Early Educators' Knowledge of Early Language Pedagogy:

How can it be measured and does it matter for child language outcomes?

Dr Sandra Mathers, Professor Lars-Erik Malmberg, Dr Ariel Lindorff and Dr Julian Gardiner

October 2022

www.nuffieldfoundation.org

Authors

Dr Sandra Mathers, Department of Education, University of Oxford

Professor Lars-Erik Malmberg, Department of Education, University of Oxford

Dr Ariel Lindorff, Department of Education, University of Oxford

Dr Julian Gardiner, Department of Education, University of Oxford

Suggested citation

Mathers, S., Malmberg, L.E., Lindorff, A.M. & Gardiner, J. (2022) *Early educators' knowledge of early language pedagogy: How can it be measured and does it matter for child language outcomes?* Oxford: University of Oxford.

CONTENTS

Acknowledgements

The research funder

The study has been funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily of the Foundation. Visit www.nuffieldfoundation.org.

We would like to thank Eleanor Ireland, the Education Program Manager, in particular.

The study advisory group

We are very grateful to members of our advisory group for helping us to refine our research questions and methodology, providing valuable critical feedback on our research findings and supporting our dissemination activities:

- Professor Kathy Sylva, University of Oxford
- Professor Edward Melhuish, University of Oxford
- Professor Steve Strand, University of Oxford
- Beatrice Merrick, Early Education
- Dr Lesley Curtis, Everton Nursery School and Family Centre

Colleagues who have supported the study

We are very grateful to Keeley Dobinson for support with data management and administration, and to Professor Iram Siraj, without whom the original OLP validation study could not have taken place.

The research participants

The study would have not been possible without the contribution of the teachers and schools who took part in the original study from which our dataset was drawn. We extend our sincere thanks.

Data and video providers

We are grateful to the Education Endowment Foundation for funding the original randomised controlled trial from which our dataset was drawn, and for allowing us to use it for the purposes of this study. Thanks also to the Department of Education and Office for National Statistics for enabling access to data from the National Pupil Database; and to Siren Films for allowing us to use an extract from one of their videos as part of the OLP measure.

List of tables

List of figures

Referencing conventions

Citations are numbered (1, 2, 3…) and full references included as endnotes

Explanatory notes are marked alphabetically (a, b, c…) and included as footnotes

EXECUTIVE SUMMARY

Theory and practice tell us that teachers need knowledge of content (*what children need to learn*) and of pedagogy (*how to support the learning of that content*). This study explores relationships between teachers' **pedagogical** knowledge and children's language progress between ages 3 and 5. It forms the second stage in validation of a new measure designed to capture such knowledge: the **Observing Language Pedagogy (OLP)** instrument.

The OLP uses video to capture early educators' oral-language-related **procedural** knowledge of pedagogy. In contrast to theoretical knowledge, procedural knowledge means knowing how *in practice*. It is the dynamic, flexible pedagogical knowledge which underpins educators' real-time decision-making in the classroom.

The first stage in validating the OLP took place within a randomised controlled trial (RCT) designed to evaluate a preschool professional development programme. Findings showed that the OLP is a reliable measure of procedural knowledge and that it predicts classroom quality: teachers with greater knowledge led classrooms with higher-quality practice[1].

The aim of the current study is to explore whether teachers' procedural knowledge (as measured by the OLP) also predicts children's language outcomes at the end of the Early Years Foundation Stage (EYFS). It uses data from the original RCT on 70 teachers, and 797 children in their classes, drawn largely from schools in disadvantaged areas of England. The study investigates whether children with more knowledgeable teachers in their reception year (age 4-5) had better outcomes at age 5 than children with less knowledgeable teachers, controlling for child language at age 3 and a range of other child, teacher and school factors.

Completing the OLP involves watching two short videos of adult-child interactions and recording strategies being used which are likely to support children's oral language. The OLP assesses educators' ability to *identify* and *interpret* language-supporting strategies in context, capturing knowledge which can be unlocked and used 'in the moment'. Responses are coded to reflect three facets (**perceiving, naming, interpreting** – see over).

Four researcher-administered measures of child language outcomes were used, reflecting children's understanding and use of vocabulary and grammar. A further three measures reflected children's achievement in teacher-completed national assessments at age 5: overall level of development, communication/ language and literacy. The final measure was children's standardised score from the national phonics check completed in Year 1 (age 6).

PERCEIVING: the extent to which educators can **identify** language-supporting strategies in the OLP videos. Since respondents have a choice of which strategies to report, responses reflect their ability to identify the most *salient* strategies – those likely to be most important for children's language learning in that specific context.

NAMING: the extent to which respondents use **'expert vocabulary'** to describe the strategies they identified, e.g., using the specific term '*open questions'* rather than the more informal '*how and why questions'*. This reflects more formal and explicit knowledge of the relevant concepts. In the initial validation study[1], teachers' use of expert vocabulary (*naming)* predicted classroom quality (Figure 1).

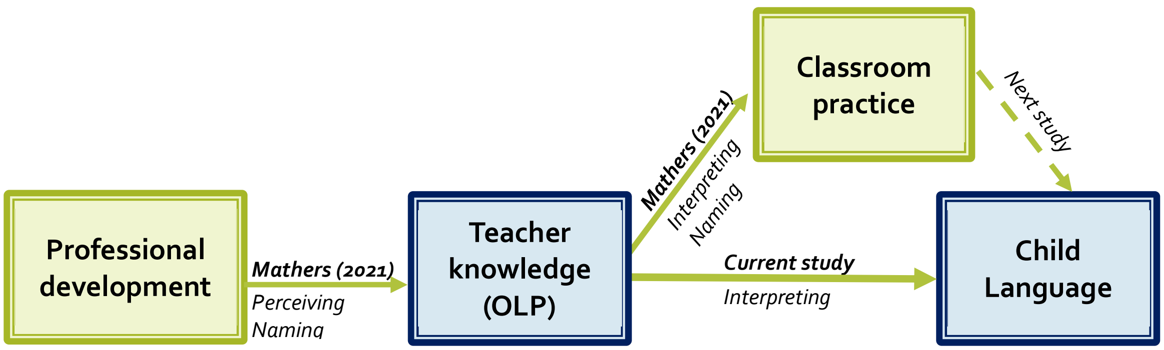
INTERPRETING: the extent to which respondents offer an **explanation or analysis** of the strategy they identified, e.g., suggesting what the adult's pedagogical goal might have been, or noting an observed effect on the child. This reflects higher-level knowledge, specifically, the ability to connect knowledge about child development and pedagogy (understanding **why** as well as **how**). In the initial validation study, teachers' *interpreting* predicted classroom quality (Figure 1).

Findings

Reception class teachers' knowledge predicted gains in children's understanding of vocabulary and sentence structure between the ages of 3 and 5.

While teachers' ability to identify salient strategies in videos (*perceiving*) was associated with child language outcomes, when all three facets of knowledge were included in the same statistical model, *interpreting* was a stronger predictor (Figure 1). Therefore – although knowing which strategy is appropriate in a specific context matters – understanding *why* such strategies might be effective matters more for child language outcomes. Given findings from the initial validation study[0] that knowledge also predicts observed quality of classroom practice (Figure 1), it is plausible that this association between teacher knowledge and child language is mediated by improvements in practice – that is, teachers who can connect their knowledge of pedagogy and child development in order to *interpret* video interactions are also more likely to use language-supporting strategies intentionally when interacting with children, with resultant benefits for child language outcomes. This mediation pathway will be examined as the next stage of work.

Figure 1: Findings from the first two stages of OLP validation (the current study and Mathers, 2021)



Conclusions and Implications

1. To ensure early years educators are prepared to support young children's oral language, we need to invest in nurturing their early-language-related pedagogical knowledge (the dynamic, procedural knowledge which informs real-time classroom decision-making and action).

2. Specifically, educators should be supported in connecting their knowledge of language-supporting strategies with knowledge of child development, in order to understand **why** as well as **how** specific techniques support oral language development.

Promising avenues for attention might include:

For government and providers of qualifications and professional development for early years educators:

Ensuring that pre- and in-service professional development opportunities are available which explicitly nurture oral-language-related pedagogical knowledge and, specifically, those which promote pedagogical *reasoning*. This might include structured opportunities to reflect on and analyse practice in a way which makes the relationships between pedagogy and child outcomes explicit (e.g., observing adult-child interactions, either live or on video, to analyse child learning needs, strategies which might support learning, and the success of strategies employed).[2]

For early years staff teams:

Setting aside regular time for professional conversations about language-supporting practice which make explicit and/or analyse relationships between pedagogy and young children's outcomes. This might range from informal end-of-day discussions to more structured activities (e.g., video-recording and analysing practice).

For researchers:

- Including measures of procedural knowledge in studies of early years educator qualifications and professional development, in order to facilitate understanding of how such knowledge can be enhanced; and also how such knowledge translates into action;
- Conducting further research into how pedagogical reasoning (*interpreting*) can be supported and enhanced through professional development.

1. Introduction

In this report we present findings from a study exploring relationships between teachers' language-related pedagogical knowledge and children's oral language progress across the Early Years Foundation Stage (EYFS). The study forms the second stage in the validation of a new measure designed to capture such knowledge: the **Observing Language Pedagogy (OLP)** instrument.

Having an early years workforce capable of supporting young children's oral language skills is vital. Oral language at school entry underpins children's later learning[3] and is one of the strongest predictors of academic progress and later life outcomes[4]. However, many children start school without the language skills they need. Research shows that the vocabulary of disadvantaged children at age 5 is almost a year behind their more advantaged peers on average,[5] with serious consequences for long-term life trajectories. High-quality preschool provision offers a powerful means of narrowing this gap.[6] However, recent research in disadvantaged areas of England shows that preschool quality may not always be sufficient to nurture children's oral language[7] and that many disadvantaged children remain below expected levels in their early language skills[8]. The effects of the Covid pandemic on young children's language development make it even more pressing that the preschool workforce is equipped to nurture these foundational skills.

The question then becomes how best to prepare early educators to become expert in supporting early language. Although we know that good quality practice is important, we actually know very little about the knowledge and skills which early educators need in order to *provide* that high quality, and to ensure children's language progress. Evidence shows that both qualifications[9] and in-service professional development[10] matter for quality and child outcomes. However, effects are small to moderate at best[9,10] and available programmes are highly variable in their evidence-base, content, quality and effects on child outcomes.[10,11] To improve the design of qualifications and professional development, we need to know more about the processes which underpin professional growth.[12]

Theory and practice tell us that *knowledge* is fundamental – both content knowledge (*the content to be learned*) and pedagogical knowledge (*how to support the learning of that content*).[13] Pedagogical knowledge can be theoretical (*knowing how in theory*) or procedural (*knowing how in practice*).[14] For example, a preschool educator with expertise in supporting oral language will know how to form the past tense of irregular verbs (*content knowledge*), know that children tend to overgeneralise regular verb endings and that 'recasting' these errors in conversation can help children grasp irregular forms (*theoretical knowledge of pedagogy*); and also know how to *use* recasting appropriately during classroom interactions (*e.g. "I runned fast"* → *"You ran fast!"*) reflecting *procedural* knowledge of pedagogy.

Procedural knowledge is both generated and applied in the classroom context through practising the craft of teaching[15]. Closely related to skill, it reflects the knowledge a teacher needs in order to perform a skill and, at the same time, it *develops* through rehearsal of specific skills in the classroom. Procedural knowledge is understood to develop both tacitly (without conscious awareness) and explicitly (whereby theoretical knowledge is transformed into procedural knowledge via processing practice).[16] In this latter context, procedural knowledge forms a 'bridge' between theory and practice.

Puzzlingly, while the language-and-literacy-related content knowledge of early educators has been linked to improved practice and child outcomes[17], studies of *pedagogical* knowledge have shown minimal effects.[18] One explanation lies in the way it has traditionally been measured.[19] Pedagogical knowledge – particularly *procedural* pedagogical knowledge – is by nature dynamic and flexible. There is no one 'correct' language-supporting strategy which applies in every situation. The choice a preschool teacher makes in a specific classroom interaction will depend on many factors: their knowledge of the abilities of individual children; their goals for teaching and learning; their knowledge and beliefs about how children learn and effective practice; their understanding of group dynamics; children's actual responses and so on.[20] Educators make 'in-the-moment' decisions, balancing multiple sources of knowledge to support them in using their strategies *flexibly* to maximise child learning in different situations.[21] However, many research studies have measure pedagogical knowledge using surveys which offer multiple choice options and assume a single correct answer (e.g., "in this scenario, would you do a, b, c or d?"). These kinds of questions may be so far removed from the complexities of real-life classroom interactions that they fail to capture the dynamic, flexible pedagogical knowledge which underpins educators' real-time decision-making within them.

Video assessment is gaining ground as an alternative method of measuring pedagogical knowledge, offering a more authentic and situated context which preserves the complexity of classroom interactions.[22] Practitioners are shown videos of classroom interactions and asked to identify or analyse pedagogical practice, or children's responses and thinking. This allows assessment, not just of what they know, but of the knowledge they can *activate* and *use* in interactions with children.[23] The approach is based on the notion that being able to notice the features of an interaction which are important for child learning underpins effective decision-making, which in turn underpins effective practice.[24] Measuring practitioners' ability to identify the important features of a video interaction can, therefore, give a window onto their pedagogical knowledge and – specifically – onto the dynamic procedural knowledge which supports in-the-moment decision-making during real interactions with children.[25]

The **Observing Language Pedagogy (OLP)** instrument uses video to capture early educators' language-related procedural knowledge. It was recently piloted within a randomised controlled trial (RCT) designed to evaluate a preschool professional development programme, using a sample of 104 teachers in 72 schools within disadvantaged areas of England. Findings showed that the OLP is a reliable measure of procedural knowledge and that it predicts observed classroom quality: teachers with greater knowledge led classrooms with higher-quality practice, with analytical models explaining around 30% of the variation in observed quality.[1]

The aim of the current research is to explore whether teachers' procedural knowledge (as measured by the OLP) also predicts children's language progress across the EYFS. It uses data from the same RCT on children's language skills at the start of the nursery year (age 3) and the end of the reception year (age 5). Using data on 70 teachers, and 797 children in their classes, it explores whether children who had more knowledgeable teachers in their reception year had better oral language outcomes at the end of the EYFS than children whose teachers were less knowledgeable – controlling for children's language skills at the start of the EYFS and a range of other child, teacher and school characteristics.

2. The Observing Language Pedagogy (OLP) instrument

The OLP instrument aims to elicit and measure preschool educators' knowledge of language-supporting strategies, specifically, their ability to **identify** and **interpret** strategies in video interactions. It is designed to be brief so it can be used in research studies without placing a large time burden on respondents.

Educators are asked to watch two short (2-3 min) videos of adult-child interactions and prompted to identify up to eight strategies used which are likely to support children's language development. The first shows a nursery teacher interacting with a single child in the block area, and the second a small group interaction. Both videos show informal child-led contexts. Respondents are asked to focus on strategies which support children to make meaning through language, rather than code-related skills such as phonological awareness. Some sample responses are shown in the box:

Example OLP responses

Uses rich vocabulary (slower, faster, steeper, further). Child soaks up the new language and starts to use it herself.

Offers a running commentary on the child's actions.

Listening to child and valuing all her ideas.

Responses are coded to reflect three facets of knowledge:

PERCEIVING reflects the extent to which educators can **identify** language-supporting strategies in the videos. Answers are coded against a list of strategies derived from prior research as being important for oral language. These include **linguistic strategies** *(e.g., modelling vocabulary, extending children's own speech)*, **interactive strategies** *(e.g., inviting communication, responding to children)*, **cognitive strategies** *(e.g., predicting, inferring)*, **relational strategies** *(e.g., giving children individual attention, responding warmly)* and **contextual strategies** *(e.g., introducing language in ways likely to be meaningful for children).* Since educators have a choice of which strategies to report, responses reflect their ability to identify the most *salient* strategies – those likely to be most important for children's language learning in that specific context. Answers are weighted using a benchmark based on expert responses to the videos, with higher scores awarded for strategies which experts rated as being both present and salient in each video. In this way, the OLP measures respondents' ability to know when certain strategies are appropriate and what matters most for child language in a specific situation.

NAMING reflects the extent to which respondents use **'expert vocabulary'** to describe the strategies they identify – for example using the specific term '*open questions'* rather than the more informal '*how and why questions*'; or the term *'descriptive commentary'* rather than the more informal '*talking about what the child is doing'*. This facet reflects **formal and explicit** knowledge of the relevant concepts, and perhaps also prior attendance at language-related professional development.

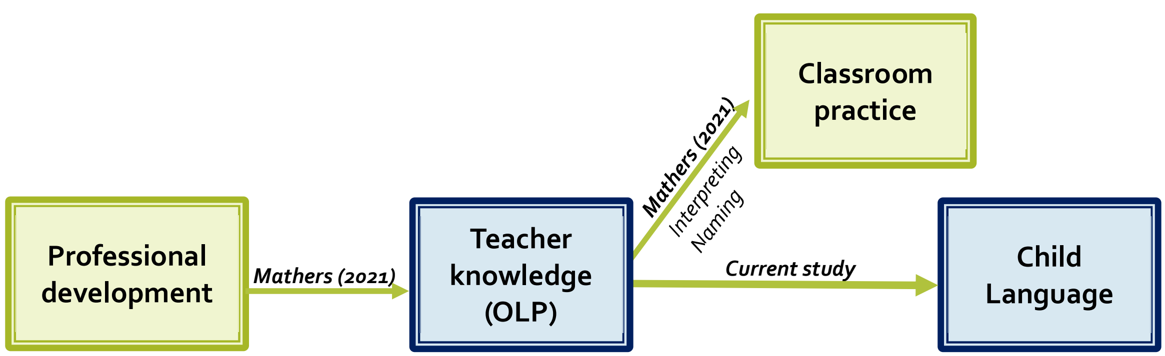
INTERPRETING reflects the extent to which respondents offer an **explanation or analysis** of the strategy they identified (e.g., suggesting what the adult's pedagogical goal might have been, noting an observed effect on the child, or suggesting an alternative strategy). *Interpreting* reflects higher-level knowledge, specifically, the ability to connect knowledge about child development and pedagogy **(understanding *why* as well as *how*).** Such knowledge is thought to underpin intentional teaching: the deliberate use of strategies to support children's oral language development in a specific way.[26]

The initial OLP validation study[1] showed that the three facets could be empirically distinguished (that is, they measure different aspects of knowledge) and were better represented by a three-factor structure than by a single combined measure. It also confirmed that the OLP is a valid and reliable measurement tool. More detail on coding and psychometric properties can be found in Appendix A.1 and in previous publications.[27,28]

3. The theoretical model

Since the OLP involves responding to videos of real adult-child interactions, we can think of it as measuring knowledge which can be unlocked and used 'in the moment'. It is based on the hypothesis that educators who are better at identifying salient language-supporting strategies in a video will also be more likely to know which strategies might be effective in a real-life interaction – and therefore more likely to *use* an appropriate strategy during that interaction. Similarly, educators who can offer a plausible *interpretation* of a video interaction (i.e., suggest why a strategy is being used, or note the effect it has on children) should also be more able to connect their knowledge of child development and pedagogy 'in the moment' to support intentional decisions about which technique/s to use. The initial validation study confirmed these hypotheses: teachers who achieved higher scores on the OLP led more language-supporting classrooms, with *naming and interpreting* proving the strongest predictors of observed high-quality practice[a] (Mathers, 2021;[1] Figure 1). It also showed that teachers' knowledge could be improved through professional development[b] (Figure 1) and that this improvement in knowledge mediated gains in classroom quality. In the current study, we examine whether teacher knowledge also predicts progress in child language outcomes.

Figure 2: Mapping the ground: the theoretical model and questions to be answered at different stages of research



[a] Unstandardised regression coefficients showed that the quality of language-supporting practice was 0.48 of a point higher on a 7-point scale (1=low, 7=high) on average for every additional expert vocabulary term used to describe the video interactions (*naming*), and 0.23 of a point higher for every interpretation provided (*interpreting*).

[b] Hedges g effect sizes in the range 0.58-0.84

4. Study methods and sample

Both the initial OLP validation study and the current study involve secondary analysis of data gathered as part of a previous randomised controlled trial (RCT), designed to evaluate the impact of the URLEY professional development programme for preschool teachers.[7] A total of 120 schools took part in the wider RCT between September 2017 and June 2018. All were state-funded primary schools recruited from disadvantaged regions of England: in the lowest 3 deciles as defined by the Indices of Multiple Deprivation 2010 [IMD].[c, 29] Participating teachers were from nursery (age 3-4) or reception (age 4-5) classes in these schools. Child language was assessed at the start of children's nursery year (age 3) and again at the end of the reception year (age 5). In England, these two years form the Early Years Foundation Stage (EYFS).

The sample for the current study comprised 70 reception class teachers[d] who completed the OLP online at the start of the children's reception year, and 797 children in the classes of these teachers. The 70 teachers were drawn from 61 schools: in nine of the schools, two reception class teachers had completed the OLP. The vast majority (n=67) held graduate level Qualified Teacher Status[e] (QTS), with three working towards QTS. Schools were largely situated in disadvantaged areas, with four fifths below the 50th IMD percentile and more than half below the 20th percentile. Characteristics of teachers, schools and children are shown in Table 1.

4.1 Educators' pedagogical knowledge

Data on teachers' knowledge were provided by the OLP instrument, specifically, factor scores for *perceiving*, *naming* and *interpreting* derived from the confirmatory factor analysis (CFA) conducted in the first validation study.[1] Descriptive data are shown in Table 1. Since the factor scores are difficult to interpret, raw sum scores are also shown in Table 1. For each OLP facet (*perceiving*, *naming*, *interpreting*) the raw scores reflect the sum of scores achieved for each video. These show that respondents achieved a mean *perceiving* sum score of 42.54 (SD 14.89); *named* the strategies they identified using one expert vocabulary term on average (SD 1.12); and provided one *interpretation* on average (SD 1.52). More information on how OLP scores were generated, and on the psychometric properties of the OLP, is shown in Appendix A1.1.

---

[c] The IMD provides a relative measure of deprivation at the small area level across England.

[d] The response rate was 50.8% (61 of 120 schools). Teachers who completed the OLP reflected the wider RCT sample on the aspects it was possible to compare (Appendix A.2). However, it was not possible to make comparisons on all variables.

[e] Of the 67 fully qualified teachers: 24 held early years QTS (3-7 years), 39 held primary QTS (3-11 years), 1 held a post-graduate teaching qualification (PCGE) and 3 held a Bachelor of Education (Hons) degree. Three taught in a shared nursery and reception class (EYFS unit).

It should be noted that the majority of classes which children attended during their reception year would have been supported by a second member of staff in addition to the lead teacher (e.g., a teaching assistant). No data were available on the knowledge of these additional staff members.

4.2 Child outcome measures

Four researcher-administered measures of child language were used, drawn from the wider RCT. More information can be found in Appendix A1.2 and the RCT evaluation report.[7]

- **The British Picture Vocabulary Scale (BPVS-3)**[30] assesses children's understanding of spoken language (receptive vocabulary). Children are shown four pictures and asked to choose the best match for each word read by the researcher administering the test.
- **The Clinical Evaluation of Language Fundamentals (CELF) Preschool 2 UK**[31] **'Sentence Structure' subtest** assesses children's acquisition of grammatical (structural) rules at the sentence level. Children are asked to interpret spoken sentences of increasing length and complexity by pointing to the appropriate picture.
- **The Renfrew Action Picture Test (RAPT)**[32] measures expressive (spoken) language. Based on children's descriptions of actions shown in a set of pictures, two scores are recorded. The *Information Score* reflects the level of information children provide, with points awarded for use of specific nouns, verbs and prepositions. The *Grammar Score* reflects children's use of different tenses, irregular past tense and plurals, simple and complex sentence constructions and the passive voice.

Four measures of children's progress on national assessments were drawn from the National Pupil Database (NPD). The Early Years Foundation Stage profile (EYFSP) is completed by teachers at the end of children's reception year (age 5). Three standardised scores for the academic year 2017-18 were used:

1. ***Good Level of Development (GLD):*** the total of children's scores for the 12 early learning goals (ELGs) in the three prime areas of learning *(communication and language; physical development; personal, social and emotional development),* literacy and mathematics. Each ELG is scored on a 3-point scale *(1=working towards, 2=at expected level, 3=above expected level)* resulting in a maximum of 36 points.
2. ***Communication and Language****:* the sum of scores for the 3 ELGs listening and attention, understanding, and speaking (maximum 9 points).
3. ***Literacy****:* the sum of scores for the 2 ELGs reading and writing (maximum 6 points).

Table 1. Sample characteristics

		N	%	Mean	SD*	Min.	Max.
TEACHERS							
OLP: *perceiving* factor score		70	-	-0.15	4.95	-9.00	11.94
OLP: *naming* factor score		70	-	0.01	0.34	-0.38	0.91
OLP: *interpreting* factor score		70	-	0.02	0.54	-0.41	1.50
OLP: *perceiving* raw sum score		67	-	42.54	14.89	17.00	80.18
OLP: *naming* raw sum score		67	-	.99	1.12	0	4.0
OLP: *interpreting* raw sum score		67	-	1.06	1.52	0	5.0
Preschool experience (years)		70	-	7.21	6.50	0.5	33.0
Teaching experience (years)		70	-	9.44	6.93	.50	26.0
Gender	Female	65	92.9%	-	-	-	-
	Male	5	7.1%	-	-	-	-
SCHOOLS							
School IMD rank (where 1=most deprived)‡		61	-	7582.0	7938.2	100	30461
% children eligible for free school meals (FSM)‡		61	-	24.3	12.5	#	#
Intervention group status	Intervention	26	42.6%	-	-	-	-
	Control	35	57.4%	-	-	-	-
CHILDREN							
Age in school year (months)♦		797	-	6.0	3.5	0.5	11.5
IDACI score (where 1=most deprived) †,‡		797	-	0.33	0.16	#	#
Gender	Female	401	50.3%	-	-	-	-
	Male	396	49.7%	-	-	-	-
Ethnic group	White	539	67.6%	-	-	-	-
	Asian	139	17.4%	-	-	-	-
	Black	57	7.2%	-	-	-	-
	Mixed / other	62	7.8%	-	-	-	-
Home language exposure	English	593	74.4%	-	-	-	-
	Not English	204	25.6%	-	-	-	-
Special Educational Need or Disability (SEND)	SEN	128	16.1%	-	-	-	-
	No SEN	669	83.9%	-	-	-	-
Attendance pattern	Part-time	< 10	< 1.3%	-	-	-	-
	Not part-time	> 787	> 98.8%	-	-	-	-
Eligible for free school meals (FSM)‡	Eligible	204	25.6%	-	-	-	-
	Not eligible	593	74.4%	-	-	-	-
Eligible for Early Years Pupil Premium (EYPP)‡	Eligible	201	25.2%	-	-	-	-
	Not eligible	596	74.8%	-	-	-	-
School intervention group status	Intervention	362	45.4%	-	-	-	-
	Control	435	54.6%	-	-	-	-
OUTCOMES AT END RECEPTION (AGE 5): POST-INTERVENTION							
BPVS		797	-	70.9	15.1	18.0	118.0
CELF Sentence Structure		797	-	15.5	3.8	0.0	22.0
RAPT Information Score		797	-	28.0	4.6	7.5	38.0
RAPT Grammar Score		797	-	22.0	5.1	2.0	34.0
EYFSP Good Level of Development		797	-	24.2	5.46	12	36
EYFSP Communication/ Language		797	-	6.24	1.62	3	9
EYFSP Literacy		797	-	3.78	1.20	2	6
OUTCOMES IN YEAR 6 (AGE 6)							
Year 1 phonics check score		797	-	34.0	8.91	#	40
LANGUAGE SKILLS ON ENTRY TO NURSERY (AGE 3): PRE-INTERVENTION							
BPVS		797	-	39.6	16.0	2.0	88.0
CELF Sentence Structure		797	-	9.0	4.7	0.0	21.0
RAPT Information Score		797	-	20.2	7.5	0.0	39.0
RAPT Grammar Score		797	-	11.9	6.6	0.0	30.0

* SD = standard deviation # Min/max figures not shown where these are shared by <10 cases to preserve confidentiality

† The IDACI is drawn from the IMD but includes only measures relevant to child income deprivation.

‡ Measures of socio-economic status/social deprivation.

♦ This variable captures the age of the child in the school year. In studies which cross academic years, it can be more relevant a variable than the child's absolute age. In this study, since all children were in the same academic year, it is essentially the same as the child's absolute age.

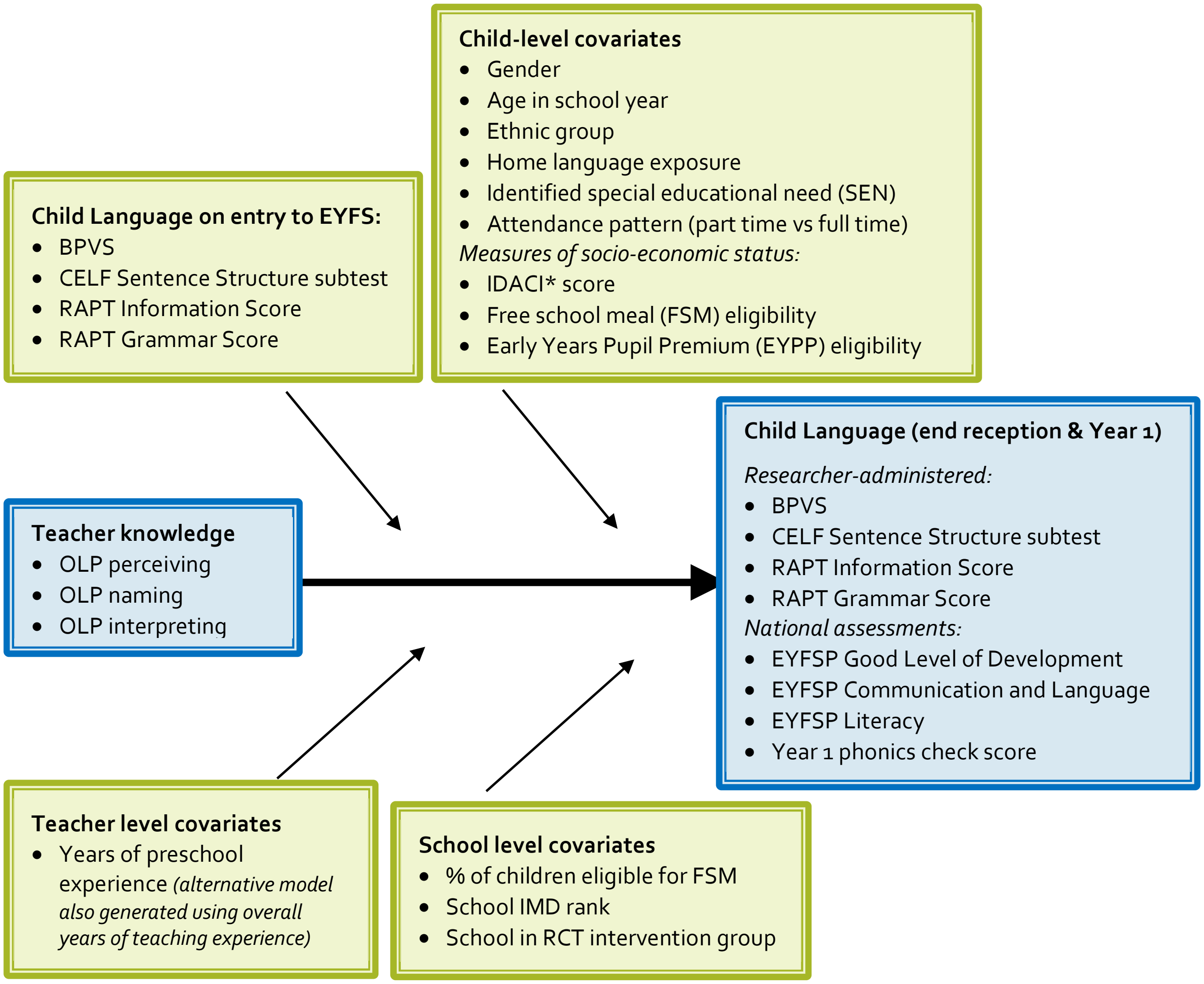
Children's standardised score from the national phonics check completed in Year 1 (age 6) was also drawn from the NPD for the academic year 2018-19 (possible range 0-40). Descriptive data for all child language variables are shown in Table 1. Appendix A2 presents a comparison of children in the OLP study sample, and children in the wider RCT but not in the study sample. Study children were largely representative of the wider RCT sample, with some differences identified. For example, OLP study children were slightly less disadvantaged on average.

4.3 Predictors of child outcomes and outline of analysis strategy

Figure 2 summarises the analytical models, with further detail provided in Appendix A4. Linear mixed-effects regression models were fitted to establish whether teachers' knowledge measured at the start of the reception year predicted child outcomes at the end of reception (and at the end of Year 1 for the national phonics assessment).

The main predictors were the three OLP factors *(perceiving, naming, interpreting)* reflecting the procedural knowledge of children's reception class teachers. The analysis controlled for children's oral language on the BPVS, CELF and RAPT at the start of the nursery year, meaning that models essentially assessed children's language *progress* across the two years of the EYFS. Baseline language skills were accounted for in all models, including those using national assessments as an outcome. A limitation of the study is the fact that most children would have had two *different* teachers during the EYFS period: one in their nursery year and one in their reception year. While OLP scores were available for all of the children's reception class teachers, data on the knowledge of nursery class teachers were only available for approximately 200 of the 797 children in the sample. As a result, it was considered more robust to include only the data on children's reception class teachers, leaving the equivalent knowledge of children's nursery class teachers (i.e., their scores on the OLP) unaccounted for.

Figure 3: Summary of child outcomes and the explanatory variables included in the models



* The IDACI is drawn from the IMD but includes only measures relevant to child income deprivation.

A range of child, teacher and school factors relevant to children's language progress were controlled for (Figure 2). For example, children's oral language skills will – on average – depend greatly on their socio-economic background or whether they speak English as an additional language (EAL). Selection effects might also come into play. If children with stronger language skills due to their home background are also more likely to attend schools with more knowledgeable teachers, then the effects of teacher knowledge might be overestimated unless we control for children's background. Children's records for 2017-18[f] within the NPD were drawn upon to provide data on their home language exposure, ethnic group, special educational needs (SEN) status, attendance pattern and a range of measures of socio-economic status (Figure 2). One school-level measure of

[f] Occasionally data from prior or later years were used in order to ensure as complete a dataset as possible (see Appendix 1.3).

deprivation was also accessed from the NPD: the proportion of children at each school who were eligible for free school meals (FSM). The wider RCT provided data on: child gender and age in the school year; teachers' years of preschool experience and overall teaching experience; and the IMD rank of the school. Further details on variables derived from the NPD are provided in Appendix 1.3.

Although Figure 2 shows all the child language measures in the same diagram, in reality a separate model was generated for each of the eight outcomes. For each child outcome, a model was first run including each of the three OLP factor scores (*perceiving etc*) alone: these were the **individual (univariate) models**. Where a significant association was identified between one or more of the OLP factors and the relevant child outcome, a fourth model was run which included all three OLP factor scores together: the **combined (multivariate) model.**

Multi-level models were used to account for the fact that multiple children per teacher were included in the analysis. The language outcomes of children in the same teacher's class may be more similar than those of children from different classes because of factors related to that teacher (teacher effects). Since some schools had responses from more than one teacher, it was also necessary to consider possible school effects – that is, the possibility that OLP scores for teachers leading different classes in the same school (or the language outcomes of their children) were more similar than those from different schools, due to some school-related characteristic. This might arise because the school has a certain ethos or draws children from the same catchment area, or because teachers in that school have all attended the same professional development course. Both two- and three-level models were fitted: the former controlling for clustering of children within classes (teacher effects), the latter also controlling for the clustering of classes within schools (school effects), in schools where more than one teacher completed the OLP. The results were very similar, likely because there were so few schools with multiple teacher respondents. Results from the two-level models are summarised below. Detailed results from both the two- and three-level models are shown in Appendix A5. As a final robustness check, linear models were fitted using the same variables as the mixed-effects models, with cluster- robust standard errors employed to allow for teacher and school effects. These made no difference to the statistical significance of the results. Further information can be found in the Appendix on procedures for dealing with missing data (A3) and on the analytical models (A4), including a number of sensitivity analyses conducted.

5. Findings

Relationships between teacher knowledge *(perceiving, naming, interpreting)* and children's language outcomes are summarised in Figures 3-5. The histograms show the size of the effects and the error bars show a 95% confidence interval.[9] The full models are shown in Appendix A5.

Figure 3 shows findings from the individual (univariate) models for each researcher-administered language assessment (i.e., separate models for each of the three OLP factors). Significant associations were identified for the BPVS *(perceiving and interpreting)* and for the CELF *(interpreting)*. This means that children whose teachers tended to be better at identifying salient strategies in videos (*perceiving*) made more progress in their receptive vocabulary across the EYFS. Children whose reception class teachers were more likely to *interpret* the interactions they observed in the videos made more progress in their receptive vocabulary and their receptive grammar. All associations identified were small (standardised betas = 0.19-0.23)[h], with a change of 1 standard deviation (SD) in the relevant OLP factor corresponding to a change of 0.1 (10%) of a SD in the relevant outcome measure in each case.

The association between BPVS scores and *naming* was on the borderline of significance in the individual (univariate) model but non-significant for all other outcomes. This suggests that children whose teachers used specialist vocabulary to describe the strategies they noticed did no better than children whose teachers used informal vocabulary.

When all three OLP factors were included in the same model (Figure 4), only the significant association between *interpreting* and the BPVS remained (standardised beta = .21). This means that, when all facets of teachers' knowledge were considered together, *interpretation* was more important than either *perceiving* or *naming* in relation to children's language progress.

Figure 5 shows findings from the individual models for the teacher-administered national assessments: the EYFSP and the Year 1 phonics score. No significant associations were identified between teacher knowledge, as measured by the OLP, and any of the teacher-administered measures.

[9] The confidence interval (CI) means that, if the study were repeated infinitely (in theory) and a 95% CI calculated each time, then 95% of these intervals would contain the population effect, and 5% would not. If the 95% CI does not include the value zero, the effect is described as being statistically significant; that is, we can be reasonably confident that the association between teachers' knowledge and the relevant child outcome is not simply a chance association.

[h] Following the guidelines proposed in Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences,* Routledge Academic, New York. As the standardised betas are defined here, they correspond to twice the value of the Cohen's d effect size statistic, giving approximate thresholds for small, medium and large effects of 0.1, 0.25 and 0.4 units.

Figure 4: Associations between OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*, *RAPT*)
*Standardised betas from the individual (univariate) models. Significant effects marked #(p<.10) *(p<.05) **(p<.01) ***(p<.001)*

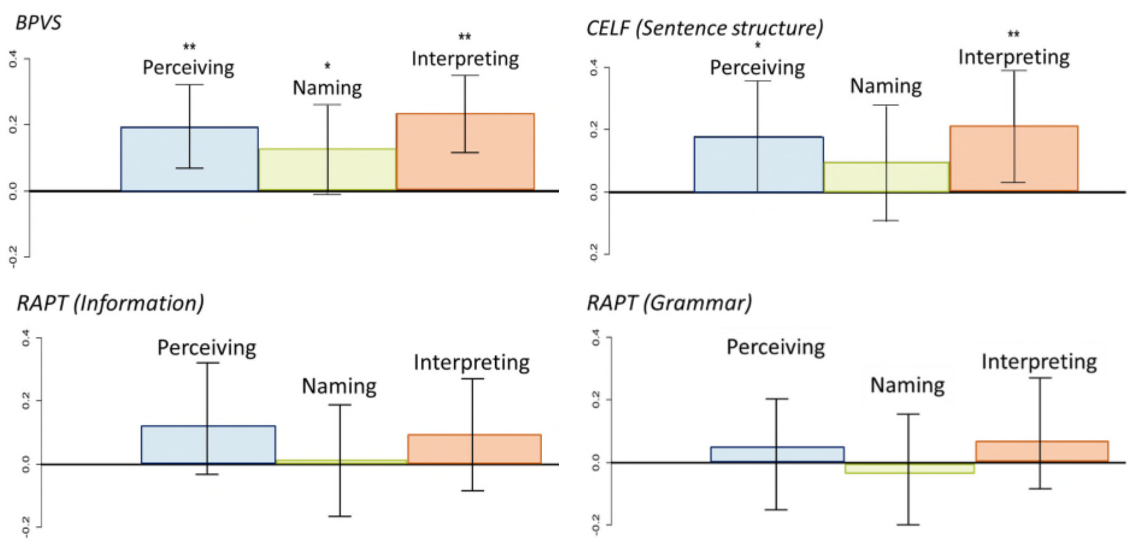


Figure 5: Associations between OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*)
*Standardised betas from the combined (multivariate) models. Significant effects marked #(p<.10) *(p<.05) **(p<.01)*

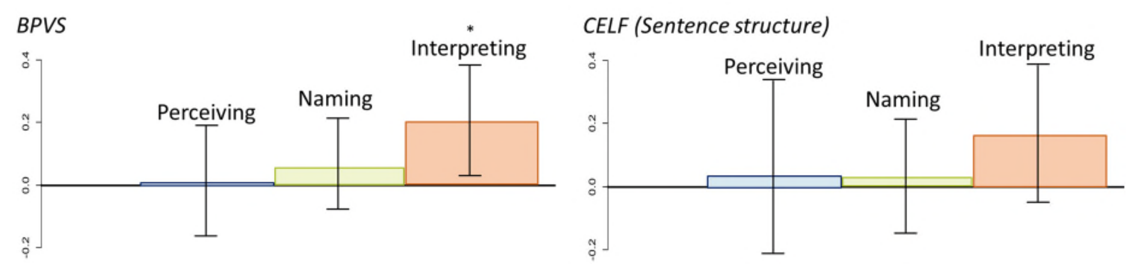
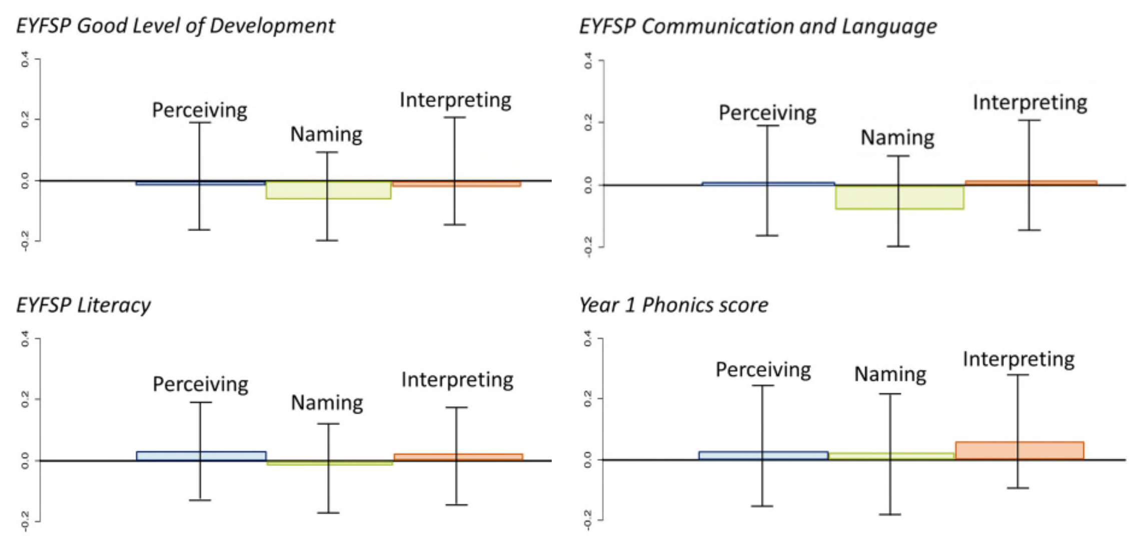


Figure 6: Associations between OLP factors and national assessment scores (EYFSP, Year 1 phonics score)
*Standardised betas from the individual (univariate) models. Significant effects marked #(p<.10) *(p<.05) **(p<.01) ***(p<.001)*



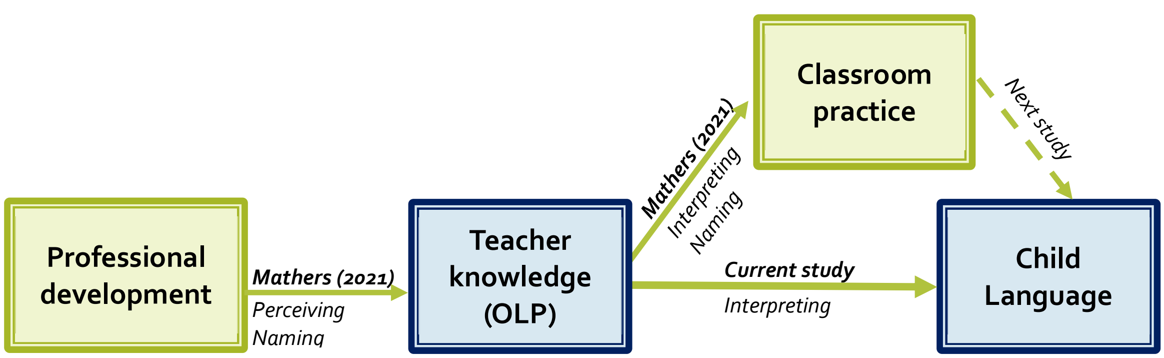
6. Discussion and conclusions

This analysis shows that early educators' knowledge of oral language strategies matters for young children's language outcomes. Reception class teachers' knowledge predicted gains in children's understanding of vocabulary and, to a lesser extent, their sentence structure between ages three and five.

Since the OLP involves responding to videos of real adult-child interactions, it represents a measure of dynamic knowledge which can be unlocked and used 'in the moment'. The first validation study[1] had already established that such knowledge predicts higher-quality classroom practice (Figure 6). Teachers who were better at identifying salient language-supporting strategies in videos (specifically, at *naming* these using expert terminology) led more language-supporting classrooms. Teachers who could explicitly connect their knowledge of pedagogy and child development (*interpret*) also led higher-quality classrooms, potentially because they were more likely to use strategies intentionally when interacting with children in support of specific language goals.

The current study confirms that pedagogical knowledge (as measured by the OLP) also predicts children's oral language progress. While educators' ability to identify salient strategies in videos (*perceiving*) predicted child language outcomes when entered alone into the statistical model, *interpreting* was the strongest predictor when all three knowledge facets were entered together into the same model. Therefore, although knowing which strategy is appropriate in a particular context matters, understanding *why* such strategies might be effective matters more for child language outcomes. Given findings from the initial validation (Mathers 2021, Figure 6), it is plausible that this association is mediated

Figure 7: Findings from the first two stages of OLP validation (the current study and Mathers, 2021)



by classroom quality – that is, teachers who can explicitly connect their knowledge of pedagogy and child development to interpret video interactions are also more likely to use language-supporting strategies intentionally when interacting with children, with resultant benefits for child language outcomes. This mediation pathway has not been tested as part of the current study but will be examined as the next stage of work.

All associations identified were small, with a change of 1 standard deviation (SD) in the relevant OLP factor corresponding to a change of 0.1 (10%) of a SD in the relevant outcome measure. This modest relationship is not surprising when one considers the context. Of the two years each child spent in the Early Years Foundation Stage, this analysis captures only the knowledge of the class teacher in their second (reception) year. It does not reflect the knowledge of others shaping the child's language environment during this period: their nursery class teacher or any teaching assistants in the nursery or reception class. Secondly, although covariates were included in the models to account for children's own characteristics and family background, these will be imperfect and cannot fully eliminate the effects of children's wider environment on their oral language progress. And finally, while knowledge is important, there is a substantial theoretical and practical gap between knowledge and practice. Even when a teacher has the required procedural knowledge, the degree to which they actually enact this knowledge in the classroom to the benefit of child language outcomes will depend on myriad other factors (e.g., the school context, staff team dynamics, teachers' own beliefs and feelings of self-efficacy). With these limitations in mind, the fact that an association between teacher knowledge and child language progress across the EYFS has been identified is worthy of attention.

It is also worth noting that teacher knowledge was more strongly associated with receptive language (understanding) than expressive language (speaking). This may stem from the fact that children's expressive language is more challenging to influence, or from differences between the receptive and expressive language measures used in the study. The RAPT involves potentially greater subjectivity in scoring than either the BPVS or the CELF, in which scores are based simply on whether children point to the correct picture. Although inter-rater agreement procedures were carried out as part of the original RCT to ensure consistency in RAPT scores,[7] it is likely that more 'noise' existed than in the BPVS or CELF data which may have made it more challenging to detect effects.

No relationships were identified between teacher knowledge and children's achievement on national assessments, as measured by the EYFSP and Year 1 phonics test. It is likely that this also reflects a measurement issue. Although teacher-completed EYFSP scores have been found to predict children's performance on national tests at age 7 and to correlate with researcher-administered tests[i], they are nonetheless recognised as being less reliable than standardised researcher-administered assessments[33]. Although moderation between teachers is built into the system, differences will exist in the way individual teachers complete the assessments. There is also some evidence that teachers can under-estimate the achievement of children from minority backgrounds.[34] This variability in EYFSP scores may make it harder to cleanly detect associations with teacher knowledge.

It is also worth noting that no significant effects were identified for the third OLP facet: *naming.* Children whose teachers used specialist vocabulary *(e.g., open questions, descriptive commentary)* to describe the strategies they noticed did no better than children whose teachers used more informal vocabulary. To a large degree this makes sense, since use of specialist terms will vary greatly – with educators using different terminology depending on where and when they trained, on which professional development courses they have attended, and on informal language conventions in schools and professional communities. However, in the first stage of the OLP validation (Figure 6), *naming* was the strongest predictor of language-supporting quality in teachers' classrooms. It was reasoned that having formal vocabulary with which to name a concept might help teachers to engage in explicit reflection and discussion on that concept, supporting deeper understanding and intentional practice. This contradiction is somewhat puzzling. One explanation may lie in the fact that the measures of quality used in the RCT captured *classroom* quality (i.e., the practice of the whole team) rather than focusing solely on the practice of the individual teachers whose knowledge had been measured. It may be that having a professional vocabulary relating to oral language is less important in shaping individual practice; and instead helps teachers to articulate their knowledge to the wider team and explain why certain practices are important. This may support overall classroom quality in a way which is measurable but not 'strong' enough to feed through into measurable benefits for children. We can see, for example, in Figure 3 that the associations between *naming* and the BPVS (+.13) and CELF (+.09) were positive but smaller than those found for *perceiving* and *interpreting*. To conclude, it is clear that teachers' professional vocabulary, while relevant for classroom quality, is not a prime driver of child language outcomes.

[i] Correlations in the region of 0.6 for this study, on a scale running from 0 (no association) to 1 (perfect correlation).

The final issue which must be considered is how pedagogical knowledge might be improved. The initial validation study concluded that teaching experience (i.e., more years in the classroom) did not necessarily mean greater pedagogical knowledge. The professional development programme which was evaluated within the wider RCT did bring benefits for knowledge, with gains in both *perceiving* and *naming.* However, no gains were identified in *interpreting:* the facet identified as being most strongly associated with children's language progress. The lack of observed gains is in some ways unsurprising, since the professional development programme did not have an explicit focus on classroom reasoning or analysis. However, it illustrates that such learning cannot be assumed, even in a formal professional development programme focused explicitly on oral language pedagogy and shown to develop other facets of pedagogical knowledge (*perceiving, naming*). It suggests that pedagogical reasoning may need to be explicitly nurtured. Although some valuable research exists in this area (e.g., using video to develop the analytical skills of primary mathematics teachers[35]), more work is needed to establish how pedagogical reasoning might be promoted among early educators, and in relation to early language.

6.1 Implications

1. To ensure early years educators are prepared to support young children's oral language, we need to invest in nurturing their early-language-related pedagogical knowledge (the dynamic, procedural knowledge which informs real-time classroom decision-making and action).

2. Specifically, educators should be supported in connecting their knowledge of language-supporting strategies with knowledge of child development, in order to understand **why** as well as **how** specific techniques support oral language development.

Promising avenues for attention might include:

For government and providers of qualifications and professional development for early years educators:

Ensuring that pre- and in-service professional development opportunities are available which explicitly nurture oral-language-related pedagogical knowledge and, specifically, those which promote pedagogical *reasoning*. This might include structured opportunities to reflect on and analyse practice in a way which makes the relationships between pedagogy and child outcomes explicit (e.g., observing adult-child interactions, either live or on video, to analyse child learning needs, strategies which might support learning and the success of strategies employed).[36]

For early years staff teams:

Setting aside regular time for professional conversations about language-supporting practice which make explicit and/or analyse relationships between pedagogy and young children's outcomes. This might range from informal end-of-day discussions to more structured activities (e.g., video-recording and analysing practice).

For researchers:

- Including measures of procedural knowledge in studies of early years educator qualifications and professional development, in order to facilitate understanding of how such knowledge can be enhanced; and also how such knowledge translates into action;
- Conducting further research into how pedagogical reasoning (*interpreting*) can be supported and enhanced through professional development.

TECHNICAL APPENDIX

A1. Further information on variables used for analysis

A1.1 Observing Language Pedagogy (OLP) instrument

The information in this section is drawn from the initial OLP validation study.[1,28] To assess ***perceiving***, teachers were prompted to report strategies (≤8) in each video which might support children's oral language. Responses were considered valid if they matched at least one strategy in the OLP framework (Table A1.1). Strategies were derived from a literature review and affirmed through expert review.[28] A single response could be credited as reflecting multiple strategies, and strategies could be described using informal language. A *perceiving* score was generated by multiplying each valid strategy by the relevant expert rating for that video, resulting in higher scores being awarded for strategies which experts rated as being both present and salient in each video (1=low, 5=high). In this way, the OLP measures respondents' ability to know when certain strategies are appropriate and what matters most for child language in a specific situation. Coding was conducted by the first author, with a proportion independently coded by a second researcher, trained to reliability on a proportion (35%) of actual responses. Independent coding was conducted on a further 35% of responses, with high levels of exact agreement (82-89%) and discrepancies resolved through discussion. Table A1.2 presents an illustrative set of coded responses. Descriptive statistics for the raw *perceiving* scores are shown in Table 1 within the main body of the report. These are represented as sum scores (i.e. the total of each respondent's score for videos 1 and 2). Strategy totals were normally distributed with a broad range. The variables actually used for analysis were factor scores generated via confirmatory factor analysis (CFA): see below for further details.

Informed by the work of Kersting and colleagues[37], the higher-order facets of *naming* and *interpreting* were not directly prompted. Although this risked under-representing teachers who could have offered an interpretation if prompted, it was reasoned that spontaneous use would reflect the knowledge most likely to be mobilised in real classroom situations — and thus most closely associated with actual practice. All valid responses (≤8 per video) were coded to reflect instances of professional vocabulary ***(naming)*** using a vocabulary list derived through expert review. Terms credited were defined prior to coding, refined following coding and then subjected to expert review. Multiple vocabulary terms could be credited within an individual response. Table A1.2 shows a coded example for one set of responses.

50% of responses were double-coded, with 100% exact agreement for all videos. Descriptives for the raw sum score are shown in Table 1.

Valid responses (≤8 per video) were also coded to reflect instances of ***interpretation***: either a possible pedagogical intention, an observed effect on the child or a possible alternative approach (Table A1.2). Half of responses were double-coded, with high levels of exact agreement (*naming*=100%; *interpreting*: 96%- 98%). The interpretation score is based on the number of responses for which a valid interpretation was provided. Outliers were replaced with a value at a defined upper threshold based on the interquartile range – a process known as Winsorisation.[38] Descriptives for the raw sum score are shown in Table 1.

Due to a high proportion of zero responses, both *naming* and *interpreting* displayed narrow ranges, low means (Table 1) and a positive skew. Nonetheless, scores did discriminate between respondents: 55.2% used at least one expert term (*naming*) and 43.3% gave at least one interpretation.

Confirmatory factor analysis (CFA) using full maximum likelihood estimation showed that the three facets could be empirically distinguished and were best represented by a three-factor structure. Loadings of observed variables onto latent variables were all ≥.6 and model fit was excellent [χ^2= 3.10 (df=4, p=.54), RMSEA = .00; CFI = 1.00; TLI = 1.03].[j] Factor scores based on the CFA were used for all analyses.

[j] χ^2 = Chi squared, RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index. Cut-off criteria (Hu & Bentler, 1999; Schreiber et al, 2006): χ^2 df<2, p>.05, RMSEA ≤.06, CFI and TLI ≥.90-.95.

Table A1.1 The 30 strategies underpinning the OLP coding framework

MODELLING LANGUAGE
1. Modelling diverse, rich or specific vocabulary
2. Modelling diverse, rich or specific grammar
3. Linguistic expansion or recasting of children's language
4. Emphasising, repeating or reinforcing language
5. Using descriptive, informative, narrative language in concrete contexts
6. General language modelling
WORD MEANINGS
7. Providing explicit definitions of words
8. Providing concrete clues to meaning
COMMUNICATION AND CONVERSATION
9. Engaging in conversation with children
10. Non-verbal strategies to invite communication
11. Verbal strategies to invite communication
12. Prompting children to use new vocabulary
13. Being a responsive conversation partner
14. Extending conversational or narrative content
15. Supporting mutual understanding/adapting language to child's level
16. Supporting children to attend and participate
17. Affirming the child's language by repeating it
HIGHER-ORDER LANGUAGE AND THINKING
18. Promoting children's thinking
19. Prediction, speculation, reasoning, explanation and inference
20. Modelling fictional narrative/pretending
21. Use of open questions
RELATIONSHIPS AND THE CHILD
22. Positive affect or communication
23. Individual attention and sensitive responding
24. Using a non-directive approach
25. Facilitating peer communication
26. Facilitating peer interactions and relationships
27. Promoting children's self-worth
MEANINGFUL AND ENGAGING CONTEXTS
28. Joint attention/following children's lead and interests
29. Meaningful/engaging contexts and activities for language
30. Deepening learning

Table A1.2. Generating the OLP perceiving, naming and interpreting scores (a worked example)

Perceiving (1 score per video):	**Naming (1 score per video):**	**Interpreting (1 score per video):**
each valid strategy identified is awarded the relevant 'expert example' rating for that video. Responses can match multiple strategies and be described using informal language (e.g., *how/why questions*). **Range:** no limit	1 mark awarded for each response which uses specific professional terms (defined by the manual) to describe the video interaction e.g., open questions, recasting, meta-cognition, descriptive commentary. **Range:** 0-8 per video	1 mark awarded for each response which includes a valid interpretation of the interaction, as defined by the manual. Valid interpretations include: a possible pedagogical intention, the observed effect on the child, possible alternative approaches. **Range:** 0-8 per video

Illustrative teacher response to Video 1	***Perceiving*** *(expert example rating)*	***Naming***	***Interpreting***
1. *"..The teacher uses key vocabulary e.g. slower, faster.. further. She continually repeats this vocabulary <u>and the child soaks up the language and begins to use it herself e.g. 'Further…yeah further'</u>"*	Strategy 1 – (4.00) Strategy 4 – (5.00)		1 (observed effect on child)
2. *"Using correct mathematical vocabulary – <u>cuboid not brick</u>"*	Strategy 1 – (4.00)		1 (alternative approaches)
3. *"Using prompts and extending children's responses"*	Strategy 11 – (4.33) Strategy 3 – (2.67)		
4. *"Giving a <u>running commentary</u> on the child's actions"*	Strategy 5 – (4.67)	1	
5. *"Questioning <u>to provide opportunities for child to apply words and show understanding linked to meaningful experience</u>"*	Strategy 12 – (4.67) Strategy 15 – (4.67) Strategy 29 – (4.33)		1 (pedagogical intention)
6. *"Using gestures – when saying steeper slope"*	Strategy 8 – (4.00)		
8. *"Listening carefully and valuing all the child's ideas"*	Strategy 13 – (5.00) Strategy 27 – (4.33)		
8. *"<u>Open ended questions</u> – how did you do that?"*	Strategy 21 – (4.67)	1	
Total score for Video 1	**56.34**	**2**	**3**

A1.2 Child outcome variables derived from the wider RCT

Researcher-administered language outcome variables

The following paragraphs provide additional information on the psychometric properties of the four researcher-administered language outcome measures used in this analysis, drawn from the Education Endowment Foundation (EEF)'s Early Years Measures Database.[39]

The **British Picture Vocabulary Scale (BPVS-3)**[30] measure of receptive vocabulary is appropriate for children aged 3 years to 16 years 11 months. It was standardised on a sample of 3,278 children in England, Wales, Scotland and Northern Ireland across 8 age ranges. *Reliability:* built into the confidence bands (confidence intervals 95%). *Construct validity:* correlations with WISC (-0.76) and Schonell (0.80). *Criterion validity*: correlated with the CATS verbal battery (0.72) and overall CATS scores (0.61). *Concurrent validity*: CDI (0.32-0.41).

The **Clinical Evaluation of Language Fundamentals (CELF-Preschool 2)**[31] is a battery of assessments designed to assess expressive and receptive language skills, designed for children aged 3-6 years. The sentence structure subtest was used in this study. It assesses children's understanding of spoken sentences of increasing length and complexity. The CELF-Preschool 2 was standardised on a sample of 486 children in the UK. *Reliability*: Test-retest reliability ranged from 0.77 to 0.96 (for ages 3 - 3 years 11 months) and 0.74-0.95 (for ages 4 - 4 years 11 months). Cronbach's alphas ranged from 0.77 to 0.95. Inter-rater reliability ranged from 0.95 to 0.97. *Criterion validity*: Co-normed with PLS-4 and CELF-4. *Construct validity*: the core language score (within which the sentence structure subtest sits) has a high correlation with other composites ranging from 0.85 to 0.93. Receptive and expressive language correlate 0.76. *Concurrent validity*: The study conducted to compare the CELF-5 with CELF-4 consisted of 1000 typically developing students between ages 5-16. Correlations between overall scores and index scores were high and ranged from 0.78 - 0.92.

The **Renfrew Action Picture Test (RAPT)**[32] is a measure of expressive vocabulary and grammatical features, appropriate for children aged 3-9 years. It is norm-referenced (norms collected in 1987). *Reliability:* Assessment of test-retest reliability revealed little difference in responses if the re-test was given within a month from the original. 3% of each set's score was affected by scoring discrepancy. *Criterion validity*: insufficient data in the public domain to evaluate. *Construct validity*: insufficient data in the public domain to evaluate. *Concurrent validity*: highly correlated with the Carrow Language Inventory for children aged 5 and 6 with moderate learning disabilities.

Inter-rater agreement was calculated for the RAPT scoring within the wider RCT to ensure consistency.[7] 10% of each rater's assessments were double-marked by a gold-standard moderator. Means and standard deviations (SDs) were compared, correlations calculated and scatterplots observed. Means and SDs were very similar: Information Score (rating team 28.27/ 5.08; gold-standard rater 28.73/ 4.97); Grammar Score (rating team=21.78/ 5.40; gold-standard rater=22.93/ 5.36). Correlations were high (Information Score=.96, Grammar Score=.90), confirmed by the linear patterns of both scatterplots.

Other variables considered but not used

Data on teacher-reported socio-behavioural development were also gathered as part of the wider RCT, using the Adaptive Social Behaviour Inventory (ASBI).[40] Post-test data were only available for 673 of the 797 children (84.4%). Exploratory analyses indicated that teacher knowledge was not associated with children's scores on the ABSI. On the basis that the theoretical case for examining relationships between teachers' oral-language-related pedagogical knowledge and children's socio-behavioural outcomes is less strong than that for children's oral language outcomes, a full analysis using the ABSI data was not completed.

A1.3 National Pupil Database (NPD) variables

Variable	Description	Structure
Child-level variables		
Gender	-	Male Female
Age in school year	Derived from age in months and calculated as follows: September birthday => age in school year = 11.5 (months) October birthday => age in school year = 10.5 (months)…… …..July birthday => age in school year = 1.5 (months) August birthday => age in school year = 0.5 (months)	Takes values 0.5, 1.5, … 11.5
Ethnic group	Derived from the *Ethnicity* variable, drawn from Spring 2018 school census, supplemented by the Spring 2017 and Spring 2019 censuses where the variable was missing in the Spring 2018 census.	White Asian Black Mixed / Other
First language	Derived from the *Language Group Major* variables in the 5 school censuses between Spring '17 and Summer '18 (LanguageGroupMajor_AUT17, SPR17, SUM17, SPR18 and SUM18). Where there were differences between censuses, the outcome recorded in the majority of the 5 censuses was used. The NPD *First Language* variable from which the *Language Group Major* is derived is defined as '*The language to which the child was exposed during early development and continues to use in the home or in the community. If a child acquires English subsequent to early development, then English is not their first language no matter how proficient in it they become'.* A possible alternative variable from the RCT dataset was considered, also recording language exposure. Due to small amounts of missing data in the RCT variable (2.6%) the NPD data were used. The correlation between the NPD and RCT variables was .73.	English Not English
Identified special educational need or disability (SEND)	Derived from the *SEN provision Major* variable. A child was taken to have Special Education Needs (SEN) if a SEN was recorded in any school census up to Summer '18. This included children with a record of a SEN whether or not a statement had been issued.	SEN No SEN
Attendance pattern (part time vs full time)	Derived from the *Part Time* variable Spring 2018 school census.	Part time Not part-time
IDACI score	Children's Income Deprivation Affecting Children Index (IDACI) scores reflect the mean *IDACI scores* recorded in the Autumn '17, Spring '18 and Summer '18 school censuses.	Continuous
Free school meal (FSM) eligibility	Two variables were derived: • Child eligible for FSM (*FSM eligible):* child recorded as eligible for FSM in at least one of the Autumn '17, Spring '18 or Summer '18 censuses. • Child ever eligible for FSM (*EVERFSM_6_P*): child recorded as eligible for FSM in any termly or annual census in the last 6 years up to the pupil's current year (in this case, up to Spring '19). These variables were strongly correlated (r = 0.93), likely because children were only in the reception year (i.e. had not attended for many years). The former was used in the models.	Eligible Not eligible
Early Years Pupil Premium (EYPP) eligibility	Data derived from the *EYPP eligible* variables from the Spring 2017 and Spring '18 censuses. A child was considered eligible if recorded as eligible in either census.	Eligible Not eligible
Absence from school	Data on children's absence from school was requested with the intention of using this in the analysis. However, data for children's reception year (2017-18) was very incomplete. It was not used in the models.	NA
School-level variables		
% of children at school eligible for FSM	Derived from the *LEA(yy)_Pct_Pupils_FSM_Eligible* variable, reflecting the % of pupils known to be eligible for FSM in the Spring '18 census. Two alternative variables were considered: % of children eligible (from the RCT dataset) and % of children taking up FSM (*LEA(yy)_Pct_Pupils_Taking_FSM.* Correlations between all three were strong (r≥ 0.79)	Continuous

A2. Assessing whether the study sample is representative of the wider RCT sample

Relevant teacher and child variables were compared to investigate whether there were differences between those in the study sample, and those in the wider RCT but not in the study sample. Teacher variables were compared between the 70 teachers in the current sample, and the 213 teachers in the wider RCT sample but not in the current sample. The means of continuous variables were compared using t-tests. The proportion of each category of the categorical variables was compared using a test for equality of proportions. No significant differences were identified for the aspects compared (Tables A2.1 & 2.2), suggesting that the sub-sample adequately represents the wider population of teachers within the RCT. It was not possible to make comparisons for gender, years of teaching experience or years of preschool experience.

Children's variables were compared for the 797 children in the current sample, and the 1181 children in the wider RCT sample but not in the current sample. Children in the current sample were somewhat more advantaged than non-study children, as reflected in lower mean IDACI scores (0.33 vs 0.35) and lower proportions eligible for Free School Meals (25.6 vs 30.3%) and the Early Years Pupil Premium (25.2 vs 29.8%). They also had higher mean RAPT Grammar scores at post-test (21.91 vs 21.37).

Table A2.1: Teacher comparison: continuous variables

TEACHER	Teacher in current sample			Teacher not in current sample			p-value from t test
	N	Mean	SD	N	Mean	SD	
School IMD rank (where 1 = most deprived)	70	8275.51	8211.53	213	7149.47	7819.19	.30
School IDACI rank (where 1 = most deprived)	70	9563.27	9339.82	213	7609.79	7995.26	.09
Days of training attended (intervention group only)	31	4.48	2.16	113	3.99	2.32	.29

Table A2.2: Teacher comparison: categorical variables

Variable	Level	Teacher in current sample		Teacher not in current sample		p-value for test of equality of proportions
		N	%	N	%	
Intervention group status	Intervention	31	44.3	114	53.5	.18
	Control	39	55.7	99	46.5	.18

Table A2.3: Child comparison: continuous variables

Variable	Children in current sample			Children not in current sample			p-value from t test
	N	Mean	SD	N	Mean	SD	
Age in school year	797	5.96	3.48	1180	5.73	3.47	0.166
IDACI score (where 1= most deprived)	795	0.33	0.16	1175	0.35	0.15	**0.004****
BPVS (post-intervention)	797	70.93	15.14	1181	69.63	14.92	0.059
CELF Sentence Structure (post-intervention)	797	15.50	3.76	1181	15.18	3.77	0.060
RAPT Information Score (post-intervention)	797	27.98	4.62	1181	27.66	5.28	0.151
RAPT Grammar Score (post-intervention)	797	21.91	5.11	1181	21.37	5.63	**0.026***
BPVS (pre-intervention)	687	39.56	16.01	970	38.47	16.12	0.173
CELF Sentence Structure (pre-intervention)	670	8.99	4.68	945	8.79	4.56	0.383
RAPT Information Score (pre-intervention)	679	20.21	7.48	954	20.35	7.23	0.708
RAPT Grammar Score (pre-intervention)	679	11.94	6.57	954	11.86	6.24	0.811
EYFSP Good Level of Development	797	24.18	5.46	1175	24.05	5.55	0.619
EYFSP Communication/ Language	797	6.24	1.62	1175	6.22	1.65	0.842
EYFSP Literacy	797	3.78	1.20	1175	3.73	1.18	0.418
Year 1 phonics check score	784	33.95	8.91	1127	33.48	9.19	0.266

Table A2.4: Child comparison: categorical variables

Variable	Level	Children in current sample		Children not in current sample		p-value for test of equality of proportions
		N	%	N	%	
Gender	Female	401	50.3	596	50.5	0.984
	Male	396	49.7	585	49.5	0.984
	Total	797	100.0	1181	100.0	
Ethnic group	White	539	67.6	797	67.7	1.000
	Asian	139	17.4	211	17.9	0.828
	Black	57	7.2	67	5.7	0.224
	Mixed / other	62	7.8	102	8.7	0.537
	Total	797	100.0	1177	100.0	
Home language exposure	English	593	74.4	902	76.5	0.310
	Not English	204	25.6	277	23.5	0.310
	Total	797	100.0	1179	100.0	
Special Educational Need or Disability (SEND)	SEN	128	16.1	165	14.0	0.223
	No SEN	669	83.9	1016	86.0	0.223
	Total	797	100.0	1181	100.0	
Attendance pattern	Full time	> 787	> 98.75	1095	93.0	**<0.001*****
	Part time	< 10	< 1.25	82	7.0	**<0.001*****
	Total	797	100.0	1177	100.0	
Eligible for free school meals (FSM)	Eligible	204	25.6	357	30.3	**0.025***
	Not eligible	593	74.4	820	69.7	**0.025***
	Total	797	100.0	1177	100.0	
Eligible for Early Years Pupil Premium (EYPP)	Eligible	201	25.2	351	29.8	**0.031***
	Not eligible	596	74.8	828	70.2	**0.031***
	Total	797	100.0	1179	100.0	
School intervention group status	Intervention	362	45.4	623	52.8	**0.002****
	Control	435	54.6	558	47.2	**0.002****
	Total	797	100.0	1181	100.0	

A3. Procedures for dealing with missing data

Five of the model covariates had missing data. The four pre-intervention language measures (BPVS, CELF, RAPT Information Score, RAPT Grammar Score) were missing for 14.8-15.9% of children. The IDACI covariate was missing for 0.25% of the children.

The presence of missing data on these variables was controlled for by using multiple imputation. Subject to the assumption that the data are missing at random, the analysis of multiply imputed data produces unbiased results, and is therefore generally to be preferred to the analysis of complete cases.[41] Missing data were imputed using the Amelia II package.[42] The imputation model assumes a multivariate normal distribution for the complete data (missing and observed). Binary, categorical and ordinal variables are incorporated into this distribution using appropriate transformations.[43] Whilst the use of the multivariate normal distribution is inevitably an approximation, its effectiveness in missing data problems is well established.[44] An imputation model was fitted including all outcomes and covariates. Ten imputations were generated, and models fitted to each imputed data set. Model results were consolidated using Rubin's Rules,[45] with degrees of freedom calculated using the method of Hesterberg.[46]

The age 6 phonics outcome was missing for 13 children. This variable was not imputed as it is not considered statistically advantageous to impute the value of outcome variables.[47]

Models were fitted either to multiply imputed data or using complete cases data.

A4. Further information on analytical models

A4.1 Intra-cluster correlations and design effects

The intra-cluster correlations (ICCs) for the outcome variables were calculated with respect to the clustering within teachers. From these, Design Effects (DEs) were found as follows:

$$DE = 1 + ICC(n - 1)$$

where *n* is the mean size of the clusters (i.e. mean number of children per teacher).

The Design Effect corresponds to the following ratio:[48]

$$\frac{squared\ standard\ error\ under\ multilevel\ model\ design}{squared\ standard\ error\ under\ standard\ design}$$

Values of the Design Effect close to 1 indicate that a multilevel model may not be necessary to model the data. Values of 2 or more indicate that a multilevel model is required.

Table A4.1: Intra-cluster correlations and design effects

Outcome	ICC	Design Effect
BPVS	0.118	2.22
CELF Sentence Structure	0.133	2.38
RAPT Information Score	0.075	1.78
RAPT Grammar Score	0.061	1.63
EYFSP Good Level of Development	0.075	1.78
EYFSP Communication/ Language	0.057	1.59
EYFSP Literacy	0.056	1.58
Year 1 phonics check score	0.078	1.81

A4.2 Two-level model specification

For teachers

$j = 1, 2, \ldots T$

And for pupils within teachers

$k = 1, 2, \ldots N_j$

The outcome variable is given by

$y_{jk} = \text{fixed effects} + \beta_j + \epsilon_{jk}$

Where β_j and ϵ_{jk} are independent and Normally distributed.

Full results for the two-level models are shown in Appendix A5 (Tables A5.1 and A5.2).

A4.3 Three-level model specification

A separate random effect for school was fitted where there was more than one teacher at the school.

For schools
$i = 1, 2, \ldots S'$
where $i = 1$ for all schools with only a single teacher.
And teachers within schools
$j = 1, 2, \ldots T_i$

And for pupils within teachers
$k = 1, 2, \ldots N_{ij}$
The outcome variable is given by
$y_{ijk} = \textit{fixed effects} + \alpha_i + \beta_{ij} + \epsilon_{ijk}$
Where α_i, β_{ij} and ϵ_{ijk} are independent and normally distributed

Full results for the three-level models are shown in Appendix A5 (Tables A5.3 and A5.4).

A4.4 Standardisation

Models were fitted with standardised coefficients which allow the value of model coefficients to be compared between different covariates and between different models. The following method was used:

1. Outcome variables were standardised to have mean 0 and standard deviation 1
2. Continuous covariates were standardised to have mean 0 and standard deviation 0.5
3. Categorical (factor) variables were coded as a series of 0 / 1 valued dummy variables

This gives the following interpretation to the model coefficients (beta):

1. For continuous covariates, beta gives the change in the outcome in units of its standard deviation corresponding to a change of two standard deviations in the model covariate.
2. For the factor variables, beta gives the change in the outcome in units of its standard deviation corresponding to the difference between a given level of the factor and the reference level.

The reason for standardising the continuous covariates to have standard deviation 0.5 rather than standard deviation 1 is that this produces beta values which are more comparable between the continuous and factor covariates.

A4.5 Percentage of variance explained

The percentage of variance explained by the models was found using the R^2 statistic. There are two versions of the R^2 statistic for linear mixed-effects models: (a) marginal R^2, which gives the percentage of variance explained by the fixed-effects only, and (b) conditional R^2, which gives the percentage of variance explained by the entire model (fixed- and random-effects). Both versions of the R^2 statistic are reported.

A4.6 Sensitivity analyses

Of the 70 teachers, 11 job-shared, with 148 of the 797 study children taught by job sharing teachers. Two of the teachers in the sample also gave a joint response to the OLP, with 19 children in the sample taught by these teachers. It is plausible that either of these situations may dilute or otherwise influence the association between teacher knowledge and child outcomes. In order to investigate possible bias, the individual (univariate) models based on multiply imputed data were fitted to three different data sets (see Appendix A5, Table A5.5):

I = All children (n = 797)

II = Sample removing teachers who gave a joint response (n = 778)

III = Sample removing teachers who gave a joint response and job-sharing teachers (n = 630)

A second sensitivity analysis was conducted to examine the effect of using an alternative measure of teacher experience. An alternative model was fitted using the covariate 'total years of teaching experience' instead of 'years of preschool experience', and the two sets of models were compared. Results are shown in Appendix 5 (Table A5.6). Switching the covariate made no difference to the statistical significance of the results.

A4.7 Software

Analyses were performed in R 4.0.2.

A5. Detailed analyses and results

Table A5.1: Associations between the OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*, *RAPT*) – two-level individual (univariate) models fitted to multiply imputed data

Outcome	OLP Factor	Beta	SE	95% CI	p-value	R2a	R2b
BPVS	Perceiving	+0.193	0.063	(+0.067,+0.319)	0.004**	49.9%	50.9%
	Naming	+0.125	0.067	(-0.012,+0.262)	0.073	49.4%	50.9%
	Interpreting	+0.233	0.057	(+0.117,+0.350)	<0.001***	50.6%	50.7%
CELF Sentence Structure	Perceiving	+0.177	0.089	(-0.000,+0.354)	0.050	36.5%	41.8%
	Naming	+0.094	0.092	(-0.091,+0.280)	0.312	35.9%	41.9%
	Interpreting	+0.210	0.089	(+0.032,+0.389)	0.022*	36.7%	41.8%
RAPT (Information)	Perceiving	+0.131	0.089	(-0.047,+0.310)	0.147	26.8%	31.0%
	Naming	+0.011	0.091	(-0.172,+0.194)	0.906	26.4%	31.1%
	Interpreting	+0.107	0.090	(-0.074,+0.288)	0.240	26.7%	30.9%
RAPT (Grammar)	Perceiving	+0.053	0.096	(-0.140,+0.246)	0.584	29.5%	35.9%
	Naming	-0.045	0.097	(-0.240,+0.150)	0.645	29.4%	35.8%
	Interpreting	+0.079	0.097	(-0.115,+0.272)	0.420	29.6%	35.8%
EYFSP Good Level of Development	Perceiving	-0.008	0.090	(-0.189,+0.173)	0.928	42.7%	49.1%
	Naming	-0.079	0.090	(-0.259,+0.102)	0.385	42.8%	49.1%
	Interpreting	-0.013	0.092	(-0.198,+0.171)	0.885	42.6%	49.1%
EYFSP Communication and Language	Perceiving	+0.011	0.087	(-0.163,+0.186)	0.898	40.6%	45.9%
	Naming	-0.094	0.087	(-0.268,+0.080)	0.282	40.7%	45.8%
	Interpreting	+0.011	0.088	(-0.166,+0.188)	0.901	40.6%	45.9%
EYFSP Literacy	Perceiving	+0.027	0.090	(-0.153,+0.208)	0.765	32.7%	38.1%
	Naming	-0.015	0.090	(-0.195,+0.166)	0.872	32.7%	38.1%
	Interpreting	+0.022	0.091	(-0.160,+0.205)	0.806	32.7%	38.1%
Year 1 phonics check score	Perceiving	+0.038	0.102	(-0.165,+0.241)	0.710	22.2%	29.5%
	Naming	+0.025	0.101	(-0.178,+0.227)	0.808	22.2%	29.5%
	Interpreting	+0.075	0.101	(-0.127,+0.277)	0.462	22.4%	29.5%

Beta = point estimate of model coefficient, SE = standard error of model coefficient, CI = 95% confidence interval for model coefficient
p-value = p-value of model coefficient, with statistical significance indicated by stars: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.
R2a = % of variance explained (fixed effects only), R2b = % of variance explained (fixed and random effects)

Table A5.2: Associations between the OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*, *RAPT*) – two-level combined (multivariate) models fitted to multiply imputed data

Outcome	OLP Factor	Beta	SE	95% CI	p-value	R2a	R2b
BPVS	Perceiving	+0.013	0.091	(-0.173,+0.199)	0.886	50.6%	50.7%
	Naming	+0.069	0.066	(-0.071,+0.209)	0.313		
	Interpreting	+0.208	0.083	(+0.035,+0.380)	0.020*		
CELF Sentence Structure	Perceiving	+0.041	0.138	(-0.236,+0.318)	0.769	36.7%	42.0%
	Naming	+0.031	0.102	(-0.176,+0.237)	0.766		
	Interpreting	+0.173	0.128	(-0.084,+0.430)	0.183		
RAPT (Information)	Perceiving	+0.146	0.140	(-0.134,+0.426)	0.302	26.8%	31.3%
	Naming	-0.061	0.103	(-0.267,+0.145)	0.556		
	Interpreting	+0.018	0.129	(-0.242,+0.278)	0.888		
RAPT (Grammar)	Perceiving	+0.044	0.151	(-0.258,+0.345)	0.774	29.6%	36.1%
	Naming	-0.083	0.111	(-0.306,+0.139)	0.457		
	Interpreting	+0.069	0.139	(-0.210,+0.347)	0.622		
EYFSP Good Level of Development	Perceiving	+0.060	0.142	(-0.224,+0.343)	0.675	42.6%	49.3%
	Naming	-0.099	0.104	(-0.309,+0.111)	0.347		
	Interpreting	-0.030	0.131	(-0.293,+0.234)	0.822		
EYFSP Communication and Language	Perceiving	+0.080	0.136	(-0.192,+0.352)	0.557	40.7%	46.1%
	Naming	-0.129	0.100	(-0.330,+0.073)	0.206		
	Interpreting	-0.012	0.126	(-0.265,+0.240)	0.923		
EYFSP Literacy	Perceiving	+0.045	0.142	(-0.240,+0.329)	0.755	32.6%	38.4%
	Naming	-0.035	0.104	(-0.245,+0.174)	0.738		
	Interpreting	-0.001	0.131	(-0.264,+0.262)	0.994		
Year 1 phonics check score	Perceiving	-0.038	0.158	(-0.354,+0.279)	0.814	22.3%	29.7%
	Naming	+0.017	0.116	(-0.215,+0.250)	0.882		
	Interpreting	+0.096	0.145	(-0.195,+0.387)	0.513		

See Table A5.1 for explanatory notes

Table A5.3: Associations between the OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*, *RAPT*) – three-level individual (univariate) models fitted to multiply imputed data

Outcome	Covariate	Beta	SE	95% CI	p-value	DF	R2a	R2b	SS
BPVS	OLP perceiving	+0.193	0.063	(+0.066,+0.320)	0.004**	36.28	49.9%	50.9%	797
	OLP naming	+0.125	0.067	(-0.014,+0.264)	0.077	22.79	49.4%	50.9%	797
	OLP interpreting	+0.233	0.057	(+0.115,+0.352)	<0.001***	19.39	50.6%	50.7%	797
CELF Sentence Structure	OLP perceiving	+0.177	0.089	(-0.001,+0.355)	0.051	50.96	36.5%	41.8%	797
	OLP naming	+0.094	0.092	(-0.093,+0.281)	0.314	37.07	35.9%	41.9%	797
	OLP interpreting	+0.210	0.089	(+0.031,+0.390)	0.023*	39.87	36.7%	41.8%	797
RAPT (Information)	OLP perceiving	+0.136	0.088	(-0.042,+0.314)	0.131	43.63	26.7%	32.8%	797
	OLP naming	+0.034	0.093	(-0.154,+0.222)	0.718	41.51	26.3%	32.8%	797
	OLP interpreting	+0.113	0.089	(-0.068,+0.294)	0.214	34.62	26.5%	32.8%	797
RAPT (Grammar)	OLP perceiving	+0.050	0.096	(-0.143,+0.243)	0.604	52.37	29.4%	36.2%	797
	OLP naming	-0.046	0.098	(-0.244,+0.152)	0.644	42.36	29.0%	36.7%	797
	OLP interpreting	+0.076	0.097	(-0.118,+0.271)	0.433	45.16	29.5%	36.1%	797
EYFSP Good Level of Development	OLP perceiving	-0.008	0.090	(-0.190,+0.173)	0.928	50.73	42.7%	49.1%	797
	OLP naming	-0.079	0.090	(-0.260,+0.103)	0.386	43.14	42.8%	49.1%	797
	OLP interpreting	-0.013	0.092	(-0.199,+0.172)	0.885	39.67	42.6%	49.1%	797
EYFSP Communication and Language	OLP perceiving	+0.011	0.087	(-0.164,+0.186)	0.898	50.65	40.6%	45.9%	797
	OLP naming	-0.094	0.087	(-0.269,+0.081)	0.283	39.84	40.7%	45.8%	797
	OLP interpreting	+0.011	0.088	(-0.167,+0.189)	0.901	40.32	40.6%	45.9%	797
EYFSP Literacy	OLP perceiving	+0.026	0.090	(-0.154,+0.206)	0.772	49.69	32.6%	38.5%	797
	OLP naming	-0.008	0.091	(-0.193,+0.178)	0.934	35.18	32.5%	38.5%	797
	OLP interpreting	+0.019	0.090	(-0.164,+0.202)	0.836	39.36	32.6%	38.4%	797
Year 1 phonics check score	OLP perceiving	+0.038	0.102	(-0.166,+0.242)	0.711	54.17	22.2%	29.5%	784
	OLP naming	+0.025	0.101	(-0.179,+0.228)	0.808	51.82	22.2%	29.5%	784
	OLP interpreting	+0.075	0.101	(-0.128,+0.278)	0.463	52.97	22.4%	29.5%	784

Beta = point estimate of model coefficient
SE = standard error of model coefficient
CI = 95% confidence interval for model coefficient
p-value = p-value of model coefficient, with statistical significance indicated by stars: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.
DF = degrees of freedom for model coefficient
R2a = % of variance explained (fixed effects only)
R2b = % of variance explained (fixed and random effects)
SS = model sample size

Table A5.4: Associations between the OLP factors (*perceiving*, *naming*, *interpreting*) and child language (*BPVS*, *CELF*, *RAPT*) – three-level combined (multivariate) models fitted to multiply imputed data

Outcome	Covariate	Beta	SE	95% CI	p-value	DF	R2a	R2b	SS
BPVS	OLP perceiving	+0.013	0.091	(-0.176,+0.203)	0.887	20.63	50.6%	50.7%	797
	OLP naming	+0.069	0.066	(-0.089,+0.227)	0.333	6.74			
	OLP interpreting	+0.208	0.083	(+0.029,+0.387)	0.026*	13.89			
CELF Sentence Structure	OLP perceiving	+0.041	0.138	(-0.237,+0.319)	0.769	46.73	36.7%	42.0%	797
	OLP naming	+0.031	0.102	(-0.178,+0.239)	0.766	31.73			
	OLP interpreting	+0.173	0.128	(-0.086,+0.432)	0.184	37.11			
RAPT (Information)	OLP perceiving	+0.134	0.139	(-0.145,+0.414)	0.339	49.92	26.6%	33.1%	797
	OLP naming	-0.038	0.105	(-0.249,+0.174)	0.721	43.35			
	OLP interpreting	+0.026	0.129	(-0.234,+0.286)	0.841	39.32			
RAPT (Grammar)	OLP perceiving	+0.042	0.150	(-0.260,+0.344)	0.782	48.90	29.5%	36.3%	797
	OLP naming	-0.082	0.111	(-0.308,+0.143)	0.463	39.47			
	OLP interpreting	+0.068	0.139	(-0.212,+0.347)	0.627	43.61			
EYFSP Good Level of Development	OLP perceiving	+0.060	0.142	(-0.225,+0.344)	0.675	48.03	42.6%	49.3%	797
	OLP naming	-0.099	0.104	(-0.310,+0.112)	0.348	39.65			
	OLP interpreting	-0.030	0.131	(-0.295,+0.236)	0.823	38.53			
EYFSP Communication and Language	OLP perceiving	+0.080	0.136	(-0.193,+0.353)	0.558	47.61	40.7%	46.1%	797
	OLP naming	-0.129	0.100	(-0.332,+0.075)	0.208	36.01			
	OLP interpreting	-0.012	0.126	(-0.266,+0.242)	0.923	38.42			
EYFSP Literacy	OLP perceiving	+0.044	0.142	(-0.240,+0.329)	0.755	49.24	32.5%	38.6%	797
	OLP naming	-0.031	0.106	(-0.246,+0.185)	0.774	30.68			
	OLP interpreting	-0.005	0.131	(-0.269,+0.260)	0.971	40.56			
Year 1 phonics check score	OLP perceiving	-0.038	0.158	(-0.355,+0.280)	0.814	51.90	22.3%	29.7%	784
	OLP naming	+0.017	0.116	(-0.216,+0.250)	0.882	49.73			
	OLP interpreting	+0.096	0.145	(-0.196,+0.388)	0.514	50.90			

Beta = point estimate of model coefficient
SE = standard error of model coefficient
CI = 95% confidence interval for model coefficient
p-value = p-value of model coefficient, with statistical significance indicated by stars: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.
DF = degrees of freedom for model coefficient
R2a = % of variance explained (fixed effects only)
R2b = % of variance explained (fixed and random effects)
SS = model sample size

Table A5.5: Sensitivity analysis examining the effects of teacher job-shares and joint responses to the OLP survey for the two-level individual (univariate) models fitted to multiply imputed data

I = model fitted to full data set
II = model fitted to data set omitting teachers who gave a joint response
III = model fitted to data set omitting teachers who gave a joint response and job-sharing teachers

Outcome	OLP Factor								
	OLP perceiving			OLP naming			OLP interpreting		
	I	II	III	I	II	III	I	II	III
BPVS	+0.193**	+0.201**	+0.174*	+0.125	+0.128	+0.146	+0.233***	+0.238***	+0.211**
CELF Sentence Structure	+0.177	+0.154	+0.046	+0.094	+0.057	+0.050	+0.210*	+0.190*	+0.076
RAPT (Information)	+0.131	+0.128	+0.045	+0.011	-0.005	-0.044	+0.107	+0.099	-0.008
Rapt (Grammar)	+0.053	+0.044	-0.050	-0.045	-0.066	-0.036	+0.079	+0.069	-0.091
EYFSP Good Level of Development	-0.008	-0.001	+0.004	-0.079	-0.077	-0.019	-0.013	-0.009	-0.055
EYFSP Communication and Language	+0.011	+0.015	+0.018	-0.094	-0.099	-0.047	+0.011	+0.013	-0.020
EYFSP Literacy	+0.027	+0.032	+0.057	-0.015	-0.014	+0.020	+0.022	+0.027	+0.003
Year 1 phonics check score	+0.038	+0.013	+0.047	+0.025	-0.015	+0.053	+0.075	+0.052	+0.104

Standardised model coefficients are given. Statistical significance is shown by stars: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Table A5.6: Analysis testing whether varying the 'teacher experience' covariate influences the findings for the two-level individual (univariate) models fitted to complete cases data

COVARIATE A: teacher's years of preschool experience

Outcome	Covariate	Beta	SE	95% CI	p-value	DF	R2a	R2b	SS
BPVS	OLP perceiving	+0.127	0.068	(-0.010,+0.263)	0.068	65.00	46.4%	48.2%	665
	OLP naming	+0.116	0.069	(-0.021,+0.253)	0.097	65.00	46.4%	48.3%	665
	OLP interpreting	+0.194	0.062	(+0.070,+0.318)	0.003**	65.00	47.1%	47.8%	665
CELF Sentence Structure	OLP perceiving	+0.134	0.093	(-0.053,+0.321)	0.156	65.00	33.9%	41.1%	665
	OLP naming	+0.076	0.094	(-0.112,+0.264)	0.424	65.00	33.6%	41.2%	665
	OLP interpreting	+0.185	0.092	(+0.001,+0.370)	0.049*	65.00	34.2%	41.0%	665
RAPT (Information)	OLP perceiving	+0.046	0.093	(-0.140,+0.232)	0.624	65.00	21.8%	26.9%	665
	OLP naming	-0.087	0.093	(-0.272,+0.099)	0.354	65.00	21.8%	26.9%	665
	OLP interpreting	+0.053	0.093	(-0.132,+0.238)	0.570	65.00	21.8%	26.8%	665
RAPT (Grammar)	OLP perceiving	-0.021	0.102	(-0.225,+0.182)	0.837	65.00	24.3%	32.1%	665
	OLP naming	-0.135	0.100	(-0.335,+0.064)	0.180	65.00	24.5%	31.9%	665
	OLP interpreting	+0.049	0.101	(-0.154,+0.251)	0.633	65.00	24.4%	32.0%	665
EYFSP Good Level of Development	OLP perceiving	-0.017	0.093	(-0.203,+0.169)	0.859	65.00	41.3%	47.8%	665
	OLP naming	-0.065	0.092	(-0.249,+0.120)	0.487	65.00	41.4%	47.8%	665
	OLP interpreting	-0.025	0.094	(-0.212,+0.162)	0.791	65.00	41.3%	47.8%	665
EYFSP Communication and Language	OLP perceiving	-0.013	0.087	(-0.188,+0.162)	0.882	65.00	40.3%	45.3%	665
	OLP naming	-0.069	0.087	(-0.242,+0.105)	0.432	65.00	40.5%	45.3%	665
	OLP interpreting	-0.008	0.088	(-0.183,+0.167)	0.930	65.00	40.3%	45.3%	665
EYFSP Literacy	OLP perceiving	+0.020	0.095	(-0.168,+0.209)	0.831	65.00	32.1%	37.5%	665
	OLP naming	-0.010	0.094	(-0.198,+0.178)	0.919	65.00	32.1%	37.5%	665
	OLP interpreting	-0.008	0.094	(-0.196,+0.181)	0.934	65.00	32.1%	37.5%	665
Year 1 phonics check score	OLP perceiving	+0.005	0.093	(-0.180,+0.190)	0.956	65.00	22.2%	27.6%	656
	OLP naming	+0.025	0.092	(-0.159,+0.209)	0.790	65.00	22.2%	27.5%	656
	OLP interpreting	+0.035	0.092	(-0.149,+0.219)	0.704	65.00	22.3%	27.6%	656

COVARIATE B: teacher's total years of teaching experience

Outcome	Covariate	Beta	SE	95% CI	p-value	DF	R2a	R2b	SS
BPVS	OLP perceiving	+0.125	0.068	(-0.010,+0.259)	0.069	65.00	46.4%	48.2%	665
	OLP naming	+0.113	0.069	(-0.024,+0.251)	0.105	65.00	46.4%	48.2%	665
	OLP interpreting	+0.187	0.061	(+0.066,+0.309)	0.003**	65.00	47.1%	47.8%	665
CELF Sentence Structure	OLP perceiving	+0.151	0.093	(-0.035,+0.337)	0.110	65.00	34.0%	41.2%	665
	OLP naming	+0.081	0.095	(-0.109,+0.271)	0.398	65.00	33.6%	41.3%	665
	OLP interpreting	+0.200	0.091	(+0.018,+0.382)	0.032*	65.00	34.3%	41.0%	665
RAPT (Information)	OLP perceiving	+0.084	0.094	(-0.104,+0.272)	0.375	65.00	21.8%	27.3%	665
	OLP naming	-0.080	0.096	(-0.271,+0.111)	0.405	65.00	21.7%	27.4%	665
	OLP interpreting	+0.093	0.093	(-0.092,+0.277)	0.320	65.00	21.8%	27.1%	665
RAPT (Grammar)	OLP perceiving	-0.007	0.103	(-0.213,+0.199)	0.945	65.00	24.0%	32.3%	665
	OLP naming	-0.140	0.102	(-0.343,+0.063)	0.173	65.00	24.2%	32.0%	665
	OLP interpreting	+0.066	0.101	(-0.137,+0.269)	0.518	65.00	24.1%	32.1%	665
EYFSP Good Level of Development	OLP perceiving	-0.018	0.093	(-0.204,+0.167)	0.843	65.00	41.3%	47.8%	665
	OLP naming	-0.065	0.093	(-0.250,+0.121)	0.488	65.00	41.4%	47.8%	665
	OLP interpreting	-0.027	0.093	(-0.212,+0.158)	0.773	65.00	41.3%	47.9%	665
EYFSP Communication and Language	OLP perceiving	-0.012	0.087	(-0.187,+0.162)	0.887	65.00	40.4%	45.3%	665
	OLP naming	-0.067	0.087	(-0.241,+0.107)	0.447	65.00	40.5%	45.3%	665
	OLP interpreting	-0.008	0.087	(-0.181,+0.165)	0.926	65.00	40.3%	45.4%	665
EYFSP Literacy	OLP perceiving	+0.009	0.094	(-0.179,+0.198)	0.920	65.00	32.1%	37.5%	665
	OLP naming	-0.011	0.095	(-0.200,+0.178)	0.908	65.00	32.1%	37.5%	665
	OLP interpreting	-0.020	0.093	(-0.206,+0.167)	0.834	65.00	32.1%	37.5%	665
Year 1 phonics check score	OLP perceiving	+0.010	0.092	(-0.174,+0.195)	0.911	65.00	22.2%	27.6%	656
	OLP naming	+0.025	0.093	(-0.160,+0.210)	0.788	65.00	22.2%	27.5%	656
	OLP interpreting	+0.040	0.091	(-0.141,+0.222)	0.660	65.00	22.3%	27.6%	656

Beta = point estimate of model coefficient
SE = standard error of model coefficient
CI = 95% confidence interval for model coefficient
p-value = p-value of model coefficient, with statistical significance indicated by stars: * = p < 0.05, ** = p < 0.01, *** = p < 0.001.
DF = degrees of freedom for model coefficient
R2a = % of variance explained (fixed effects only)
R2b = % of variance explained (fixed and random effects)
SS = model sample size

The additional analyses and tables in the Appendix confirm the broad overview of findings presented in the main body of this report.

References

[1] Mathers, S. (2021) Using video to assess preschool teachers' pedagogical knowledge: explicit and higher-order knowledge predicts quality, *Early Childhood Research Quarterly*, 55, 64-78.

[2] For example: Van Es, E. A. & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *J Tech Teach Educ*, 10(4), 571–596.

[3] Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov et al (2007) School readiness and later achievement. *Developmental Psych.*, 43, 1428 – 1446.

[4] Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900–1902.

[5] Waldfogel, J. & Washbrook, E.V. (2010). *Low income and early cognitive development in the UK. A report for the Sutton Trust*. London: Sutton Trust.

[6] Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. eds. (2010). *Early childhood matters: Evidence from the effective pre-school and primary education project*. London: Routledge.

[7] Wright, H., Carr, D., Wiese, J., Stokes, L., Runge, J., Dorsett, R., Heal, J. & Anders, J. (2020) *URLEY Evaluation Report.* London: Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/URLEY_Report.pdf?v=1630925607

[8] Forrest C.F., Charlton, J., Mathers, S.J., Law, J. & Dockrell, J. (in press) Social disadvantage, oral language and social, emotional and mental health in nursery children in England. Submitted *to Early Childhood Research Quarterly*.

Ofsted (2022) Education Recovery in Early Years Providers: summer 2022. Available at: Education recovery in early years providers: summer 2022 - GOV.UK (www.gov.uk)

[9] Bonetti, S. & Blanden, J. (2020) *Early years workforce qualifications and children's outcomes: an analysis using administrative data*. London: Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2020/12/Early-years-workforce-qualifications-and-childrens-outcomes.pdf,

Melhuish, E. & Gardiner, J. (2017, revised 2018) *Study of Early Education and Development (SEED): Study of Quality of Early Years Provision in England (Revised) Research Report.* London: Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/723736/Study_of_quality_of_early_years_provision_in_England.pdf

[10] Markussen-Brown, J., Juhl, C.B, Piasta, S.B., Bleses, D.B., Højen, A. & Justice, L.M. (2017) The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Res. Quarterly,* 38, 97–115.

[11] Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B. et al (2015) *Developing Great Teaching: Lessons from the International Reviews into Effective Professional Development*. London: Teacher Development Trust.

Campbell-Barr, V., Bonetti, S., Bunting, F., & Gulliver, K. (2020). *A systematic review of early years degrees and employment pathways.* Plymouth: University of Plymouth.

[12] Sheridan, S. M., Edwards, C. P., Marcin, C. A., Knoche, L. L. (2009) Professional development in early childhood programs: Process issues and research needs. *Early Education and Development,* 20, 377–401.

[13] Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* 57, 1-22.

[14] Anderson, L.W. & Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

König, J. (2013) First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education,* 11, 999-1028.

[15] Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26(2), 247–27

Stürmer, K., Seidel, T. & Schäfer, S. (2013). Changes in professional vision in the context of practice. *Gruppendynamik und Organisationsberatung, 44*(3), 339–355.

Cochran-Smith, M., & Lytle, S. I. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, 24, 249–305.

[16] Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

[17] Piasta, S.B., Connor, C.D., Fishman, B.J. & Morrison, F.J. (2009) Teachers' Knowledge of Literacy Concepts, Classroom Practices, and Student Reading Growth. *Scientific Studies of Reading,* 13(3), 224-248.

Schachter, R.E., Spear, C.F., Piasta, S.F., Justice, L.M. & Logan, J.A.R. (2016). Early childhood educators' knowledge, beliefs, education, experiences and children's language- and literacy-learning opportunities: What is the connection? *Early Childhood Research Quarterly,* 36, 281–294.

[18] Phillips, B.M., Oliver, F., Tabulda, G., Wood, C. & Funari, C. (2020) Preschool teachers' language and vocabulary knowledge: Development and predictive associations for a new measure. *Dyslexia*, 26, 153-172.

Schachter, R.E., Spear, C.F., Piasta, S.F., Justice, L.M. & Logan, J.A.R. (2016). Early childhood educators' knowledge, beliefs, education, experiences and children's language- and literacy-learning opportunities: What is the connection? *Early Childhood Research Quarterly,* 36, 281–294.

Spear, C.F., Piasta, S.B., Yeomans-Maldonado, G., Ottley, J.R., Justice, L.M. & O'Connell, A.A. (2018) Early Childhood General and Special Educators: An Examination of Similarities and Differences in Beliefs, Knowledge, and Practice. *Journal of Teacher Ed*., 69(3), 263-277

[19] Schachter, R.E., Spear, C.F., Piasta, S.F., Justice, L.M. & Logan, J.A.R. (2016). Early childhood educators' knowledge, beliefs, education, experiences and children's language- and literacy-learning opportunities: What is the connection? *Early Childhood Research Quarterly,* 36, 281–294.

[20] Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher,* 15, 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* 57, 1-22.

Dwyer, J. & Schachter, R.E. (2020) Going beyond defining: Preschool educators' use of knowledge in their pedagogical reasoning about vocabulary instruction. *Dyslexia,* (26), 173-199.

[21] Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* 57, 1-22.

Kersting, N., Givvin, K.B., Sotelo, F.L., & Stigler, J.W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61(1–2), 172–181.

[22] Sherin, M.G. & van Es, E. (2009) Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60, 20-37.

Bruckmaier, G., Krauss, S., Blum, W. & Leiss, D. (2016) Measuring mathematics teachers' professional competence by using video clips (COACTIV video*) ZDM Math. Ed.*, 48, 111–124.

Kersting, N., Givvin, K.B., Sotelo, F.L., & Stigler, J.W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61(1–2), 172–181.

Wiens, P.D., Hessberg, K., LoCasale-Crouch, J. & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers' knowledge related to effective teaching. *Teaching and Teacher Education, 33*, 24–33.

[23] Kersting, N.B., Givvin, K.B., Thompson, B.J., Santagata, R. & Stigler, J.W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, 49(3), 568–589.

[24] Goodwin, C. (1994) Professional vision. *American Anthropologist*, 96 (3), 606-633.

Sherin, M.G. & van Es, E. (2009) Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60, 20-37.

Wiens, P.D., LoCasale-Crouch, J. Cash, A.H. & Romo-Escudero, F. (2020) Preservice Teachers' Skills to Identify Effective Teaching Interactions: Does It Relate to Their Ability to Implement Them? *Journal of Teacher Education*, 1-15.

[25] Jamil, F., Sabol, T., Hamre, B. & Pianta, R. (2015). Assessing Teachers' Skill in Detecting and Identifying Effective Interactions in the Classroom. *Elem Sch Journal*, 115 (3), 407-432.

[26] Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, 45, 169–204.

[27] Mathers (2020) Observing Language Pedagogy (OLP): *Developing and piloting a contexualised video-based measure of early childhood teachers' pedagogical language knowledge (doctoral dissertation).* https://discovery.ucl.ac.uk/id/eprint/10094735/

[28] Mathers, S. & Siraj, I. (2021, November). Researching preschool teachers' knowledge of oral language pedagogy using video. In *Frontiers in Education* (Vol. 6). Frontiers Media.

[29] Department for Communities & Local Government (2011) *The English Indices of Deprivation 2010.* Available at: http://www.communities.gov.uk/documents/statistics/pdf/1871208.pdf

[30] Dunn L & Dunn D. *The British Picture Vocabulary Scales.* 3. London: GL Assessment; 2009.

[31] Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals—Preschool, (CELF Preschool-2).* Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

[32] Renfrew C E. *Renfrew Language Tests*. C. E. Renfrew, Oxford 1980. Revised Edition (2010) used for this study.

[33] Sammons, P., Sylva, K., Hall, J., Siraj, I., Melhuish, E., Taggart, B., & Mathers, S. (2017). Establishing the effects of quality in early childhood: comparing evidence from England. Available at: Establishing the Effects of Quality in Early Childhood: Comparing Evidence from England (exeter.ac.uk)

[34] Burgess, S. & Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities, *Journal of Labor Economics*, 31(3) 535-576.

[35] For example: Santagata, R. & Yeh, C. (2014) Learning to teach mathematics and to analyse teaching effectiveness: evidence from a video-and practice-based approach. *J Math Teacher Educ*, 17, 491-514.

[36] For example: Van Es, E. A. & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *J Tech Teach Educ*, 10(4), 571–596.

[37] Kersting, N., Givvin, K.B., Sotelo, F.L., & Stigler, J.W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61(1–2), 172–181.

Kersting, N.B., Givvin, K.B., Thompson, B.J., Santagata, R. & Stigler, J.W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, 49(3), 568–589.

[38] Blaine, B. (2018) Winsorizing. In Frey, B. B. (Ed.). (2018). *The SAGE Encyclopaedia of Educational Research, Measurement, and Evaluation*. California: SAGE Publications.

Dixon, W.J. (1960). Simplified Estimation from Censored Normal Samples. *The Annals of Mathematical Statistics*, 31, 385–391.

[39] Education Endowment Foundation Early Years Measures Database: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/early-years-measures-database-2

[40] Hogan A.E., Scott K.G. & Baurer C.R. (1992) The Adaptive Social Behaviour Inventory (ASBI): A new assessment of social competence in high risk three year olds. *Journal of Psychoeducational Assessments* 10: 230-239.

[41] Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., ... & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

[42] Honaker J., Joseph A., King G., et al. *Amelia II (R Package).* 2011. Available from: http://cran.r-project.org/web/packages/Amelia/index.html.

[43] Honaker J., King G., Blackwell M. *Amelia II: a program for missing data*. Available from: http://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf.

[44] Schafer, J.L. (1997) *Analysis of incomplete multivariate data.* Boca Raton: CRC Press.

[45] Rubin, D.B. (1987) *Multiple imputation for nonresponse in surveys*. Philadelphia: Wiley & Sons.

[46] Hesterberg, T. (1998) *Combining multiple imputation t, chi-square, and F inferences.* Seattle, WA: MathSoft.

[47] von Hippel, P. T. (2007). Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. *Sociological Methodology*, 37: 83-117.

[48] Snijders, T.A.B. *Power and Sample Size in Multilevel Linear Models*. In: B.S. Everitt and D.C. Howell (eds.), Encyclopedia of Statistics in Behavioral Science. Volume 3, 1570–1573. Chicester (etc.): Wiley, 2005.