

**Using video to assess preschool teachers' pedagogical knowledge:  
explicit and higher-order knowledge predicts quality**

**Mathers, S. (2021) Department of Education, University of Oxford, UK**

**Sandea.mathers@education.ox.ac.uk**

**Abstract:** Theory suggests that effective real-time decision-making in classrooms requires teachers to have flexible access to rich and well-organised knowledge of effective teaching practices. Yet prior research on the role and importance of procedural knowledge has been equivocal. This exploratory study used a new video measure of procedural knowledge to examine relationships with observed classroom quality, and establish which opportunities to learn (qualifications, professional development, classroom experience) predict greater knowledge. It focused on preschool teachers' knowledge of oral language pedagogy, on the basis that early language provides the foundation for children's later learning.

The sample comprised 104 teachers participating in a wider RCT, designed to evaluate a professional development intervention. Teachers were shown two short videos of classroom interactions and asked to identify instances of effective practice. Responses were coded to capture three facets: ***perceiving*** (the ability to identify salient language-supporting strategies); ***naming*** (the use of specific professional vocabulary to describe interactions); and ***interpreting*** (the ability to interpret the interactions observed). The three facets could be empirically distinguished. Explicit and higher-order procedural knowledge (*naming*, *interpreting*) most strongly predicted classroom quality. Formal learning opportunities were stronger predictors of procedural knowledge than classroom experience. Intervention effects on classroom quality were mediated by knowledge. Implications for workforce development are discussed.

## Introduction

### The importance of high-quality support for early language

Early language provides a bedrock for later learning (Bleses, Makransky, Dale, Højen & Ari, 2016; Hjetland et al, 2019). Language in preschool predicts literacy and wider outcomes at school entry (Morgan, Farkas, Hillemeier, Hammer & Maczuga, 2015; Roulstone, Law, Rush, Clegg & Peters, 2011), which in turn predict later school achievement (Duncan et al, 2007). However, many children - particularly those from disadvantaged homes - start school without the language skills they need (Waldfoegel & Washbrook, 2010), with consequences for their long-term life trajectories (Heckman, 2006). While high-quality preschool offers a potential means of narrowing this gap (Sylva, Melhuish, Sammons, Siraj & Taggart, 2010), evidence suggests that current practice may be inadequate to nurture children's language skills, particularly in the disadvantaged areas where it is most needed (Mathers & Smees, 2014).

Although in-service professional development shows promise in improving practice and child outcomes, it is not uniformly effective (Markussen-Brown et al., 2017). A recent review concluded that teacher professional development in England is insufficiently evidence-based and inconsistent in quality (Cordingley et al, 2015). To inform intelligent design of future training we must understand the '*dynamic and transactional teacher and learning processes*' which form the mechanism of change (Sheridan, Edwards, Marcin & Knoche, 2009, p.378).

### The role of teacher knowledge and the limits of current research

There is an implicit assumption in much professional development that teacher *knowledge* is fundamental to improved practice and child outcomes (Ben-Peretz, 2011). Effective teaching is understood to require knowledge of both subject *content* (what is to be taught) and *pedagogy* (how to represent the content for learners) (Shulman, 1986, 1987;

Ball & Bass, 2000), with evidence suggesting that these facets can be distinguished empirically (Baumert et al, 2010; Voss, Kunter & Baumert, 2011). Knowledge of pedagogy can be further divided into declarative (*knowing that*) and procedural (*knowing how*), with both required for expert practice (Anderson & Krathwohl, 2001; König, 2013). To illustrate, a preschool teacher with expertise in supporting oral language will likely know how to form the past tense of irregular verbs (*content knowledge*); know that children tend to overgeneralise regular suffixes and that recasting such errors in conversation supports mastery of irregular forms (*theoretical knowledge of pedagogy*); and know how to use recasting appropriately within classroom interactions (*procedural knowledge of pedagogy*).

While the language-and-literacy *content* knowledge of preschool teachers has been linked to practice and child outcomes (Piasta, Connor, Fishman & Morrison, 2009; Schachter, Spear, Piasta, Justice & Logan, 2016), the same cannot be said for *pedagogical* knowledge. Much is inferred, for example, in studies of professional development which aim to improve knowledge and have subsequently identified effects on practice (Dickinson & Caswell, 2007; Wasik, Bond & Hindman, 2006). Studies which explicitly assess pedagogical language-and-literacy knowledge have identified small or null effects (Piasta, Park, Farley, Justice & O'Connell, 2020; Phillips, Oliver, Tabulda, Wood & Funari, 2020; Spear et al, 2018; Schachter et al, 2016). Such equivocality is puzzling, given the theoretical value of domain-specific pedagogical knowledge and evidence of its importance in later education phases, for example, in the domain of mathematics teaching (Baumert et al, 2010; McCray, 2008).

### **Capturing procedural knowledge of pedagogy - and the case for video assessment**

It is plausible that many null results identified for pedagogical knowledge stem from failings in its measurement (Markussen-Brown et al, 2017). The studies cited above all used written questionnaires, primarily multiple-choice tests. Though appropriate for measuring

explicit and relatively static knowledge (content, theory), questionnaires are ill-suited to capture the more dynamic and situated *procedural* knowledge of pedagogy (Schachter et al, 2016) which develops and is applied in the classroom context (Brouwer & Korthagen, 2005; Eraut, 2004; Stürmer, Seidel & Schäfer, 2013). While multiple-choice tests assume a 'correct' answer in response to a hypothetical scenario, the best course of action in a real classroom situation depends on multiple factors, including the traits and responses of the children and the pedagogical beliefs or goals of the teacher (Putnam & Borko, 2000; Van Driel & Berry, 2012). The dynamic procedural knowledge of an expert teacher allows them to orchestrate these factors 'in the moment', applying strategies flexibly to maximise child learning (Knievel, Lindmeier & Heinze, 2015; Shulman, 1987). Studies of preschool teachers' language-related pedagogical reasoning shows that it is multi-faceted and highly context-specific (Dwyer & Schachter, 2020). Static questionnaire scenarios are simply too remote from the complexity of real classroom interactions to capture the contextually-codified knowledge which shapes decision-making within them (Alonzo & Kim, 2016).

This study uses videos of classroom interactions to provide a more sensitive measure of preschool teachers' procedural knowledge of oral language pedagogy. Video assessment involves teachers watching video clips and identifying or interpreting teaching practice or children's responses (Jamil, Sabol, Hamre & Pianta, 2015; Kersting, Givvin, Sotelo & Stigler, 2010). This preserves the complexity of the classroom and offers an authentic context for assessing situated knowledge (Alonzo & Kim, 2016). It enables assessment, not just of what teachers know, but of the knowledge they are able to access, activate and use in a classroom situation - their 'useable knowledge' (Kersting, Givvin, Thompson, Santagata & Stigler, 2012). The approach is underpinned by theory and empirical evidence (elaborated below) that expert and novice teachers differ in their observational abilities, that these differences stem

from variations in knowledge codification, and that effective observation supports effective practice. By developing a more discerning assessment, this study aims to illuminate the role of procedural knowledge in ensuring effective support for oral language teaching, offering immediate and practical guidance for workforce development and enabling future research.

### **Using video observations ('professional vision') to elicit procedural knowledge**

At the heart of video assessment lies the centrality of classroom observation. Procedural knowledge is understood to develop through observation. Teachers notice their own and their colleagues' actions, and the consequences of those actions for children, and internalise these observations to adapt their practice (Bandura, 1986; Jamil et al, 2015). The cognitive scripts generated (Moskowitz, 2005; Schank, 1997) become more deeply and holistically codified through repeated rehearsal and observation, supporting efficient knowledge retrieval (Sabers, Cushing, & Berliner, 1991). They enable experts to assimilate, understand and interpret new situations more quickly, accurately and comprehensively than novices (Berliner, 1988; Bransford, Brown & Cocking, 2000; Glaser & Chi, 2014; Sabers, Cushing & Berliner, 1991; van Es & Sherin, 2002). Expert teachers become 'expert observers', able to filter out irrelevant information and focus on the features of an interaction which are salient for child learning. This ability to perceive and interpret classroom situations (termed '*noticing*' or '*professional vision*') (Goodwin, 1994; Seidel & Stürmer, 2014; Sherin & Van Es, 2009) supports effective real-time decision-making (van Es & Sherin, 2002). In the context of video assessment, it can also offer a window onto teachers' procedural knowledge because - in order to perceive and interpret strategies in a video - teachers must have cognitive scripts to represent those strategies (Alonzo & Kim, 2016; Jamil et al, 2015).

Emerging evidence indicates promise. Video assessments of domain-specific professional vision in middle and secondary mathematics teachers have predicted student-reported

instructional quality (König & Kramer, 2016), the cognitive quality of tests set by teachers (Bruckmaier, Krauss, Blum & Leiss, 2016) and student learning gains (Kersting et al, 2010, 2012), albeit often in small samples. Similar tools – largely focused on mathematics and non-domain-specific professional vision in primary, middle and secondary education - have yet to be validated but show encouraging psychometric properties (e.g. Kaiser, Busse, Hoth, König & Blömeke, 2015; Knievel, Lindmeier & Heinze, 2015; Keppens, Consuegra, Goossens, De Maeyer & Vanderlinde, 2019; Seidel & Stürmer, 2014). To our knowledge, the only validated preschool measure is the Video Assessment of Interactions and Learning (VAIL). The VAIL, which assesses teachers' ability to identify effective (non-domain-specific) adult-child interactions, has been shown to predict observed quality of instructional and emotional support (Jamil et al, 2015; Wiens, LoCasale-Crouch, Cash & Romo-Escudero, 2020) and may play a mediating role in the improvement of quality (Hamre et al, 2012).

This study builds on the successful methodology of the VAIL, adapting it to capture the professional vision of preschool teachers in the domain of oral language pedagogy, with the aim of eliciting their underlying dynamic procedural knowledge – and, more broadly, of enabling research which can inform language-specific professional development. The Observing Language Pedagogy (OLP) tool further extends the VAIL by taking exploratory steps in examining the relevance of higher-order knowledge facets such as interpretation, informed by video studies conducted with teachers of older children.

### **The cognitive facets of professional vision and underlying procedural knowledge**

The cognitive facets of interest can be determined by considering pathways to knowledge development. Procedural knowledge may develop *implicitly*, through unconscious classroom observation, or through 'reactive learning' - where observation is accompanied by brief, near-spontaneous, in-the-moment reflection (Eraut, 2004). In this way, teachers develop

procedural scripts for a range of teaching strategies, enabling them to *recognise* these strategies and *recall* their knowledge in specific classroom contexts, in order to apply an appropriate strategy (Anderson & Krathwohl, 2001). The ability to *perceive* salient strategies (revealing the existence of associated knowledge scripts) is thus the first facet of interest, with evidence of its importance offered by the predictive validity of the VAIL.

Procedural knowledge may also develop *explicitly*. In this conception, theoretical knowledge gained through formal training (qualifications, professional development, etc.) is transformed into procedural knowledge through cycles of application and observation in the classroom (Anderson, 1982). Explicit procedural knowledge is believed to be more deeply codified than tacit knowledge, and to support intentional practice through its accessibility for deliberate manipulation (Kind, 2009). Examples include explicit knowledge of teaching strategies and how these relate to the broader principles of teaching and learning, and the ability to reason about how a specific strategy might support child learning (Anderson & Krathwohl, 2001; Van Es & Sherin, 2002). To illustrate - alongside a tacit script governing when to use questions beginning with '*how*' or '*why*', a teacher might also possess an explicit, overarching script identifying these as open questions, understand when and why open questions might be appropriate (based on theory but refined by experience) and be able to analyse how effective their open question was in eliciting a specific child's thinking.

One indicator of such explicit knowledge might be the use of specific pedagogical terms (e.g. 'open questions') to describe strategies observed in video interactions. Whilst naming does not guarantee understanding, it is hard to engage in explicit discussion or reflection on a concept until one can name it - and learning the language of a discipline is part of learning the discipline itself. Although efforts have been made to detail the professional lexicon of mathematics teachers (Clarke, Mesiti, Cao & Novotná, 2017), no such work exists for

preschool language pedagogy and this facet has not yet been considered in video studies.

The use of professional vocabulary to describe video interactions is theorised to reflect both explicit procedural knowledge and the ability to articulate this knowledge to others.

The ability to *interpret* video interactions offers a further indicator of explicit, higher-order procedural knowledge. Expert teachers are better equipped than novices to reason about possible pedagogical intentions, alternative approaches, or children's thinking and outcomes (Berliner, 1988; Blömeke, Gustafsson & Shavelson, 2015; Klein & Hoffman, 1993; Sabers, Cushing & Berliner, 1991; Seidel & Stürmer, 2014; van Es & Sherin, 2002). Studies using one of the most promising video tools - the Classroom Video Analysis (CVA) – shows that the ability of 5<sup>th</sup>-7<sup>th</sup> grade teachers to analyse video interactions and suggest alternative strategies predicts children's mathematical learning (Kersting et al, 2010, 2012). An interesting reflection is the fact that only teachers' *spontaneous* suggestions for improvement predicted learning gains, with no effects found for the three prompted indicators of reasoning. The authors suggest that spontaneous reasoning reflects expertise which has become embedded through repeated use and is thus 'close to the surface' and routinely applied – these were the teachers *most* likely to use reasoning in live situations. The ability to interpret video interactions forms the third and final facet of interest, theorised to reflect the ability to reason during actual classroom situations to support intentional teaching. Other higher-order dimensions such as decision-making, evaluation and generation (Anderson & Krathwohl, 2001; König et al, 2014; Knievel et al, 2015) are acknowledged, but not reflected, in the current study. The three facets of professional vision (and underlying procedural knowledge) captured by the OLP tool are shown in Table 1 below.

**The pedagogical facets of professional vision and underlying procedural knowledge**



This study focuses on procedural knowledge of how to represent content effectively for learners, and adapt teaching for specific contexts (Shulman, 1986; Kind, 2009). Specifically, it aims to capture knowledge of the teaching strategies which support oral language in the early years (vocabulary, grammar/morphology, narrative, pragmatics) and which in turn underpin later reading comprehension (Hulme & Snowling, 2014; Gardner-Neblett & Iruka, 2015). The specific strategies included within the OLP tool were informed by a thorough review of the literature (Mathers, 2020) and are shown in Table 2. There are, of course, many other facets of pedagogical knowledge – most importantly, knowledge of learners and learning (Shulman, 1986; Kind, 2009). This study does not claim to capture all domains, but rather to test the water in exploring the potential of video assessment.

### **The current study**

The current study sat within a wider randomised controlled trial (RCT), designed to evaluate a professional development intervention for preschool teachers targeting their language-supporting practice (Mathers, 2020). 120 schools participated in the RCT between September 2017 and June 2018, with 61 receiving the intervention and 59 in a business-as-usual control group. The sample for the current study comprised 104 teachers (from 72 schools) who responded to an online survey at the post-test stage of the RCT, designed to gather data on procedural knowledge, teaching qualifications and experience. These data were combined with data on teacher characteristics and observed classroom quality to answer four questions:

1. What is the structure of preschool teachers' language-related professional vision, and underlying procedural pedagogical knowledge?
2. Does procedural knowledge predict observed quality of practice?
3. Is procedural knowledge predicted by teachers' formal and informal opportunities to

learn (qualifications, professional development, classroom experience)?

#### 4. Does procedural knowledge mediate intervention effects on classroom quality?

This is the first preschool video study to examine knowledge of oral language pedagogy and to examine the roles played by *naming* and *interpreting*. In relation to structure, prior evidence suggests that *perceiving* and *interpreting* can be empirically distinguished (König et al, 2014; Seidel & Stürmer, 2014) but the role played by professional vocabulary (*naming*) is unknown. Both a three-factor model (*perceiving* + *naming* + *interpreting*) and a two-factor model [lower order knowledge (*perceiving*) + explicit, higher-order knowledge (*naming/interpreting*)] are hypothesised and assessed. It is also possible that *interpreting* is more akin to *perceiving* than to *naming*. Although interpretation is higher-order in nature – implying formal learning – like *perceiving*, it depends on informal opportunities to observe strategies ‘in action’ and their effects on children. A third possible model [observation-based knowledge (*perceiving/interpreting*) + formal knowledge (*naming*)] is therefore also assessed. Positive associations are expected between classroom experience and both *perceiving* and *interpreting*, with no association expected for *naming*. Formal opportunities to learn (qualifications, professional development) are anticipated to predict all three knowledge facets. Procedural knowledge is expected to predict classroom quality, with stronger associations anticipated for the explicit, higher-order facets of *naming* and *interpreting* than for *perceiving*. In the event that intervention effects are identified on knowledge, these are expected to partially mediate intervention effects on practice quality.

## Method

### Sampling context and participants

The 120 schools participating in the wider RCT were all state funded, and recruited from disadvantaged regions in the Midlands and North West of England [in the lowest 3 deciles

defined by the Indices of Multiple Deprivation (IMD) (DCLG, 2011)]. Participating teachers were from nursery (age 3) or reception (age 4) classes. After attrition, the post-test sample included 115 schools and 283 teachers, 67% of whom had participated since inception.

Of the 104 teachers who responded to the knowledge survey – and so comprise the current sample - almost two-thirds ( $n=68$ ) taught in reception, 30 taught in nursery and 6 were coded as 'other' (teachers in combined classes, leaders with no direct teaching role). All held or were working towards the graduate level 'Qualified Teacher Status' (QTS) [primary specialism 3-11 years= $61$ ; early childhood specialism 3-7 years= $36$ ; unspecified= $5$ ; working towards QTS= $2$ ]. On average, participants had 10.7 years of teaching experience (range=.5-33,  $SD=7.6$ ) and 8.2 years of preschool teaching experience (range .5-33,  $SD=7.0$ ). Most ( $n=99$ ) were female. Three quarters ( $n=78$ ) had participated in the RCT since inception. Just under half ( $n=50$ ) were from schools ( $N=35$ ) receiving the intervention. Of these, two thirds ( $n=33$ ) had attended  $\geq 5$  days of the 6-day training and one quarter ( $n=13$ ) had attended  $\leq 2$  days. Mean attendance was 4.5 days. Participants' schools tended to be located in disadvantaged areas, with four fifths ( $N=84$ ) below the 50<sup>th</sup> IMD percentile and more than half ( $N=59$ ) below the 20<sup>th</sup> percentile.

The modest response rate (63% school response, 37% teacher response) must be noted. Although study teachers reflected the RCT sample on many dimensions (area disadvantage, intervention group status, training attendance and observed quality) they were more likely than RCT participants to be teaching in reception [ $F(1,281)=4.16$ ,  $p=.04$ ], and to have participated in the RCT since inception [ $F(1,279)=4.23$ ,  $p=.04$ ]; and it was not possible to compare the samples on qualifications and experience. As such, the possibility of response bias cannot be ruled out.

## **Procedures**

Both the online survey (assessing procedural knowledge, teaching qualifications and experience) and the observations of classroom quality took place post-intervention, between October 2017 and February 2018. The quality assessments were conducted in one reception class per school and were available for 55 of the 104 study teachers. Ethical approval was provided by the Education Research Ethics Committee of the University of Oxford (for the RCT) and the UCL Institute of Education (for the current study).

### ***Measures of procedural knowledge***

Procedural knowledge was assessed using the Observing Language Pedagogy (OLP) tool, designed to capture the facets of *perceiving*, *naming* and *interpreting*. An overview is provided here, and further detail on rationale and development in Mathers (2020). The OLP is based on a framework of 30 pedagogical strategies, generated through a review of evidence on benefits for child language and literacy outcomes and affirmed through expert review (Table 2). These focus primarily on adult-child interactions as the most powerful vehicle for language learning (Bruner, 1975; Bronfenbrenner & Morris, 2006).

Teachers were asked to watch 2 short (2-3 minute) videos of authentic preschool classroom interactions. In Video 1, a teacher and child interact in the block area. In Video 2, a teacher supports a small group to talk about and draw a shopping list. Respondents could download a transcript and watch videos more than once. The videos were selected from a wider pool, via piloting and expert review, as reflecting good-quality examples of many of the 30 strategies (Table 2). A third video showing a whole-class activity was dropped during development (Mathers, 2020). Thus, although the use of videos of this length and number is supported by previous studies (Bruckmeier et al, 2016; Jamil et al, 2015), the OLP cannot be said to reflect the full range of classroom contexts.

Informed by the validated VAIL and CVA tools, the OLP employed open prompts rather

than the closed rating items used in other video studies (e.g. Seidel & Stürmer, 2014), with the aim of eliciting knowledge which teachers are able and likely to access, activate and use in a classroom situation (Kersting et al, 2012). Following the VAIL methodology, *perceiving* was directly prompted: teachers were asked to identify up to 8 strategies reflected in each video which might support children's oral language (Appendix A.1). The decision to prompt for 8 strategies rather than 5 (as in the VAIL) was made following piloting, to enable a more fine-grained analysis of strategy identification. Teachers' responses ( $\leq 8$  per video) were considered valid if they matched at least one of the 30 strategies in the OLP framework. Coding was conducted by the author, with a proportion independently coded by a second researcher who had been 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 [Video 1: 89%, Video 2: 88%) and discrepancies resolved through discussion. Each valid strategy identified was awarded the relevant 'expert example' rating for the appropriate video (Table 2) to generate two *perceiving* scores – one for each video (Table 3).

Informed by the findings of Kersting and colleagues (2010, 2012), 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, we reasoned that spontaneous use would reflect knowledge most likely to be mobilised in real classroom situations - and thus most closely associated with actual practice. Prior piloting had indicated that this approach would elicit sufficient examples, and also that direct prompting to interpret all 8 strategies required a longer completion time than was possible within the constraints of the wider RCT. Though undeniably light-touch and exploratory as a means of examining pedagogical reasoning, the resulting responses did reflect awareness of pedagogical intention, observed effects on the children and possible alternative approaches

(Table 3). All valid responses ( $\leq 8$  per video) were coded as shown in Table 3 to generate two *naming* and two *interpreting* scores: one for each video. 50% of responses were double-coded, with high levels of exact agreement [*naming*: 1=100%, 2=100%; *interpreting*: 1=96%, 2=98%).

All respondents identified at least one valid language-supporting strategy (Video 1 mean = 5.5; Video 2 mean = 5.0). *Perceiving* scores were normally distributed with a broad range. *Naming* and *interpreting* scores displayed narrower ranges, lower means and a positive skew due to a high proportion of zero responses (Table 4). As noted, some zero responses may reflect teachers who could have named or interpreted if prompted. *Naming* and *interpreting* can therefore be considered as ‘censored’ continuous variables: they reflect a continuous underlying construct but include values which may not reflect a true position on the number line, due to limitations on measurement opportunity (McDonald & Moffitt, 1980). Despite the high proportion of zeros, scores did discriminate between respondents: 57.4% used at least one professional term (*naming*) and 44.2% provided at least one interpretation. Additional reliability and validity data are presented in the results section.

### ***Data on opportunities for learning***

Data on qualifications and teaching experience were gathered via the online survey, and data on participation in the professional development intervention from the RCT dataset.

### ***Measures of practice quality***

Three live-observation measures of classroom quality were used in the main RCT: the Early Childhood Environment Rating Scale Third Edition (ECERS-3; Harms, Clifford & Cryer, 2014), the literacy subscale from the curricular extension to the ECERS (ECERS-E; Sylva, Siraj & Taggart, 2003) and the Sustained Shared Thinking and Emotional Wellbeing (SSTEWS) scale (Siraj, Kingston & Melhuish, 2015). All 3 scales are known to predict children’s development

(e.g. Sylva, Melhuish, Sammons, Siraj & Taggart 2010; Howard et al, 2018). The ECERS-3 comprises 35 items assessing the global education and care environment, including interactions, activities, the physical environment and care routines. The 6 items of the ECERS-E literacy subscale assess support for language and emergent literacy. The 14 items of the SSTEW assess adult-child interactions supporting children's thinking, language, emotional well-being and self-regulation. All items are scored on a 7-point scale from 1 (*inadequate*) to 7 (*excellent*). Data collectors were experienced observers blind to treatment allocation, with training and reliability conducted according to author guidance. All 3 scales were completed one day: ECERS-3 over 3 hours; STEW and ECERS-E over the full day.

For the purposes of the current study, an Oral Language Quality score was generated using relevant items from all 3 rating scales. Although this departs from traditional use of these tools, it was necessary because they capture many practices not directly relevant to language pedagogy. The 10 items most closely aligned with the OLP pedagogical categories were selected by the author, based on a mapping of item content. These items, illustrative indicators, and links with the OLP are shown in Table 5. Confirmatory factor analysis (CFA) was used to establish whether they formed a coherent latent construct (Schreiber et al, 2006) in both the study (n=55) and full RCT (n=115) datasets. Two items with factor loadings  $\leq .60$  were removed, with the aim of identifying the smallest number of coherent items reflecting all six OLP categories (Table 5). Model fit for the final 8-item model was good in both datasets, and Cronbach's alphas high ( $>.90$ ). The Oral Language Quality score was created using factor loadings from the 8-item CFA in the study sample (Table 5).

In addition to the Oral Language Quality score, the overall ECERS-3 and SSTEW scores were used as broader measures of classroom quality. Descriptive data for the three quality measures are shown in Table 6. Quality was low on average (although there was variation

between schools) and was representative of wider RCT sample. Associations between the quality measures were high ( $r=.81-.95$ ). All 3 measures displayed normal distributions.

### **The intervention**

The intervention aimed to improve children's oral language skills (pragmatics, vocabulary, grammar and narrative) via improvements in language-supporting teaching practice (Mathers, 2020). In intervention schools, at least one nursery and reception class teacher participated, with two reception teachers participating in larger schools. The programme comprised 6 days of training and up to 3 days of on-site mentoring per school, spanning just under one year, with design features informed by evidence on the characteristics of effective professional development (e.g. Desimone, 2009). Teachers attended the training and were supported to engage their wider class team in implementation. The intervention was designed to develop theoretical and procedural knowledge and to provide direct examples of effective practice. During training, teachers were introduced to a set of language-learning principles (e.g. *'be a magnet for communication'*, *'be a language radiator'*) which reflected all six of the OLP pedagogical categories (Table 2). They were taught how to use these principles and a range of research tools (including ECERS-3 and SSTEW) to observe practice seen in videos, and were provided with strategies and resources for refining practice. Between training days, they were supported to use these resources to evaluate their practice and make changes.

### **Data preparation and analysis**

Three responses (2.9%) were missing for Video 2 of the OLP assessment. Little's MCAR Test indicated no statistically significant patterns in the missing data ( $p=.22$ ). Nevertheless, to avoid bias associated with listwise deletion (Enders, 2010), analyses were conducted using full information maximum likelihood (FIML) estimators (Allison, 2012) where appropriate.



Twelve outliers were identified across the two *interpreting* scores. Given the low incidence of interpretations, removing outliers would exclude valuable data and potentially introduce bias. Instead, Winsorisation (Blaine, 2018; Dixon, 1960) was used, in which outliers are replaced with a value at a defined upper threshold based on the interquartile range<sup>1</sup>.

Descriptives for the Winsorised scores are shown in Table 4.

All analyses were conducted using Stata 15 (StataCorp, 2017) and are described below. School-level clustered robust standard errors were used to allow for within-school associations in schools with respondents from more than one classroom (of the 72 schools, 44 had two respondents and 15 had three respondents). For analyses exploring relationships between teacher knowledge and classroom quality, the sample was restricted to teachers whose classes were observed (n=55 reception teachers) and robust standard errors used.

## Findings

### Descriptive statistics

Descriptive statistics for the six OLP scores are shown in Table 4 (full sample) and Appendix A.2 (intervention group/control group). All mean scores were higher in the intervention than the control group, with the largest differences seen for *perceiving* and *naming*.

### Reliability and validity of the OLP and the structure of professional vision/knowledge

Bivariate correlations between *perceiving*, *naming* and *interpreting* scores are shown in Table 7. Despite commonality within score type, associations *between* score types were only small-to-medium, and the Cronbach's alpha statistic for all six scores was modest ( $\alpha=.50$ ).

Next, the different hypothesised models of procedural knowledge (Table 8) were tested using CFA in a Structural Equation Modelling (SEM) framework. Theoretically justifiable

---

<sup>1</sup> Twelve values over 2.0 were replaced with 2.5: upper threshold defined as  $Q_3 + 1.5 \times IQR$ , where  $Q_3$  is the 75<sup>th</sup> percentile.

refinements with a modification index (MI)  $\geq 3.84$  were adopted, specifically, method factors reflecting associations between scores derived from the same video (Marsh & Bailey, 1991). The 3-factor model (*perceiving + naming + interpreting*) provided the best initial fit to the data (Figure 1, Table 8). MIs highlighted three possible additions of which only one - a correlation between residuals for Video 2 *perceiving* and *interpreting* scores - was accepted as a method factor. The corresponding Video 1 covariance was included based on theory.

The final 3-factor model is shown in Figure 1. Loadings of observed variables onto latent variables were all  $\geq .6$  and model fit was excellent (Table 8). Although correlations between *perceiving* and the higher-order facets of *naming* ( $r=.44$ ) and *interpreting* ( $r=.57$ ) appeared stronger than associations between the two higher-order facets themselves ( $r=.24$ ), this may stem from the dependence of *naming* and *interpreting* on the reporting of a strategy. What can be confidently concluded is that - as predicted - the two facets theorised to depend on classroom experience (i.e. *perceiving* and *interpreting*) were more strongly associated than were *perceiving* and the more formal *naming*. Factor scores for *perceiving*, *naming* and *interpreting* were generated using the final 3-factor model, and used for all later analyses.

### **Does procedural knowledge predict observed quality of practice?**

The next analysis examined relationships between teacher knowledge (measured by the OLP) and classroom quality as measured by the Oral Language Quality score, the SSTEW (*quality of adult-child interactions*) and the ECERS-3 (*global quality of education and care*). Bivariate correlations are shown in Appendix A.3. Figure 2 shows the multivariate path model assessing relationships between the three OLP knowledge factors (*perceiving*, *naming*, *interpreting*) and the three quality measures, for teachers whose classes were observed ( $n=55$ ). Covariates were selected based on significant correlations with quality.

As expected, the explicit, higher-order knowledge facets proved the strongest predictors

of quality. *Naming* uniquely predicted all three quality measures, with large positive associations ( $>.5$ ). *Interpreting* predicted ECERS-3 scores and fell just below significance for the SSTEWS and Oral Language Quality scores. Relationships between *naming*, *interpreting* and quality are presented visually in Figure 3, for associations identified as significant in Figure 2. On average, quality increased with every additional professional term or interpretation used. This suggests that, despite the high proportion of zero scores, *naming* and *interpreting* are better represented as scales than as dichotomous variables (i.e. 0 terms vs  $\leq 1$  terms).

Although *perceiving* showed small positive correlations with quality in bivariate tests ( $p \leq .10$ ; Appendix A.3), effects did not persist in the multivariate models (Figure 2). In fact, once *naming* and *interpreting* were accounted for, *perceiving* showed significant *negative* associations with classroom quality. Both covariates showed weak positive associations with quality [*intervention participation*: Oral Language Quality ( $\beta = .24$ ,  $p \leq .05$ ), ECERS-3 ( $\beta = .18$ ,  $p \leq .10$ ); *teaching experience*: Oral Language Quality ( $\beta = .18$ ,  $p \leq .10$ ), SSTEWS ( $\beta = .19$ ,  $p \leq .10$ )].

The full path model (Figure 2) explained 37% of variation in the Oral Language Quality score and 27-31% of the variation in wider quality (SSTEWS, ECERS-3). Without covariates, the OLP knowledge factors explained 25-30% of variation in classroom quality.

Ideally, all analyses would have been replicated in the control group, as representing a population unaffected by intervention. Sample sizes were too small to support robust subgroup analysis, so results are not reported here. However, exploratory analysis (Mathers, 2020) indicates that effects for *naming* and *interpreting* are maintained for this group.

### **Is procedural knowledge predicted by teachers' opportunities to learn?**

The next stage tested associations between knowledge and teachers' opportunities to learn: type of teaching qualification, intervention participation and years of classroom

experience. A path model (Figure 4) was estimated to predict the three OLP knowledge factors (*perceiving*, *naming*, *interpreting*), with independent variables selected based on bivariate correlations (Appendix A.4) and theory. On the basis that expertise develops during the first 5-10 years in the classroom (Palmer, Stough, Burdenski & Gonzales, 2005), the model was rerun excluding respondents with >7 years of teaching experience (n=41).

As hypothesised, access to formal learning opportunities predicted higher levels of pedagogical knowledge. Teachers from intervention group schools were significantly better at *perceiving* and *naming* than those in the control group, though associations were small. Effect sizes (*g*) were .58 and .65 respectively. To test the effects of dose, 17 intervention group teachers who had attended <5 days of the 6-day training were excluded, and the analysis rerun (n=87). Most of those excluded (n=15) had joined the study (and programme) after inception. Effects on knowledge were larger for the 'full attenders' group (*perceiving* *g*=.73, *naming* *g*=.84). Significant effects were confirmed using robustness checks (Appendix A.5). The lack of data on teachers' baseline OLP scores means that we cannot ascribe effects on knowledge to the intervention with certainty. However, the fact that effect sizes were greater for full attenders - alongside the random allocation of schools – allows for relative confidence.

Options for testing qualification effects were limited, since all respondents were graduate-qualified teachers. There were indications that primary-trained teachers had greater knowledge than those with a preschool qualification, but the only significant association (for *interpreting*) was weak and did not stand up to robustness checks (Appendix A.5).

Turning now to informal learning opportunities, the hypothesised relationships between *perceiving*, *interpreting* and years of classroom experience were not confirmed by the path

model (Figure 4). Associations were larger for teachers in their first 7 years of practice but remained weak and below conventional statistical significance thresholds [preschool experience & *perceiving* ( $\beta=.26$ ,  $p=.11$ ); preschool experience & *interpreting* ( $\beta=.19$ ,  $p=.43$ )]. Further research in a larger sample is needed to establish more conclusively whether such skills develop during the first years of hands-on practice with young children.

### **Does change in knowledge mediate intervention effects on quality?**

Previous models (Figures 2 & 4) indicated a possible indirect effect of the intervention on the Oral Language Quality Score, via the *naming* knowledge of the participant. A reduced path model was generated to test direct and indirect effects (Figure 5). Intervention effects on the Oral Language Quality score were fully mediated by *naming*, with the indirect pathway accounting for 42% of total effects. Model fit was fair but not perfect (Figure 5).

## **Discussion**

The OLP tool captures teachers' professional vision as means of eliciting the 'dynamic' procedural knowledge they are able to access, activate and use in classroom situations (Kersting et al, 2012) in order to support real-time decision-making. It is the first preschool video measure to focus on knowledge of oral language pedagogy, and to take exploratory steps in considering the role of higher order knowledge for preschool teachers.

The current study examines the structure of teachers' procedural knowledge and considers whether knowledge predicts classroom quality, whether participation in different learning experiences (qualifications, professional development, classroom experience) predicts teacher knowledge – and whether change in knowledge through professional development (PD) mediates the effects of that PD on classroom quality. This section presents conclusions, interpretations and implications for each question in turn, considers measurement issues and study limitations, and highlights next steps for future research.

**Can *perceiving*, *naming* and *interpreting* be empirically distinguished?**

Three cognitive facets were examined. *Perceiving* (the ability to identify salient language-supporting strategies in video interactions) is theorised to reflect the existence of codified strategy scripts which can be recalled to support use of appropriate strategies in classroom interactions. *Naming* (use of professional vocabulary to describe interactions) and *interpreting* (the ability to interpret interactions) are theorised to reflect explicit and higher-order knowledge, accessible for deliberate manipulation to support pedagogical reasoning, intentional teaching and the articulation of knowledge to others. Findings indicate that the three facets can be empirically distinguished, with moderate associations between facets indicating that valuable information is provided by each, independent of the others. The distinction between *perceiving* and *interpreting* has been shown for secondary teachers (König et al, 2014; Seidel & Stürmer, 2014) but this study offers the first evidence for preschool teachers. The identification of *naming* as a third facet further extends existing research.

**Does dynamic pedagogical knowledge predict teaching quality?**

Greater knowledge of language-supporting strategies was associated with higher quality classroom practice. Together, the three OLP factors and covariates explained 27-37% of variation in quality. This confirms the theoretical importance of pedagogical content knowledge (Schulman, 1986), and extends empirical studies of mathematics and science teachers (Baumert et al, 2010; Bruckmaier, Krauss, Blum & Leiss, 2016; McCray, 2008; Kersting 2010, 2012) in showing that such knowledge also matters for preschool teachers. Findings offer confirmatory evidence that procedural knowledge is important for effective teaching, that prior null results may have stemmed from measurement failings, and that video assessment offers a more sensitive means of capturing dynamic knowledge than questionnaires (Jamil et al, 2015; Seidel & Stürmer, 2014; Sherin & van Es, 2009; Kersting,

2010).

Despite the light-touch way in which they were captured, the explicit and higher-order facets of *naming* and *interpreting* were - as hypothesised - stronger predictors of classroom quality than lower-order knowledge (*perceiving*). Although there is value in knowing 'how to do', explicit knowledge of 'how and why' matters more. Unstandardised regression coefficients show that Oral Language Quality Scores were .48 of a point higher on the 7-point scale for every additional vocabulary term used to describe the video interactions, and .23 of a point higher for every interpretation provided. Given the low levels of quality seen overall (mean scores all <3), these differences can be considered meaningful.

This study confirms prior research suggesting that pedagogical reasoning is an indicator of teaching expertise (Kersting et al, 2012; Seidel & Sturmer, 2014; van Es & Sherin, 2002) and offers indicative evidence of its relevance for preschool teachers. It also highlights a potential role for professional vocabulary in expert practice. Having the vocabulary with which to name a concept may allow teachers to engage in professional reflection and discussion upon that concept, supporting deeper understanding and intentional teaching. Use of specific vocabulary may also indicate that teachers have attended prior language-related professional development, and thus be a proxy for wider formal knowledge. Being able to reason about classroom situations may further support teachers in intentional practice – that is, the deliberate promotion of specific language content or child outcomes (Kind, 2009). *Naming* and *interpreting* may also aid teachers in pedagogical leadership, for example in articulating their knowledge to others and explaining why certain practices are important. This notion is supported by the fact that the SSTEWS and ECERS-3 scales reflect *global* classroom quality, including the physical environment and the practice of adults other than the lead teacher. Thus, teachers with greater knowledge led classes which were of

higher quality rather than (or as well as) displaying individually more effective practice.

### **How can procedural knowledge best be promoted?**

The key question for policy and practice is how best to nurture teachers' procedural knowledge. In line with previous research (Jamil et al, 2015; Seidel, Königs & Stürmer, 2014), findings suggest that 'informal' learning gained through classroom experience is not sufficient to ensure a knowledgeable workforce. Teachers also need access to formal learning opportunities, such as pre-service qualifications and in-service professional development, in order to develop the explicit and higher-order knowledge which underpins effective practice.

The strongest predictor of pedagogical knowledge was participation in the professional development intervention, with moderate effects for *perceiving* ( $g=.58$ ) and *naming* ( $g=.65$ ) and larger effects ( $.73/.84$ ) for teachers who had attended the majority of the training. This supports previous research in showing that preschool teachers' professional competence can be improved through carefully-designed programmes which offer a mix of theory and classroom-based support (Hamre et al, 2012). The path analysis offers tentative evidence of a mediating role for knowledge in improvement to practice, with 42% of intervention effects on language-supporting quality explained by growth in professional vocabulary (*naming*). Finally, it is notable that the professional development had no effect on *interpreting*. This is in some ways unsurprising, since interpreting was not an explicit focus of the programme. However, it illustrates that such learning cannot be taken for granted, even in formal programmes, and that pedagogical reasoning must be explicitly nurtured (e.g. Santagata & Yeh, 2014).

Although options for exploring pre-service qualification effects were limited, findings highlight a possible advantage for primary-trained teachers as compared with preschool specialists in relation to *interpreting*. It is impossible to establish whether primary programmes offer more effective preparation, or whether differences stem from the



characteristics of teachers who elect to undertake this pathway. However, prior evidence from Germany indicates that pre-service graduate programmes combining theory and practical internships *can* promote pedagogical reasoning in secondary school teachers (Stürmer, Seidel & Schäfer, 2013). Similar research is now needed within the UK preschool context to explore the extent to which the diverse qualifications held by the preschool workforce are successful in promoting procedural knowledge, and identify where refinements can be made. The creation of the OLP tool makes such research more feasible.

A final reflection on the value of classroom experience: although anticipated relationships with procedural knowledge were not confirmed, small associations *were* seen with classroom quality (Oral Language Quality  $\beta=.18$ ,  $p=.09$ ; SSTEWS  $\beta=.19$ ,  $p=.07$ ) which may have been statistically significant in a larger sample. Perhaps the informal learning which occurs through practice is too deeply tacit for even a contextualised video measure to detect.

### **Is higher-order knowledge domain specific?**

Ironically, having worked to capture knowledge relating explicitly to oral language, the study indicates that such knowledge may not be domain-specific. Although the proportion of variance explained was highest for the Oral Language Quality score, the OLP (alongside covariates) also explained 31% of variance in SSTEWS scores and 27% of variance in ECERS-3 scores. Teachers with a richer professional vocabulary and better reasoning skills in relation to *language pedagogy* led classrooms which offered better quality learning environments *overall*. This may reflect the fact that higher-order knowledge is inherently universal rather than domain-specific. Alternatively, it may reflect the relevance of oral language to child learning across all domains and the interconnected nature of early childhood pedagogy, in which effective adult-child interactions support development in multiple domains. This does not argue against domain-specific professional development but suggests that such training

should also explicitly nurture universal, higher-order knowledge and skills.

### **Issues of measurement and study limitations**

This study builds on the sound methodological foundations of the first validated video measure of preschool teachers' professional vision (the VAIL), which predicted 12% of variation in classroom quality (Jamil et al, 2015) by assessing teacher's ability to identify effective teaching strategies (*perceiving*). The superior predictive power of the OLP (37% for the Oral Language Quality score) can plausibly be explained by its inclusion of higher-order cognitive facets such as *naming and interpreting*. Nonetheless, findings can only be regarded as exploratory. *Naming* and *interpreting* were captured in a light-touch manner. Scores reflected teachers' spontaneous use of professional vocabulary and pedagogical reasoning, and the resulting scales were blunt (0-4, 0-5) with many zero responses. Teachers who offered interpretations are likely the most 'expert' - those for whom classroom reasoning is embedded and routine - and relationships with quality may be weaker when analysis is directly prompted. The study suffers other important measurement limitations. Designed for completion in  $\leq 30$  minutes, the OLP is based on responses to 2 short videos, both reflecting child-initiated play contexts with small numbers of children. It focuses on knowledge of teaching strategies and omits other facets (e.g. knowledge of child development and learning). As such, it cannot be said to reflect all domains of pedagogical knowledge, or assess knowledge across a representative range of early childhood contexts. In addition, both the initial study response rate and the resulting sample size were modest, particularly for analyses of relationships between knowledge and classroom quality ( $n=55$ ), limiting both the generalisability of results and the opportunities for robust sub-group analysis.

There is clearly much scope for broadening and deepening the OLP following this exploratory study. However, the fact that even this relatively shallow measure proved a

strong predictor of quality, and that findings align with hypothesised outcomes and previous studies, offers promise for more nuanced approach in yielding further insight. Confirmatory research is now needed in a larger sample, extending the OLP's scope to capture a more holistic view of pedagogical knowledge (e.g. using a wider range of videos) and drawing on the studies of Kersting and others to elicit *interpreting* and *naming* more richly, so that associations with classroom quality and child learning can be more robustly examined.

Future research should also delve more deeply into the apparent failure of *perceiving* to predict classroom quality. The importance of strategy knowledge is supported by previous studies (e.g. Jamil et al, 2015) and should not be dismissed. It may be that the expert benchmarking process failed to capture strategy salience (Mathers, 2020), meaning that a high *perceiving* score reflected indiscriminate rather than targeted noticing (i.e. a tendency to simply list as many strategies as possible). It is also possible that effects were limited by the use of a global classroom quality measure. While informal strategy knowledge may not promote global quality because it does not enable teachers to articulate knowledge to others, it may support individuals' in their *own* teaching. Stronger positive effects may have been seen, had a measure of individual practice been available.

### **Final thoughts**

This research offers new evidence on the importance of practical classroom knowledge in ensuring high-quality teaching, but also illustrates its higher-order nature and the need for it to be explicitly nurtured. The dynamic procedural knowledge of the most effective preschool classroom leaders is multi-dimensional, complex and deeply cognitive – and does not 'come naturally'. To ensure a workforce capable of supporting young children's language and wider development, we must offer learning opportunities which are both formal and situated, and which support practitioners in transforming theory into explicit, higher-order procedural

understanding (Eraut, 2004) – including the development of pedagogical reasoning and a rich professional lexicon. Careful attention must be paid to the optimal integration of theory and practice components within pre- and in-service training (Brouwer & Korthagen, 2005), and to the ways in which deliberative learning can be promoted in the day-to-day life of the classroom. Further empirical work is needed to examine how this can best be achieved, using the new opportunities provided by video tools to assess knowledge growth. Much would also be gained from routinely including these more sensitive measures of procedural knowledge in studies of professional development, to inform intelligent programme design.

## References

- Allison, P.D. (2012). Handling Missing Data by Maximum Likelihood. *SAS Global Forum, paper 312-2012*. <https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf>
- Alonzo, A.C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, 53, 1259-1286.
- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
- 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.
- Ball, D.L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. New Jersey: Prentice-Hall.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133-180.
- Ben-Peretz, M. (2011). Teacher knowledge: What is it? How do we uncover it? What are its implications for schooling? *Teaching and Teacher Education*, 27(1), 3-9.
- Berliner, D.C. (1988, February). The development of expertise in pedagogy. *Charles W. Hunt Memorial Lecture presented at the Annual Meeting of the American Association of Colleges for Teacher Education* (New Orleans, LA, February 17-20, 1988) ED 298 122.
- Blaine, B. (2018) Winsorizing. In Frey, B. B. (Ed.). (2018). *The SAGE Encyclopaedia of Educational Research, Measurement, and Evaluation*. California: SAGE Publications.
- 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.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13.
- Bransford, J.D., Brown, A.L. & Cocking, R.R.(eds.) (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academies Press.
- Bronfenbrenner, U. & Morris, P.A. (2006). The biological model of human development. In W. Damon & R.M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). New York: Wiley.

Brouwer, N., & Korthagen, F. (2005). Can Teacher Education Make a Difference? *American Educational Research Journal*, 42(1), 153–224.

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.

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2(1), 1-19.

Clarke, D., Mesiti, C., Cao, Y., & Novotna, J. (2017). The lexicon project: examining the consequences for international comparative research of pedagogical naming systems from different cultures, In T. Dooley, & G. Gueudet (Eds.) *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10, February 1 – 5, 2017)*. Dublin, Ireland: DCU Institute of Education and ERME.

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.

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

Desimone, L. (2009). Improving impact studies of teachers' professional development: towards better conceptualisations and measures. *Educational Researcher*, 39(3), 181-199.

DiCiccio, T.J. & Efron B. (1996) Bootstrap confidence intervals. *Statistical Sc.*, 11, 189-228.

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: effects of the Literacy Environment Enrichment Program. *Early Childhood Research Quarterly*, 22(2), 243–260.

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

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.

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.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guildford Press.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26(2), 247–273.

Gardner-Neblett, N. & Iruka, I. U. (2015). Oral narrative skills: Explaining the language-emergent literacy link by race/ethnicity and SES. *Developmental Psychology*, 51(7), 889–904.

Glaser, R. & Chi, M.T.H. (2014) Overview. In Chi, M.T.H, Glaser, R & Farr, M.J. (2014) *The Nature of Expertise*. New York/London: Psychology Press.

- Goodwin, C. (1994) Professional vision. *American Anthropologist*, 96 (3), 606-633.
- Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., et al. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49(1), 88-123.
- Harms, T., Clifford, R.M. & Cryer, D., (2014) *Early childhood environment rating scale, Third Edition*. New York: Teachers College Press.
- Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900–1902.
- Hjetland, H.N., Lervåg, A., Lyster, S.-A.H., Hagtvet, B.E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, 111(5), 751-763.
- Howard, S.J., Siraj, I., Melhuish, E., Kingston, D., Neilsen-Hewett, C., de Rosnay, M. et al (2018). Measuring interactional quality in pre-school settings: introduction and validation of the SSTEWS scale. *Early Child Development and Care*, 1–14.
- Hulme C. & Snowling M.J. (2014) The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society B* 369: 20120395.
- 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.
- Kaiser, G., Busse, A., Hoth, J., König, J. & Blömeke, S. (2015) About the Complexities of Video-Based Assessments: Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, 13, 369–387.
- Keppens, K., Consuegra, E., Goossens, M., De Maeyer, S. & Vanderlinde, R. (2019) Measuring pre-service teachers' professional vision of inclusive classrooms: A video-based comparative judgement instrument. *Teaching and Teacher Education* 78, 1-14.
- 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.
- Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, 45, 169–204.
- Klein, G.A. & Hoffman, R.R. (1993) Seeing the invisible: perceptual-cognitive aspects of expertise. In M. Rabinowitz (ed.) *Cognitive science foundations of instruction* (pp.203-226). Mahwah, NJ: Erlbaum.

Kniesel, I., Lindmeier, A.M. & Heinze, A. (2015). Beyond Knowledge: Measuring Primary Teachers' Subject-Specific Competences in and for Teaching Mathematics with Items Based on Video Vignettes. *International Journal of Science and Maths Education*, 13, 309–329.

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.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A & Kaiser, G., (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education* 38, 76-88.

König, J. & Kramer, C. (2016) Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Mathematics Education*, 48, 139-151.

Marsh, H.W. & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15(1), 47–70.

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.

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

Mathers, S. & Smees, R. (2014) *Quality and Inequality - Do three and four year olds in deprived areas experience lower quality early years provision*. London: Nuffield Foundation

Morgan, P.L., Farkas, G., Hillemeier, M.M., Hammer, C.S. & Maczuga, S. (2015) 24-Month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, 86, 1351-1370.

Moskowitz, G.B. (2005) *Social cognition: Understanding self and others*. New York: Guilford.

McCray, J. (2008) *Pedagogical content knowledge for preschool mathematics: relationships to teaching practices and child outcomes (Doctoral dissertation)*. Available at: <https://search.proquest.com/docview/304557941/8860865A1E7D4848PQ/3?accountid=13042>

McDonald, J.F. & Moffitt, R.A. (1980) The Uses of Tobit Analysis. *The Review of Economics and Statistics*, 62(2), 318-321.

Palmer, D. J., Stough, L. M., Burdinski, J., Thomas, K., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Ed. Psych.*, 40(1), 13–25.

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.



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.

Piasta, S.B., Park, S., Farley, K.S., Justice, L.M. & O'Connell, A.A. (2019) Early childhood educators' knowledge about language and literacy: Associations with practice and children's learning. *Dyslexia*, 26, 137–152.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4–15.

Roulstone, S., Law, J., Rush, R., Clegg, J., & Peters, T. (2011). *Investigating the role of language in children's early educational outcomes. Research Report DFE-RR134*. London: Department for Education.

Sabers, D.S., Cushing, K.S., & Berliner, D.C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, 28(1), 63–88.

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.

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.

Schank, R. (1997). *Virtual learning*. New York: McGraw-Hill.

Schreiber, J.B., Stage, F.K., King, J., Nora, A. & Barlow, E.A. (2006) Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6), 323-337.

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in pre-service teachers. *American Educational Research Journal*, 51(4), 739–771.

Seidel, T., Königs, K.D. & Stürmer, K. (2014) Factors In University-Based Teacher Education Relating to Preservice Teachers' Professional Vision. *Vocations & Learning*, 8, 35–54.

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.

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

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.

Siraj, I., Kingston, D., & Melhuish, E.C. (2015) *Assessing Quality in Early Childhood Education and Care: Sustained Shared Thinking and Emotional Well-Being (SSTEW) Scale for 2–5-Year-Olds Provision*. Stoke-on-Trent: Trentham Books.

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

StataCorp (2017) *Stata Statistical Software 15*. College Station, TX: StataCorp LLC.

Stürmer, K., Seidel, T. & Schäfer, S. (2013). Preservice teachers’ professional vision changes following practical experience: a video-based approach in university-based teacher education. *Gruppendynamik und Organisationsberatung*, 44(3), 339-355.

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.

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing Quality in the Early Years Early Childhood Environment Rating Scales Extension (ECERS-E). Four Curricular Subscales*. Stoke on Trent, UK and Stirling: USA Trentham Books.

Van Driel, J.H. & Berry, A. (2012) Teacher Professional Development Focusing on Pedagogical Content Knowledge. *Educational Researcher*, 41(1), 26-28.

Van Es, E. A. & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers’ interpretations of classroom interactions. *J of Technology and Teacher Ed.*, 10(4), 571–596.

Voss T., D., Kunter, M. & Baumert, J. (2011). Assessing Teacher Candidates' General Pedagogical/Psychological Knowledge: Test Construction and Validation. *Journal of Educational Psychology*, 103 (4), 952-969.

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

Wasik, B.A., Bond, M.A. & Hindman, A. (2006) The Effects of a Language and Literacy Intervention on Head Start Children and Teachers. *Journal of Ed. Psychology*, 98(1), 63–74.

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.

**Table 1** *Three facets of professional vision and underlying procedural pedagogical knowledge*

Cognitive facet of professional vision	<b>Perceiving:</b> the ability to perceive salient pedagogical strategies being used in a specific video interaction	<b>Naming:</b> the ability to describe the interaction observed using specific professional vocabulary	<b>Interpreting:</b> the ability to interpret the interaction observed (e.g. the teacher's pedagogical intention, the observed effect on the child, possible alternative approaches)
Underlying procedural knowledge	<b>Lower-order and potentially-tacit knowledge:</b> teaching strategy scripts which can be recalled to support application of an appropriate strategy in a specific classroom situation.	<b>Higher-order and explicit knowledge</b> of pedagogical strategies, which can be deliberately manipulated to support reasoning about classroom situations, intentional teaching and the articulation of knowledge to others.	

**Table 2.** *The OLP coding framework: 30 items within six pedagogical categories*

Category	Strategies	Expert ratings (1=low, 5=high)		
		Import- ance	Expert example of strategy	
			Video 1	Video 2
1: Modelling language	1. Modelling diverse, rich or specific vocabulary	5.00	4.00	2.33
	2. Modelling diverse, rich or specific grammar	5.00	3.67	3.67
	3. Linguistic expansion or recasting of children's language	4.75	2.67	4.33
	4. Emphasising, repeating or reinforcing language	4.75	5.00	4.00
	5. Using descriptive, informative, narrative language in concrete contexts	4.50	4.67	4.00
	6. General language modelling (i.e. appropriate but not reflecting above features)	3.50	4.00	4.67
2: Word meanings	7. Providing explicit definitions of words	4.00	2.33	1.00
	8. Providing concrete clues to meaning (e.g. gestures, pictures, props)	4.00	4.00	4.00
3: Communication & conversation	9. Engaging in conversation with children	5.00	4.00	4.67
	10. Non-verbal strategies to invite communication	4.75	3.67	4.67
	11. Verbal strategies to invite communication	4.75	4.33	4.67
	12. Prompting children to use new vocabulary	4.00	4.67	1.67
	13. Being a responsive conversation partner	4.50	5.00	5.00
	14. Extending conversational or narrative content (semantic extension)	4.25	4.67	4.33
	15. Supporting mutual understanding/adapting language to child's level	4.50	4.67	4.67
	16. Supporting children to attend and participate	4.50	3.67	5.00
4: Higher order language & thinking	17. Affirming the child's language by repeating it	3.25	4.67	4.67
	18. Promoting children's thinking	4.75	3.67	3.33
	19. Prediction, speculation, reasoning, explanation and inference	3.75	4.67	3.67
	20. Modelling fictional narrative/pretending	3.75	2.33	2.67
	21. Use of open questions	4.50	4.67	4.33
5: Relationships & the child	22. Positive affect or communication	4.00	4.67	5.00
	23. Individual attention and sensitive responding	3.75	4.33	5.00
	24. Using a non-directive approach	3.75	4.33	4.67
	25. Facilitating peer communication	4.25	1.00	3.67
	26. Facilitating peer interactions and relationships	3.25	1.00	1.67
	27. Promoting children's self-worth	3.25	4.33	4.33
6: Meaningful & engaging contexts	28. Joint attention/following children's lead and interests	4.75	4.67	4.33
	29. Providing meaningful/engaging contexts and activities for language	4.25	4.33	4.67
	30. Deepening learning (e.g. revisiting language across contexts, activities etc)	4.00	1.33	1.67

<b>Perceiving (1 score per video):</b> each valid strategy identified is awarded the relevant 'expert example' rating for that video (Table 2). Responses can match multiple strategies and be described using informal language (e.g. <i>how/why questions</i> ). <b>Range:</b> no limit	<b>Naming (1 score per video):</b> 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. <b>Range:</b> 0-8 per video	<b>Interpreting (1 score per video):</b> 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. <b>Range:</b> 0-8 per video	
<b>Illustrative teacher response to Video 1</b>	<b>Perceiving</b> (expert example rating)	<b>Naming</b>	<b>Interpreting</b>
1. "...The teacher uses key vocabulary eg. <i>slower, faster.. further. She continually repeats this vocabulary and the child soaks up the language and begins to use it herself eg. 'Further...yeah further'</i> "	Strategy 1 – (4.00) Strategy 4 – (5.00)		1 (observed effect on child)
2. "Using correct mathematical vocabulary - <i>cubeoid not brick</i> "	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 <i>running commentary</i> on the child's actions"	Strategy 5 – (4.67)	1	
5. "Questioning <i>to provide opportunities for child to apply words and show understanding linked to meaningful experience</i> "	Strategy 12 – (4.67) Strategy 15 – (4.67) Strategy 29 – (4.33)		1 (pedagogical intention)
6. "Using gestures - when saying <i>steeper slope</i> "	Strategy 8 – (4.00)		
8. "Listening carefully and valuing all the child's ideas"	Strategy 13 – (5.00) Strategy 27 – (4.33)		
8. "Open ended questions – how did you do that?"	Strategy 21 – (4.67)	1	
<b>Total score for Video 1</b>	<b>56.34</b>	<b>2</b>	<b>3</b>

	Video 1 (n=104)					Video 2 (n=101)*				
	%>0	Min.	Max.	Mean	S.D.	%>0	Min.	Max.	Mean	S.D.
<b>Perceiving</b> (no limit)	98	0	55.0	23.0	10.3	99	0	41.7	20.7	9.0
<b>Naming</b> (0-8)	48	0	2.0	.57	.65	31	0	2.0	.40	.65
<b>Interpreting</b> (0-8)	30	0	6.0	.60	1.2	31	0	6.0	.53	1.0
<b>Interpreting</b> (Winsorised) <sup>#</sup>	30	0	2.5	.50	.85	31	0	2.5	.46	.76

# Winsorised to address outliers (scores >2) – see ‘Data Preparation’

Factor loadings are standardised coefficients from CFAs conducted in the study sample of observed teachers (n=55) using Stata 15. To address the low case-to-item ratio (5.5:1) the same CFAs were run in the full RCT dataset (n=115; case-to-item ratio=11.5:1). Results were almost identical. The two items scored through in the table below were dropped to form the final 8-item model. Correlations between residuals for items were included in the model where modification indices were  $\leq 3.84$ .

[illegible]

E3 12. Helping children expand vocabulary	Staff frequently use specific words for people, places, things, actions and descriptive words as children experience routines and play	✓	✓			.71	.72
E3 30. Staff-child interactions	Frequent positive staff-child interaction observed, no long periods of no interaction			✓		.90	.89
SS2. Encouraging choices and independent play	<del>If invited, the staff will join in with children's play while allowing the children to lead, respecting the level of play and rules established by the children</del>			✓		.56	
SS5. Encouraging children to talk with others	Children are encouraged to choose and lead interactions, conversations and play	✓	✓	✓	✓	.87	.87
SS6. Staff actively listen to children & encourage children to listen	Staff allow long pauses, so children have time to think and respond. They ..allow different length of pauses with different children.		✓	✓		.85	.85
SS7. Staff support children's language use	The staff scaffold & model language with individual children...slightly above the child's current level.	✓	✓	✓		.82	.84
SS8. Sensitive responsiveness	Staff ensure that most children receive extended individual attention at least once during the session.		✓	✓		.81	.81
SS10. Encouraging sustained shared thinking through storytelling, books, singing, rhymes	<del>Staff encourage children to hold and 'read' books or retell familiar stories, including their own 'stories', sing songs, or join in with rhymes &amp; word games.</del>		✓	✓	✓	.43	
SS12. Supporting concept development & higher-order thinking	Staff support children in thinking through what they are doing and extending it through modelling, asking simple open and closed questions, and providing additional resources			✓	✓	.78	.69
EE6. Talking and listening	Children are often encouraged to talk to each other in small groups and adults encourage peers to listen		✓	✓	✓	.82	.81
# Goodness of fit for the 8-item model in the study dataset: $\chi^2$ (df, p) = 13.44 (17, .71); RMSEA = .00; CFI = 1.00; TLI = 1.02.							

**Table 6** Descriptive statistics and Cronbach's alpha ( $\alpha$ ) for the three quality measures (n=55 observed teachers)

	N	Min.*	Max.*	Mean*	S.D.	$\alpha$
<b>Oral Language Quality score</b> (support for oral language)	55	1.08	5.10	2.90	1.08	.94
<b>SSTEW mean score</b> (quality of adult-child interactions)	55	1.21	5.43	2.76	1.06	.94
<b>ECERS-3 mean score</b> (global education/care environment)	55	1.54	4.50	2.86	.72	.93

\* 7 point scale (1=inadequate, 3=minimal, 5=good, 7=excellent)

**Table 7** Spearman's rank order correlations between OLP scores (Video 1 n=104, Video 2 n=101)

	Perceiving. 1	Perceiving 2	Naming 1	Naming 2	Interpreting 1
<b>Perceiving 2</b>	<b>.54***</b>				
<b>Naming 1</b>	.28**	.17			
<b>Naming 2</b>	.20*	.27**	<b>.42***</b>		
<b>Interpreting 1</b>	.40***	.30**	.15	.07	
<b>Interpreting 2</b>	.35***	.47***	.12	.06	<b>.50***</b>

**Table 8** Fit indices for the hypothesised 1-, 2- and 3-factor CFA models of procedural pedagogical knowledge

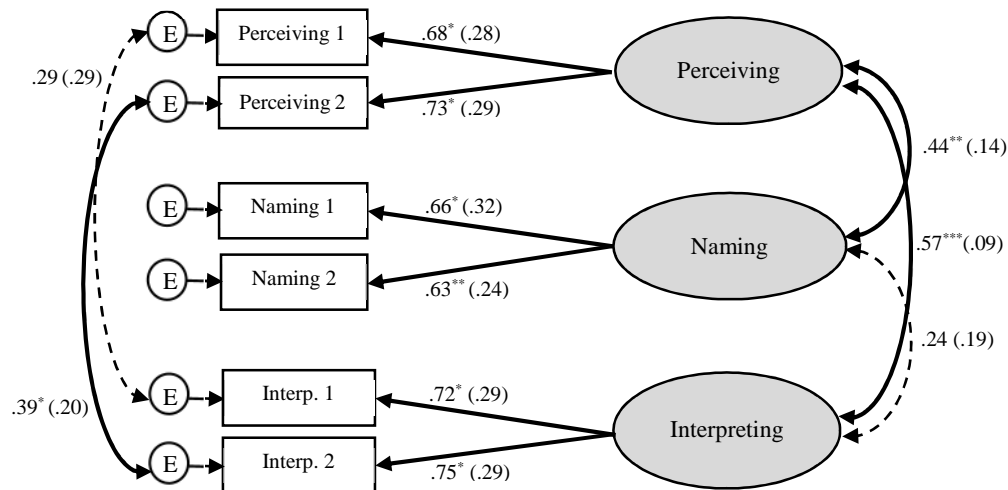
All analyses used FIML estimation (n=104), with clustered robust standard errors adjusted for 72 school-level clusters

	$\chi^2$ (df, p)	RMSEA	CFI	TLI
<b>1-factor model:</b> (perceiving/naming/interpreting)	40.21 (9, .00)	.18	.75	.58
<b>2-factor model:</b> perceiving + (naming/interpreting)	35.55 (8, .00)	.18	.78	.58
<b>2-factor model:</b> (perceiving/interpreting) + naming	24.65 (8, .00)	.14	.87	.75
<b>3-factor model:</b> perceiving + naming + interpreting	15.41 (6, .02)	.12	.92	.81
<b>Final 3-factor model after modification</b>	3.10 (4, .54)	.00	1.00	1.03
<b>Cut-off criteria</b> (Hu & Bentler, 1999; Schreiber et al, 2006)	$\chi^2$ :df<2, p>.05	≤.06	≥.90	≥.95

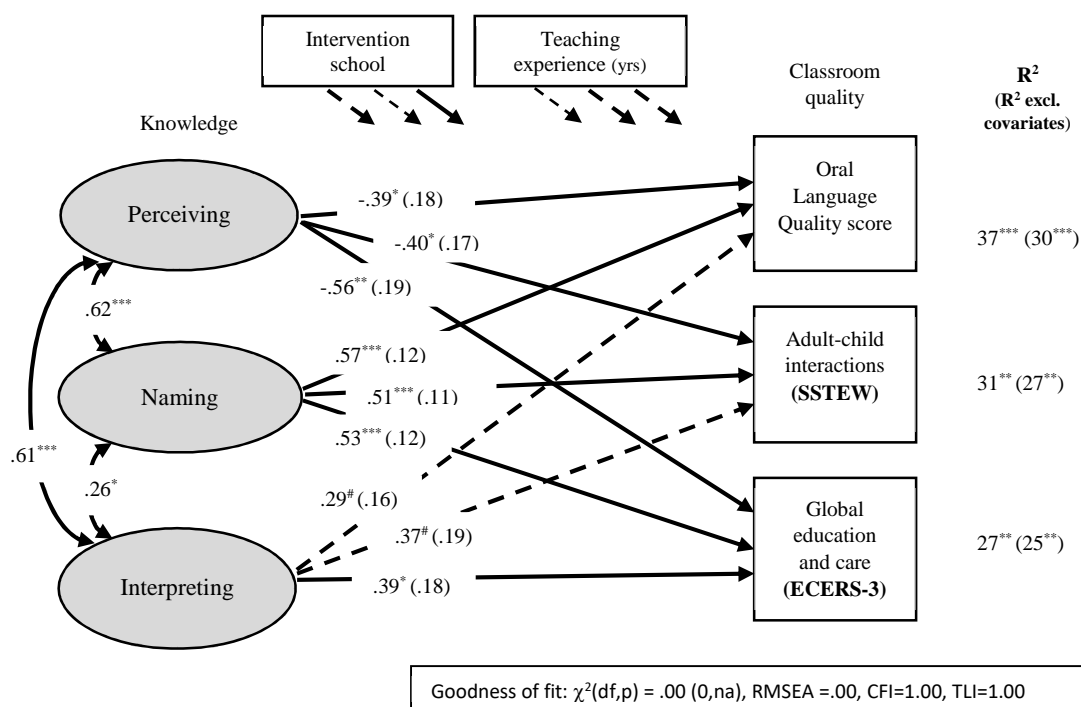
$\chi^2$  = Chi squared, RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index

**Figure 1** Final three-factor CFA model, including modifications

FIML estimation (n=104). All estimates are standardised, with clustered robust standard errors (adjusted for 72 school-level clusters) shown in brackets. Twin headed arrows represent correlations between latent variables/residuals. Significant pathways are represented by solid lines (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .00$ ) and non-significant pathways by dashed lines.

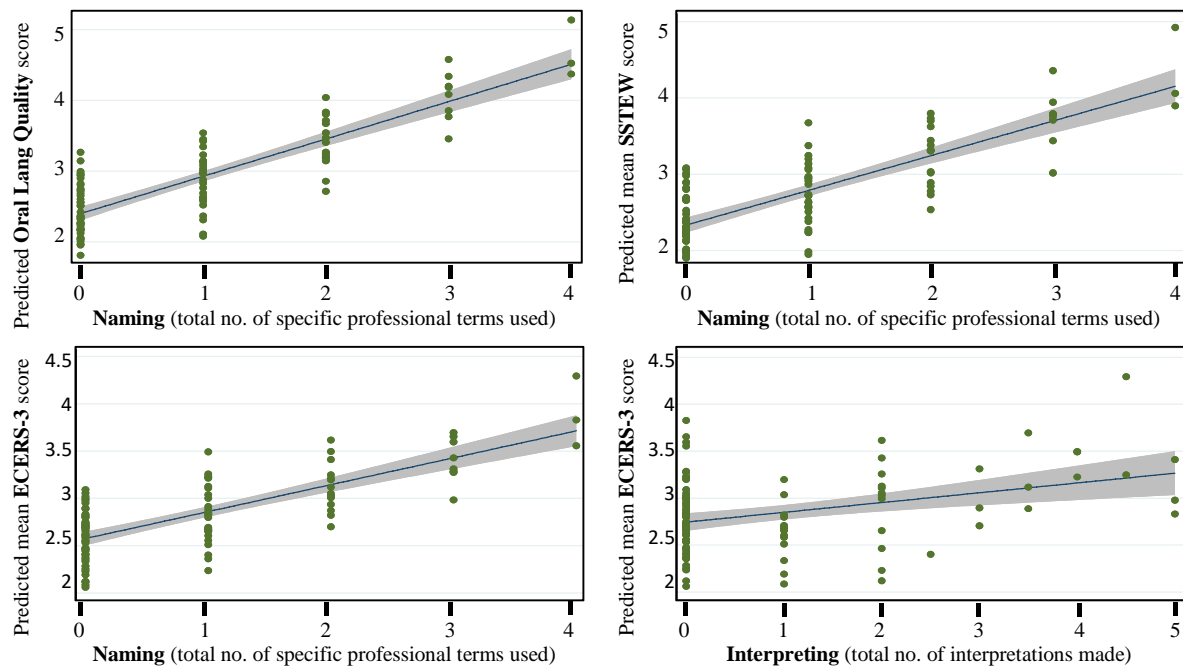
**Figure 2** Path model testing relationships between knowledge and classroom quality

Maximum Likelihood estimation for observed teachers only (n=55). Twin-headed arrows represent correlations between variables. One-headed arrows represent regression paths. All estimates are standardised, with robust standard errors in brackets. Correlations between residuals for the three quality variables were included in the model but are not shown. Significant pathways are represented by solid lines (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .00$ ), non-significant pathways by dashed lines, and pathways close to the threshold of significance (#  $p < 0.10$ ) by bold dashed lines. R squared values were generated using multiple linear regression.

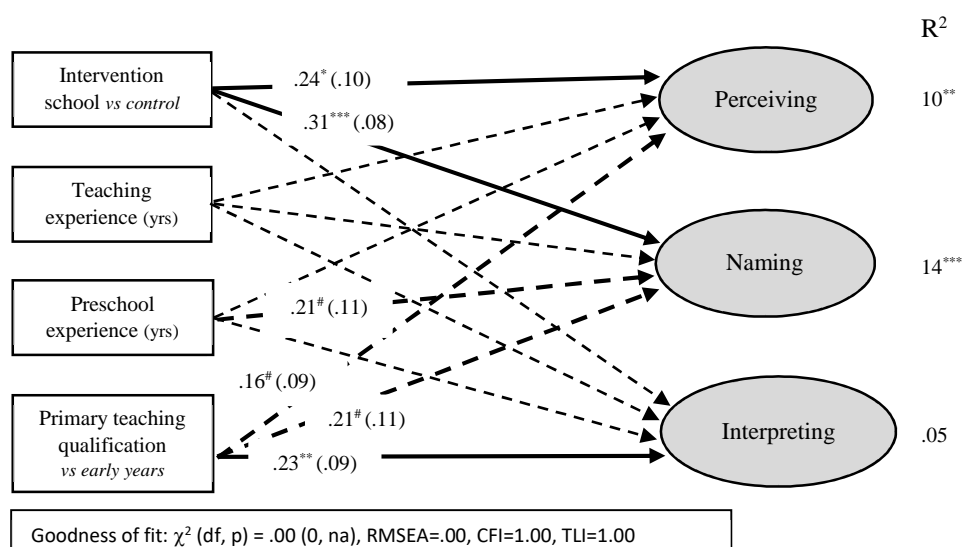


**Figure 3** Scatterplots showing relationships between *naming/interpreting* scores and classroom quality

Predicted classroom quality scores were generated using multiple linear regression, mirroring the path model shown in Figure 2 (n=55 observed teachers) for significant associations only. In place of factor scores, *naming* and *interpreting* (on the x axes) are represented by their sum score across both videos (correlated  $r=.98$  with the relevant latent factor score). The solid line represents the regression line, and the shading the 95% confidence interval for the mean.

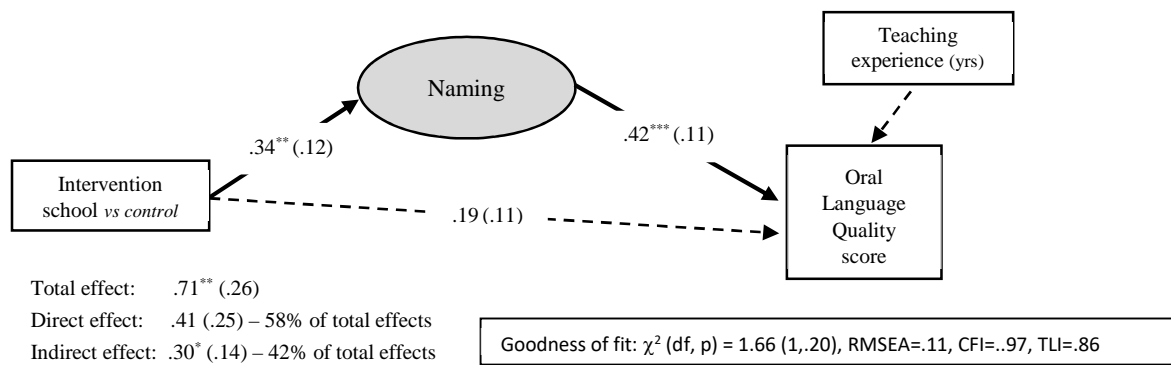
**Figure 4** Path model testing relationships between teachers' opportunities to learn and knowledge (n=97)

Maximum Likelihood estimation in the full sample, excluding 7 respondents with an unspecified teaching qualification in the interests of interpretability. All estimates are standardised, with clustered robust standard errors (adjusted for 66 school-level clusters) in brackets. Correlations between residuals for *perceiving*, *naming*, *interpreting* were included in the model but are not shown. Significant regression pathways are represented by solid lines (\*  $p<.05$ , \*\*  $p<.01$ , \*\*\*  $p<.00$ ), non-significant pathways by dashed lines, and pathways close to the significance threshold (#  $p<.10$ ) by bold dashed lines. R squared values were generated using multiple linear regression.



**Figure 5** Direct and indirect effects of the intervention on classroom quality (Oral Language Quality Score)

Maximum Likelihood estimation (observed teachers only, n=55). All estimates are standardised, with robust standard errors in brackets. Significance levels were identical using bootstrapped standard errors. Significant pathways are shown by solid lines (\* p<.05, \*\* p<.01, \*\*\* p<.00), non-significant pathways by dashed lines, and pathways close to the threshold of significance (# p<0.10) by bold dashed lines.





## Appendix

### Appendix A.1 OLP video clip prompts

Watch the clip and identify the strategies this practitioner is using which might support children's language development. This could include children's understanding of language and/or their expression. If you can, try to record strategies which support a range of language skills (e.g. communication, narrative, vocabulary and/or grammatical skills). You can watch the clip again if you need to, or refer to the transcript. Use the eight boxes below to record your answers. List as many strategies as you can, but if you identify fewer than eight you can leave some boxes blank. [Note: the introductory text had already specified a focus on skills related to making meaning through language rather than code-related skills.]

### Appendix A.2 Descriptive statistics for perceiving, naming and interpreting scores in intervention/control groups

INTERVENTION	Video 1 (n=50)					Video 2 (n=47)				
	%>0	Min.	Max.	Mean	S.D.	%>0	Min.	Max.	Mean	S.D.
Perceiving (no limit)	100	11.67	45.01	24.94	9.00	100	6.33	41.66	23.25	8.49
Naming (0-8)	52	0.0	2.00	.66	.72	45	0.0	2.00	.64	.79
Interp. (0-8) Winsorised	32	0.0	2.50	.59	.94	31	0.0	2.50	.47	.77
CONTROL	Video 1 (n=54)					Video 2 (n=54)				
	%>0	Min.	Max.	Mean	S.D.	%>0	Min.	Max.	Mean	S.D.
Perceiving (no limit)	96	0.0	55.02	21.30	11.12	98	0.0	38.01	18.54	8.96
Naming (0-8)	44	0.0	2.00	.48	.57	19	0.0	1.00	.19	.39
Interp. (0-8) Winsorised	27	0.0	2.50	.42	.76	30	0.0	2.50	.44	.76

### Appendix A.3 Spearman's rank order correlations: OLP scores and classroom quality for observed teachers

	OLP (Observed) Scores (n=55/52)						OLP (Factor) Scores (n=55)		
	Perceiving		Naming		Interp.		Perceiving	Naming	Interp.
	Vid 1	Vid 2	Vid 1	Vid 2	Vid 1	Vid 2			
Oral Language Quality	.19	.13	.39**	.42**	.13	.22	.26 <sup>#</sup>	.48***	.35**
SSTEW mean score	.17	.15	.35**	.43**	.15	.24 <sup>#</sup>	.26 <sup>#</sup>	.46***	.37**
ECERS-3 mean score	.12	-.04	.33*	.29*	.16	.14	.11	.36**	.23 <sup>#</sup>

### Appendix A.4 Associations between OLP factor scores and teachers' opportunities to learn

Parametric tests (Pearson's *r* correlations and analysis of variance) were used for *perceiving*, and non-parametric tests (Spearman's rank order correlations and Kruskal-Wallis *H* tests) for *naming* and *interpreting*. Significance levels: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .00$ , #  $p < .10$ .

	Perceiving	Naming	Interpreting
Teaching experience (years) (n=104)	.02	-.05	-.00
Preschool experience (years) (n=104)	.07	-.01	-.05
Intervention vs control school (n=104)	$F(1,102)=8.81^{**}$	$\chi^2(1)=9.43^{**}$	$\chi^2(1)=2.62$
Primary (3-11) QTS qualification vs early years QTS (3-7) (n=97)	$F(1,95)=2.87^{\#}$	$\chi^2(1)=3.28^{\#}$	$\chi^2(1)=4.05^*$

### Appendix A.5 Robustness checks completed

Robustness checks were completed to ensure that the censored nature of the *naming* and *interpreting* scores had not introduced bias. Tobit regression is designed specifically for censored dependent variables (McDonald & Moffitt, 1980). *Naming* and *interpreting* were regressed on the four opportunities to learn (Figure 4), comparing equivalent tobit and multiple linear regression models for consistency. Since factor scores are by nature centred at 0, *naming* and *interpreting* sum scores for Videos 1 and 2 were used in place of the factor scores, and the lower censoring level set at 0. The effect size for *naming* was also confirmed using bootstrapping (DiCiccio & Efron, 1996), to allow for violation of assumptions of normality.