Word learning in children with developmental language disorder: a meta-analysis testing the encoding hypothesis

Journal of Memory and Language 145 (2025) 104678

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Word learning in children with developmental language disorder: A meta-analysis testing the encoding hypothesis

Paola Calabrese *, Nicholas Hedger, Katherine Pritchard, Vesna Stojanovik, Emma Pagnamenta

University of Reading, UK

ARTICLE INFO

Keywords:
Word learning
Development language disorder
Encoding
Working memory
Meta-analysis

ABSTRACT

Many children with Developmental Language Disorder (DLD) find learning new words difficult, which negatively affects their educational and psycho-social outcomes. Word learning involves encoding, consolidation and reconsolidation of words, but the most challenging phase and factors which moderate word learning remain unclear.

We conducted a systematic review and meta-analysis to determine which phase is most challenging and which factors predict oral word learning success in children with DLD. The search including PsycINFO, PubMed, Web of Science, and LLBA identified forty-six studies published before April 2024 comparing children with DLD and typically developing (TD) age-matched peers in word learning tasks. Seventy-eight effect sizes were calculated for encoding (n DLD = 1462, n TD = 2161), eight for consolidation (n DLD = 107, n TD = 112), and 19 for reconsolidation (n DLD = 296, n TD = 278).

The random effect model identified an effect for encoding (k = 78, d = 0.82, [0.66, 0.98], p < .001) but not consolidation (k = 8, d = −0.2, [−0.68, 0.29], p = .43) or reconsolidation (k = 19, d = 0.23, [−0.14, 0.59], p = .22) of new words. The moderator analysis via random effects models identified verbal short-term memory and lexical knowledge as significant moderators of encoding, while word length was the most important task characteristic.

Despite limited data for consolidation and reconsolidation, our findings provide new insights into oral word learning difficulties in children with DLD. These insights help clinicians and teachers identify support strategies while also highlighting gaps in existing research, driving future studies forward.

Introduction

Developmental Language Disorder (previously known as Specific Language Impairment; SLI[1]) is a neurodevelopmental condition characterised by persistent difficulties in different linguistic domains (e.g., syntax, vocabulary, discourse) that do not have a biomedical aetiology (Bishop et al., 2017). One of the difficulties often associated with DLD is word learning (Kan & Windsor, 2010). Word learning is the process by which a newly encountered word is stored in memory, becoming available for future understanding and production. Given the importance of word learning for children's social and academic success (Bleses et al., 2016; Rantalainen et al., 2021; Westrupp et al., 2020), it is vital to understand what stage of the process may prevent children with DLD from acquiring new words as easily as their typically developing (TD) peers. This in turn will help to inform clinical practice and to develop evidence-based interventions specifically tailored to address this difficulty.

Word learning unfolds through different phases (Gupta, 2005). When a new word is encountered for the first time, the acoustic information is perceived creating a first memory trace (Munro et al., 2012). This stage, referred to as encoding, involves the recognition of the phonological and semantic characteristics of the new word (Craik et al., 2007), and it is associated with increased activation in the hippocampus (Davis & Gaskell, 2009). Initially encoded words can be retained in long term memory or forgotten. Throughout the process of consolidation, the memory trace becomes stable, it is transferred in long term memory and

* Corresponding author.
E-mail address: paola.calabrese@student.reading.ac.uk (P. Calabrese).

[1] In line with CATALISE-2 (Bishop et al., 2017), we use DLD in place of SLI. As DLD has broader criteria, some children now identified may not have met SLI criteria. For clarity, we report both terms where relevant, reflecting each study's original terminology.

https://doi.org/10.1016/j.jml.2025.104678
Received 28 November 2024; Received in revised form 21 July 2025; Accepted 23 July 2025
Available online 6 August 2025

integrated with other words in the lexicon (Storkel, 2015). Consolidation takes place over time independently from experience, it is supported by overnight sleep (Henderson et al., 2012, James et al., 2020) and is associated with a decreased hippocampal activity (Johnson et al., 2021) and increased activity in the middle temporal cortex (Takashima et al., 2019). Finally, once a word is retrieved, it becomes susceptible to modification and updating. This process, referred to as reconsolidation (Stickgold & Walker, 2005), provides the learner with the opportunity to add and integrate information to the first encoding of the word (Buckner et al., 2001) and to strengthen the memory trace (Laurino et al. 2022).

Word learning is related to working memory and lexical knowledge through dynamic and complex interactions. The encoding of a new word is closely related to the working memory system (Archibald & Gathercole, 2006), which is the cognitive construct that holds and manipulates verbal or visual information necessary to complete a task. Working Memory (WM) is a multi-component system composed of the Central Executive (CE), responsible for directing attention, and two slave mechanisms serving for the temporary storage of visual (visual sketchpad) or verbal (phonological loop) information (Baddeley & Logie, 1999). The components of the WM system supporting the serial and temporary storage of information can also be referred as Verbal Short-Term Memory and Visual Short-Term Memory (Baddeley, 2012). In particular, when a new word is encountered, verbal short-term memory holds the strings of phonemes (Magro et al., 2018) while verbal working memory supports the chunking and elaboration of the word sounds and the transfer of the word into long term memory through rehearsal processes (Munro et al., 2012).

Previously stored lexical knowledge is also key for word encoding (Archibald, 2018). As children develop their lexicon, they build a growing repertoire of lexical (whole words) and sublexical representations (consisting of single phonemes but also sequences of sounds). During the encoding stage, children recognise the chunks already stored in their repertoire, reducing the number of units they need to hold in memory and therefore reducing the WM load (Szewczyk et al., 2018). Lexical and sublexical representations are also thought to support the process of redintegration. If a word in encoded inaccurately before it is transferred to long term memory and embedded into the lexicon, previously stored words can be used to repair inaccurate encoding, filling the missing parts with likely sequences of based on those already stored in long term memory (Jones & Witherstone, 2011).

Word learning is a complex process and better understanding of its stages and the interplay between working memory and lexical knowledge during new word acquisition, is essential for identifying the specific mechanisms underlying word learning challenges in children with DLD. The following section further explores the evidence for challenges in different phases of word learning faced by children with DLD.

Word learning in DLD: Where is the challenge?

The most recent meta-analysis on word learning in children with DLD confirmed significant difficulties in this domain. However, the analysis was limited to fast mapping, a specific aspect of word learning characterized by quickly associating a word form with its meaning after minimal exposure (Kan & Windsor, 2010). Fast mapping represents a preliminary stage of word learning, and it is part of the encoding phase (Weismer & Evans, 2002). Importantly, Kan & Windsor did not examine subsequent stages beyond fast mapping, leaving open the question of which phases pose the greatest challenges for children with DLD.

The first empirical motivation for the encoding hypothesis is presented in the Nichols and colleagues' (2004) study in which they compared patterns of word list learning in 29 children with SLI, aged 6 to 14 years, with 28 typically developing peers of similar age. Their findings indicated that children with SLI were less effective at encoding words, which prevented them from learning the lists. Consistently with this observation, Gray (2003, 2004) found that children with SLI required significantly more trials to achieve comprehension and production of novel words compared to their typically developing counterparts.

Building on this initial evidence of encoding difficulties in children with DLD, more recent studies investigated whether encoding is the critical stage preventing word learning in children with DLD, or whether later stages of the process are also impacted by the disorder. Bishop and Hsu (2015) compared 28 children with DLD aged 7–11 years with age-matched and language-matched children, finding that the DLD group was significantly less accurate in recognising the new words on the first day of training, but over the following three days, improved to the same extent as age-matched controls. Jackson et al. (2021) tested 50 children with DLD and 54 age-matched TD controls in a 4-day word learning task and observed that the two groups only differed during the first day of the task. There were no differences between groups in their learning pattern during the following days of training and in their ability to retain the words after the end of the training. Together, these studies suggested that children with DLD experience difficulties in initially encoding new words but, their ability to consolidate and reconsolidate previously encoded information, is comparable to that of their typically developing peers.

The evidence of challenges with the encoding of words for individuals with DLD has been strengthened by studies conducted in adults (McGregor et al., 2013, 2017, 2020). Furthermore, this hypothesis is consistent with psycholinguistic processes required by word encoding and the profile of relative strengths and weaknesses associated with DLD. Children with DLD have limited lexical knowledge compared to their peers, both in terms of the number of words stored into the mental lexicon and the quality of the mental representation of each word (see Jones & Brandt, 2018 for a meta-analysis). These inefficient representations are hypothesised to impede the effective processing of new words (Archibald, 2018). Consistent with this hypothesis, Kan and Windsor's (2010) meta-analysis showed that wider differences in fast mapping between children with DLD and TD peers corresponded to significant differences in receptive vocabulary between the two groups.

Furthermore, DLD is often linked to limited working memory capacity (Archibald & Gathercole, 2007; Jackson et al., 2021), and reduced verbal short-term memory (Jackson et al., 2020, Talli & Stavrakaki, 2020), both of which were found to mediate children's difficulty in acquiring new words. Children scoring lower on Verbal Short-Term Memory (VSTM) tasks (Bishop & Hsu, 2015) and global measures of working memory (Jackson et al., 2021) seemed to find word learning the most difficult. In addition, longer words are more challenging than shorter words for children with DLD (Jackson et al., 2019).

There have been alternative views using computational simulations which suggest that WM in DLD is not under-resourced but may be overloaded due to low level auditory perceptual deficits resulting in word learning difficulties (Jones & Westermann, 2022; Jones et al., 2024). While it is acknowledged that low level auditory processing may contribute to word encoding difficulties in children with DLD, a detailed examination of auditory processing falls outside the scope of this study. Different theoretical perspectives have shaped research in working memory over the past five decades and the reader is referred to a comprehensive review by Cowan (2022).

Despite the theoretical rationale, word encoding difficulties in children with DLD have not been documented in every study exploring word learning in this population. Gray and Brinkley (2011) tested encoding and consolidation over time of nonwords in pre-school children and found no difference between children with SLI and language-matched controls in the number of targets learnt. It should be noted however that, consistent with the diagnostic criteria of SLI, the participants in the study had cognitive non-verbal skills and vocabulary scores within the average expected for their age. This means that the study might not have identified difficulties in word encoding because of the cognitive strengths of the participants. Similarly, Adlof et al. (2021) found that after 30 min of training, the ability to name, comprehend and recognise novel words was similar in school-aged children with DLD and age-matched TD controls, and the DLD group performed at a lower level on the description of targets only. The authors concluded that children with DLD experience difficulties in the elaboration of meanings but not in the encoding of word forms. It is worth noting that task characteristics might have had an impact on these results. For example, the repeated exposure to the target might have facilitated the encoding as the quantity of input received may drive differences between children with DLD and controls (Riches et al., 2005). In addition, two-syllable targets might have not been long enough to identify differences between children with DLD and controls (Jackson et al., 2019).

Difficulties with the consolidation of new words in children with DLD have also been reported, arguing against the existence of a selective difficulty with the encoding of new words. Malins et al. (2021) observed that children with DLD and dyslexia successfully consolidated significantly fewer words 24 h after training, compared to children with dyslexia without comorbid DLD. Importantly, the difference between the groups in consolidation remained significant even after controlling for participants' performance on day one, confirming that the difference specifically reflected differences in consolidation rates rather than poor initial encoding. It has been suggested that children with DLD exhibit poor lexical organization (Esbensen & Thomsen, 2021) which might explain the difficulty in integrating new lexical entries during consolidation. Nevertheless, following a subgroup analysis, the authors argued that the difficulties in offline consolidation might have resulted from individual differences in their sample, such as age, rather than the presence of DLD. In addition, the targets presented in the study were highly phonologically similar. Therefore, an alternative interpretation would be that ineffective encoding did not allow participants to create distinct and accurate representations of similar words.

Following the study by Malins and colleagues, Gordon et al. (2021) conducted research to examine how nine children with DLD encoded, consolidated, and reconsolidated words over time compared to nine typically developing controls, when provided with highly supportive training. The results indicated that the initial encoding of novel words was less effective in children with DLD, although both groups demonstrated similar abilities to consolidate the words after a month. Importantly, the authors acknowledged the limited sample size, suggesting that this may have hindered the detection of group differences in consolidation. Nevertheless, they argued that based on their results the disparity in encoding between groups was likely more pronounced than any potential differences in consolidation. Additionally, the study did not yield conclusive results regarding whether reconsolidation posed a specific challenge for children with DLD. However, the authors suggested that this stage may be less critical to overall word learning success, aligning with the perspective of Bishop and Hsu (2015).

The assumption that consolidation remains unaffected in children with DLD aligns with the procedural/declarative memory hypothesis. This hypothesis suggests that children with DLD experience selective impairments in procedural memory (which underlies implicit learning, storage and retrieval) while their declarative memory system (which underlies conscious recall of facts, events, experiences) remains relatively intact (Lum & Conti-Ramsden, 2013). Both consolidation and reconsolidation are linked to declarative memory. During consolidation, a newly learned word is transferred into semantic memory, while reconsolidation involves the conscious retrieval and updating of that word from episodic memory. These processes are associated with activation in the middle temporal cortex, a brain region linked to declarative memory (Squire & Zola, 1996). Given that the declarative memory system is presumed to be preserved in children with DLD, it follows that consolidation, and reconsolidation may not pose significant difficulties for them. However, it is important to recognize that declarative and procedural memory systems interact and often share neural substrates (Brown & Robertson, 2007). Therefore, disentangling their roles is complex, and the implications of consolidation and reconsolidation in word learning difficulties among children with DLD cannot be fully understood based solely on this hypothesis.

In summary, while the evidence suggests that encoding may be particularly challenging for children with DLD, synthesizing findings related to all stages of word learning may be key to identifying where the core difficulties lie.

The current study

Word learning difficulties in DLD have been widely documented in children and adults. Nevertheless, it is still unclear what phase of this complex process is most challenging for this clinical population. The most recent meta-analysis to our knowledge confirmed the association between DLD and fast mapping difficulties (Kan & Windsor, 2010). However, in the past decade, a growing body of research contribute with insights into the nature of word learning difficulties in DLD by exploring word learning beyond fast mapping and including consolidation and reconsolidation in the investigation (Bishop and Hsu 2015, Gordon et al., 2021, Jackson et al., 2021). Furthermore, whilst previous work has explored the contribution of children's lexical knowledge, the potential role of working memory has not been explored. Given the importance of working memory for word learning and the evidence of weakness in this domain in children with DLD, this aspect should be carefully considered when exploring word learning in children with DLD.

Finally, there has been great inconsistency in the approaches used to test word learning in children with DLD. Tasks characteristics might contribute to differences between typically developing children and children with DLD. The literature suggests that word length (Jackson et al., 2019; Jackson et al., 2021), the number of items (Kapa & Erikson, 2020), and the quantity of exposure (Gray, 2003) have an impact on children's ability to learn words in both children with typical development and children with DLD. Furthermore, the outcome measures selected are also crucial for detecting differences between children with DLD and controls, for example children with DLD seem to be less accurate in the recognition of new phonological forms compared to production (Kan & Windsor, 2010). This variability has important implications for the selection of assessments in future research and in clinical practice.

Overall, this background provides rationale for an updated meta-analysis aimed to address the following research questions:

- Do word learning difficulties in children with DLD result from poor encoding, consolidation or reconsolidation?
- Are word learning difficulties in DLD associated with children' working memory and/or lexical knowledge?
- Do the task characteristics (word length, number of targets, level of exposure, outcome measure) contribute to the magnitude of the gap in performance on a word learning task between children with DLD and TD?

Methods

This systematic review and meta-analysis followed PRISMA reporting guidelines (Page et al., 2021) and has been registered on the OSF registry.

Data sources

The search strategy was applied to four electronic databases (PsycINFO, MEDLINE/PubMed, Web of Science, the Linguistics and Language Behaviour Abstracts) to identify relevant papers, while doctoral dissertations were retrieved through ProQuest. In addition, the following key journals were consulted individually through hand searches: American Speech-Language-Hearing Association journals, the International Journal of Language and Communication Disorders, Child Language Teaching and Therapy, International Journal of Speech-Language Pathology. Finally, the results were integrated with reference checking and forward citation searching for key articles (Jackson et al., 2021; Kan & Windsor, 2010). After completing the screening, authors of the papers were contacted directly if studies met the eligibility criteria, but relevant data was missing.

Search strategy

The search strategy was based on Kan and Windsor's (2010) search which combined terms for word learning and SLI. This list of keywords was integrated with additional terms introduced in the literature in the past ten years following the change of diagnostic label for the clinical population of interest, now called Developmental Language Disorder (DLD). The key words are reported in Appendix A.

Eligibility criteria

The eligibility criteria applied to the studies in this meta-analysis are listed below.

- *Design*: group comparison studies (DLD or SLI vs TD) where the typically developing controls and the children in the clinical sample were of the same chronological age.
- *Participants*: monolingual children (age < 18) with a diagnosis of DLD or SLI, and no history of sensory impairments, brain injuries, severe deprivation or neglect, and comorbidity with autism spectrum disorder or profound learning disability (IQ < 70). Studies testing cildren with a language impairment that did not meet the criteria for DLD as described by Bishop et al. (2017) were excluded. Studies involving bilingual children were excluded as limited familiarity with the phonological sequences of the targets may affect participants' ability to learn new words (Storkel, 2001). This suggests that linguistic experience could act as a confounding factor, particularly if the targets were developed based on the phonological rules of the language of which participants had limited exposure. However, studies that focused on bilingual children were included if there was also a monolingual group for which the data was available separately.
- *Measures*: The ability to acquire new words had to be tested through an experimental word learning paradigm. The stimuli needed to be presented verbally, excluding the papers where the novel words were presented in written form. Although orthographic forms may facilitate word learning in children with DLD, it remains unclear whether they provide the same level of support as they do for typically developing children (Colenbrander et al., 2019). Therefore, including written forms could influence group differences. Since word learning is a multifaceted process, a variety of measures can be used to assess different aspects of the newly acquired linguistic knowledge. Comprehension, production, recognition, and definition of novel words were accepted as outcome measures because they all allow the observation of children's ability to map word forms to the corresponding meaning. Each of these outcomes is defined in Table 1 Studies which assessed exclusively semantic (e.g., visual word association or categorization) or phonological (e.g., nonword repetition task) knowledge were excluded. Different paradigms were accepted (e.g. fast mapping, cross-situational learning) as long as they required an association between word form and meaning.
- Studies were required to have a baseline assessment that reflected the encoding phase. A study was considered as including encoding data if children's performance immediately after the end of the training was reported. Papers that explored consolidation and reconsolidation beside encoding needed to have assessed each data point independently. Reconsolidation data reflected children's performance after additional training, conducted at least one day after the initial session. Consolidation data represented children's performance after a time gap (at least 24 h) from the last training phase see Table 2 for definitions of each period.. Studies could differ in the metric used to quantify children's performance (e.g., percentage of correct words, number of correct responses) if they presented a total score for each outcome measure.
- *Data*: To be included in the meta-analysis the studies needed to report the effect size or mean and standard deviation of word learning task performance for each group (DLD vs controls). This information was necessary to address the first research question. If the studies described participants' working memory level and/or lexical knowledge, this data was extracted to answer the second research question. Papers that did not report data on working memory and language but met the criteria for the first research question were included.

Language: Studies could be conducted in any language; however, the full text of the paper needed to be available in English to be included in the present study.

Screening

The search strategy applied to all the sources in February 2022 (Search 1) found 1410 references. Of these, 593 were excluded because of duplicates and five were excluded because they were only citations. The abstract and title of the remaining 833 papers were screened by two authors independently. As not all the information necessary to define whether each article met all the inclusion criteria were included in the abstract, the following criteria were applied to determine inclusion at this stage:

- A group of monolingual children diagnosed with DLD or SLI
- A control group of TD children of the same age
- An experimental task of word learning

One hundred and sixty papers met all the criteria above and qualified for full text screening. Before each stage of the screening (titles and abstracts, full texts) the inclusion and exclusion criteria were piloted and amended to ensure consistency across the reviewers. The full-text screening applied the full inclusion and exclusion criteria to each paper to identify the final set of records eligible to answer the research questions. At the end of each of the two stages of screening, the results were compared and discussed until the two reviewers reached 100 % agreement. The screening of the 160 full texts led to 72 records being

Table 1
Outcomes accepted and description

Outcome measures	Description
Comprehension	The participants are asked to point the correct picture/object upon hearing a word.
Naming	The participants are presented with an object, a visual representation or a verbal description of the target and they need to name it
Description	The participants need to describe the meaning or what the target looks like
Recognition	The participants need to judge whether the name assigned verbally to a picture or object is correct

Table 2
Phases of word learning explored and description.

Phase	Description
Encoding	Initial phase of word learning, tested immediately after the initial training.
Consolidation	Offline mechanism allowing the transfer of new words into long term memory, tested at least 24 h after the end of the training
Reconsolidation	Mechanism enabling the updating and refinement of memory traces after retrieval, tested following additional training occurring at least 24 h after the initial session

identified which met all the inclusion criteria and qualified for data extraction. However, 29 additional studies were discarded during data extraction because the data reported was not compatible with the meta-analysis (e.g. did not report standard deviations, or data points were not independent), these are listed in Appendix B with relevant justifications. Forty-three studies met all the criteria and were included in the meta-analysis. This search was updated in April 2024 (Search 2). All the stages outlined above were carried out with papers published between February 2022 and April 2024 leading to the identification of three additional papers to be included in the meta-analysis. See Fig. 1 for details.

Critical appraisal

The Newcastle Ottawa Scale (NOS) for case-control studies was used to appraise the quality of the included papers (Wells et al., 2011). This

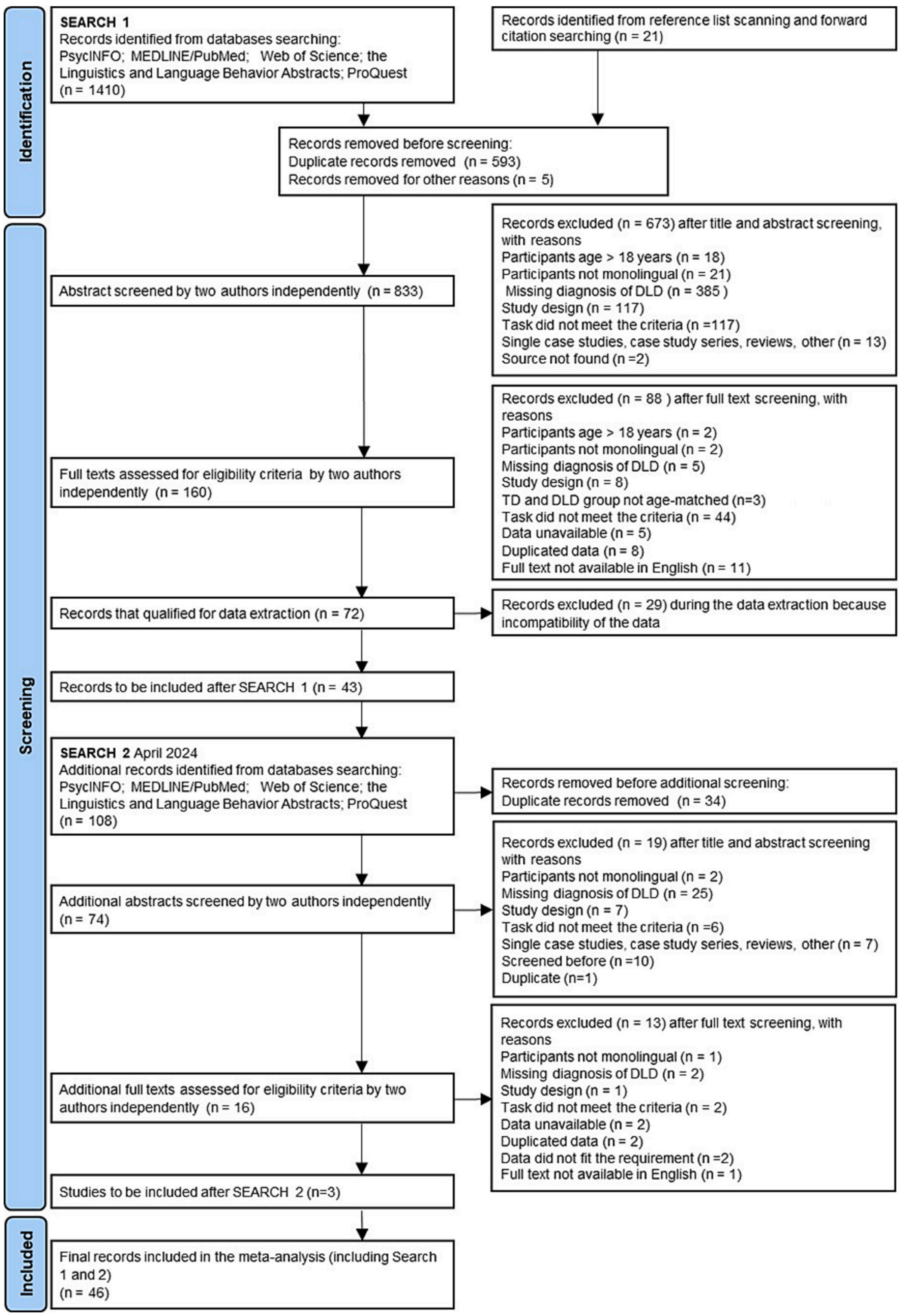

Fig. 1. Selection. Flow chart describing studies' selection consistent with the PRISMA statement (PRISMA, 2020).

version of the validated scale includes eight items grouped into three domains: 'selection of the participants', 'comparability of the groups', and 'outcome'. The items and scoring criteria indicated in the NOS manual were adapted to the current metanalysis and are fully described in Appendix C.

Studies obtaining a score between 6 and 10 points are considered to have a high-quality design and low risk of bias, studies with scores of between 3 and 5 are considered to be of fair quality with a moderate risk of bias, whereas a score below 3 suggests poor quality and a high risk of bias. The quality of the design did not constitute an exclusion criterion for the current meta-analysis. All forty-six studies included in the meta-analysis were appraised by two authors independently. Any differences were discussed until full agreement was obtained for all the records.

Data extraction

The coding strategy was piloted across reviewers prior the data extraction. One reviewer extracted the data from all the studies that met inclusion criteria, while a second reviewer verified that the data extracted from each study was accurate. The data was extracted manually from each paper and added into a data frame created in Microsoft Excel.

The mean and standard deviation of task performance by each group were retrieved from each paper to calculate the effect sizes and compare the results across studies. If the data for the main outcomes was not reported in the paper, the first author was contacted via Additionally, in papers where the relevant data was illustrated in a figure but not reported in the text, group means, and standard deviations were extracted directly from the figure using the software WebPlotDigitizer. If the data was presented disaggregated across non-relevant variables, such as phonotactic structure (e.g. separate scores for CV, CVC, CCVC words), the means and standard deviations for each variable were pooled to obtain the mean and standard deviation of the averaged data.

To analyse the effect of verbal working memory, two variables were created according to Baddeley 's model of working memory (Baddeley and Logie,1999): verbal short memory, tested by digit (or words) span forward, and central executive tested by complex spans (e.g. digit (or words) span backward). The Non-Word Repetition task (NWRT) was treated as a separate variable because, although this task recruits the working memory system, it also taps into other linguistic domains such as perception, phonology and articulation (see Estes et al., 2007 for a meta-analysis). Standardised test scores of expressive and receptive vocabularies were extracted as measures of lexical knowledge. Finally, the following task characteristics were retrieved from each study: maximum number of syllables of the target words (word length), number of targets, number of times of exposures (exposure) to each target and type of outcome measure (comprehension, naming, recognition, description). The task characteristics are reported in Appendix D.

The forty-six studies from which the data were extracted are reported and briefly described in Table 3. Outcome measures, phase of word learning, if and how participants' working memory and lexical knowledge was assessed, and demographic characteristics of the samples are reported for each study.

Effect size metric

The effect size metric was Cohen's d, calculated as the difference between TD and DLD group means divided by the pooled standard deviation. In all cases, a positive value indicates greater performance in the TD relative to the DLD group.

$$d_S = \frac{\overline{X}_1(TD) - \overline{X}_2(DLD)}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-2)SD_2^2}{n_1+n_2-2}}}$$

For *encoding* effects, the group average performance for each outcome measure tested immediately after the end of the training was considered (see Table 1). If the study involved multiple days of training, the data from post training assessments conducted on the days following the first session, was used to calculate the effect size for reconsolidation. In particular for *reconsolidation* effects, the effect size was calculated as the group difference obtained during after the final training period, corrected by subtraction of the corresponding effect size calculated for encoding. This was deemed to be the most relevant effect, since it reflected the maximal effect of reconsolidation permitted by the study design.

When the study included a delayed post training assessment conducted at least 24 h after the last training session this data was used to calculate effect sizes for consolidation. *Consolidation* effects were similarly calculated by adjusting the group difference during the consolidation period by the corresponding encoding effect.

An a-priori decision was made to analyse the effect size outcomes in a random effects model, due to its tolerance of heterogenous effect sizes. Unless reported otherwise, parameter estimates were obtained via restricted maximum likelihood estimation, owing to its superior accuracy given the smaller numbers of studies (Lopez-Lopez et al., 2014). Statistical tests of model coefficients were computed via Wald-type chi squared tests. All analyses were conducted with the 'metafor' package (Viechtbauer, 2010) implemented in the R programming language.

For each analysis the number of included effects (nested within samples) and samples (independent groups of participants, nested within studies) were coded. Many of the samples contributed to multiple effects, this is because in some studies, participants' word learning was tested by more than one outcome measure (e.g., naming and comprehension). In this case it would have not been possible to aggregate the effects as they reflect different aspects of word knowledge. Thus, to minimize this information loss and increase statistical power, conditions were used, rather than samples as the unit of analysis in the models ($k = 105$). However, when samples contribute multiple effect sizes, the assumption of independence could be violated and bias the outcome of the meta-analysis, particularly if there is anything unrepresentative about these samples (Matt & Cook, 2009; Rosenthal, 1991). To model the influence of dependency on the outcomes, multi-level models were created (see Cheung, 2014) wherein effects (level 2) were hierarchically nested within their samples (level 3), thereby estimating random effects at both the effect and sample level. The random effects structure used was '~1 — sample_id/effect_id'. Using this approach, it was possible to partition the heterogeneity between effect sizes into heterogeneity occurring at level 2 (between conditions) and heterogeneity at level 3 (between samples), statistically examining whether there was a significant amount of effect size dependency (i.e. whether a 3-level model provides a significantly better fit than a 2-level model).

Analysis

The analysis strategy consisted of three main steps.

Global meta-analytic outcomes for word learning phases

The first phase of the analysis was a global assessment that compared the pooled meta-analytic performances of children with DLD or SLI and TD children in word learning tasks across studies. The aim of this analysis was to quantify the magnitude of the difficulties in each phase of word learning (encoding, consolidation, reconsolidation) in this clinical population. This step also included other assessments of the global model, including publication bias and potential outliers.

Sample-level moderators

A further analysis estimated whether the sample characteristics, in terms of working memory and lexical knowledge, have a significant moderating effect on word learning difficulties in DLD.

Table 3
Demographic characteristics of the samples, tests, effect sizes of the included studies.

Paper	Diagnosis	TD Age Mean (SD)	DLD Age Mean (SD)	Tests of lexicon	Tests of WM/ STM	Period	Outcome	TD (n =)	DLD (n =)	Effect n	d	CI
Adlof et al 2021	DLD	95.75 (4.75)	95.75 (4.75)	EVT, PPVT	–	Encoding	Naming	90	53	1	0.28	[-0.06, 0.62]
						Encoding	Comprehension	90	53	2	0.34	[0, 0.69]
						Encoding	Description	90	53	3	0.61	[0.27, 0.96]
						Encoding	Recognition	90	53	4	0.43	[0.08, 0.77]
	DLD + DYS	95.75 (4.75)	95.75 (4.75)	EVT, PPVT	–	Encoding	Naming	90	69	5	0.56	[0.24, 0.87]
						Encoding	Comprehension	90	69	6	0.88	[0.55, 1.2]
						Encoding	Description	90	69	7	0.9	[0.57, 1.23]
						Encoding	Recognition	90	69	8	1.23	[0.89, 1.57]
Ahufinger et al 2021	DLD	105.47 (21.95)	103.15 (21.82)	–	–	Encoding	Comprehension	38	38	9	−0.05	[-0.5, 0.4]
Alt 2002	SLI	58.2 (8.5)	57.9 (7.7)	–	–	Encoding	Recognition	23	23	10	0.85	[0.25, 1.46]
Alt 2011	SLI	93.95 (6.45)	91.25 (5.5)	PPVT	–	Encoding	Naming	20	20	11	1.26	[0.58, 1.94]
						Encoding	Recognition	20	20	12	1.34	[0.66, 2.03]
Alt et al 2004	SLI	61.4 (7.8)	60.5 (7.7)	PPVT	–	Encoding	Recognition	26	26	13	1.06	[0.48, 1.64]
Alt et al 2019	DLD + DYS	92.82 (4.96)	94.61 (5.66)	EVT	–	Encoding	Naming	167	44	14	1.01	[0.67, 1.36]
Alt & Plante 2006	SLI	58.2 (5)	57.9 (5.75)	–	–	Encoding	Comprehension	23	23	15	0.85	[0.25, 1.46]
Archibald & Joanisse 2013	SLI	100.8 (13.2)	102 (12)	–	AWMA	Encoding	Naming	27	23	16	2.39	[1.66, 3.12]
Barak 2019	DLD	67.17 (7.02)	63.15 (6.53)	Goralnik,	–	Encoding	Comprehension	24	26	17	0.83	[0.25, 1.41]
Benham & Goffman 2020	DLD	59 (6.24)	60 (5.16)	EVT, PPVT	–	Encoding	Comprehension	21	21	18	−0.2	[-0.8, 0.41]
Bishop & Hsu 2015	SLI	106.8 (9.24)	103.2 (15.84)	ACE 6–11, BVSII	Word span	Encoding	Comprehension	20	28	19	0.93	[0.33, 1.54]
						Reconsolidation	Comprehension	20	28	20	−0.06	[0.36,1.51]
Busker 2010	SLI	69.83 (10.5)	75.75 (10.75)	–	–	Encoding	Comprehension	23	10	21	0.83	[0.06, 1.59]
						Encoding	Naming	23	10	22	0.25	[-0.5, 0.99]
Chen & Liu 2014	SLI	65.6 (6.7)	65.4 (6.5)	PPVT	,	Encoding	Comprehension	33	37	23	0.49	[0.01, 0.96]
Chung & Yim 2020	SLI	61.9 (9.1)	61.4 (7.79)	–	–	Encoding	Comprehension	20	10	24	1.1	[0.29, 1.91]
Gordon et al 2021	DLD	58.78 (9.05)	59.33 (8.25)	PPVT	–	Encoding	Naming	9	9	25	0.81	[-0.15, 1.77]
						Consolidation	Naming	9	9	26	0.19	[-9.12,1.73]
						Reconsolidation	Naming	9	9	27	−0.32	[-0.12,1.74]
						Consolidation	Naming	9	9	28	−0.38	[-0.13,1.74]
Gray 2003	SLI	55.2 (6)	54.36 (6.12)	EVT, PPVT	–	Encoding	Comprehension	30	30	29	0.29	[-0.22, 0.8]
						Encoding	Naming	30	30	30	0.31	[-0.2, 0.82]
						Reconsolidation	Comprehension	30	30	31	1.14	[-0.26,0.83]
						Reconsolidation	Naming	30	30	32	1.27	[-0.25,0.86]
Gray 2004	SLI	58.55 (7.08)	57.85 (6.85)	PPVT	–	Encoding	Comprehension	20	20	33	0.81	[0.16, 1.45]
						Encoding	Naming	20	20	34	−0.45	[-1.08, 0.17]
Gray 2006	SLI	42.67 (2.35)	43.33 (3.09)	PPVT	–	Encoding	Comprehension	15	15	35	0.11	[-0.6, 0.83]
						Encoding	Naming	15	15	36	0.77	[0.02, 1.51]
Gray & Brinkley 2011	SLI	54.76 (5.24)	56.67 (6.07)	EVT, PPVT	–	Encoding	Comprehension	42	42	37	2.75	[2.16, 3.35]
						Encoding	Naming	42	42	38	0.55	[0.12, 0.99]
Gray et al 2012	SLI	55.95 (4.43)	56.58 (5.02)	EVT, PPVT	–	Encoding	Comprehension	39	40	39	0.01	[-0.43, 0.45]
						Reconsolidation	Comprehension	39	40	40	0.4	[-0.2,0.69]
						Encoding	Naming	39	40	41	0.65	[-1.1, −0.2]
						Reconsolidation	Naming	39	40	42	1.65	[-1.16,-0.14]
Gray et al. 2014	SLI	55.89 (6.08)	56.96 (5.77)	EVT, PPVT		Encoding	Naming	44	48	43	0.22	[-0.19, 0.63]
						Reconsolidation	Naming	44	48	44	0.19	[-0.19,0.63]

(continued on next page)

Table 3 (*continued*)

Paper	Diagnosis	TD Age Mean (SD)	DLD Age Mean (SD)	Tests of lexicon	Tests of WM/ STM	Period	Outcome	TD (n =)	DLD (n =)	Effect n	d	CI
Gul et al 2023	DLD	129.27 (15)	125.5 (19.27)	–	–	Encoding	Comprehension	15	14	45	0.94	[0.17, 1.71]
Haebig et al 2017	SLI	124.56 (15.36)	123.36 (14.16)	PPVT	–	Encoding	Comprehension	26	23	46	0.64	[0.07, 1.22]
Hansson et al 2004	SLI	126 (8.75)	120 (8.5)	PPVT	NWRT[b]	Encoding	Naming	38	27	47	0.8	[0.29, 1.31]
Horohov & Oetting 2004	SLI	71.72 (4.59)	74.77 (5.99)	PPVT	–	Encoding	Comprehension	18	18	48	1.02	[0.33, 1.72]
Jackson et al 2016	SLI	65.92 (2.98)	64.39 (4.1)	PPVT, CELF-core	NWRT	Encoding	Naming	26	23	49	2.51	[1.76, 3.26]
Jackson et al 2021	DLD	82.04 (7.56)	83.54 (7.59)	CELF-core	DSF, DSB, NWRT	Encoding	Naming	54	50	50	1.46	[1.02, 1.89]
						Reconsolidation	Naming	54	50	51	−0.23	[1.07,1.84]
						Consolidation	Naming	54	50	52	−1.36	[1.03,1.88]
						Encoding	Comprehension	54	50	53	0.5	[0.11, 0.89]
						Reconsolidation	Comprehension	54	50	54	−0.17	[0.11,0.88]
						Consolidation	Comprehension	54	50	55	−0.13	[0.11.0.88]
						Encoding	Description	54	50	56	0.75	[0.35, 1.15]
						Reconsolidation	Description	54	50	57	0.09	[0.36,1.13]
						Encoding	Recognition	54	50	58	1.5	[1.06, 1.93]
						Reconsolidation	Recognition	54	50	59	0.02	[1.11,1.88]
						Consolidation	Recognition	54	50	60	0.03	[1.11, 1.88]]
Johnson & de Villiers 2009	LI	73.65a (17.18)	76.39a (18.43)	–	–	Encoding	Comprehension	78	33	61	1.46	[1.01, 1.91]
Kapa & Erikson 2020	DLD	59.61 (5.38)	59.44 (5.22)	PPVT	DSF, DSB	Encoding	Naming	41	41	62	0.7	[0.25, 1.14]
						Reconsolidation	Naming	41	41	63	0.26	[0.26,1.13]
						Encoding	Comprehension	41	41	64	1.1	[0.64, 1.57]
						Reconsolidation	Comprehension	41	41	65	0.01	[0.67,1.53]
						Encoding	Recognition	41	41	66	0.73	[0.28, 1.18]
						Reconsolidation	Recognition	41	41	67	0.15	[0.3,1.16]
Malins et al 2021	DLD + Reading diff	110.4 (7.2)	115.2 (8.4)	PPVT	–	Encoding	Comprehension	25	34	68	0.75	[0.21, 1.28]
						Reconsolidation	Comprehension	25	34	69	−0.12	[0.23,1.26]
Matrat et al 2023	DLD	79.3 (18)	80.1 (15.7)	EVALO	–	Encoding	Naming	23	9	70	1.79	[0.9, 2.67]
						Encoding	Description	23	9	71	0.6	[-0.19, 1.38]
McGregor et al 2022	DLD	86.59 (4.63)	86.68 (6.2)	NIH	–	Encoding	Recognition	44	28	72	0.85	[0.35, 1.34]
						Encoding	Comprehension	44	28	73	0.57	[0.08, 1.05]
McKean et al. 2014	DLD	51.68 (7.33)	55.42 (15.26)	EOWPVT, ROWPVT	–	Encoding	Comprehension	38	12	74	0.74	[0.07, 1.4]
Moav-Scheff et al 2015	SLI	62 (7)	70 (8)	TEV	Syllable span	Encoding	Comprehension	54	30	75	0.88	[0.41, 1.34]
						Encoding	Recognition	54	30	76	0.59	[0.14, 1.05]
Nash & Donaldson 2005 (implicit task)	SLI	83.62 (13.62)	83.94 (13.7)	BPVS	–	Encoding	Naming	16	16	77	1.74	[0.93, 2.56]
						Reconsolidation	Naming	16	16	78	1.03	[1.01,2.48]
Nash & Donaldson 2005 (explicit task)	SLI	83.62 (13.62)	83.94 (13.7)			Encoding	Naming	16	16	79	1.54	[0.75, 2.32]
						Reconsolidation	Naming	16	16	80	0.71	[0.82,2.25]
Nash & Donaldson 2005 (implicit task)	SLI	83.62 (13.62)	83.94 (13.7)			Encoding	Recognition	16	16	81	2.33	[1.44, 3.23]
						Reconsolidation	Recognition	16	16	82	−1.83	[1.51,3.16]
Nash & Donaldson 2005 (explicit task)						Encoding	Recognition	16	16	83	2.39	[1.48, 3.3]

(*continued on next page*)

Table 3 (*continued*)

Paper	Diagnosis	TD Age Mean (SD)	DLD Age Mean (SD)	Tests of lexicon	Tests of WM/ STM	Period	Outcome	TD (n =)	DLD (n =)	Effect n	d	CI
						Reconsolidation	Recognition	16	16	84	−0.58	[1.68,3.1]
Oetting 1999		75.5 (5.75)	77.15 (7.51)			Encoding	Comprehension	20	20	85	1.26	[0.58, 1.94]
Pomper et al 2022 (explicit task)	DLD	86.6 (4.58)	86.81 (5.64)	PPVT	DSB, NWRT	Encoding	Comprehension	36	45	86	0.59	[0.14, 1.03]
Pomper et al 2022 (implicit task)	DLD	86.6 (4.58)	86.81 (5.64)			Encoding	Comprehension	36	45	87	0.52	[0.08, 0.97]
Pomper et al 2022 (explicit task)	DLD	86.6 (4.58)	86.81 (5.64)			Encoding	Recognition	36	45	88	2.25	[1.69, 2.8]
Pomper et al 2022 (implicit task)	DLD	86.6 (4.58)	86.81 (5.64)			Encoding	Recognition	36	45	89	0.55	[0.11, 1]
Pomper et al 2022 (explicit task)	DLD	86.6 (4.58)	86.81 (5.64)			Consolidation	Recognition	36	45	90	0.38	[1.8, 2.69]
Pomper et al 2022 (implicit task)	DLD	86.6 (4.58)	86.81 (5.64)			Consolidation	Recognition	36	45	91	0.32	[0.11, 0.99]
Rice et al 2000	SLI	60.91 (3.82)	59.85 (4.4)	PPVT	–	Encoding	Comprehension	22	20	92	0.29	[-0.32, 0.9]
Ricketts et al 2015	SLI	136.48 (20.4)	135.72 (20.28)	–	–	Encoding	Comprehension	27	27	93	0.32	[-0.22, 0.86]
Rohlfing et al 2018	SLI	41.3 (3.6)	39.7 (3.5)	–	–	Encoding	Comprehension	8	8	94	1.46	[0.36, 2.56]
						Consolidation	Comprehension	8	8	95	−0.83	[0.44, 2.48]
Southwood & White 2021	SLI	82 (19)	82 (19)	–	–	Encoding	Comprehension	253	36	96	0.45	[0.1, 0.8]
Thomas 2013	SLI + CAS	117 (30)	126 (22)	–	–	Encoding	Comprehension	9	9	97	1.07	[0.08, 2.06]
Weismer & Hesketh 1993	SLI	70.6 (4.7)	71.6 (4.7)	PPVT	–	Encoding	Comprehension	8	8	98	0.74	[-0.27, 1.75]
						Encoding	Naming	8	8	99	0.97	[-0.07, 2.01]
Weismer & Hesketh 1996	SLI	86 (13)	86 (12)	PPVT	–	Encoding	Comprehension	16	16	100	0.71	[0, 1.43]
						Encoding	Naming	16	16	101	1.12	[0.37, 1.86]
						Encoding	Recognition	16	16	102	0.11	[-0.58, 0.81]
Weismer and Hesketh 1998	SLI	96 (12)	98 (11)	PPVT	–	Encoding	Comprehension	20	20	103	0.82	[0.18, 1.47]
						Encoding	Naming	20	20	104	0.79	[0.15, 1.44]
Yim & Yang 2021	VD	70.43 (8.85)	67.27 (10.15)	–	–	Encoding	Comprehension	35	15	105	0.69	[0.07, 1.31]

AWMA = Automated Working Memory Assessment, BPVS = British Picture Vocabulary Scale, CAS = Childhood Apraxia of Speech, DLD = Developmental language disorder, DSB = Digit span Backward, DSF = Digit span forward, DYS = Dyslexia, EOWPVT = Expressive One Word Picture Vocabulary Test, EVALO = French Battery for the Evaluation of Language, EVT = Expressive Vocabulary Test, LI = Language Impairment, NIH = the vocabulary subtest from the NIH Toolbox, NWRT = Non Word Repetition Task, PPVT = Peabody Picture Vocabulary Test, PRES = Preschool. eceptive-Expressive Language Scale, ROWPVT = Receptive One Word Picture Vocabulary Test, SLI = Specific Language Impairment, STM = Short term memory, TEV = Tavor Expressive Vocabulary Test, VD = Vocabulary Delay, WMI = Working Memory Impairment. [a] Global mean calculated from partial means in paper,.[b] Only reported for one group

Task-level moderators

A separate analysis evaluated whether the difference between TD and DLD (or SLI) depended on the following characteristics of the experimental word learning paradigm: word length (in syllables), number of targets, level of exposure, outcome measure (comprehension, production, definition, recognition).

Moderator analyses

Our analysis of moderators followed different stages and proceeded as follows:

First pass phase

Given the low number of studies with complete cases for all moderators, attempting to evaluate all candidate moderators within a single model entailed a low number of observations per coefficient. Therefore, in the initial ‘first pass’ phase, a series of independent ‘single moderator’ models that contained each one of the moderators individually were fitted. This phase determined a subset of potentially explanatory variables for further exploration. Statistical tests of model coefficients were computed via likelihood ratio tests, comparing a model including the moderator to an empty (intercept only) model, using maximum likelihood estimation.

‘Model comparison’ phase

To test multiple combinations of moderators, the following strategies were employed.

If the number of effects with complete cases exceeded 75 %, automated model selection was employed and all combinations of models were fitted to these complete-case effects, comparing them on the basis of Akaike’s information criterion (AIC). This was the case of task-level moderators because most of the studies reported basic details of the paradigm. The reasoning behind this approach was that the model with the lowest AIC provides the optimal fit to most of the data, taking into account the wider population of models. This information-theoretic approach (Burnham & Anderson, 2002) allowed the determination of the ‘importance’ of each coefficient on the basis of their summed Akaike weights across the population of models. This approach, implemented in the ‘MuMin’ R package (Bartoń, 2020), depends on repeated evaluations of a ‘full model’ formula that includes all moderators. For the sample-level moderators, the information reported varied significantly across studies on account of the heterogenous sample characteristics recorded. This meant that the distribution of moderators was sparse and there were very few complete cases. Therefore, the approach taken was to proceed via forward selection and examine the impact of adding additional moderators to the best-fitting single moderator model identified in the first pass phase. This choice was justified by the fact that the automated strategy would have resulted in fitting and comparing models based on small and unrepresentative subsets of effects.

Results

Critical Appraisal

Quality appraisal indicated a high-quality design with low risk of bias in 59 % (n = 27) of the studies (NOS score 10–6), 47 % (n = 17) had a fair quality with moderate risk of bias, and only 4 % (n = 2) were of poor quality with high risk of bias. Overall the studies showed a high degree of bias in the representativeness of cases with DLD, as the sample of children with DLD did not represent all the possible eligible cases in a community in any of the studies, even if often they were recruited within the same community. Only twelve studies recruited children with DLD and typically developing controls from the same community, while 74 % (n = 34) of themrecruited children from different schools or services in broad geographic areas. The majority of studies defined precise criteria for DLD (91 %, n = 42) and reported the scores of standardised non-verbal IQ and language tests for TD (87 %, n = 40).

Eighteen papers compared the mean age of the two groups without matching the participants systematically, eight of them matched children with DLD and TD controls on chronological age, while 20 matched the two groups on an additional characteristic beside age.

In terms of outcome, 100 % of the studies used the same task and outcome measures to test word learning in children with DLD and TD; however only six papers (13 %)reported the response rate of the two groups.

The NOS score obtained for each study is reported in Appendix C.

Meta-analysis of encoding

Effect size of encoding

All 46 studies included data on encoding yielding 78 effect sizes from 57 independent samples. The overall sample sizes were 1462 children for the DLD group and 2161children for the control group of typically developing peers. Analysis of Cook’s distances highlighted one influential case as potential outlier.

The random effect model indicated a large effect size for encoding (k = 78, d = 0.82, [0.66, 0.98], p < .001) confirming that children with DLD encode fewer words than typically developing peers, as observed in Fig. 2A. Framed in terms of the probability of superiority (Ruscio, 2008), this implies that a randomly sampled individual from the TD population has a 72 % chance of having higher encoding performance than a randomly sampled individual from the DLD group. Further inspection revealed no coding errors, and since the effect for encoding was still significant after the outlier was removed (k = 77, d = 0.79, [0.63, 0.94], p < .001), they were considered non influential and retained in the analysis.

The studies were highly heterogeneous (Q (78) = 346.09, p < .001) and the I^2 statistics indicated that this heterogeneity accounted for 81 % of the total variance. The heterogeneity between samples was estimated to explain 49 % of the total variance; however, the heterogeneity within samples accounted for 32 % of the total variance. The 3- level random effect model had a superior fit to the 2-level model, justifying the nested model specification (LRT = 4.54, p = .03).

Publication bias: Encoding

An association between effect size and sampling variance was detected via both the rank sum test (Kendall’s tau = 0.24, p = .002) and Egger’s Regression Test (z = 3.31, p < .001), consistent with the possible presence of publication bias. However, Fail-Safe N calculation indicated that 2495 additional studies with null results would be needed to reduce the overall effect size to non-significance (p > .05). This suggests that the observed effects are relatively robust. The funnel plot is represented in Fig. 2B.

Moderator analysis of participants characteristics: First pass

Having established a significant effect for encoding during the first phase of the analysis, the next step of the analysis focused on whether participants’ working memory and/or lexical knowledge moderated the differences in the ability to encode new words between children with DLD and TD children. The only component of the working memory system included in the moderator analysis was Verbal Short-Term Memory (VSTM) as the number of studies that tested the participants’ Central Executive component of working memory was not sufficient (n = 2). Frequencies and co-occurrences of these moderator variables are shown below in Table 4.

For VSTM, NWRT, receptive, and expressive vocabulary, effect sizes were calculated that reflected the group difference between these scores. A series of random effect models were then run with a single moderator corresponding to one and each of these variables. The effect sizes for VSTM had an effect on the effect sizes for encoding (k = 15, β = 1.41, [0.56, 2.25] p = .001) suggesting that lower performance in encoding new words in children with DLD was associated with lower VSTM levels (Fig. 2C). Furthermore, an effect of participants ‘receptive vocabulary was detected on encoding (k = 55, β = 0.28, [0.03, 0.53], p = 0.03), while expressive vocabulary (k = 25, β = 0.18, [- 0.22, 0.59], p = 0.37) and NWRT (k = 22, β = 0.02, [- 0.33, 0.37], p = 0.9) did not have an effect.

Moderator analysis of sample characteristics: Model comparison

The first pass phase of the moderator analysis highlighted VSTM and receptive vocabulary as potential moderators. To determine which of these characteristics had the most significant impact on encoding, it was necessary to compare the models and identify the one with the best fit to the data.

The tendency of different papers to report different sample

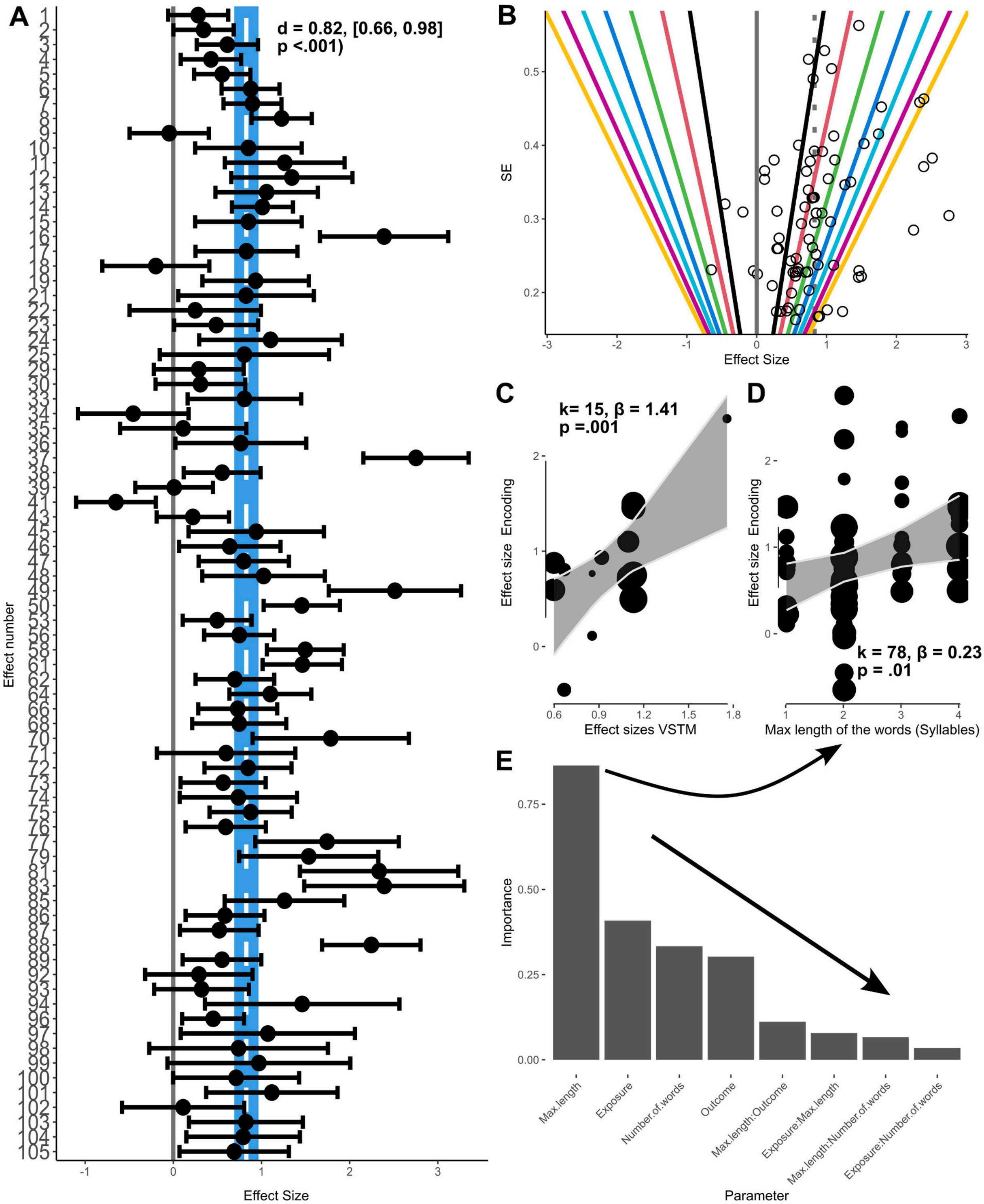

Fig. 2. Meta-analysis of encoding. (A) Forest plot of effects for the encoding phase: error bars are 95 % confidence intervals (CI). Dotted white line is the pooled summary effect, blue region is the 95 % CI. (B) Funnel plot for encoding. Dotted line is the pooled effect size. Solid lines represent p values (From central, vertical line outward: 1,.05,.01,.001,.0001,.00001,.000001). (C) Relationship between effect sizes of encoding and Verbal Short Term Memory (VSTM). Size of points is inversely proportional to the standard error of the effect (larger = more precise). Shaded region represents the 95 % confidence intervals of the values estimated by the rando effect model with VSTM as only predictor. (D) Relationship between the length of the words (syllables) and effect sizes for encoding. Size of points is inversely proportional to the standard error of the effect (larger = more precise). Shaded region represents the 95 % confidence intervals of the values estimated by the random effect model with word length as only predictor. (E) Box plot representing the importance of task related moderators. The X axis reports the variables considered in the automated model selection for the moderator analysis of task characteristics. The Y axis shows the sum of the Akaike weights for the models with the variable.

Table 4
Frequencies and co-occurrences of data available for participants characteristics.

Variable	Number of effects
Receptive vocabulary	55
Expressive vocabulary	25
Non word repetition task	22
Verbal short term memory	15
Receptive vocabulary X Non word repetition task	18
Receptive vocabulary X Expressive vocabulary	19
Receptive vocabulary X Verbal short term memory	8
Expressive vocabulary X Non word repetition task	5
Expressive vocabulary X Verbal short term memory	3
Verbal short term memory X Non word repetition task	9
Receptive vocabulary X Verbal short term memory X Non word repetition task	5
Receptive vocabulary X Expressive vocabulary X Non word repetition task	5
Receptive vocabulary X Expressive vocabulary X Verbal short term memory	1
Expressive vocabulary X Non word repetition task X Verbal short term memory	1
Expressive vocabulary X Non word repetition task X Non word repetition task X Verbal short term memory	1

characteristics led to a very sparse distribution of these moderators (see Table 3), meaning that it was not possible to perform automated model selection based on AIC. Accordingly, the analysis proceeded via forward selection. This approach involved first selecting the model with the most promising moderator identified during the initial phase, which was VSTM (k = 15), and subsequently adding the other potential moderator, which was the receptive vocabulary (k = 55). The aim was to observe whether adding the further moderator the model fit improved. The full model estimated the effect of both VSM and receptive vocabulary on encoding in the subset of data which included information on both these participants' characteristics.

The full model did not improve the models with VSTM and receptive vocabulary included as single predictors (LRT = 0.94, p = .33), therefore it was not possible to establish which of these two moderators was more relevant. This might be the result of the limited number of effect sizes (n = 8) considered in this more complex model, as only four studies reported information of both participants' receptive vocabulary and short-term memory.

Moderator analysis: First pass, task characteristics

After exploring the potential impact of participants profiles on the encoding of new words, the analysis focused on whether the difference in encoding between groups was moderated by the characteristics of the experimental task used in each study. The variables considered were the outcome measure (naming, comprehension, recognition or description), the number of targets, the maximum length of the words in syllables, and the times of exposure to the target. Table 5 describes the frequencies and co-occurrences of these moderator variables. The variable 'Number of words' and 'Outcome measure' were present for all the effect sizes therefore the combinations for these variables are not reported.

Table 5
Frequencies and co-occurrences of data available for task characteristics.

Variable	Number of effects
Outcome measure	78
Number of words	78
Maximum length of the words (Syllables)	74
Exposure to the targets (Times)	70
Exposure to the targets X Maximum length of the words	66
Number of words X Maximum length of the words X Exposure to the targets	66

The effect of each variable was firstly explored individually by a series of models with each task characteristic as single moderator. The number of words taught in the task (k = 78, β = 0.02, [- 0.02, 0.05], p = .39) and the number of times of exposure to the target (k = 78, β = -0.005, [- 0.02, 0.005], p = .32), did not contribute to highlight a difference in encoding between TD children and children with DLD in the present dataset. None of the outcome measures used had an effect on encoding: Description (k = 78, β =- 0.02, [-0.58, 0.53], p = .93), Naming (k = 78, β = 0.05, [- 0.25, 0.35], p = .76), Recognition (k = 78, β = 0.3, [-0.04, 0.65], p = .09). The maximum length of the words was the only one of the task characteristics explored with an effect on encoding (k = 78, β = 0.23, [0.05, 0.41], p = .01).

Moderator analysis: Model comparison, task characteristics

The first pass phase of the moderator analysis exploring the effect of the task characteristics on encoding pointed to maximum word length as potential moderator. The model comparison was necessary to further understand the importance and contribution of the task characteristics on encoding.

Sixty-six (84.6 %) effects had complete cases for all task-related moderators. As such, it was possible to consider only those effects with complete cases and perform automated model selection of all 167 possible models based on AIC. The importance of all model terms, as defined by summed Akaike weights, is illustrated below in Fig. 2E.

The optimal model, identified by the automated model selection according to AIC, included just the maximum length of the words as the moderator. This model confirmed an effect of word length on encoding (k = 66, β = 0.23, [0.05,0.4], p = .01). This result suggests that difficulties in encoding new words in children with DLD are particularly evident when the assessment task involves longer words. The relationship between the length of the targets and the difference in word encoding ability in the two groups is represented in Fig. 2D.

Meta-analysis of consolidation

Effect size of consolidation

The meta-analysis for consolidation included 8 effect sizes across 4 studies. The sample sizes were 107 and 112 children for the typically developing and DLD groups respectively. Consolidation was measured by subtracting the score obtained for encoding, which represented the baseline, from the score obtained by each group at the assessment of consolidation. In this way it was possible to rule out that potential difference in the rate of consolidation would depend on gaps during the initial encoding of words.

The random effect model did not detect a significant effect of consolidation (k = 8, d = − 0.2, [-0.68,0.29] p = .43), showing that, in the studies included, children with DLD consolidate the same proportion of words as children with typical development (Fig. 3A). The length of the gap from the last training session did not have an impact on the effect sizes for consolidation (p = .98). There was a high heterogeneity in the data for consolidation (Q(7) = 44.02p < .001); however, ANOVA showed that the three-level model including the random effect of the effects size did not have a better fit than the two-level model (LRT = 0.05, p = .83).

Publication bias of consolidation

The rank-correlation (Kendall's tau = − 0.21, p = .55) and Egger's Regression (z =- 0.38; p = .7) tests for asymmetry of the funnel plot were not significant.

Meta-analysis of reconsolidation

Effect size of reconsolidation

Nine studies tested reconsolidation of words during word learning. Overall, the meta-analysis of reconsolidation included 19 effect sizes

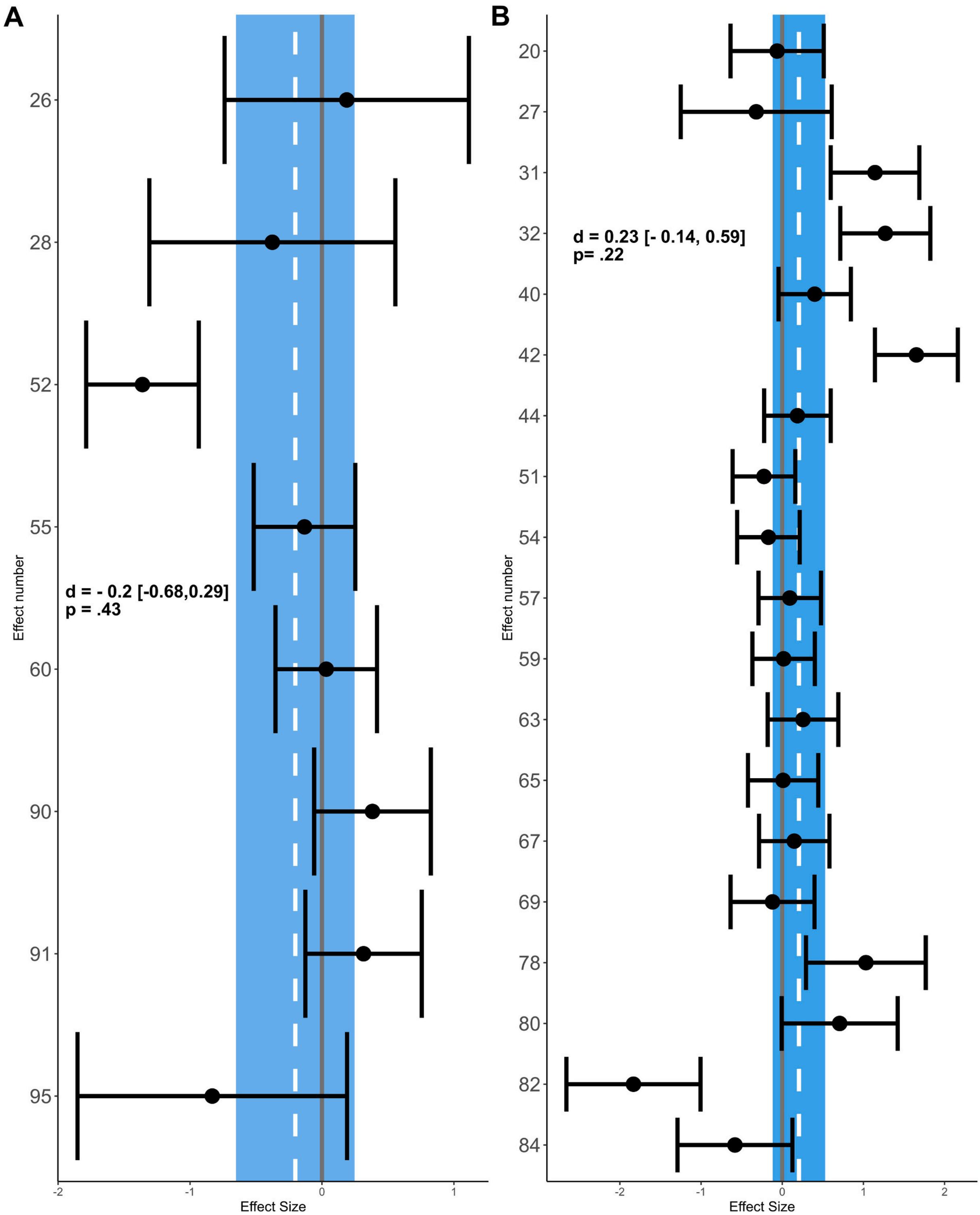

Fig. 3. Meta-analysis of consolidation and retention. Forest plot of effects for the consolidation (A) and reconsolidation (B) phases: error bars are 95% confidence intervals (CI). Dotted white line is the pooled summary effect, blue region is the 95% CI.

and compared 278 children with typical development to 296 children with DLD taken from 9 independent samples.

Consistent with the approach followed for consolidation, reconsolidation was measured by subtracting the score obtained at the end of the first day of training, which represents encoding, from the score on the last day of training. This provided a measure of reconsolidation that only reflected the learning pattern during the additional days of training, without being influenced by the baseline.

The effect of reconsolidation estimated by the random effect model was not significant ($k = 19$, $d = 0.23$, [- 0.14, 0.59], $p = .22$) (Fig. 3B) and the number of days of training did not have an effect on the ability to reconsolidate words in the two groups ($p = .65$). There was a high heterogeneity in the data ($Q(18) = 106.05$, $p < .0001$); however, ANOVA showed that the three-level model including the random effect of the effects size did not have a better fit than the two-level model (LRT $= 0.3$, $p = .58$).

These results confirmed that children with DLD and TD included in the dataset did not differ in their ability to reconsolidate words across multiple days of training regardless of the number of days of training.

Publication bias of reconsolidation

The rank-correlation test (Kendall's tau = 0.11, p = .53) and Egger's Regression Test (z =- 0.78; p = .44) for the funnel plot asymmetry were not significant.

Discussion

The current meta-analysis synthesised the existing evidence on word learning in children with DLD in terms of encoding, consolidation and reconsolidation, associations with lexical knowledge and working memory and task characteristics. The first aim was to identify which stage of word learning is less effective in children with DLD compared to TD age-matched peers. The results identified encoding as the most critical challenge during word learning in children with DLD. These findings support the hypothesis that poor encoding abilities act as bottleneck preventing later consolidation and reconsolidation of words in this population (Bishop and Hsu, 2015; Gordon et al., 2021; Jackson et al., 2021). A large number of effect sizes were calculated for encoding (k = 78), and therefore, while acknowledging the possible presence of publication bias, the evidence of poor word encoding can be considered robust.

The meta-analysis of encoding not only confirmed the extent of the difficulty in children with DLD but also clarified the processes contributing to this difficulty. Verbal short-term memory was a significant predictor of encoding in the studies that included this information (k = 15). This means that the greater the difference in verbal short-term memory between groups, the larger the gap in their ability to encode new words. The findings argue in favour of the importance of verbal short-term money for effective encoding of new word and strengthen the evidence for the impact that short term memory difficulties might have for language learning in children with DLD (Archibald, 2018). Receptive vocabulary also moderated encoding in the studies that reported this information (k = 55). A smaller vocabulary leads to weaker mental representations, making it harder to process new words effectively (Archibald, 2018). This explains why greater vocabulary gaps between TD children and those with DLD result in more pronounced differences in their ability to encode new words.

It has been suggested that the verbal working memory system, of which verbal short term memory is a subcomponent, plays a greater role in early word learning than lexical knowledge. (Gray et al., 2022). However, due to the limited data on participant characteristics, it was not possible to compare the relative contributions of short-term memory and vocabulary knowledge to word encoding in children with typical development and those with DLD. This meta-analysis indicates that both verbal short-term memory and previously stored lexical knowledge are essential for encoding new words in children with DLD. However, the findings are insufficient to determine which aspect of their cognitive profile (verbal short-term memory or lexical knowledge) plays the dominant role. Still, given the well-documented challenges these children face with both working memory and lexical knowledge, examining how these processes influence word encoding offers further evidence that encoding is likely the most difficult stage of word learning for this population.

In contrast, consolidation and reconsolidation appeared to be relatively intact, in line with the literature reviewed in the Introduction. Although more research is needed to better understand the mechanisms of offline consolidation of new words in typically developing children and those with DLD (Henderson et al., 2013; Lukács et al., 2017), the studies included in this analysis suggest comparable consolidation abilities between the two groups. This finding supports the view that declarative memory may be a relative strength in children with DLD. However, this conclusion should be interpreted with caution, as it is based on a limited number of effect sizes. For instance, in a study examining word consolidation over time (Gordon et al., 2021), participants underwent a retrieval-based training protocol, which has shown to enhance word consolidation (Haebig et al., 2019). Consequently, the lack of of group differences in this case may have been influenced by the highly supportive nature of the training procedure.

Further research is needed to better understand not only whether children with DLD experience difficulties with consolidation of words, but also how factors such as the nature of the learning targets and the training schedule may influence outcomes. For example, one factor that was not explored in the present meta-analysis was the impact of type of training. Analysis of differences across incidental tasks, retrieval-based practice, fast mapping, and cross-situational learning at different stages might provide further clarification on the possible dissociation between procedural and declarative memory in children with DLD.

Reconsolidation also appeared to be a relative strength in the sample included in this meta-analysis. It seems that once information is encoded, children with DLD are able to retrieve and update it successfully. During reconsolidation, memories are susceptible to modification and updating upon retrieval which can be triggered either by active recall or by further exposure to the target. Unlike initial encoding, reconsolidation does not require the acquisition of entirely new linguistic information but rather the updating of previously encoded and consolidated material. Therefore, it is plausible that children with DLD might experience difficulties in encoding but not in consolidation. Additionally, like consolidation, reconsolidation relies on the declarative memory system. The idea that reconsolidation may be a relative strength is further supported by evidence suggesting that retrieval practice can enhance word learning in children with DLD. If the retrieval and updating processes underlying reconsolidation were inefficient in these children, they would be unlikely to benefit from retrieval practice.

In terms of tasks characteristics, this meta-analysis pointed to word length as the most important task characteristic for the encoding of new words, in line with previous studies, suggesting that children with DLD have difficulties with longer words (Jackson et al., 2019; Jackson et al., 2021). The moderator analysis found a larger discrepancy in encoding between TD children and age-matched children with DLD in studies that used longer targets. This finding further strengthens the evidence for the contribution of verbal short-term memory in word encoding. Long words require processing and temporary storage of a longer sequences of sounds, therefore specific difficulties with longer items might reflect a limited capacity in the serial storage of verbal information. Surprisingly, the analysis did not detect an effect of exposure. Previous studies have suggested that children with DLD need to be exposed to the target word a greater number of times than TD children to effectively encode new words (Gray, 2003). Therefore, it was expected that the difference between groups would have been more significant when the task involved limited exposure to the target. Nevertheless, many of the studies included used a fast-mapping paradigm that relies on a single exposure to each target, which might have had an impact on the estimation of the effect of exposure on encoding.

Overall, the findings from this meta-analysis provide important insights into the nature of word learning, which are crucial for informing future research, as well as clinical and educational practice. Understanding the specific stages where children with DLD encounter difficulties in learning new words allows clinicians and teachers to implement evidence-based strategies grounded in a theoretical understanding of the children's profiles of strengths and weaknesses.

Based on these results teachers and clinicians should focus on the encoding phase of word learning, applying facilitating strategies and creating supportive learning environments. For example, using shorter and simpler words when explaining new topics, providing repeated opportunities for retrieval, practice and exposure, and giving direct instruction could enhance encoding (Leonard et al., 2019; Storkel et al., 2017; Pomper et al., 2022). Direct instruction not only helps focus the child's attention but also makes the learning process more explicit. Additionally, the assumption that consolidation and reconsolidation are preserved mechanisms of word learning in DLD also has significant implications for clinical practice. This perspective suggests that repeated training sessions could be beneficial because reconsolidation may serve as a compensatory process for initial encoding. Ineffective encoding might lead to inaccurate representations, which can be then consolidated. Reconsolidation through further training may offer an opportunity to correct the inaccurate representations of words (Storkel et al., 2019).

Furthermore, the meta-analysis highlighted that individual differences in children's vocabulary knowledge and short-term memory are related to their word learning success. Therefore, when planning interventions, it is important to take into account each child's unique strengths in these areas. Finally, although this meta-analysis did not find a significant effect of the type of outcome measure, and therefore did not point to specific difficulties with phonological form, semantic representation, or the form-referent link, it remains important to recognize that children may require support in developing one or more of these aspects of lexical knowledge during the word encoding process.

Limitations

The limited data available on consolidation and reconsolidation a limitation for the robustness of our results regarding these stages of word learning. The fact that the analysis did not observe a difference in the ability to consolidate words over time between children with TD and DLD might reflect the small number of studies that explored these stages, and the specific methodologies employed. Therefore, general conclusions on the consolidation and reconsolidation of new words in DLD should not be drawn. Nonetheless, despite being based on limited data, the findings synthetised in the current paper are in line with a series of studies conducted by McGregor, which tested both encoding and consolidation of words in adults with DLD (McGregor et al., 2013, 2017, 2020).

This meta-analysis not only synthesised evidence on word learning difficulties in children with DLD but also exposed critical gaps in the literature. The findings revealed a stark imbalance between studies examining the initial encoding of new words and those investigating how children with DLD consolidate and reconsolidate over time. To advance understanding, future research must focus on better understanding of the ability to consolidate words over time in children with DLD. Moreover, future meta-analyses on the topic should incorporate studies involving adults to provide a more comprehensive understanding of word learning in DLD across the lifespan.

A further limitation was the high variability in the linguistic and cognitive aspects of the samples reported across studies. Only a few studies tested both participants' working memory and vocabulary, hence the moderator analysis estimated the effect of each variable on encoding on different samples of participants, making it impossible to have a comparison between moderators. In addition, it was only possible to observe the verbal short-term memory and not the central executive component of working memory given the restricted number of studies that tested the latter. Therefore, we are unable to draw more general conclusions on the whole working memory system. Future research exploring the role of working memory in language learning should refer to the models of working memory and use different tasks to assess different components of the system.

In addition, the present study did not account for the type of training and instructions received by participants. These factors may influence the gap between children with DLD and their typically developing peers. They are also relevant to broader theoretical frameworks, such as the procedural/declarative deficit hypothesis. Therefore, the omission of this parameter is acknowledged as a limitation and future studies should consider this important factor to better understand its potential impact.

Finally, studies involving bilingual children were not included in this meta-analysis. While this decision was justified, future research could broaden the sample by including studies on bilingual populations, carefully controlling for factors such as the amount of linguistic input received in each language and the characteristics of the target words.

Conclusions

This meta-analysis confirmed word encoding difficulties in children with DLD and provided evidence suggesting that children with DLD consolidate and reconsolidate new words as successfully as typically developing children. The present findings also provide important insights into the potential contribution of working memory in the acquisition of new words for children with DLD: when reported, participants' verbal short-term memory mediated their encoding difficulties. This finding has implications for clinical and educational practice for children with DLD. The results also highlighted gaps in the existing knowledge and literature on this topic thus providing a new direction for future research which should consider all the components of the working memory system and explore word learning beyond fast mapping.

CRediT authorship contribution statement

Paola Calabrese: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nicholas Hedger:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Formal analysis. **Katherine Pritchard:** Writing – review & editing, Investigation. **Vesna Stojanovik:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization. **Emma Pagnamenta:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research team would like to thank the authors who made the data available for this meta-analysis, especially those who provided additional unpublished data. The study was funded by the University of Reading.

Appendix A–D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2025.104678.

Data availability

The data was made available via link: All data including the full dataset, the R code, the registration and the PRISMA checklist are available via OSF. https://osf.io/pr7dw/?view_only=3eb39c98a60b4cc382fbc7818cfb1864.

References

Adlof, S. M., Baron, L. S., Bell, B. A., & Scoggins, J. (2021). Spoken word learning in children with developmental language disorder or dyslexia. *Journal of Speech, Language, and Hearing Research, 64*(7), 2734–2749. https://doi.org/10.1044/2021_JSLHR-20-0021

Ahufinger, N., Guerra, E., Ferinu, L., Andreu, L., & Sanz-Torrent, M. (2021). Cross-situational statistical learning in children with developmental language disorder.

Language, Cognition and Neuroscience, 36(9), 1180–1200. https://doi.org/10.1080/23273798.2021.1922723
Alt, M. (2002). *Semantic attributes and aural encoding: A study of young children*. The University of Arizona.
Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders, 44*(2), 173–185. https://doi.org/10.1016/j.jcomdis.2010.09.003
Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research,* 49(5), 941-954. https://doi.org/1044/1092-4388(2006/068.
Alt, M., Gray, S., Hogan, T. P., Schlesinger, N., & Cowan, N. (2019). Spoken word learning differences among children with dyslexia, concomitant dyslexia and developmental language disorder, and typical development. *Language, Speech, and Hearing Services in Schools, 50*(4), 540–561. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0138
Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping. *Journal of Speech, Language, and Hearing Research, 47*(2), 407–442. https://doi.org/10.1044/1092-4388(2004/033)
Archibald, L. M. (2018). The reciprocal influences of working memory and linguistic knowledge on language performance: Considerations for the assessment of children with developmental language disorder. *Language, Speech, and Hearing Services in Schools, 49*(3), 424–433. https://doi.org/10.1044/2018_LSHSS-17-0094
Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language and Communication Disorders, 41*(6), 675–693. https://doi.org/10.1080/13682820500442602
Archibald, L. M. D., & Gathercole, S. E. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language, 57*(2), 177–194. https://doi.org/10.1016/j.jml.2006.11.004
Archibald, L. M., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory and Cognition, 41*, 268–280. https://doi.org/10.3758/s13421-012-0259-4
Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422
Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005.
Barak, L. (2019). Aluntit & Kurnas: Influences of bilingualism and developmental language disorder on novel word learning in children (Master's thesis, University of Haifa (Israel).
Bartoń, K. (2023). *MuMIn: Multi-model inference* (R package version 1.43.17). https://CRAN.R-project.org/package=MuMIn.
Benham, S., & Goffman, L. (2020). Lexical–semantic cues induce sound pattern stability in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(12), 4109–4126. https://doi.org/10.1044/2020_JSLHR-20-00244
Bishop, D. V., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology, 3*, 1–12. https://doi.org/10.1186/s40359-015-0062-7
Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Catalise-2 Consortium, Adams, C., ... & house, A. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068-1080. https://doi.org/10.1111/jcpp.12721.
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. https://doi.org/10.1017/S0142716416000060
Brown, R. M., & Robertson, E. M. (2007). Off-line processing: Reciprocal interactions between declarative and procedural memories. *Journal of Neuroscience, 27*(39), 10468–10475. https://doi.org/10.1523/JNEUROSCI.2799-07.2007
Buckner, R. L., Wheeler, M. E., & Sheridan, M. A. (2001). Encoding processes during retrieval tasks. *Journal of cognitive neuroscience, 13*(3), 406–415.
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edition). New York: Springer.
Busker, K. (2010). *The effects of gestures on the acquisition of novel words in children with specific language impairments and typically developing children*. The William Paterson University of New Jersey). Master's thesis.
Chen, Y., & Liu, H. M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35*(1), 10–20. https://doi.org/10.1016/j.ridd.2013.10.010
Cheung, M. W. L. (2014). Modelling dependent effect sizes with three-level meta-analyses: A structural equation modelling approach. *Psychological Methods, 19*(2), 211. https://doi.org/10.1037/a0032968
Cowan, N. (2022). Working memory development: A 50-year assessment of research and underlying theories. *Cognition, 224*, 1–19. https://doi.org/10.1016/j.cognition.2022.105075
Chung, H., & Yim, D. (2020). Quick incidental learning of words by children with and without specific language impairment: An eye-tracking study. *Communication Sciences & Disorders, 25*(3), 499–516. https://doi.org/10.12963/csd.20715.
Colenbrander, D., Miles, K. P., & Ricketts, J. (2019). To see or not to see: How does seeing spellings support vocabulary learning? *Language, speech, and hearing services in schools, 50*(4), 609–628. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0135
Craik, F. I. M., Hasselmo, M. E., & Davachi, L. (2007). Encoding. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 122–144). Oxford: Oxford University Press.
Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111
Esbensen, A., & Thomsen, P. (2021). Word retrieval and lexical organization in children with hearing loss and developmental language disorder. *Communication Disorders Quarterly, 42*(4), 238–248. https://doi.org/10.1177/1525740120932531
Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)
Gordon, K. R., Storkel, H. L., Lowry, S. L., & Ohlmann, N. B. (2021). Word Learning by preschool-age children with Developmental Language Disorder: Impaired encoding and robust consolidation during slow mapping. *Journal of Speech, Language, and Hearing Research, 64*(11), 4250–4270. https://doi.org/10.1044/2021_JSLHR-21-00046
Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research, 46*(1), 56–67. https://doi.org/10.1044/1092-4388(2003/005)
Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47* (5), 1117–1132. https://doi.org/10.1044/1092-4388(2004/083)
Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 955–969. https://doi.org/10.1044/1092-4388(2006/069)
Gray, S. I., Levy, R., Alt, M., Hogan, T. P., & Cowan, N. (2022). Working memory predicts new word learning over and above existing vocabulary and nonverbal IQ. *Journal of Speech, Language, and Hearing Research, 65*(3), 1044–1069. https://doi.org/10.1044/2021
Gray, S., & Brinkley, S. (2011). Fast mapping and word learning by preschoolers with SLI in a supported learning context: Effect of encoding cues, phonotactic probability and object familiarity. *Journal of Speech, Language, and Hearing Research, 54*(3), 870. https://doi.org/10.1044/1092-4388(2010/09-0285)
Gray, S., Brinkley, S., & Svetina, D. (2012). Word learning by preschoolers with SLI: Effect of phonotactic probability and object familiarity. *Journal of Speech, Language, and Hearing Research, 55*(5), 1289–1300. https://doi.org/10.1044/1092-4388(2012/11-0095)
Gray, S., Pittman, A., & Weinhold, J. (2014). Effect of phonotactic probability and neighbourhood density on word-learning configuration by preschoolers with typical development and specific language impairment. *Journal of Speech, Language, and Hearing Research, 57*(3), 1011–1025. https://doi.org/10.1044/2014_JSLHR-L-12-0282
Gul, A., Baron, L. S., Black, K. B., Schafer, A. L., & Arbel, Y. (2023). Declarative learning mechanisms support but not probabilistic feedback-based learning in children with Developmental Language Disorder (DLD). *Brain Sciences, 13*(12), 1649. https://doi.org/10.3390/brainsci13121649
Gupta, P. (2005). Clinical issues: What's in a word? A functional analysis of word learning. *Perspectives on Language Learning and Education, 12*(3), 4–8. https://doi.org/10.1044/lle12.3.4
Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with development language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research, 62*(4), 944–964. https://doi.org/10.1044/2018_jslhr-l-18-0071
Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry, 58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734
Hansson, K., Forsberg, J., Löfqvist, A., Mäki-Torkko, E., & Sahlén, B. (2004). Working memory and novel word learning in children with hearing impairment and children with specific language impairment. *International Journal of Language and Communication Disorders, 39*(3), 401–422. https://doi.org/10.1080/13682820410001669887
Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674-68. https://doi.org/10.1111/j.1467-7687.2012.01172.
Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of experimental child psychology, 116*(3), 572–592. https://doi.org/10.1016/j.jecp.2013.07.004
Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25*(1), 43–65. https://doi.org/10.1017/S0142716404001031
Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language and Communication Disorders, 51*(1), 61–73. https://doi.org/10.1111/1460-6984.12185
Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders, 79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2020). Working, declarative, and procedural memory in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(12), 4162–4178. https://doi.org/10.1044/2020_JSLHR-20-00135

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2021). Word learning and verbal working memory in children with developmental language disorder. *Autism and Developmental Language Impairments, 6*. https://doi.org/10.1177/2396941521100410

James, E., Gareth Gaskell, M., & Henderson, L. M. (2020). Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: It'll be alright on the night. *Journal of Child Psychology and Psychiatry, 61*(10), 1104–1115. https://doi.org/10.1111/jcpp.13253

Johnson, E. G., Mooney, L., Estes, K. G., Nordahl, C. W., & Ghetti, S. (2021). Activation for newly learned words in left medial-temporal lobe during toddlers' sleep is associated with memory for words. *Current Biology, 31*(24), 5429–5438. https://doi.org/10.1016/j.cub.2021.09.058

Johnson, V. E., & de Villiers, J. G. (2009). Syntactic frames in fast mapping verbs: Effect of age, dialect, and clinical status. *Journal of Speech, Language, and Hearing Research, 52*(3), 610–622. https://doi.org/10.1044/1092-4388(2008/07-0135)

Jones, G., & Witherstone, H. L. (2011). Lexical and sublexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition, 39*, 588–599. https://doi.org/10.3758/s13421-010-0045-0

Jones, S. D., & Brandt, S. (2018). Auditory lexical decisions in developmental language disorder: A meta-analysis of behavioural studies. *Journal of Speech, Language, and Hearing Research, 61*(7), 1766–1783. https://doi.org/10.1044/2018_JSLHR-L-17-0447

Jones, S. D., & Westermann, G. (2022). Under-resourced or overloaded? Rethinking working memory deficits in developmental language disorder.*Psychological. Review, 129*(6), 1358–1372. https://doi.org/10.1037/rev0000338

Jones, S. D., Stewart, H. J., & Westermann, G. (2024). A maturational frequency discrimination deficit may explain developmental language disorder. *Psychological Review, 131*(3), 695.

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 53* (2), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248

Kapa, L. L., & Erikson, J. A. (2020). The relationship between word learning and executive function in preschoolers with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(7), 2293–2307. https://doi.org/10.1044/2020_JSLHR-19-00342

Laurino, J., Forcato, C., Coaker, N., Pedreira, M. E., & Kaczer, L. (2022). Learning new words: Memory reactivation as a mechanism for strengthening and updating a novel word's meaning. *Memory & Cognition, 1–17*. https://doi.org/10.3758/s13421-021-01247-1

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., & Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder I: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research, 62*(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology, 67*(1), 30–48. https://doi.org/10.1111/bmsp.12002

Lukács, Á., Kemény, F., Lum, J. A., & Ullman, M. T. (2017). Learning and overnight retention in declarative memory in specific language impairment. *PloS one, 12*(1), Article e0169474. https://doi.org/10.1371/journal.pone.0169474

Lum, J. A., & Conti-Ramsden, G. (2013). Long-term memory: A review and meta-analysis of studies of declarative and procedural memory in specific language impairment. *Topics in language disorders*, *33*(4), 282-297. ttps://doi.org/10.1097/01.TLD.0000437939.01237.6a.

Magro, L. O., Attout, L., Majerus, S., & Szmalec, A. (2018). Short-and long-term memory determinants of novel word form learning. *Cognitive Development, 47*, 146–157. https://doi.org/10.1016/j.cogdev.2018.06.002

Malins, J. G., Landi, N., Ryherd, K., Frijters, J. C., Magnuson, J. S., Rueckl, J. G., & Morris, R. (2021). Is that a pibu or a pibo? Children with reading and language deficits show difficulties in learning and overnight consolidation of phonologically similar pseudowords. *Developmental Science, 24*(2), Article e13023. https://doi.org/10.1111/desc.13023

Matrat, M., Delage, H., & Kehoe, M. (2023). A new dynamic word learning task to diagnose language disorder in French-speaking monolingual and bilingual children. *Frontiers in Rehabilitation Sciences, 3*, Article 1095023. https://doi.org/10.3389/fresc.2022.1095023

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 537–560). Russell Sage Foundation.

McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism and Developmental Language Impairments, 5*, Article 2396941519899311. https://doi.org/10.1177/2396941519899311

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research, 60*(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research, 56*(6), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Smolak, E., Jones, M., Oleson, J., Eden, N., Arbisi-Kelm, T., & Pomper, R. (2022). What children with developmental language disorder teach us about cross-situational word learning. *Cognitive Science, 46*(2), 13094. https://doi.org/10.1111/cogs.13094

Mckean, C., Letts, C., & Howard, D. (2014). Triggering word learning in children with language impairment: The effect of phonotactic probability and neighbourhood density. *Journal of Child Language, 41*(6), 1224–1248. https://doi.org/10.1017/S0305000913000445

Moav-Scheff, R., Yifat, R., & Banai, K. (2015). Phonological memory and word learning deficits in children with specific language impairment: A role for perceptual context. *Research in Developmental Disabilities, 45*, 384–399. https://doi.org/10.1016/j.ridd.2015.08.010

Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology, 3*, 41. https://doi.org/10.3389/fpsyg.2012.00041

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48*(2), 439–458. https://doi.org/10.1044/1092-4388(2005/030)

Nichols, S., Jones, W., Roman, M. J., Wulfeck, B., Delis, D. C., Reilly, J., & Bellugi, U. (2004). Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain and language, 88*(2), 180–189. https://doi.org/10.1016/S0093-934X(03)00097-X

Oetting, J. B. (1999). Children with SLI use argument structure cues to learn verbs. *Journal of Speech, Language, and Hearing Research, 42*(5), 1261–1274. https://doi.org/10.1044/jslhr.4205.1261

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews., bmj*, 372. https://doi.org/10.1136/bmj.n71

Pomper, R., McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Ohlmann, N. (2022). Direct instruction improves word learning for children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 65*(11), 4228–4249. https://doi.org/10.1044/2022_JSLHR-22-0030

Rantalainen, K., Paavola-Ruotsalainen, L., Alakortes, J., Carter, A. S., Ebeling, H. E., & Kunnari, S. (2021). Early vocabulary development: Relationships with prelinguistic skills and early social-emotional/behavioral problems and competencies. *Infant Behavior and Development, 62*, Article 101525. https://doi.org/10.1016/j.infbeh.2020.101525

Rice, M. L., Cleave, P. L., & Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(3), 582–594. https://doi.org/10.1044/jslhr.4303.582

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48*(6), 1397–1411. https://doi.org/10.1044/1092-4388(2005/097)

Ricketts, J., Dockrell, J. E., Patel, N., Charman, T., & Lindsay, G. (2015). Do children with specific language impairment and autism spectrum disorders benefit from the presence of orthography when learning new spoken words? *Journal of Experimental Child Psychology, 134*, 43–61. https://doi.org/10.1016/j.jecp.2015.01.015

Rohlfing, K. J., Ceurremans, J., & Horst, J. S. (2018). Benefits of repeated book readings in children with SLI. *Communication Disorders Quarterly, 39*(2), 367–370. https://doi.org/10.1177/15257401176924

Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine, 53*(3), 247–271.

Southwood, F., & White, M. J. (2021). Fast mapping of verbs in Afrikaans-speaking children from low and mid socioeconomic backgrounds and children with language impairment. *Clinical Linguistics & Phonetics, 35*(9), 891–908. https://doi.org/10.1080/02699206.2020.1839968

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93* (24), 13515–13522. https://doi.org/10.1073/pnas.93.24.13515

Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in neurosciences, 28*(8), 408–415.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L. (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology, 17*(1), 1–12. https://doi.org/10.3109/17549507.2014.987818

Storkel, H. L., Komesidou, R., Pezold, M. J., Pitt, A. R., Fleming, K. K., & Romine, R. S. (2019). The impact of dose and dose frequency on word learning by kindergarten children with developmental language disorder during interactive book reading. *Language, speech, and hearing services in schools, 50*(4), 518–539. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0131

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, speech, and hearing services in schools, 48*(1), 16–30. https://doi.org/10.1044/2016_LSHSS-16-0014

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition, 179*, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Takashima, A., Bakker-Marshall, I., Van Hell, J. G., McQueen, J. M., & Janzen, G. (2019). Neural correlates of word learning in children. *Developmental Cognitive Neuroscience, 37*, Article 100649. https://doi.org/10.1016/j.dcn.2019.100649

Talli, I., & Stavrakaki, S. (2020). Short-term memory, working memory and linguistic abilities in bilingual children with Developmental Language Disorder. *First Language, 40*(4), 437–460. https://doi.org/10.1177/0142723719886954

Thomas, A. (2013). *Word learning effects on speech motor performance in children with CAS. The University of*. Nebraska-Lincoln.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36(3), 1–48*, 8. https://doi.org/10.18637/jss.v036.i03.

Weismer, S. E., & Evans, J. L. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in language Disorders, 22*(3), 15–29.

Weismer, S. E., & Hesketh, L. J. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 36*(5), 1013–1025. https://doi.org/10.1044/jshr.3605.1013

Weismer, S. E., & Hesketh, L. J. (1996). Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech, Language, and Hearing Research, 39*(1), 177–190. https://doi.org/10.1044/jshr.3901.177

Weismer, S. E., & Hesketh, L. J. (1998). The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1444–1458. https://doi.org/10.1044/jslhr.4106.1444

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2011). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

Westrupp, E. M., Reilly, S., McKean, C., Law, J., Mensah, F., & Nicholson, J. M. (2020). Vocabulary development and trajectories of behavioral and emotional difficulties via academic ability and peer problems. *Child Development*, *91*(2), e365-e382. https://doi.org/10.1111/cdev.13219.

Yim, D., & Yang, Y. (2021). Can we enhance statistical learning? Exploring statistical learning improvement in children with vocabulary delay. *Communication Sciences and Disorders, 26*(3), 558–567. https://doi.org/10.12963/csd.21804.