

Case Study 1: An Evidence-Based Practice Review Report

Theme: School/setting Based Interventions for Learning.

How effective is guided play following shared book reading in improving language development in the Early Years?

Summary

What play-based interventions help young children effectively learn? Research on play-based learning has sought to answer this question. This systematic review moves beyond conventional approaches of free and instructional play, to a play-based intervention known as guided play. Guided play incorporates elements of child exploration and light adult scaffolding and guidance to support learning goals. Evidence suggests that student-led learning supported by teachers is more successful in comparison to direct instruction and independent free play in a range of skills (Honomichl & Chen, 2012). Thus, a systematic literature review was conducted to investigate the effectiveness of guided play in improving language development in the Early Years

Three databases were searched (Web of Science, PsychINFO and ERIC) and seven studies from the USA were found. The researcher, using the Weight of Evidence Framework (Gough, 2007) and Gersten's Quality Indicators for Group Experimental Protocol assessed the quality of each study. All seven studies revealed small to large effect sizes, indicating that shared book reading followed by guided play improved children's vocabulary in Early Years settings. The implications of the review alongside limitations of the studies and further investigation are discussed.

Introduction

The way children effectively learn has been a long-lasting debate in educational research. One method for encouraging literacy and language development in children is using shared book reading. When children engage in shared book reading they are provided with the opportunity to discuss word meaning whilst being supported by an adult. In the classroom, research indicates that during shared book reading, teachers are most likely to discuss definitions of words and their meanings (Dickinson et al., 2018). Therefore, this adult-child interaction provides an ideal situation for meaningful language learning. Reviews on preschool interventions indicate shared book reading is consistently associated with vocabulary growth (National Early Literacy Panel, 2008), and provides a deeper processing of sophisticated words in “semantically rich contexts” (Kleeck, 2008). More recently, literature has explored shared book reading with activities such as play.

Play allows children to tap into creativity and develop curiosity and imagination. The ability to create and explore ideas is the basis for problem-solving skills later on in life. By incorporating play within the classroom setting the fun, dynamic and engaging activities help to increase a child’s enthusiasm and motivation to learn (Garris, Alhers & Driskall, 2002). Evidence indicates play is an essential element of a child’s development, strengthening their understanding of the world around them and learning a range of skills such as oral language skills (Singer, Golinkoff & Hirsh-Pasek, 2006), social skills (Ramani, Siegler & Hitti 2012), executive functioning (Savina, 2014), self-regulation (Becker, McClelland, Loprinzi, & Trost 2014), maths and science learning (Cohen & Emmons, 2017).

A potential reason why play is considered an effective pedagogical approach is that it features the ideal environment to facilitate learning; active attention, engagement, interaction and the ability to build connections to real life (Hirsh-Pasek, Golinkoff, Berk & Singer, 2009). Within the context of the classroom, play can take numerous

forms involving child-led free play and adult-led play, where teachers use scaffolding to demonstrate and support learning. The debate between direct instructions versus a more discovery based free play has received attention (Roskos & Christie, 2001). Some researchers argue independent learning should be dismissed due to the lack of evidence to show its effectiveness (Mayer, 2004). Mayer suggests that the lack of structure and support in unassisted learning could result in a lack of understanding in the classroom. In this review, these two approaches are integrated, with a child-initiated adult supported form of play known as guided play (Weisberg, Hirsh-Pasek, Golinkoff, Kittredge & Klahr 2016).

What is Guided Play?

Guided play is a form of play where children explore an environment or activity whilst receiving adult scaffolding to directly teach (Massey, 2013). This play is child directed combined with gentle guidance where teacher's use suggestions and questioning techniques to support learning goals. The light scaffolding when required may help to engage children in activities for longer periods of time and prevent frustration (Story, 2017).

Some evidence indicates that student-led learning supported by teachers is more successful in comparison to direct instruction and independent free play (Honomichl & Chen, 2012). Likewise, there is evidence to suggest memory and understanding improves when learning is in some way created by the student. This is commonly known as the generation effect (Slamecka & Graf, 1978). It might be suggested that learning is improved in approaches such as guided play, where students have the ability to generate their own ideas. Furthermore, guided play provides the opportunity to receive immediate meaningful feedback. Alfieri and colleagues (Alfieri, Brooks, Aldrich & Tenenbaum, 2011) conducted two meta-analyses on the effects of discovery-based learning and concluded that for "optimal approaches", one of the following must be present: learning that incorporates scaffolding; learning

that requires children to describe ideas and receive feedback; learning that includes successful examples (Alfieri et al., 2011).

Guided play often involves toys and props that are related to the content of lesson to promote conversation (Weisberg, Hirsh-Pasek, Golinkoff & McCandliss., 2014). The use of toys in learning can further develop semantic representations and word association, for example using a cup to learn the word goblet.

Psychological Background

Guided play is grounded in Vygotskian theory (1978) that suggests students are more likely to be effective learners with scaffolding and guidance from an adult with more subject knowledge. When a child is in the Zone of Proximal Development, providing this assistance and experience allows children to achieve their learning outcomes. The role of the teacher is to mediate student learning and collaborate with learners to facilitate meaning and understanding. Vygotsky also believed children's understanding of the world develops through playful experiences and adults should inspire this development (Han, Moore, Vukelich & Buell, 2010). In addition, Vygotsky also emphasised the importance of peer interaction during play for children to acquire new skills. According to this theory, teachers should provide children with opportunities to play to develop imagination and generate understanding. Adult scaffolding through play not only improves skills such as language, literacy, emotional development, social development, self-regulation, memory and cognitive skills (Bodrova & Leong, 2001), but also introduces skills required for future such as problem-solving skills.

Rationale for review

Evidence shows that what happens during the early years of a child's life is important for a range of cognitions and life skills. For example, early language acquisition is found to predict a child's later reading skills and academic attainment

(Singer et al., 2006). It is therefore essential that research investigates strategies to make sure children are receiving the most effective teaching methods in their learning environment.

Furthermore, the statutory framework for the Early Years Foundation Stage (DfE, 2014) reports educational programmes must contain elements of “planned, purposeful play” and a “mix of adult-led and child-initiated activity”. There is also an ongoing debate about balance between child-led and adult-guided activities and so this review aims to be the first to review studies on guided play and language development. With regards to Educational Psychology practice, there is an emphasis on early years in the Code of Practice (Department for Education and Department of Health, 2015), stating it is a statutory requirement to support children and make sure they achieve the best possible educational outcomes. Therefore, it is essential for Educational Psychologists to be aware of effective evidence-based strategies to support children from 0-25, including children in the Early Years.

More recently, literature has begun exploring the effects of guided play on learning in the classroom. Classroom based interventions that incorporate reading as the key element are shown to have positive effects on language (Dickinson et al., 2018). Shared book reading is an extensively researched evidence based strategy to enhance language development. Therefore, research on guided play often involves shared book reading prior to the play sessions. Key words and terminology from the books are reintroduced in the play sessions to support the child’s language. Thus, this review aims to explore guided play following shared book reading systematically.

Review question

How effective is guided play following shared book reading in improving language development in the Early Years?

Critical Review

Systematic Literature search

A systematic literature search was conducted in January 2020 using three electronic databases: Web of Science, ERIC (EBSCO) and PsycINFO. The combination of databases were investigated as they provide a strong overview of research from education, medical and psychological backgrounds. The search terms used to identify the studies in this review are presented in Table 1.

Table 1

Databases and search terms

Database Searches	Search Terms
Web of Science	“Guided play” OR instructional play
ERIC	AND
PsycINFO	learning OR reading OR "reading fluency" OR language OR "language acquisition" OR "social skills" OR "critical thinking" OR cognition OR "executive functioning" OR vocabulary OR literacy OR awareness OR attention OR development
	AND
	“early years" OR early childhood education OR preschool OR kindergarten OR nursery OR infants OR infancy OR children OR toddlers

Screening of articles

The initial search yielded 211 studies in total. From these studies 40 were duplicates and were therefore removed. The remaining studies were screened by title and abstract using the inclusion and exclusion criteria illustrated in Table 2. An ancestral search was performed and a further 9 studies were identified. Following a full text screening, 18 studies were excluded and the remaining 7 studies qualified for the

review (Table 3). Figure 1 shows the study selection process, and the list of excluded studies and the rationale for exclusion can be found in Appendix A.

Table 2

Inclusion and exclusion criteria

	Criteria Reference	Inclusion Criteria	Exclusion Criteria	Rationale
Type of publication	1	Articles from peer reviewed journals	Articles published in non-peer reviewed journals for example dissertations or theses.	Studies that have been peer reviewed have received higher level of inspection and therefore seen as higher quality research.
Design	2	Study must have collected primary quantitative data	Systematic literature reviews, meta-analyses and qualitative data.	Only including original data.
Location	3	Organisation for Economic Co-operation and Development (OECD) countries	Countries not considered OECD status.	To be able to generalise findings to countries similar to the UK.
Participants	4	Children in Early Years Foundation Stage (0-5)	Children 6 years of age or older.	Guided play is more commonly used in Early Years settings.
Intervention	5	Study must include guided play and shared book reading	Study does not include any form of guided play or shared book reading.	The topic in review is to assess the effectiveness of the intervention following shared book reading.
Outcomes	6	The study must have pre and post intervention data for children’s language	Study does not included a measure for language.	The review question is evaluating guided play and its effects on language development.

Figure 1

Flow diagram showing study selection procedure

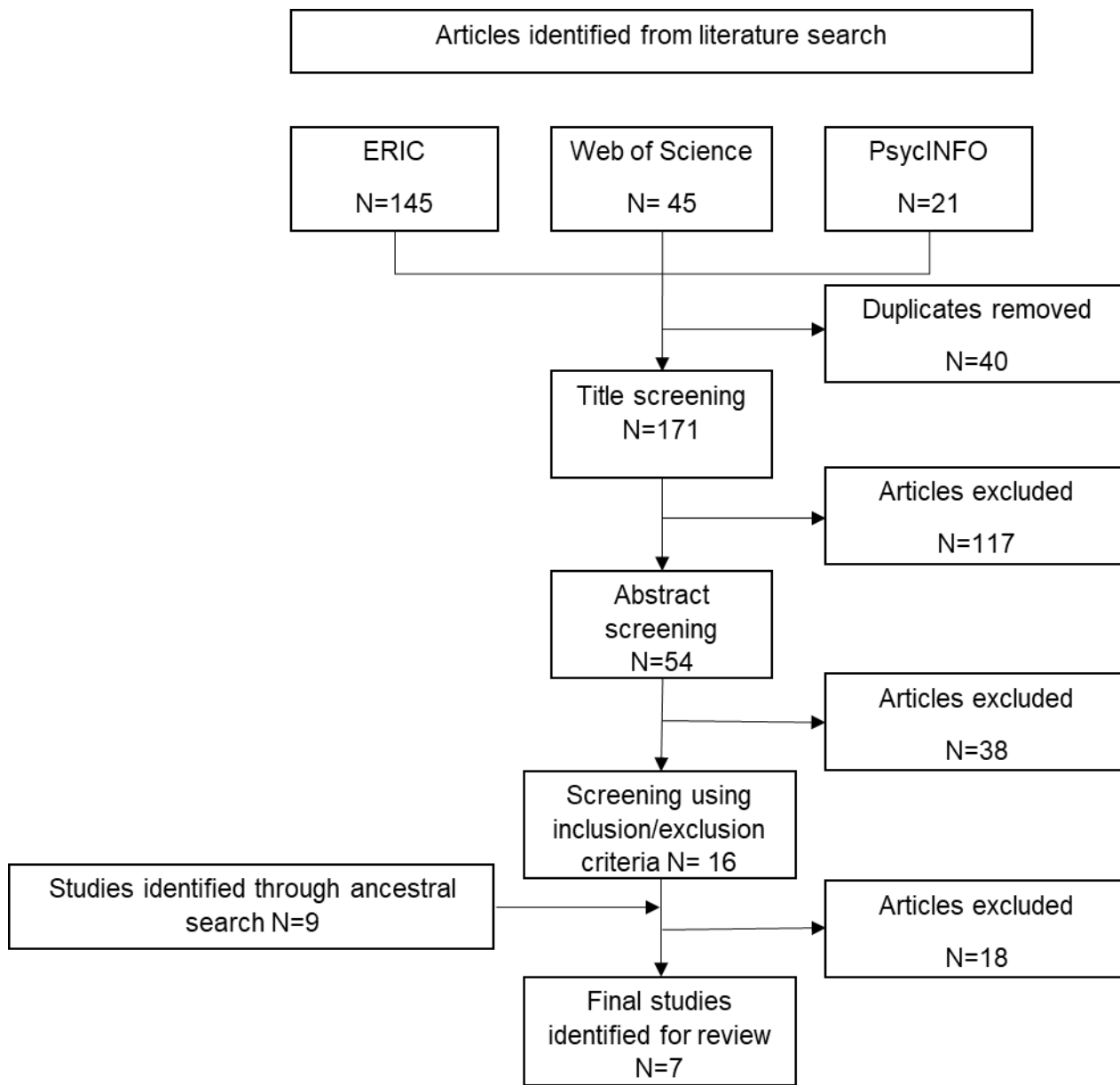


Table 3

List of studies included in this review

Eligible Studies	
1	Cavanaugh, D. M., Clemence, K. J., Teale, M. M., Rule, A. C., & Montgomery, S. E. (2017). Kindergarten scores, storytelling, executive function, and motivation improved through literacy-rich guided play. <i>Early Childhood Education Journal</i> , 45(6), 831-843.
2	Dickinson, D. K., Collins, M. F., Nesbitt, K., Toub, T. S., Hassinger-Das, B., Hadley, E. B., ... & Golinkoff, R. M. (2018). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. <i>Journal of Cognition and Development</i> , 20(2), 136-164.
3	Hadley, E. B., & Dickinson, D. K. (2019). Cues for word-learning during shared book-reading and guided play in preschool. <i>Journal of child language</i> , 46(6), 1202-1227.
4	Han, M., Moore, N., Vukelich, C., & Buell, M. (2010). Does play make a difference? How play intervention affects the vocabulary learning of at-risk preschoolers. <i>American Journal of Play</i> , 3(1), 82-105.
5	Hassinger-Das, B., Ridge, K., Parker, A., Golinkoff, R. M., Hirsh-Pasek, K., & Dickinson, D. K. (2016). Building vocabulary knowledge in preschoolers through shared book reading and gameplay. <i>Mind, Brain, and Education</i> , 10(2), 71-80.
6 ^a	Spiewak Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., ... & Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. <i>Early Childhood Research Quarterly</i> , 45, 1-17.
7	Spiewak Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., ... & Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. <i>Early Childhood Research Quarterly</i> , 45, 1-17.

^a Toub et al., (2018) conducted two studies in the same journal article: Study 1 guided play implemented by researchers; Study 2 guided play implemented by teaching staff.

Weight of Evidence (WoE)

The seven studies included in this review were assessed for relevance and quality using Gough's (2007) Weight of Evidence (WoE) framework. The framework features three key dimensions to appraise research and allows for an overall weighting of each study to be calculated. It consists of quality of methodology (WoE A), methodological relevance (WoE B) and the relevance of the topic to the review question (WoE C).

For WoE A an adapted version of the Gersten's (2005) quality indicators for group experimental and quasi-experimental research in special education was used. Adaptions made to the coding protocol can be seen in Appendix C. WoE B of each study was calculated based on Petticrew and Roberts' (2003) Hierarchy of Evidence. WoE C criteria was based on the relevance to these studies in answering the review question. Criteria and the rationale for the criteria is detailed in Appendix C. These weightings averaged together gave an overall study weighting (WoE D). Table 4 provides an outline of all WoE scores for each study and a description of their weighting. Table 5 shows the classification and cut-off points for scores.

Table 4

Summary of WoE Scores

	Studies	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall Weighting	WoD
1	Cavanaugh et al. (2017)	2	2	2.5	2.2	Medium
2	Dickinson et al. (2018)	3	3	2.7	2.9	High
3	Hadley & Dickinson (2019)	2	1	2.3	1.8	Medium
4	Han et al. (2010)	3	3	2.3	2.8	High
5	Hassinger-Das et al. (2016)	3	3	2.2	2.7	High
6	Toub et al. (2018a)	3	3	2.5	2.8	High
7	Toub et al. (2018b)	3	2	2.5	2.5	High

Table 5

Classification and cut-off points for scores

WoE D	
Overall weight of evidence	Scores
High	≥ 2.5
Medium	1.5 – 2.4
Low	≤ 1.4

Participant Characteristics

A total of 741 participants were obtained from the seven studies that were reviewed. All children were aged between 2 and 5 years of age and were either in preschool or reception year at school (kindergarten) to meet the inclusion criteria. The percentage of female and male children was documented in all studies, and an average percentage found approximately 52% of all children were female. Thus,

similar percentages of female and male participants were described across the seven studies and this was reflected in WoE C scores.

All reviewed studies were conducted in the USA. Three studies were carried out in Tennessee and Pennsylvania (Dickinson et al., 2018; Toub et al., 2018a; Toub et al., 2018b), two in Mid-Atlantic states (Han, Moore, Vukelich, & Buell, 2010; Hassinger-Das et al., 2016), one from “Central US states” (Cavanaugh, Clemence, Teale, Rule & Montgomery, 2017) and one from a South-Eastern US city (Hadley & Dickinson, 2019). Demographics revealed that all but one study (Hassinger-Das et al., 2016) recruited children from a state-funded program for low-income families. Furthermore, Hassinger-Das and colleagues (2016) recruited children from 9 preschool classrooms affiliated with a university in the Mid-Atlantic region. Geographical location of the studies were reflected in WoE C scores, with studies scoring higher if they were conducted in the United Kingdom or a country with a comparable education system.

With regards to languages spoken at home, six studies reported the percentage of English Language Learners amongst their sample ranged from 14% (Toub et al., 2018a) to 60% (Han et al., 2010). In comparison, Hassinger-Das and colleagues (2016) reported English was the primary language spoken at home. Likewise, three studies reported that the participants did not have referrals or Individualised Education Plans relating to their language, cognition speech or hearing (Dickinson et al., 2018; Hadley & Dickinson, 2019; Toub et al., 2018a). Information on participant demographics is reflected in the WoE C scores, with studies providing full demographic information (age, gender, socioeconomic status, ethnicity, languages spoken at home, information on individualised education plan) receiving a higher WoE C score.

Research design

In support of Petticrew and Roberts' (2003) hierarchy of evidence, Randomised Control Trials (RCTs) were deemed to be the 'highest quality' for studies examining the effectiveness of interventions.. Four studies in this review were RCTs (Dickinson et al., 2018; Han et al., 2010; Hassinger-Das et al., 2016; Toub et al., 2018a) and this is reflected in WoE B scores. In Toub and colleagues' study (2018a), three groups were outlined; guided play, free play and directed play in order to identify which play approach is the most successful for supporting vocabulary development. Cavanaugh and colleagues (2017) used a repeated measures, counterbalanced design with all participants experiencing both conditions at different times of the study. Thus, receiving a 'Medium' WoE B score. Similarly, one study used a within-subjects design to compare guided play and non-play activities (Toub et al., 2018b), and therefore received a 'Medium' WoE B score. This study ensured that group equivalence was recognised with regards to the selection and distribution of taught words between conditions. Hadley and Dickinson (2019) used a mixed-methods design whereby children were randomly assigned to groups of 3 children. All groups received the guided play intervention, and so this study received a 'Medium' WoE B score.

All studies were group designs ranging from dyads (Han et al., 2010; Hassinger-Das et al., 2016) to small groups of 3 to 5 children (Cavanaugh et al., 2017; Dickinson et al., 2018; Hadley & Dickinson, 2019 & Toub et al., 2018). All studies included a form of shared book reading, with the experimental group participating in some form of guided play. This was reflected in their WOE A scores. In addition, one study in this review included a follow-up following the immediate post-test (Toub et al., 2018a). One study included multiple assessments over 5 time points to assess monthly growth of expressive and receptive vocabulary (Han et al., 2010).

All seven studies were conducted in a classroom setting familiar to the child. It is easier to generalise the effects of the intervention to a school setting if the conditions of the research is in a naturalistic environment the participant is familiar with. Thus, all studies scored highest in this criteria for WoE C.

Intervention Content and Fidelity

Guided play was explicitly stated in all studies but specific details of the intervention differed between studies. Six studies started with shared book reading with selected target word incorporated into the books. In one study, Cavanaugh and colleagues (2016) used a standard literacy lesson incorporating shared book reading. This was followed with guided play whereby children in the experimental groups were provided with toys and props to play with that were relevant to the stories. Adults asked open-ended questions and provided some scaffolding when required. Target vocabulary words were integrated into the play sessions with child friendly definitions.

In one study, a vocabulary board game was introduced to the intervention group following shared book reading instead of toys (Hassinger-Das et al., 2016). Certain spaces on the board contained questions ranging from low demand (recalling elements from the story) to high demand (making predictions and inferences about the story). In comparison, Cavanaugh and colleagues (2016) incorporated literacy activities to practice letter sounds and consonant-vowel-consonant rhyming word families using material such as toys and letter tiles. In the experimental condition, children developed their own games with the materials to practice these literacy skills compared to the teacher-led control condition.

Interventions were delivered by teaching staff (Cavanaugh et al., 2016; Dickinson et al., 2018; Toub et al., 2018b), tutors (Han et al., 2010), intervention specialists (Toub et al., 2018a) or researchers (Hadley & Dickinson, 2019; Hassinger-Das et al., 2016). This was reflected in WoE C scores with teacher led interventions scoring

higher as this review aims to investigate the effectiveness of a school-based intervention.

Fidelity of implementation was considered a significant factor in Gersten's (2005) WoE protocol. When implementing an intervention in the classroom, it is important to apply the intervention as planned to increase the likelihood of the intervention being successful. If teaching staff are provided with high quality supervision and training, they are more likely to implement the intervention as it was intended. Therefore, studies that included information on the fidelity of implementation received a higher WoE score. Studies where teaching staff implemented interventions received assistance and guidance from intervention specialists (Dickinson et al., 2018; Toub et al., 2018b). Tutors and intervention specialists received training ranging from one hour (Han et al., 2010) to four hours (Toub et al., 2018a) and supervision from researchers. One study used a fidelity check list (Hassinger-Das et al., 2016) and five studies used video recordings (Dickinson et al., 2018; Hadley & Dickinson, 2019; Han et al., 2010; Hassinger-Das et al., 2016; Toub et al., 2018). One study did not refer to fidelity and this was reflected in their WoE A score (Cavanaugh et al., 2017).

Outcome measures

All studies used appropriate methods to measure children's language growth and this was reflected in their WoE A scores. Six studies used measures to explore children's receptive and expressive vocabulary (Dickinson et al., 2018; Hadley & Dickinson, 2019; Han et al., 2010; Hassinger-Das et al., 2016; Toub et al., 2018a; Toub et al., 2018b). Five studies used the New Word Definition Test and the Peabody Picture Vocabulary Test (PPVT) to assess expressive language and children's receptive and general vocabulary respectively. These established tests are considered reliable and effective tools to measure children's language skills (Beres et al., 2000; Blewitt et al., 2009). In comparison, Han and colleagues

measured expressive vocabulary with the Individual Growth and Development Indicator. Two studies used experimenter designed measures to assess vocabulary breadth (Hadley & Dickinson, 2019) and receptive vocabulary (Hassinger-Das et al., 2016). One study measured language using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) to evaluate participant's sound fluency, reading fluency and phonemic awareness (Cavanaugh et al., 2016).

All studies used multiple measures to collect data on language development including standardised assessments, observations (Cavanaugh et al., 2016), teacher and parent questionnaires (Dickinson et al., 2018; Toub et al., 2018a; Toub et al., 2018b). Cavanaugh (2016) assessed vocabulary understanding using journals and calculated the amount of times target sounds and letters were applied. Using multiple sources encourages more comprehensive and accurate measurement and provides a more holistic view of students' learning. This is reflected in WoE A scores.

Findings and Effect Sizes

Table 6 provides a summary of the outcome measures and effect sizes of the seven studies in this review. Effect sizes were reported for all studies excluding Han et al. (2010). The reviewer used an effect size calculator to determine the effect size using the reported t-test results. Effect sizes fluctuated from small to large, and descriptors were based on Cohen's (1992) thresholds (described in Table 6).

Overall, the findings reveal positive effects and support the use of guided play to help vocabulary development in the Early Years. All six studies exploring expressive vocabulary found large (Dickinson et al., 2016), medium (Hadley & Dickinson, 2019; Han et al., 2010; Hassinger-Das et al., 2016) and small (Toub et al., 2018a; Toub et al., 2018b) effect sizes for vocabulary gains. In addition, Dickinson et al. (2016), Hassinger-Das et al. (2016) and Toub et al. (2018a) found an increase in receptive

vocabulary with large and small effect sizes. Cavanaugh et al. (2016) found a significant increase in literacy skills for the experimental condition.

Although Dickinson et al. (2019) found large effects on receptive and expressive language, there was no significant difference between conditions for general vocabulary growth. Researchers highlighted that larger effects were found for the intervention group (receptive $d = .30$; expressive $d = .20$) compared to the control group, although non-significant. Similarly, Toub et al. (2018b) found no significant group differences for receptive knowledge gains. In addition, one study investigated language across three different types of play; guided, directed and free play (Toub et al., 2018a). Results revealed that children in the conditions that involved a form of adult support (direct and guided play) were more successful for developing vocabulary than unsupported free play. No significant differences were found between guided play and directed play conditions for vocabulary growth, and this was also observed in follow-up scores.

Han et al. (2010) found no significant differences between conditions for receptive vocabulary, although both conditions improved. With regards to the PPVT, children who receive a standard score between 85 and 115 are said to be within their appropriate age range. Pre-test scores revealed children in both the control and intervention condition did not meet this benchmark. Interestingly, post-test results showed 62.5% of children in the guided play condition met this benchmark compared to the control group (44%).

Overall, the effect sizes for the reviewed studies suggest that guided play following shared book reading can help with areas of language development for children in the Early Years. Particularly, studies evaluating expressive vocabulary and receptive vocabulary (Dickinson et al., 2016; Hadley & Dickinson, 2019; Han et al., 2010; Hassinger-Das et al., 2016) found gains that are more significant in expressive

vocabulary than receptive vocabulary, and this is discussed further in the next section of this review.

Table 6 Summary of Effect Sizes and Outcome Measures

Study	N	Outcome Measures	Effect Size	Follow-up	Effect size descriptor ^a	WoE D
Cavanaugh et al. (2016)	41	DIBELS <i>Sound fluency, phonemic awareness components, and oral reading fluency</i>	Cohen's d .57	N/A	Medium	Medium
Dickinson et al. (2018)	227	Picture Vocabulary Test- modified <i>Receptive vocabulary</i>	Cohen's d 1.32	N/A	Large	High
		New Word Definition Test-modified <i>Expressive language</i>	Cohen's d 1.12		Large	
		The PPVT-Fourth Edition <i>Growth in general vocabulary between groups</i>	Cohen's d .04	-	-	
Hadley & Dickinson (2019)	30	Experimenter designed vocabulary breadth measure <i>Vocabulary breadth</i>	Cohen's d .859	N/A	Large	Medium
		New Word Definition Test-Modified <i>Expressive vocabulary and depth of knowledge</i>	Cohen's d .622		-	Medium
Han et al. (2010)	49	Individual Growth and Development Indicators: Picture Naming <i>Expressive vocabulary</i>	Cohen's d .7	No data ^b	Medium	High
		Peabody Picture Vocabulary Test (PPVT III) <i>Growth in receptive vocabulary</i>	Cohen's d .02		-	-

Study	N	Outcome Measures	Effect Size	Follow-up		Effect size descriptor ^a	WoE D	
Hassinger-Das et al. (2016)	44	Author Generated Assessment <i>Receptive vocabulary.</i>	Cohen's d 1.23	N/A		Large	High	
		New Word Definition Test-Modified <i>Expressive vocabulary knowledge</i>	Cohen's d 0.54			Medium		
Toub et al. (2018a)	249	The PPVT-Fourth Edition <i>Receptive language</i>	Guided Play vs Free Play: Cohen's d -0.39	Guided Play vs Free Play: Cohen's d -0.28		Small	High	
			Guided Play vs Directed Play: Cohen's d -0.06	Guided Play vs Directed Play: Cohen's d .004	-	-		
			New Word Definition Test-Modified <i>Expressive language</i>	Guided Play vs Free Play: Cohen's d -0.26	Guided Play vs Free Play: Cohen's d -0.28	-	-	Small
			Guided Play vs Directed Play: Cohen's d .007	Guided Play vs Directed Play: Cohen's d .005				
Toub et al. (2018b)	101	The PPVT-Fourth Edition <i>Receptive language</i>	Cohen's d .01	N/A	-	-	High	
		New Word Definition Test-Modified <i>Expressive language</i>	Cohen's d .41			Small		

^a Cohen's d effect size descriptors based on thresholds: 0.2= low effect, 0.5= medium effect, 0.8= large effect (Cohen, 1992)

^b Not enough data presented for follow-up to calculate effect sizes. - Not statistically significant

Conclusions and Recommendations

The purpose of this review was to evaluate the effectiveness of guided play for improving language and vocabulary development in the Early Years. The seven reviewed studies had WoE scores ranging from 'medium' to 'high'. Overall, the findings highlight positive effects for guided play following shared book reading and literacy lessons.

All seven studies reported guided play improved children's expressive vocabulary therefore indicating guided play can help children to express their ideas and thoughts. Three studies found an increase in receptive vocabulary (Dickinson et al., 2018; Hassinger-Das et al., 2016; Toub et al., 2018a) linked to language understanding. Interestingly, more studies showed significant gains for expressive vocabulary than receptive vocabulary. Toub et al (2018b) proposes this could be a result of a two-stage word learning process. Children are often able to understand what has been said to them (receptive vocabulary) before understanding and communicating (expressive vocabulary). The initial sounds are therefore formed quickly and then the deeper level word processing begins, to make sense of the information. Therefore, this deeper processing for expressive vocabulary uses more attention and establishes stronger connections to long-term memory.

In addition, Toub et al. (2018a) found significant differences between guided play and free play, although no differences were found between guided play and directed instruction. Therefore suggesting play incorporating a form of scaffolding from an adult was more effective than play without guidance. Likewise, vocabulary gains were larger for Toub et al. (2018b) when teachers implemented guided play, in comparison to the researcher (Toub et al., 2018a). This therefore provides support for the implementation of guided play by teaching staff in a school setting.

Although this review suggests promising outcomes, it is not without its limitations. The aim of this review was to focus on studies conducted in the United Kingdom

(UK) so the evidence can be generalised to the UK education system. After exploring the literature it became apparent that research on guided play in the UK was non-existent and all research was conducted in the USA. However, the USA is an Organisation for Economic Co-operation and Development (OECD) country with a similar Early Years education system and so this was not perceived as a disadvantage. Furthermore, as the research only looked at children under the age of five, the effects of guided play cannot be generalised to other age groups beyond preschool settings. However, the Early Years Foundation Stage statutory framework (EYFS; DfE, 2014) suggests that as children grow older and approach Year 1, there is to be a shift from more child initiated to adult initiated activities. It may be beneficial to investigate play in older children to see if it is advantageous to continue implementing guided play in other year groups.

In addition, although the effects of guided play were found, the reviewer cannot determine the appropriate length of guided play time to see the largest effects. Therefore, further research should investigate the amount of time required for the most effective form of guided play. Researchers have also reported that even when teachers implement any form of play with scaffolding, there can be time restrictions and difficulties finding a balance between the play and teaching elements in a lesson (Meacham et al., 2014). Likewise, finding play activities that coincide with lesson plans and teacher styles can be difficult (Dickinson et al., 2018). It may be that teachers require effective time management and play strategies in order to effectively balance teaching and play in the classroom.

Overall, using shared book reading with the additional guided play sessions is an encouraging approach. The EYFS (DfE, 2014) emphasises the importance of play at building confidence and problem-solving skills, and states learning in the classroom must incorporate a form of adult and child initiated play. However, with the rise in education standards and skill development, there has been a reduction of play in

education (Roskos & Christie, 2007). This combined with fewer strategies aimed at increasing language acquisition during infancy, highlights the importance for research to continue exploring language interventions and play-based interventions in the Early Years.

References

- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning?. *Journal of Educational Psychology*, 103(1), 1-18
- Becker, D. R., McClelland, M. M., Loprinzi, P., & Trost, S. G. (2014). Physical activity, self-regulation, and early academic achievement in preschool children. *Early Education & Development*, 25(1), 56-70.
- Beres, K. A., Kaufman, A. S., & Perlman, M. D. (2000). Assessment of Child Intelligence. In *Handbook of Psychological Assessment*, 1(3), 65-96.
- Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared Book Reading: When and How Questions Affect Young Children's Word Learning. *Journal of Educational Psychology*, 101(2), 294–304.
- Bodrova, E., & Leong, D. J. (2001). Tools of the Mind: A Case Study of Implementing the Vygotskian Approach in American Early Childhood and Primary Classrooms. 7, 1095-1111.
- Cavanaugh, D. M., Clemence, K. J., Teale, M. M., Rule, A. C., & Montgomery, S. E. (2017). *Kindergarten Scores, Storytelling, Executive Function, and Motivation Improved through Literacy-Rich Guided Play*. 45, 831–843.
- Cohen, L. E., & Emmons, J. (2017). Block play: spatial language with preschool and school-aged children. *Early Child Development and Care*, 187(5-6), 967-977.
- Department for Education, D. (2014). Statutory framework for the early years foundation stage. Accessed January 2020.
- Department for Education and Department of Health (2015) Special educational needs and disability code of practice: 0 to 25 years. Accessed January 2020.
- Dickinson, D. K., Collins, M. F., Nesbitt, K., Toub, T. S., Hassinger-Das, B., Hadley,

- E. B., ... Golinkoff, R. M. (2018). Effects of Teacher-Delivered Book Reading and Play on Vocabulary Learning and Self-Regulation among Low-Income Preschool Children. *Journal of Cognition and Development, 20*(2), 136–164.
- Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test Fourth Edition. Minneapolis, MN: NCS Person. Inc. Measure used with Cohort, 3.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming, 33*(4), 441-467
- Gersten, R. M., Fuchs, L., Coyne, M. D., & Greenwood, C. R. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Council for Exceptional Children, 71*(2), 149–164.
- Gough, D. (2007). *Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence* . 22(2), 213–228.
- Hadley, E. B., & Dickinson, D. K. (2019). Cues for word-learning during shared book-reading and guided play in preschool. *Journal of Child Language, 46*(6), 1202–1227.
- Han, M., Moore, N., Vukelich, C., & Buell, M. (2010). Does Play Make a Difference? How Play Intervention Affects the Vocabulary Learning of At-Risk Preschoolers. *American Journal of Play, 3*(1), 82–105.
- Hassing-Das, B., Ridge, K., Parker, A., Golinkoff, R. M., Hirsh-Pasek, K., & Dickinson, D. K. (2016). Building Vocabulary Knowledge in Preschoolers Through Shared Book Reading and Gameplay. *Mind, Brain, and Education, 10*(2), 71–80.
- Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. (2009). *A mandate for playful learning in preschool: Applying the scientific evidence*. USA: Oxford University Press.

- Honomichl, R. D., & Chen, Z. (2012). The role of guidance in children's discovery learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(6), 615-622.
- Massey, S. L. (2013). From the reading rug to the play center: Enhancing vocabulary and comprehensive language skills by connecting storybook reading and guided play. *Early Childhood Education Journal*, 41(2), 125-131.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning?. *American Psychologist*, 59(1), 14.
- Meacham, S., Vukelich, C., Han, M., & Buell, M. (2014). Preschool teachers' questioning in sociodramatic play. *Early Childhood Research Quarterly*, 29(4), 562-573.
- NELS (National Early Literacy Panel) (2008). Developing early literacy: report of the National Early Literacy Panel. Washington, DC: National Institute for Literacy.
- Petticrew, M. & R. H. (2003). Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health*, 57, 527–529.
- Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of educational psychology*, 104(3), 661.
- Roskos, K., & Christie, J. (2001). Examining the play–literacy interface: A critical review and future directions. *Journal of Early Childhood Literacy*, 1(1), 59-89.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and Memory*, 4(6), 592
- Savina, E. (2014). Does play promote self-regulation in children?. *Early Child Development and Care*, 184(11), 1692-1705.
- Singer, D., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Play= Learning: How play

motivates and enhances children's cognitive and social-emotional growth.

Oxford University Press.

Spiewak Toub, T., Hassinger-Das, B., Turner Nesbitt, K., Ilgaz, H., Skolnick Weisberg, D., Hirsh-Pasek, K., ... Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading.. *Early Childhood Research Quarterly*, 45, 1-17.

Story, A. D. (2017). The Case of Brain Science and Guided Play. *Young Children*, 72(2), 45-50

Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., ... Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. *Early Childhood Research Quarterly*, 45, 1–17.

Kleeck, A. V. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools*, 45(7), 627-643. Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, 25(3), 177-182.

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., & McCandliss, B. D. (2014). Mise en place: Setting the stage for thought and action. *Trends in Cognitive Sciences*, 18(6), 276-278.

Appendix A: List of excluded studies

Excluded Study during full screening	Rationale for exclusion based on inclusion and exclusion criteria
<p>Aguiar, C., & McWilliam, R. A. (2013). Consistency of toddler engagement across two settings. <i>Early Childhood Research Quarterly</i>, 28(1), 102-110.</p>	5: Intervention not guided play
<p>Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning?. <i>Journal of educational psychology</i>, 103(1), 1.</p>	2: Design
<p>Basilio, M., & Rodríguez, C. (2017). How toddlers think with their hands: Social and private gestures as evidence of cognitive self-regulation in guided play with objects. <i>Early Child Development and Care</i>, 187(12), 1971-1986.</p>	6: Outcome measures
<p>Colliver, Y., & Veraksa, N. (2019). The aim of the game: A pedagogical tool to support young children's learning through play. <i>Learning, Culture and Social Interaction</i>, 21, 296-310.</p>	5: Intervention not guided play
<p>Eason, S. H., & Ramani, G. B. (2018). Parent–Child Math Talk About Fractions During Formal Learning and Guided Play Activities. <i>Child development</i>.</p>	6: Outcome measures
<p>Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. <i>Child development</i>, 84(6), 1872-1878.</p>	6: Outcome measures
<p>Gardner-Neblett, N., Holochwost, S. J., Gallagher, K. C., Iruka, I. U., Odom, S. L., & Pungello, E. P. (2016). Guided versus Independent Play: Which Better Sustains Attention among Infants and Toddlers?. <i>Society for Research on Educational Effectiveness</i>.</p>	6: Outcome measures

Excluded Study during full screening	Rationale for exclusion based on inclusion and exclusion criteria
Jensen, H., Pyle, A., Alaca, B., & Fesseha, E. (2019). Playing with a goal in mind: exploring the enactment of guided play in Canadian and South African early years classrooms. <i>Early Years</i> , 1-15.	3: Location
Jemutai, S., & Webb, P. (2019). Effects of a 6 Brick Duplo Block guided play intervention on pre-literate learners' visual perception. <i>South African Journal of Childhood Education</i> , 9(1), 1-8.	6: Outcome measures
LaGamba, E. (2018). <i>An Investigation of Read-Alouds, Classroom Interactions, and Guided Play as Supports for Vocabulary Learning in Preschool</i> (Doctoral dissertation, University of Pittsburgh).	1: Type of publication
Palma, M. S., Pereira, B. O., & Valentini, N. C. (2014). Guided play and free play in an enriched environment: Impact on motor development. <i>Motriz: Revista de Educação Física</i> , 20(2), 177-185.	6: Outcome measure
Papacek, A. M. (2015). The Role of Peer Guided Play for Children with Autism Spectrum Disorder. <i>Journal of the American Academy of Special Education Professionals</i> , 80-97.	1. Type of publication
Ramani, G. B., Zippert, E., Schweitzer, S., & Pan, S. (2014). Preschool children's joint block building during a guided play activity. <i>Journal of Applied Developmental Psychology</i> , 35(4), 326-336.	6: Outcome measures
Sawyer, J. E., & Goldstein, T. (2019). Can Guided Play and Storybook Reading Promote Children's Drawing Development?. <i>Empirical Studies of the Arts</i> , 37(1), 32-59.	6: Outcome measures

Excluded Study during full screening	Rationale for exclusion based on inclusion and exclusion criteria
Sliogeris, M., & Almeida, S. C. (2019). Young Children's Development of Scientific Knowledge Through the Combination of Teacher-Guided Play and Child-Guided Play. <i>Research in Science Education</i> , 49(6), 1569-1593.	6: Outcome measures
Sobel, D., & Sommerville, J. (2010). The importance of discovery in children's causal learning from interventions. <i>Frontiers in Psychology</i> , 1, 176	5: Intervention not guided play
Tsao, Y. L. (2008). Using Guided Play to Enhance Children's Conversation, Creativity and Competence in Literacy. <i>Education</i> , 128(3), 515-520.	6. Outcome measures
Van Schijndel, T. J., Singer, E., van der Maas, H. L., & Raijmakers, M. E. (2010). A sciencing programme and young children's exploratory play in the sandpit. <i>European Journal of Developmental Psychology</i> , 7(5), 603-617.	6: Outcome measures

Appendix B: Mapping the field

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Cavanaugh et al. (2016)	Repeated measures, counterbalanced design with all children experiencing both conditions at different times of the study.	USA	N=41 Two preschool classes; one in a middle-class neighbourhood (N=20) and one in a low-income preschool (N=21). 54% female.	All children received the same literacy class except for 15 minutes when the intervention took place. In the control condition, the teacher led activities with items and toys to practice letter sounds and rhyming word families. In the experimental condition, children received the same teacher led activity and then developed their own games to practice their sounds and word family concepts.	Children in each classroom were divided into two groups group A and B. from both classrooms, group A began with the control condition and after 3 weeks a mid-point test was administered and the groups switched. After another 3 weeks, the final post-test was administered.	<i>Dynamic Indicators of Basic Early Literacy Skills (DIBELS;</i> Kaminski & Good 1998) –Sound fluency, phonemic awareness components, and reading fluency. Children engaged in journal writing where the teacher tallied the amount of time target sounds were used. Teacher observations.	Children in the guided play condition scored significantly higher in the DIBELS than the control condition. The experimental condition also improved vocabulary application, sequencing ideas and storytelling ideas.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Dickinson et al. (2018)	RCT with two groups. Read and Play (R + P) or Read only (RO) groups	USA Tennessee and Pennsylvania	N= 227 Low-income preschool children. 52% female	Intervention conducted over 6-month period. Classroom teachers delivered the intervention. Teachers read thematically related books to both groups, and the R+P condition were put into play groups and participated in guided play sessions with toys and props relating to the books.	Children in the RO group without guided play conditions.	<i>Picture Vocabulary Test</i> similar to the PPVT (Dunn & Dunn, 2007) Receptive knowledge of taught and control words. <i>New Word Definition Test-Modified</i> (Hadley et al., 2015) – Expressive language <i>The PPVT-Fourth Edition</i> (Dunn & Dunn, 2007)- General vocabulary Self-regulation and teacher questionnaires administered.	Both RO and R+P groups made significant gains on taught vocabulary words. No differences between conditions.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Hadley & Dickinson. (2019)	Mixed-methods with participants randomly assigned to a mixed-gender play group	USA South-eastern US state	N= 30 Low-income preschool children. 57% female.	Intervention conducted over two-month period. Groups of 3 children were taught target words during shared book reading lasting 10 minutes. The target word were further reinforced during guided play with a selection of related toys and props lasting 10 minutes. All book reading and guided play sessions were video-recorded.	No comparison group. All children first received shared book reading and then guided play sessions.	<i>The PPVT-Fourth Edition</i> (Dunn & Dunn, 2007)- General vocabulary <i>Experimenter designed vocabulary breadth measure</i> - Receptive understanding of target words. <i>New Word Definition Test-Modified</i> (adapted from Blewitt et al., 2009) – Depth of knowledge for target words. Coding scheme measured target word use during the intervention.	Hearing the taught words in shared book reading and understanding meaning during guided play, was significantly associated with vocabulary growth and depth.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Han et al. (2010)	RCT with two groups, Explicit Instructional Vocabulary Protocol (EIVP) and EIVP with guided play sessions (EIVP+ Play).	USA Mid-Atlantic State	N=49 4-5 year old children from a group of 118 lowest performing students. Low-income families 47% female.	Children received 30 minutes of EIVP or EIVP + Play twice a week for four months. 16 words were selected as target words. In both conditions, tutors read the story with target words and explicit vocabulary instruction was introduced using child friendly definitions. In the EIVP+ Play condition, children were engaged in guided play with props to give target words meaning.	Children in the EIVP group without guided play conditions.	<i>The PPVT-Fourth Edition</i> (Dunn & Dunn, 2007) - Receptive language attainment. <i>Individual Growth and Development Indicators: Picture Naming</i> (Early Childhood Research Institute on Measuring Growth and Development, 2000) – Ability to name pictures quickly. <i>Curriculum-based Measurement</i> (Fuchs and Fuchs, 2002) – Attainment of the target words.	The mean gain of expressive vocabulary was significantly higher in the EIVP + Play group. Both groups receptive vocabulary increased but not group difference found. EIVP+ Play group showed steeper gains for expressive vocabulary as time progressed.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Hassinger-Das et al. (2016)	RCT with two groups, shared book reading plus vocabulary game (intervention) or shared book reading plus game that did not teach vocabulary (comparison group).	USA Mid-Atlantic State	N=44 4-year-old children, 48% female.	Children randomly assigned to conditions and then randomly paired into dyads. Two 30-minute sessions over the course of a week. Each dyad heard one of the two stories, which included 10 target words. The intervention group played a vocabulary board game with questions and information about the target words. The comparison group received after-word review and then played a game without vocabulary involvement.	Children in the book reading plus game that did not involve guided play.	<i>Author generated assessment-Receptive vocabulary.</i> <i>New Word Definition Test-Modified</i> (Blewitt et al., 2009) – Expressive vocabulary knowledge.	Children in the intervention group outperformed their peers in receptive and expressive vocabulary. Children in the intervention group that scored the lowest for receptive knowledge experienced the most gains in word knowledge.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Toub et al.(2018a)	RCT with three groups, guided play, free play or directed play.	USA Tennessee and Pennsylvania	N=249 Two low-income preschool classes. 54% female.	The intervention lasted two months and 10 target words were selected. In all conditions intervention specialists (IS), read stories with target words and there was a collection of toys for each book. In direct play, the IS followed a script to direct children to play. For the guided play group children chose what to do with the toys and the IS followed the child's lead incorporating target words. In the free play condition children played openly and the IS was not involved.	Children in the free and directed play conditions.	<i>The PPVT-Fourth Edition</i> (Dunn & Dunn, 2007) - Receptive language. <i>New Word Definition Test-Modified</i> (Hadley et al., 2015) – Expressive language and meaning of vocabulary. <i>Parent questionnaire</i> for children's home information	Children in the guided play group made significantly larger gains in expressive and receptive vocabulary than the free play condition. No differences between children in guided or directed play.

Study	Study Design	Location	Participants	Intervention	Comparison Condition	Measures	Main Findings
Toub et al.(2018b)	Within-subjects design to compare play and non-play activities.	USA Tennessee and Pennsylvania	N=101 Two low-income preschool classes. 54% female.	The intervention lasted two weeks and 16 target words were selected. All activities were led by the classroom teacher and incorporated toys relating to the books. All children examined half of the target words through a mix of guided and directed play, and half of the target words through a picture card review activity.	Both groups experienced play and non-play activities. Guided play was compared to the words examined in the picture card review activity condition.	<i>The PPVT-Fourth Edition</i> (Dunn & Dunn, 2007) - Receptive language. <i>New Word Definition Test-Modified</i> (Hadley et al., 2015) – Expressive language and meaning of vocabulary. <i>Parent questionnaire</i> for children’s home information	Children showed expressive and receptive vocabulary gains in both the play and non-play activities. Significantly greater expressive vocabulary gains were observed in the play condition.

Appendix C: Details for Weight of Evidence Ratings

WoE A: Methodological Quality

WoE A judges the methodological quality of the studies in the review in line with the study's design (Gough, 2007). The studies in this review all used group designs and so the researcher chose to use Gersten's (2005) Quality Indicators for group experimental and quasi-experimental research in special education. Modifications were made to the first question only, as the studies in the review did not include participants with special educational needs:

"Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?"

Yes

No

N/A

Unknown/Unable to Code"

Instead, the researcher included a question on participant demographics to make sure there was a quality indicator for describing participants. Scores for essential and desirable indicators were calculated and an overall score was given. For essential quality scores, a study had to receive 9 or more indicators to receive a score of 1. For desirable quality indicators, a study had to receive 4 or more quality indicators for a score of 2, and less than 4 indicators for a desirability score of 1. A total score of 3 received a 'high quality' score, a score of 2 received an 'acceptable quality' score and scores less than 2 received a 'poor quality' score. Table 1 shows WoE scores for each study.

Table 1- WoE A scores

Study	Total Essential Quality Score	Total Desirable Quality Score	Overall WoE A Score	Descriptor*
Cavanaugh et al. (2017)	1	1	2	Acceptable
Dickinson et al. (2018)	1	2	3	High
Hadley & Dickinson (2019)	1	1	2	Acceptable
Han et al. (2010)	1	2	3	High
Hassinger-Das et al. (2016)	1	2	3	High
Toub et al. (2018a)	1	2	3	High
Toub et al. (2018b)	1	2	3	High

*3= High quality, 2= Acceptable quality, 1= Poor quality

WoE B: Methodological relevance

WoE B assesses the methodological relevance of each study and whether it is appropriate to conclude if guided play is an effective intervention for improving language in infancy. WoE B of each study was calculated based on Petticrew and Roberts' (2003) Hierarchy of Evidence. This views Randomised Control Studies (RCTs) as the most effective study design to answer questions on the effectiveness of interventions. Qualitative research, surveys, non-experimental evaluations and case-control studies are deemed to be the lowest quality for effectiveness questions. Table 2 outlines the rationale for WoE B criteria and Table 3 outlines WoE B scores for each study.

Table 2- WoE B Descriptors and Rationale

Weighting	Description	Rationale
High (3)	Randomised control studies	Petticrew & Roberts (2003) suggest that these are the study designs, excluding systematic literature reviews,
Medium (2)	Cohort, quasi-experimental and single case experimental designs	that are the most suitable to answer questions on the “effectiveness” of interventions.
Low (1)	Qualitative research, survey, non-experimental evaluations and case-control studies	

Table 3- WoE B scores

Studies	WOE B Rating	WOE B Descriptor*
Cavanaugh et al. (2017)	2	Medium
Dickinson et al. (2018)	3	High
Hadley & Dickinson (2019)	2	Medium
Han et al. (2010)	3	High
Hassinger-Das et al. (2016)	3	High
Toub et al. (2018a)	3	High
Toub et al. (2018b)	2	Medium

*3= High, 2= Medium, 1= Low

WoE C: Relevance to the question

WoE C is a review specific judgement about how relevant and appropriate the studies are in answering the review question. In the case of this review this was to evaluate whether guided play is an effective intervention to increase language in infants. In order to make this judgement, the reviewer carefully considered the setting, location, participants, outcome measures, implementation and whether a follow-up study was conducted. Overall, WoE C scores for each study listed in Table 5 is an average of the ratings for criteria A-F Table 4 outlines WoE C criteria and Table 5 shows the WoE C scores for each study.

Table 4: WoE C Criteria

Criteria	Weighting*	Rationale for Criteria
A. Setting	3. Intervention delivered in classroom setting the student is familiar with	It is easier to generalise the effects of the intervention if the conditions of the research is in a naturalistic environment the participant is familiar with.
	2. Intervention carried out in another education setting (e.g. college)	
	1. No clear description of the setting or intervention carried out outside education setting (e.g. lab)	
B. Participant demographics	3. Full demographic information provided for pupils: age, gender, socioeconomic status, ethnicity, languages spoken at home, information on individualised education plan	Findings can be generalised more widely if demographic information is provided on participants.
	Information only provided for pupils: age, gender, socio-economic status and ethnicity	
	2. Basic demographic information provided	
C. Age of participants	3. Participants are in Early Years Foundation Stage	Guided play is most commonly applied in Early Years Foundation Stage and therefore this age is explored in the research question.
	2. Participants are age 6-7	
	1. Participants are age 8-9	

Criteria	Weighting*	Rationale for Criteria
D. Intervention delivery	3. Intervention delivered by academic teaching staff	This review aims to investigate the effectiveness of a school-based intervention.
	2. Intervention delivered by academic teaching staff and researchers	
	1. Intervention delivered by researchers only	
E. Location/Educational System	3. Study conducted in England	To be able to generalise study findings and see the intervention effectiveness in England and its education system.
	2. Study conducted in a country with a comparable education system to the UK	
	1. Study conducted in country with a different education system to the UK	
F. Post-test	3. The study conducted a post-test at two different time points	To see whether vocabulary and literacy gains remain over time.
	2. The study conducted a post-test at one time point	
	1. The study did not include post-test information	

*3= High, 2= Medium, 1= Low

Table 5: WoE C Scores

Study	Criteria						Overall WOE C
	A	B	C	D	E	F	
Cavanaugh et al. (2017)	3	2	3	3	2	2	2.5
Dickinson et al. (2018)	3	3	3	3	2	2	2.7
Hadley & Dickinson (2019)	3	3	3	1	2	2	2.3
Han et al. (2010)	3	2	3	1	2	3	2.3
Hassinger-Das et al. (2016)	3	2	3	1	2	2	2.2
Toub et al. (2018a)	3	3	3	1	2	3	2.5
Toub et al. (2018b)	3	2	3	3	2	2	2.5

*3= High, 2= Medium, 1= Low

WoE D: Overall WoE

WoE D is an overall weight of evidence derived from WoE A, WoE B and WoE C scores. This overall score judges the ability for the study provides evidence to answer the question being reviewed. For a study to be considered 'high' WoE D must be higher than 2.5, for a 'medium' score, WoE D must be between 1.5 and 2.4, and for a study to be considered 'low' WoE D must be less than 1.5. Table 6 outlines WoE D scores for each study.

Table 6: WoE D scores

	Studies	WoE A – Methodological Quality	WoE B – Methodological Relevance	WoE C – Relevance of Evidence to Review Question	WoE D – Overall Weighting	WoD Description*
1	Cavanaugh et al. (2017)	2	2	2.5	2.2	Medium
2	Dickinson et al. (2018)	3	3	2.7	2.9	High
3	Hadley & Dickinson (2019)	2	2	2.3	2.1	Medium
4	Han et al. (2010)	3	3	2.3	2.8	High
5	Hassinger- Das et al. (2016)	3	3	2.2	2.7	High
6	Toub et al. (2018a)	3	3	2.5	2.8	High
7	Toub et al. (2018b)	3	2	2.5	2.5	High

*High= 2.5≥, Medium= 1.5 - 2.4, Low= >1.5

Appendix D: Gersten Protocol for each study

Coding protocol: Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71,149-164.

Study: Kindergarten Scores, Storytelling, Executive Function, and Motivation Improved through Literacy-Rich Guided Play.

Author: Cavanaugh et al. (2016)

Essential Quality Indicators - Quality indicators for describing participants

~~Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?~~

Yes

No

N/A

Unknown/Unable to Code

Was sufficient information provided on participants (Age, Ethnicity, Social Economic Status, information on any disabilities)?

Yes

No

N/A

Unknown/Unable to Code

N ___ 41 ___

Did the study use appropriate participants for the research question?

Yes

No

N/A

Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Yes

No

N/A

Unknown/Unable to Code

Essential Quality Indicators - Quality indicators for Implementation of the Intervention and Description of Comparison Conditions

Was the intervention clearly described and specified?

Yes

No

N/A

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

- Yes
- No
- N/A
- Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

- Yes
- No
- N/A
- Unknown/Unable to Code

Essential Quality Indicators – Quality Indicators for Outcome Measures

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

- Yes
- No
- N/A
- Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

- Yes
- No
- N/A

Unknown/Unable to Code

Essential Quality Indicators – Quality Indicators for Data Analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the limit of analysis in the study?

Yes

No

N/A

Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

Yes

No

N/A

Unknown/Unable to Code

Essential Quality Indicators Total Score: 11

Desirable Quality Indicators

Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

Yes

No

N/A

Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

Yes

No

N/A

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

Yes

No

N/A

Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

Yes

No

N/A

Unknown/Unable to Code

Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

- Yes
- No
- N/A
- Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

- Yes
- No
- N/A
- Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

- Yes
- No
- N/A
- Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

- Yes
- No
- N/A

Unknown/Unable to Code

Desirable Quality Indicators Total Score: 2

	Total	Score
Essential Quality Indicators Total of >9 = Score 1 Total of <9 = Score 0	11	1
Desirable Quality Indicators Total of ≥4 = Score 2 Total of <4= Score 1 Total 0 = Score 0	2	1
Total Score (3 = High Quality; 2 = Acceptable Quality; <2 = Poor Quality)		2

Study Rating: Acceptable Quality