likely. Children born throughout the year have decreasing chances of being deemed 'Good,' the further their birthday falls from September.

Secondly, the current study explored associations between the mean age of a child's peers – whether the child is in a school year cohort that is relatively older or relatively younger, on average – and their chances of being attributed a 'GLD.' Being with more older children is associated with lowered chances of 'GLD' attribution, for summer-borns, spring-borns, and autumn-borns. Like findings by children's own birth month, these patterns hold regardless of controls for potential confounders and on considering children in schools of different sizes and with more or less age-skewed year-groups.

The finding of an overall negative relationship between more older peers and attribution of a 'GLD' for all children regardless of their own age contrasts with results from previous studies using validated research instruments and standardised cognitive tests. It also diverges from the wider evidence on peer effects among young children, which indicates generally that having more older school peers tends to be beneficial to children's development. What this begins to imply is that the EYFSP may not be 'a reliable, valid and accurate assessment of individual children,' as stated by the DfE, because expected benefits of more older peers are not observed through its lens – even in smaller schools, where children are more proximal and peer effects more likely to manifest. The consistent finding that, regardless of a child's own age, and regardless of whether the school year-group is more markedly young or old, higher mean peer age leads to lowered chances of a 'GLD' lends support to the influences of tendencies to grade relatively, 'on a curve,' and/or of the EYFSP 'moderation' process, in determining whether an individual child's development is denoted as 'Good.'

Along with the fact that 'GLD' attribution is strongly driven by age, this is problematic because one of the DfE's aims for the EYFSP is that it should 'inform parents about their child's development.' Parents of summer-borns are being informed that the development of their child is not 'Good' simply because of the month of their birth. It is also potentially problematic because EYFSP results are intended directly to influence teachers' practice

(DfE, 2020c), which will impact and differentiate children's trajectories. Lastly, it is problematic because the DfE states that it uses amassed EYFSP judgements to inform policy 'both nationally and locally.' If the data are questionable at the child level, can they be trusted at the aggregate level? This harks back to Calsamiglia & Loviglio's contention, that 'grading 'on a curve,' creates a 'a source of distortion' (p1) that affects the aggregate.

Therefore findings here suggest that the EYFSP does not fulfil its own criteria and its aim to 'provide a reliable, valid and accurate assessment of individual children:' firstly, because to a substantial extent it measures age, rather than development. Secondly, denotation of a 'GLD' is influenced by the composition of ages of children within a school, further diminishing the measure's accuracy in reliably or validly assessing the individual child. The EYFSP's use to 'inform parents about their child's development' should therefore be interrogated, because the development of summer-borns is disproportionately deemed not 'Good,' while even the development of autumn-borns is less likely to be deemed 'Good' if they are in a school with more older peers. Additionally, due in part to these issues regarding the information produced at the individual level, the extent to which the EYFSP provides 'an accurate national data set relating to levels of child development at the end of the EYFS' (DfE, 2017) should be scrutinised.

***What this study does not imply, and conclusions***

Given the findings and interpretations above, a next step may be conjured concluding that, as the process of teacher EYFSP assessments (and 'moderation' processes of these assessments) have proved questionable, validated standardised tests of reception children should instead be used. That is not the conclusion of this paper, for a number of reasons.

One is disagreement with the premise that reception-aged children should formally, summatively, nationally be assessed. In unpicking the workings and limitations of denotation of the EYFSP 'GLD,' this paper makes no requisite argument for summative assessment at this age, nor that the EYFSP assessment should be replaced with other tools. Formal, summative, national assessment of all children in their first year of school has yet to prove its utility or that on balance it helps children, schools, or parents.

Additionally, apart from the need to revisit and reassess the assumption that such assessment should continue, and to reconsider the best ways for schools to work with parents, for teachers to assess children, and for children's progress to be monitored in-school, there is a body of work including that by Goldstein et al (2018) indicating the difficulties around using aggregate summative statistics on young children within the current school system in England – even if they result from standardised tests rather than teacher judgements.

Therefore this paper's intention is to scrutinise the workings of the EYFSP without suggesting the necessity of its existence. Writing in 2007, in the lead up to the 2008 implementation of the new EYFS, Rogers & Rose synthesised the literature on early education for four and five year olds, and cited evidence for a system utilising 'formative diagnostic assessment' (p58). In the intervening years, this is not what the EYFSP and particularly the 'GLD' measure have become, and this deserves transparent inspection and consideration.

The latest iteration of the EYFSP continues to ignore both the dominance of age in determining attribution of the 'GLD' and the distorting effects of school contextual factors and relative judgement (DfE, online; 2020a; 2020b). Though the DfE has stated an intention to remove compulsory moderation of scores, it has not indicated any plans to cease to use and publish results at the aggregate level, nor has it removed the EYFSP's statutory status as an summative assessment whose results must be returned to the Local Authority and collected nationally, and nor has it addressed the (mis)uses of and pressures on the EYFSP resulting from the wider accountability system (DfE, 2020c).

Moreover, the fundamental principle that summative statutory assessment is useful in the reception year has not yet been addressed by the Department; nor have unintended negative consequences of the EYFSP for children, parents, or pedagogy been considered. In its recent consultation response, the DfE explicitly refuses meaningfully to address inequalities by birth month, despite 'respondents [referring] to the protected characteristic of age, specifically citing concerns that summer-born children are less likely to be assessed as having reached the expected level of development than their autumn-born peers' (DfE, 2020c, p27).

The Department instead states a commitment to 'ensure [assessment criteria] are based on the latest evidence in childhood development' and 'overarching principles that every child is a unique child and that children develop and learn at different rates' (DfE, 2020c, p7; p5). This appears to misunderstand both the existence and nature of structural and contextual effects, and of the birth month gradient: to ignore the vast systematic disparity in attribution of a 'GLD' between September-borns and August-borns, which does not in fact usefully reflect the individual child. Instead, it is based in a tendency evidenced throughout the research on child development: that for the majority of children, skills advance with age.

More change, proper consideration, and unconstrained, genuinely open, evidence-informed reform of the EYFSP is still needed.

## Acknowledgements

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. Thanks to the Department for Education for use of the data, and to LSE IT colleagues for facilitating and supporting access. This work was supported by a British Academy Postdoctoral Fellowship, under Grant [PF2 \180019]. Many thanks to Ludovica Gambaro, Polina Obolenskaya, participants in the CASE researchers' workshop, and two anonymous reviewers, for useful comments and suggestions. All analyses and interpretations are the author's.

## Data Availability

The data that support the findings of this study are available from the Department for Education. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts with the permission of Department for Education.

## Ethical approval

This work is secondary analysis of de-identified administrative data – no new data collection took place, so ethical review and approval was not required.

## Conflict of Interest

There is no conflict of interest.

## References

Ansari, A., Purtell, K. & Gershoff, E. (2016). Classroom age composition and the school readiness of 3- and 4-year olds in the Head Start Programme. Psychological Science, 27(1), 53-63.

Campbell, T. Gambaro, L. & Stewart, K. (2019). Inequalities in the experience of early education in England: Access, peer groups and transitions. https://sticerd.lse.ac.uk/case/_new/publications/abstract/?index=6266 (accessed 25.10.21)

Barbetta, G. P., Sorrenti, G. & Turati, G. (2019). Multigrading and Child Achievement. University of Zurich. https://www.jacobscenter.uzh.ch/dam/jcr:e1af546c-6e38-4d80-9652-dfcd1dbfddab/econwp275_Revised_Aug19_Final.pdf (accessed 03.04.21)

Borbelly, D., Gehrsitz, M., McIntyre, S., Rossi., G. & Roy, G. (2021). Composite classes, class size and human capital accumulation. London: Nuffield Foundation.

Bradbury, A. (2011). Rethinking assessment and inequality: the production of disparities in early years education. Journal of Education Policy, 26(5), 655-696

Bradbury, A. (2012). 'I feel absolutely incompetent': Professionalism, policy, and early childhood teachers. Contemporary Issues in Early Childhood, 13(3), 175-186

Bradbury, A. & Roberts-Holmes, G. (2017) Creating an Ofsted story: the role of early years assessment data in schools' narratives of progress. British Journal of Sociology of Education, 38(7), 943-955

Calsamiglia, C. & Loviglio, A. (2019). Grading on a curve: When having good peers is not good. Economics of Education Review, 73, 101916

Cascio, E., & Schanzenbach, D. W. (2007). First in the class? Age and the education production function. NBER Working Paper 13663.

Crawford, C., Dearden, L., & Meghir, C. (2007). When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England. IFS: London.

Crawford, C., Dearden, L., & Greaves, E. (2013a). The drivers of month of birth differences in children's cognitive and non-cognitive skills: a regression discontinuity analysis. IFS: London.

Crawford, C., Dearden, L., & Greaves, E. (2013b). Identifying the drivers of month of birth differences in educational attainment. IFS: London.

Crawford, C., Dearden, L., & Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. Journal of the Royal Statistical Society A, 177(4), 829-860

Department for Education (online). Early adopter schools: EYFS framework. https://www.gov.uk/government/publications/early-adopter-schools-eyfs-framework (accessed 09.01.21)

Department for Education (2020a). Statutory framework for the early years foundation stage. Setting the standards for learning, development and care for children from birth to five. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896810/EYFS_Early_Adopter_Framework.pdf (accessed 09.01.21)

Department for Education (2020b). Development Matters. Non-statutory curriculum guidance for the early years foundation stage. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944603/Development_Matters_-_non-statuatory_cirriculum_guidance_for_EYFS.pdf (accessed 09.01.21)

Department for Education (2020c). Early Years Foundation Stage Reforms. Government consultation response. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896872/EYFS_reforms_consultation_-_government_response.pdf (accessed 06.02.21)

Department for Education (2019). Early years foundation stage profile results in England, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/839934/EYFSP_2019_Main_Text_Oct.pdf (accessed 15.01.21)

Department for Education (2017). Early years foundation stage profile 2018 handbook. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/669079/Early_years_foundation_stage_profile_2018_handbook.pdf (accessed 09.01.21)

Department for Education (2010a). Achievement of Children in the Early Years Foundation Stage Profile. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181551/DFE-RR034.pdf (accessed 09.01.21)

Department for Education (2010b). Month of Birth and Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/182664/DFE-RR017.pdf (accessed 09.01.21)

Elder, T. & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement. The Journal of Human Resources, 4(3), 641-683.

Goldstein H, Moss G, Sammons P, Sinnott G and Stobart G (2018). A baseline without basis:.The validity and utility of the proposed reception baseline assessment in England, London:.British Educational Research Association. https://www.bera.ac.uk/researchers-resources/publications/a-baseline-without-basis (accessed 06.02.21)

Guo, Y., Tompkins, V., Justice, L. & Petscher, Y. (2014). Classroom age composition and vocabulary development among at-risk pre-schoolers. Early Education and Development, 25, 1016-1034

Meissel, K., Meyer, F., Yao, E. S., Rubie-Davies, C. M. (2017). Subjectivity of teacher judgements: Exploring student characteristics that influence teacher judgements of student ability.

Moller, A. C., Forbes-Jones, E., & Hightower, D. (2008). Classroom age composition and developmental change in 70 urban preschool classrooms. Journal of Educational Psychology, 100(4), 741-753

Moss, P. (2012). Early childhood policy in England 1997-2013: anatomy of a missed opportunity. International Journal of Early Years Education, 22(4), 346-358

Office for Statistics Regulation (online). National Statistics. https://osr.statisticsauthority.gov.uk/national-statistics/ (accessed 20.04.21)

Roberts-Holmes, G. & Bradbury, A. (2016). The datafication of early years education and its impact upon pedagogy. Improving Schools 19(2), 119-128

Rogers, S. & Rose, J. (2007) Ready for Reception? The advantages and disadvantages of single-point entry to school. Early Years 27(1), 47-63

Sharp, C., George, N., Sargent, C., O'Donnell, S., & Heron, M. (2009). International thematic probe: The influence of relative age on learner attainment and development. NFER.

## Annex A

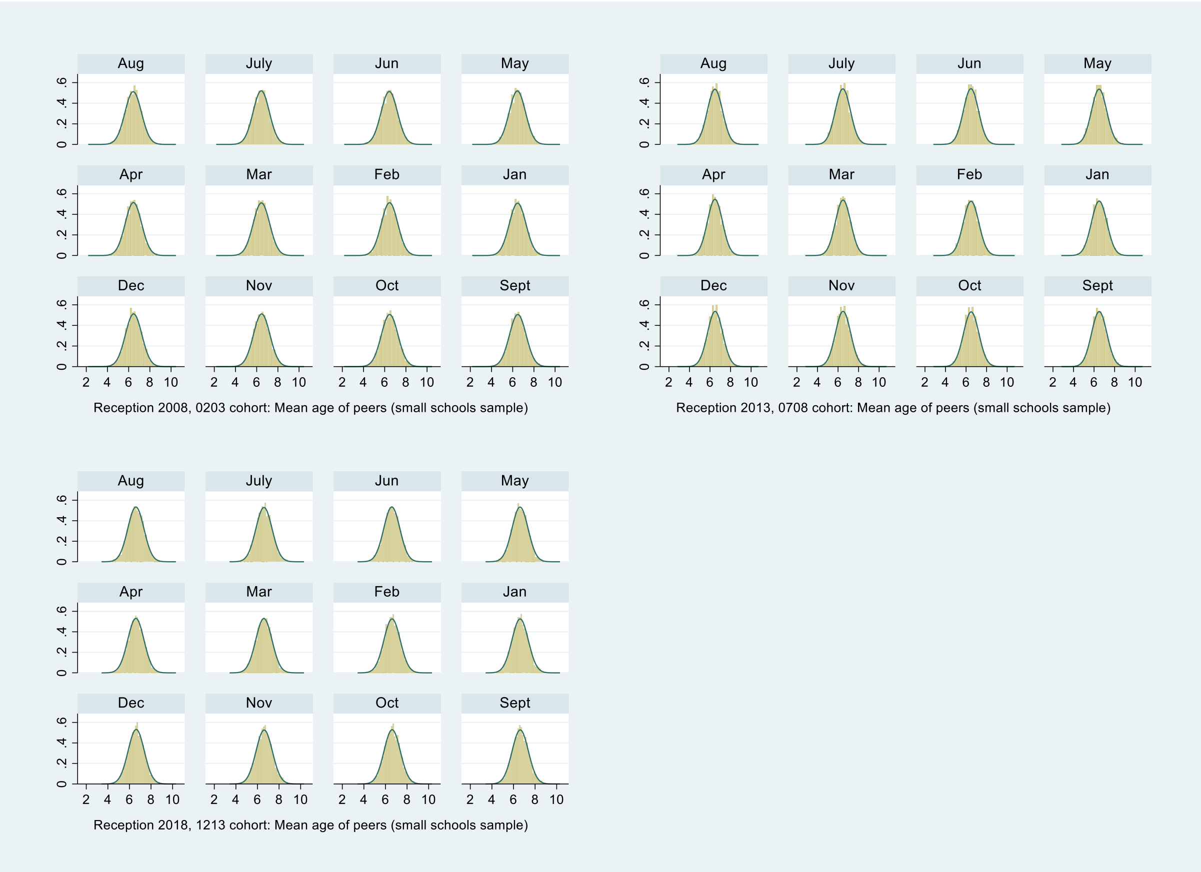


**Figure A1: Distribution of mean peer ages experienced by children born in each month: small schools sample.** 2008 N=159,071; 2013 N=152,370; 2018 N=138,933.

This is the accepted version of Campbell, T. (2021), *Relative age and the Early Years Foundation Stage Profile: How do birth month and peer group age composition determine attribution of a "Good Level of Development" – and what does this tell us about how "good" the Early Years Foundation Stage Profile is?* British Education Research Journal: https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1002/berj.3771

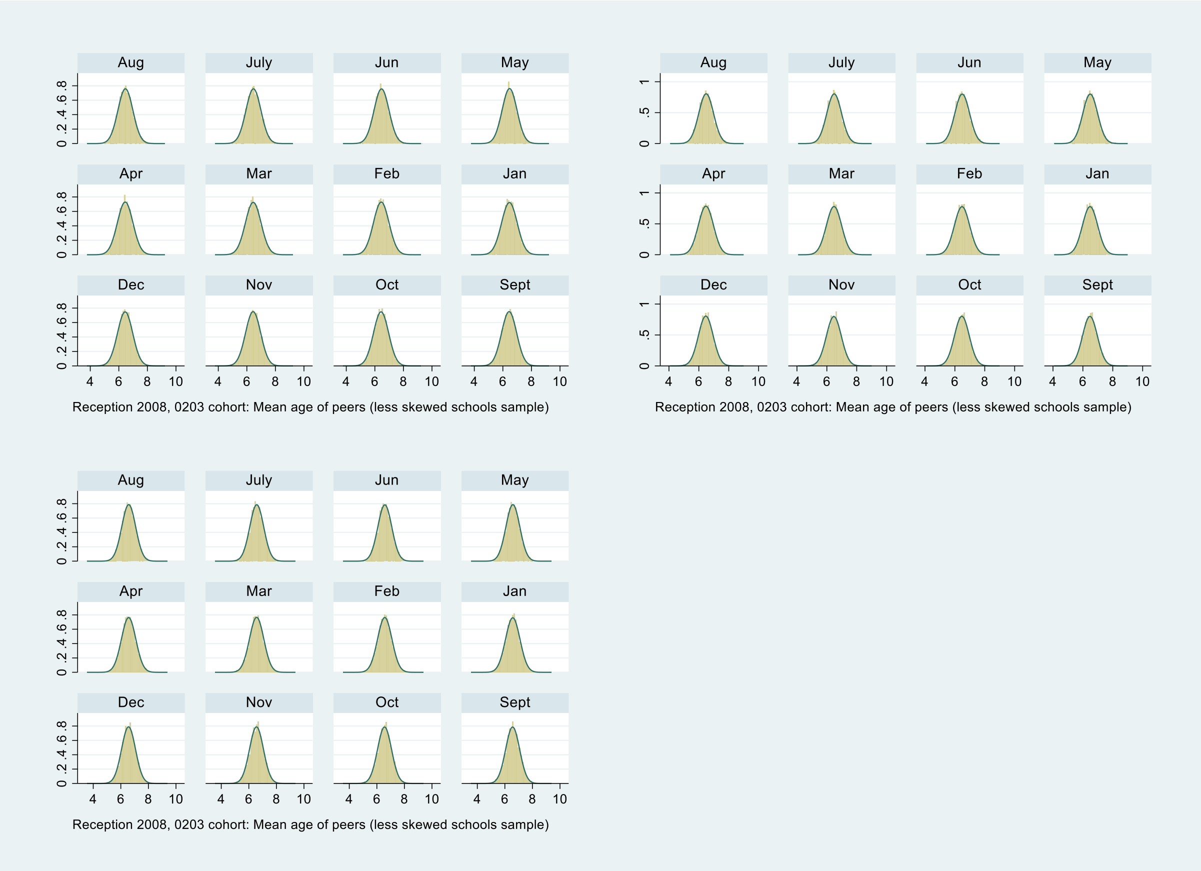


**Figure A2: Distribution of mean peer ages experienced by children born in each month: less skewed sample.** 2008 N=470,391; 2013 N=585,097; 2018 N=505,851.