LSHSS

Review Article

The Influence of Quantitative Intervention Dosage on Oral Language Outcomes for Children With Developmental Language Disorder: A Systematic Review and Narrative Synthesis

Pauline Frizelle,[a] Anna-Kaisa Tolonen,[b] Josie Tulip,[c] Carol-Anne Murphy,[d] David Saldana,[e] and Cristina McKean[c]

Purpose: The aim of this study was to examine the degree to which quantitative aspects of dosage (dose, dose frequency, and total intervention duration) have been examined in intervention studies for children with developmental language disorder (DLD). Additionally, to establish the optimal quantitative dosage characteristics for phonology, vocabulary, and morphosyntax outcomes.
Method: This registered review (PROSPERO ID CRD42017076663) adhered to PRISMA guidelines. Search terms were included in seven electronic databases. We included peer-reviewed quasi-experimental, randomized controlled trial or cohort analytical studies, published in any language between January 2006 and May 2020. Included articles reported on participants with DLD (M = 3–18 years); oral language interventions with phonology, vocabulary, or morphosyntax outcomes; and experimental manipulation or statistical analysis of any quantitative aspect of dosage. Studies were appraised using the Cochrane risk-of-bias tool.
Results: Two hundred forty-four articles reported on oral language interventions with children with DLD in the domains of interest; 13 focused on experimentally/statistically manipulating quantitative aspects of dosage. No article reported phonological outcomes, three reported vocabulary, and eight reported morphosyntax. Dose frequency was the most common characteristic manipulated.
Conclusions: Research is in its infancy, and significant further research is required to inform speech-language pathologists in practice. Dosage characteristics are rarely adequately controlled for their individual effects to be identified. Findings to date suggest that there is a point in vocabulary and morphosyntax interventions after which there are diminishing returns from additional dosage. If dose is high (number of learning opportunities within a session), then the literature suggests that session frequency can be reduced. Frequent, short sessions (2/3 × per week, approximately 2 min) and less frequent, long sessions (1 × per week, approximately 20 min) have yielded the best outcomes when composite language measures have been used; however, replication and further research are required before clinicians can confidently integrate these findings into clinical practice.
Supplemental Material: https://doi.org/10.23641/asha.13570934

[a]Department of Speech and Hearing Sciences, University College Cork, Ireland
[b]Department of Logopedics, University of Oulu, Finland
[c]School of Education, Communication and Language Sciences, Newcastle University, United Kingdom
[d]School of Allied Health, University of Limerick, Ireland
[e]Departamento de Psicología Evolutiva y de la Educación, University of Seville, Spain
Correspondence to Pauline Frizelle: p.frizelle@ucc.ie

Editor-in-Chief: Holly L. Storkel
Editor: Katie Squires
Received May 1, 2020
Revision received July 9, 2020
Accepted September 28, 2020
https://doi.org/10.1044/2020_LSHSS-20-00058

In the 33 years since the publication of the first systematic review of interventions for childhood speech and language disorders (Nye et al., 1987), there has been sustained growth in both the number and quality of intervention studies published in the field. The question at that time was whether or not interventions could have a positive effect on outcomes for children. It is clear from this and subsequent reviews, meta-analyses, and randomized controlled trials (RCTs) that interventions can and do effect meaningful change for children and young people with speech, language, and communication disorders (Broomfield &

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Dodd, 2011; Law et al., 2005, 2004; Roberts & Kaiser, 2011). Practitioners can now confidently counsel parents and advise managers and commissioners of services that effective interventions exist. However, if effective and cost-effective services for children with speech, language, and communication disorders are to be delivered and funded, more specific questions must now be addressed. Crucial to the design of evidence-based services and policy is the issue of dosage: how much intervention, in which form, and at what intensity are required for positive outcomes to be achieved. While practitioners and services strive to provide evidence-based interventions, surveys and reviews of practice demonstrate that factors other than current best evidence influence decisions regarding intervention dosage and delivery. These include available funding, service configuration, and cultures of current "custom and practice" (Brandel & Froeme Loeb, 2011; McKean et al., 2019; Ruggero et al., 2012; Sugden et al., 2018).

This study examines and synthesizes current evidence regarding optimal intervention dosage and intensity, with respect to children with developmental language disorder (DLD). DLD affects approximately 8% of children and is diagnosed in children presenting with persisting language difficulties, which affect their social and educational functioning and which is not caused by another neurobiological condition (Bishop et al., 2017). DLD is one of the most common neurodevelopmental disorders with potentially profound and long-term consequences, increasing risks of poor outcomes for mental health, education, social inclusion, and employment. Despite this, services to children with DLD are not universally available across childhood at levels sufficient to deliver interventions in the dosages found to be effective in intervention studies (Law et al., 2019).

Why Are Issues of Dosage Important?

The most obvious drivers for research regarding optimal intervention dosage are economic. More speech and language therapy (SLT) input comes with associated costs (Sciberras et al., 2014), and so there is a need to determine whether increased dosage really does lead to better outcomes; whether any such relationship is linear, such that more is always better, or curvilinear, where we begin to see diminishing returns above a certain level; and also whether there is a baseline dosage below which little or no effect can be expected. Finding the optimal dosage for intervention is also important in terms of the burden placed on children and their families. Attending SLT has implications for families' time and resources, and so intervention duration and intensity should not be more than needed to attain the goals of therapy or so minimal that they effectively waste the time and effort of those involved. Where children are pulled out of their classroom for SLT, it is essential that dosage is such that the benefits of intervention outweigh the costs of missed classroom learning and of potential stigmatization associated with SLT attendance. When considering the burden of interventions on families and children, it is hard not to conclude that delivery of interventions in dosages so low as to have no chance of effecting change is not only uneconomical but also unethical.

Finally, research regarding optimal dosage is vital for commissioners and policy makers to develop, fund, and deliver evidence-based policy and for practitioners, families, and individuals with DLD to advocate for appropriate levels of service provision.

What Is "Dosage"?

Although an intuitively simple construct, dosage in behavioral interventions is a complex phenomenon to describe and hence to measure. Warren et al. (2007) proposed a list of five dosage characteristics to describe intervention intensity. Three quantitative components are *dose, dose frequency,* and *total intervention duration,* which can be combined to quantify *cumulative intervention intensity.* There is also a qualitative component, *dose form.*

Dose form refers to "the typical tasks or activities (i.e., active ingredients) within which the teaching episodes are delivered."

- *Dose* is "the number of properly administered teaching episodes during a single intervention session" and has three subcomponents:
  - the average rate of teaching episodes per unit of time,
  - the length of the intervention session, and
  - the distribution/density of episodes over the session.
- *Dose frequency* can be defined as "the number of intervention sessions per unit of time" (i.e., a day, a week, a month).
- *Total intervention duration* is "the total period of time for which a specified intervention is provided."
- Finally, *cumulative intervention intensity* is a product of the previous three components, that is, Dose × Dose Frequency × Total Intervention Duration.

What Is Known About Optimal Intervention Dosage for Children With DLD?

Zeng et al. (2012) completed a systematic review to examine the influence of intervention intensity on outcomes for children with speech and language disorders. Study reporting hampered the review, as the authors noted that dosage data are not consistently reported in intervention studies. In particular, studies rarely included the average rate of teaching episodes per unit (dose), making it impossible to calculate cumulative intervention intensity. Using length of each session as a proxy for dose, they concluded that there is a nonlinear relationship between dosage and effect size, suggesting that intervention volume is not as important as its quality: More is not necessarily better.

There is contradictory evidence as to the "minimum" dose required to effect change, with an average of 6 hr therapy (range: 0–24, over 6 months—using an intention to treat protocol or recommendation for review) being linked to greater gains than a wait list control in a study by Broomfield

and Dodd (2011), and a similar level of input (average of 6.2 hr, range: 0–15, over 12 months) being associated with no significant difference in a study by Glogowska et al. (2000). Consideration of study methodology would suggest that Broomfield and Dodd's findings may be more robust (e.g., power: *N* of 703 vs. 159; homogeneity of participants; greater treatment fidelity). However, it is not possible from either study to determine the "optimal" dosage for clinically meaningful changes to occur; as Law and Conti-Ramsden (2000) note, it is highly unlikely that 6 hr of therapy is enough. When it comes to defining "optimal" intervention dosage, things become even less clear, as previous research has reported differing values. In their meta-analyses, Nye et al. (1987) reported that interventions of more than 13 weeks duration were not as effective as interventions with shorter durations (i.e., 1–12 weeks), with the highest effect size found for interventions lasting 4–12 weeks. However, Law et al. (2004) found that interventions lasting for more than 8 weeks seemed more effective than shorter interventions. Additionally, considering session lengths, Nye et al. reported that session lengths shorter than 90 min yielded higher effect sizes than longer sessions. Jacoby et al. (2002) studied the number of individual "treatment units" (i.e., 15-min sessions) needed to facilitate functional communication improvements in children with articulation and/or language disorders. They found that the degree of improvement was correlated with the number of treatment units (time in therapy). In this study, the odds of improvement increased when the child received at least 20 hr of therapy. There are a number of potential reasons for these differing findings. Therapy outcomes may be particularly important. The complex and interrelated nature of dosage means that studies rarely manipulate only one element at a time, making causal conclusions difficult. Furthermore, a number of theories of language acquisition and/or explanatory theories of DLD posit that vocabulary, phonology, and morphosyntax may invoke differing learning mechanisms in children, and hence, optimal dosage characteristics may vary across domains (Botting & Marshall, 2017).

Theories of Learning and Their Implications for Dosage

Theories of learning that are relevant across domains, in the context of dose and dose frequency with respect to children with DLD, pertain to how and over what time period information is encoded and consolidated. One theory posits that learning is more efficient when the same number of teaching episodes is distributed over several sessions than when they are massed/concentrated into one or a few sessions (see Janiszewski et al., 2003, for meta-analysis of 93 studies with typical language learners). If treatment sessions are distributed across different days or weeks, this allows for new information to be re-encoded during each session and consolidated between sessions. On the other hand, massed practice does not offer the same opportunity for consolidation following children's encoding of new information. Children with DLD have been shown to have encoding difficulties (Alt & Plante, 2006) and require a greater number of exposures to both vocabulary and syntactic forms than children with typical development (Cleave et al., 2015; Gray, 2003; Rice et al., 1994). They have also been shown to have poorer phonological short-term memory and working memory than their typically developing peers, thereby negatively impacting their memory consolidation. If children's primary difficulty is one of encoding, then we would expect that the dose per session or cumulative dose may be more important than the dose frequency. If children receive a high treatment dose, they have the opportunity to encode and re-encode multiple times, thereby strengthening their initial representation. On the other hand, if consolidation is the more significant impediment to learning, then we might hypothesize that dose frequency would have a greater impact on treatment outcomes. Even if the information has only been partly encoded following initial exposure, it may be that memory consolidation can work incrementally, building on the encoded representation at each time point. The processes of encoding new information and memory consolidation are also very relevant for the timing of outcome measures. Immediate testing, particularly with respect to probes during treatment, is likely to measure the child's encoding ability, whereas delayed testing (postintervention and at follow-up) is tapping the level of consolidation or decay that has occurred.

The Current Study

Since the publication of the Zeng et al. (2012) review, a number of studies that directly manipulate aspects of intervention dosage have been published. In order to inform evidence-based service delivery, commissioning, and policy, this review article presents a systematic review and narrative synthesis of intervention studies for children with DLD in which aspects of oral language intervention dosage are experimentally manipulated or retrospectively statistically analyzed. The review is the first of a pair completed with similar methodology and focuses on quantitative aspects of dosage. The focus of the other review is on the qualitative characteristic, dose form. To increase confidence in the conclusions drawn, the Oxford Centre for Evidence-Based Medicine hierarchy of evidence was applied, and only studies using designs at Levels 1, 2, and 3 were included (systematic reviews of RCTs, RCTs, nonrandomized controlled cohorts/follow-up designs). Those at Levels 4 and 5 (case series, case control and mechanism-based reasoning) were excluded (OCEBM Levels of Evidence Working Group, 2011). The review focuses on interventions in which there are outcomes in the domains of phonology, vocabulary, and morphosyntax and reports findings separately to determine whether optimal dosage characteristics differ across domains.

We addressed the following research questions (RQs):

1. To what degree have the quantitative aspects of dosage (dose, dose frequency, and total intervention duration) been specifically manipulated and compared

in intervention studies and how confident can we be in the study findings?

2. What are the optimal quantitative dosage characteristics for phonology, vocabulary, and morphosyntax outcomes? Does optimal dosage differ across these domains?
3. What gaps remain in the evidence?

Method

This systematic review was registered with PROSPERO (ID CRD42017076663; McKean et al., 2017) and is one of a series completed as part of European COST Action 1406. Action 1406 focused on understanding intervention and service delivery for children with DLD across Europe and a number of partner countries. Our methods adhere to PRISMA guidelines for systematic reviews (Moher et al., 2015). Due to the heterogeneity of the included studies, combining data in meta-analysis was not appropriate; the review is therefore presented as a narrative synthesis.

Search Procedure

Searches were conducted to identify empirical peer-reviewed articles, in any language, that related to oral language interventions with children with DLD. Due to the adoption of DLD terminology and criteria being very recent (Bishop et al., 2017), our searches included previous terminologies used to refer to this group of children or to subgroups within the umbrella of DLD, such as specific language impairment or language impairment. The exact terminologies used in each study were extracted and are presented in Supplemental Material S1. Seven electronic databases were used and included Web of Science (including MEDLINE, SSCI), MEDLINE (PubMed), ERIC, PsycINFO, Cochrane Library, Scopus, and LLBA. The initial search was limited to peer-reviewed studies, published between and inclusive of January 2006 and December 2015. Three updated searches were then completed; the first includes studies published between January 2016 and October 2017, the second includes studies published between November 2017 and May 2019, and the third includes studies published between June 2019 and May 2020. Search terms were developed through discussion between authors and consultation with a research librarian. The search string is given in Supplemental Material S2. Reference lists of all articles included on full-text and relevant systematic reviews were also hand-searched for any additional articles.

Inclusion/Exclusion Criteria

Included articles met the following criteria:

- Research design—(a) RCTs, (b) quasi-experimental designs (nonrandom assignment) with an element of control, or (c) cohort analytical designs, observational studies in which groups were assembled according to whether or not they have received the intervention, with control.
- Peer-reviewed publication in any language, published between January 2006 and May 2020.
- Participants with a mean age of ≥ 3 and ≤ 18 years,
- Participants identified as having (a) DLD or an equivalent term such as primary language impairment or specific language impairment and (b) difficulties on at least one oral language assessment (vocabulary, morphosyntax, or discourse) falling below 1 *SD* below the mean. Those with language impairment secondary to those conditions identified by CATALISE criteria as precluding a DLD diagnosis (e.g., autism spectrum condition, learning disability) were not included. Those with language difficulties and an "associated condition" allowed in CATALISE criteria (e.g., attention-deficit/hyperactivity disorder, dyslexia) were included. Children with childhood apraxia of speech were excluded on the basis that their pattern of response to phonological interventions may differ from those with other disorders (Morgan & Vogel, 2008), in particular with respect to dosage, and so their inclusion could potentially bias our findings regarding dosage effects in DLD.
- Examined an oral language intervention, which measured outcomes in the domains of phonology, vocabulary, and/or morphosyntax.
- Experimentally manipulated or statistically analyzed an aspect of either dose, dose frequency, or cumulative intervention intensity while keeping other variables constant.

More detailed definitions of our research design categories and our definition of intervention are given in our PROSPERO preregistration (ID CRD42017076663).

Article Selection and Reliability of Search Procedures

Stage 1. The initial search formed the basis of several COST Action IS1406 reviews with differing foci. The aim was to identify articles evaluating interventions for children with DLD across all language domains (vocabulary, phonology, morphosyntax, and pragmatics). These articles were initially screened on title and abstract for inclusion/exclusion based on the criteria of date, target group, level of evidence (whether there was an element of control included in the study design), or evaluation of an intervention. Twenty percent were double-screened by two independent reviewers (C. A. M. and D. S. for the initial search and C. A. M. and P. F. for the three updated searches) using specialist software supporting systematic reviews (EPPI-Reviewer 4). Reliability calculation was undertaken at each stage with an overall agreement rate of 96%. Disagreements at this and all subsequent stages were resolved through discussion. This stage yielded 1,198 articles. All non-English articles at this and subsequent stages were considered by either author A. K. T. (who is fluent in a number of languages) or by a

native speaker of the relevant language in the COST Action, and the relevant criteria discussed with P. F. after translation.

Stage 2. To identify those specifically relevant to vocabulary, phonology, or morphosyntax outcomes considered in this review, two independent reviewers (P. F. and A. K. T.) screened 100% of the articles included after Stage 1 on title and abstract. Agreement rate of 93%. This yielded 698 articles.

Stage 3. Full-text screening was completed against the inclusion/exclusion criteria by the same two independent reviewers. Agreement rate was 94%.

Stage 4. Full-text screening was then completed on the 244 articles emerging from Stage 3 to identify those with a specific focus on dosage characteristics, which were experimentally manipulated or statistically analyzed, and with research designs at Levels 1, 2, or 3 in the Oxford Centre for Evidence-Based Medicine hierarchy of evidence. Agreement rate at this stage was 97%.

Stage 5. Finally, full-text screening was completed on the 39 articles that emerged from Stage 4 and only those that focused specifically on dose, dose frequency, intervention duration, or total intervention intensity were included (n = 13). See Supplemental Material S3 for PRISMA flowchart.

Data Extraction

The first author extracted the following data from the articles and tabulated it in an Excel spreadsheet: study design (RCT, quasi-experimental, cohort analytical), participant variables (number, mean age at intervention baseline), treatment detail (intervention context, dose form, treatment/control targets, dosage manipulation, planned/received dose (both were extracted if reported), planned/received dose frequency and intervention duration, and outcome measures (the nature and timing of measures and the main findings).

Risk of Bias

The first and last authors (P. F. and C. M., respectively) appraised study quality using the Cochrane risk-of-bias tool for RCTs (Higgins et al., 2011). The tool aims to evaluate selection bias (random sequence generation and allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective reporting), and other bias deemed important by the reviewers (for which we included fidelity measures and noted whether a power calculation was completed). For studies in which the target group or items were not randomized, the two evaluation categories for selection bias were coded as not applicable. These studies were evaluated according to the remaining categories. We assigned risk of bias ratings of high, low, or unclear. Both reviewers rated each article independently, and disagreements were resolved by consensus. The risk of bias assessment for each article is shown in Figure 1.

Results

Thirty-nine articles reported on studies in which dosage was experimentally or statistically manipulated, and 13 of these focused on the quantitative aspects of dosage. These 13 articles came from eight journals, nine of which were in the English language and one was in German: three (23%)[1] from the *Journal of Speech, Language, and Hearing Research*, three (23%) from *Language, Speech, and Hearing Services in Schools*, two (15%) from the *American Journal of Speech-Language Pathology*, and one (8%) each from *International Journal of Language and Communication Disorders, International Journal of Speech-Language Pathology*, *Child Language Teaching and Therapy*, *Communication Disorders Quarterly*, and *L.O.G.O.S. Interdisziplinair*. Eleven of the 13 studies were conducted in the United States (85%), one was conducted in the United Kingdom (8%), and one was conducted in Germany (8%). A total of 481 children with DLD (M = 40.1, SD = 61.3) were represented in the 13 studies. Sample sizes varied from 12 to 233 children (Mdn = 25), and children with DLD had an average age range from 3;11 to 12;1 (years;months). See Supplemental Material S1.

Selection Criteria for Children With DLD

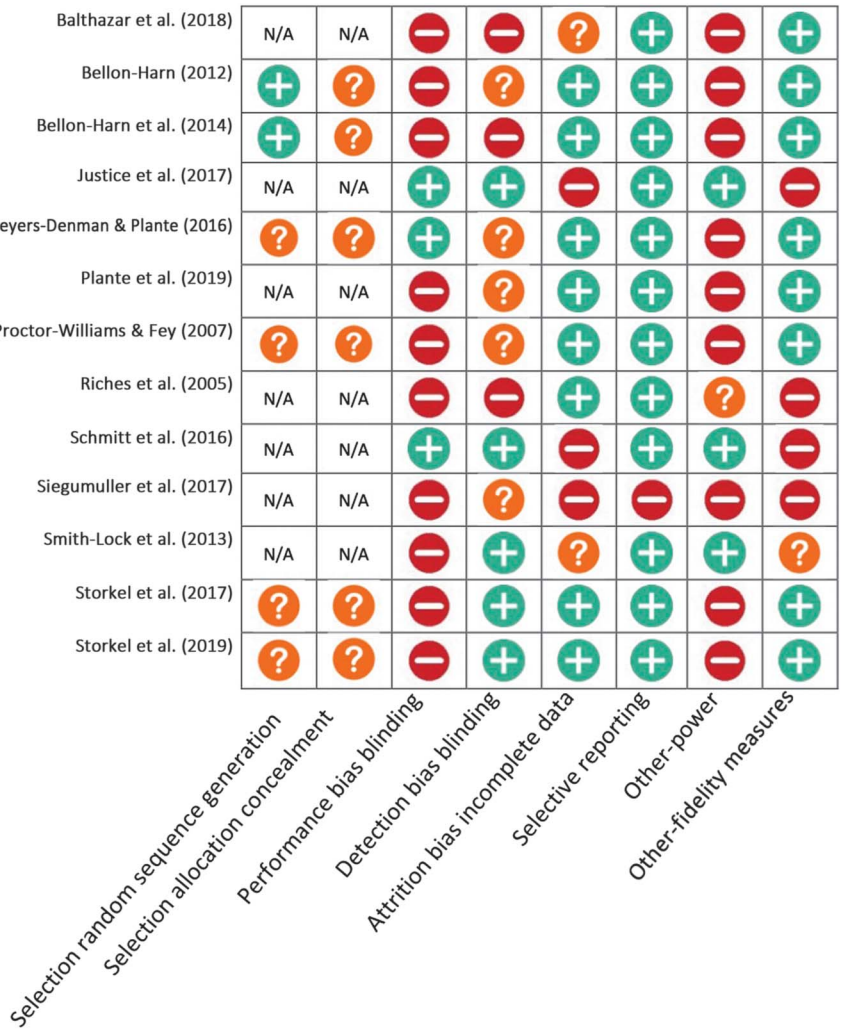
The majority of studies identified children as having DLD (or a previously used term such as *specific language impairment/language impairment*) using the following criteria: (a) a composite score of below 1 *SD* on a standardized language measure such as the Clinical Evaluation of Language Fundamentals–Fourth Edition or the Structured Photographic Expressive Language Test–Third Edition; (b) nonverbal IQ scores within 1 *SD* of the norm on a test of cognitive functioning; (c) hearing within the normal range (shown by passing a pure-tone hearing screening); and (d) no known neurological, social-emotional, or psychiatric disorders. With respect to the two studies that took place in the community [Studies 4 and 8 in Supplemental Material S1][2] although the language cut-point for inclusion was −1 *SD,* the authors point out that, on average, the included sample scored more than 2 *SD*s below the mean. The DLD diagnosis in Germany [9] was different in that it was based on medical history and the participants were required to have specific language characteristics pertinent to the intervention—such as a mean length of utterance of three words and a language sample showing no more than 15% of expressions with the verb in the second position. Three studies had a slightly lower cut-point in relation to cognitive ability, that is, a standard score of 80 [1] and 75 [6, 7]. No evidence of speech impairment was specified in three of the 13 studies.

RQ 1: To what degree have quantitative aspects of dosage been specifically manipulated and compared in interventions studies and how confident can we be in the study findings?

[1]May not sum exactly 100% due to rounding.

[2]Numbers in square brackets indicate the study number in Supplemental Material S1.

Figure 1. Critical appraisal of each included study. N/A = not applicable.



No studies manipulating quantitative dosage characteristics were identified, which focused on phonological outcomes in children with DLD. There is therefore no clear evidence regarding dose, dose frequency, total intervention duration, or cumulative treatment intensity in relation to phonology with this population. In contrast, there were three studies (23%) specific to vocabulary and eight studies (62%) specific to morphosyntax. Lastly, there were two studies (15%) with omnibus outcomes in which dosage was statistically manipulated. Figure 1 summarizes the risk of bias in each of the studies. Five studies (39%) were RCTs, Level 1 in the hierarchy of evidence, and within those RCTs, three of the five focused on morphology, only two studies explicitly described selected random sequence generation, none of the five described selection allocation concealment, and none reported on a priori power calculation. Participant numbers in RCTs were generally small (ranging from 12 to 34 children), raising concerns regarding statistical power to detect differences. In addition, although RCTs aim to control for differences across groups, this does not always work with small sample sizes. Of the five RCTs, none recruited randomly from a larger population, two recruited from a single setting, two recruited from multiple settings, and one was unclear. Randomization was always with respect to the treatment condition. Each of the trials was preliminary and included elements of Phase 1 and Phase 11 trials (Fey & Finestack, 2009). With respect to Phase 1, studies aimed to address the core treatment parameter of intensity, and in relation to Phase 11, they examined treatment benefit across children, preliminary indications of efficacy. Quasi-experimental (Level 2) studies made up 39% of the articles, with four of the five focused on morphology. In broad terms, these studies were nonequivalent group designs, although in some studies, there was an attempt to match across variables, such as nonverbal IQ and language scores. Our inclusion criteria ensured an element of control for all studies. Detection bias blinding was either not addressed or unclear in four of the five studies, and similarly,

there was no reported power calculation for four of the five studies. The cohort analytical studies (n = 3, 23%) included two with the same sample [4, 9], neither of which reported explicitly on attrition. Due to the nature of language studies, performance bias blinding is extremely challenging for all studies. Biases not present in the majority of studies were attrition bias, selective reporting, and other fidelity measures. Analysis of the publication dates for the included studies show that the majority have been published in the previous 5 years (2016–2020: n = 8, 62%; 2011–2015: n = 3, 23%; 2005–2010 inclusive: n = 2, 15%), demonstrating an increasing focus and interest in this important issue and a growing evidence resource to inform practice.

RQ 2. What are the optimal dosage characteristics for phonology, vocabulary, and morphosyntax outcomes? Does optimal dosage differ across these domains?

RQ 3. What gaps remain in the evidence?

The following provides a narrative summary of the findings of the articles identified, organized by outcome (vocabulary, morphosyntax, phonology, omnibus measure). In each section, we report on each of Warren et al.'s (2007) quantitative dosage components in turn (dose, dose frequency, total intervention duration, and cumulative treatment intensity), identifying whether evidence exists, summarizing the findings, and describing the level of confidence in those findings. Supplemental Material S1 also summarizes the data extracted from the articles.

Vocabulary

For this domain, studies manipulating dose (n = 1) [12] and dose frequency (n = 2) [8, 13] were identified, but none were found for total intervention duration or cumulative intervention intensity.

Dose: number of properly administered teaching episodes during a single session.

The issue of optimal number of exposures, with respect to new word learning, is addressed by Storkel et al. in their 2017 article [12], in the context of interactive book reading using a novel escalation design methodology. Twenty-seven children with DLD (M = 5;8) were randomly assigned to one of four word learning treatment intensities: 12, 24, 36, or 48 cumulative exposures. Children heard each target word in a shared book-reading context, followed by its definition and the use of the target word in a supportive context sentence, and lastly, they were given a synonym of each target word. Target words included nouns, verbs, and adjectives, and word learning was assessed through a definition task and a naming task. The dose per session was either 3, 4, or 6, depending on the treatment intensity. For example, in the case of 24 cumulative exposures, the target word was repeated 4 times in each book, and the book was read 6 times over the course of the intervention. Based on the word definition outcome (administered immediately postintervention), no children learned the target words following 12 exposures. At 36 exposures, 43% of children with DLD responded to treatment, while at 48 exposures, fewer children were responding (29%). Diminishing returns were also evident when using the average number of words with correct definitions in the last block as the outcome measure for each treatment intensity. Children showed the ability to define the most words (n = 5) following 36 exposures, and word learning began to diminish at 48 exposures. In addition, results from the naming task indicated 36 exposures to be the optimal dose (with 86% of children responding). A decrease in treatment response was again evident as the number of exposures increased to 48. The finding that children's optimal performance was following 36 exposures supports the theory that there is a critical minimum number of exposures required to allow adequate encoding of words to occur. On the other hand, diminishing returns at 48 exposures may be in keeping with deficient-processing theories of learning, which suggest that learning effectiveness is dependent on the degree of attention directed toward what is being learned. A reduction in attention is thought to occur as what is being learned becomes overly familiar, and while this has previously been discussed in relation to massed practice (Cepeda et al., 2006), it could also occur in the context of too many word exposures within a given time period.

While this study is highly innovative, in the application of an escalation design to the field of language learning, there are a number of points to note with respect to dosage. The number of treatment sessions ranged between 10 and 20 and were given 2–3 times a week. Therefore, the total intervention duration is a confound, as it was not constant for each dose. It is also noteworthy that children's response to treatment was very low at all exposures, when using the definition task as a measurement of learning. Only 43% of children responded at optimal dosage, and only five treatment words were correctly defined. A more optimistic result was evident using the naming task as the outcome measure, with 86% of children responding at optimal dose and 60% responding at a minimum of 12 exposures. The authors posit that semantic knowledge is measured by the definition task and that the naming task is a measure of phonology. We suggest this may be an overly conservative approach to the measurement of semantic knowledge and that word definitions are perhaps the pinnacle of semantic knowledge. More graded outcome measures, sensitive to differing levels of semantic learning, such as the children's ability to provide a synonym (a measure included in the study), could perhaps have yielded different results. It is interesting to contrast this finding with that of Aguilar et al. (2018), who manipulated dose form rather than dose in their word learning study. Aguilar et al. found that, with high variability in the referent presented, preschool children with DLD had the ability to learn three new words, having been exposed to them 18 times over three sessions and asked to name the items once per session. However, learning was measured through a comprehension probe in the Aguilar et al. study, a task significantly less challenging than the definition probes and naming tasks used by Storkel et al. (2017). In addition, in contrast to Storkel et al. where the outcomes were administered immediately postintervention, the Aguilar et al. retention outcome measure was administered at follow-up (6 weeks

postintervention), allowing for a consolidation period, which may have facilitated word learning.

Dose frequency: number of sessions over a given time frame

Riches et al. (2005) [8] investigated the effect of dose (number of word exposures) and dose frequency (spacing/period between exposures) on novel verb learning in children with DLD. The study was based on the premise that distributed learning is more efficient than a massed approach. Twenty-four children with DLD (M = 5;6) and 24 language-matched control children were taught four novel verbs, using a dual morphological frame (*Look its dacking, see it dacks*), modeled through play activities. The manipulation of the number of exposures, along with the spacing of the treatment sessions, resulted in four experimental conditions: (a) massed 12, with 12 exposures on a single day; (b) massed 18, with 18 exposures on a single day; (c) spaced 12, with 12 exposures spread over 4 days (three each day); and (d) spaced 18, with 18 exposures spread over 4 days (either four or five each day). Outcome measures were carried out directly following and 1 week postintervention and included an action probe (What does it do? Can you show me?), a production probe (What's it doing? Can you tell me?), and a comprehension probe (From a choice of three objects, which one was *verb*ing?). Posttest measures showed that children with DLD benefitted from a greater number of exposures to novel verbs with respect to comprehension. However, based on production, the spacing effect was greater and more significant than the effect of the number of exposures, that is, children had better learning after 12 presentations when the exposures were spaced than after 18 presentations when the exposures were massed. It is important to highlight a number of points in relation to this study. First, the outcome measures administered were not blind and were designed to assess comprehension and expression at a single word level. In addition, results are based on children's learning of a very small number of verbs (n = 4). Furthermore, the authors acknowledge that because each verb label was linked to a single object, we cannot assume that, following 12 or 18 exposures, the children developed a generalized representation of each verb meaning. Although the cumulative treatment intensity is equivalent across some conditions, the massed presentations differ from the spaced presentation on both dose and dose frequency, making the relative contribution of each dosage variable on children's performance difficult to extract. In addition, while highlighting the potential of manipulating spacing effects for positive gains, the study sheds little light on *optimally* spaced learning intervals or optimal number of exposures with respect to word learning in children with DLD.

Storkel et al. (2019) built on this work in their examination of whether different combinations of dose and dose frequency (while keeping treatment intensity constant) influenced the ability of kindergarten children with DLD to learn new words in an interactive book reading context. Children (between ages 5;0 and 6;2) were given 36 exposures to two word sets, 60 words in total consisting of nouns, verbs, and adjectives. For the first word set, a 6 dose × 6 dose frequency format was used with all children. For the second word set, children were randomly assigned to one of two conditions, either 4 dose × 9 dose frequency or 9 dose × 4 dose frequency, while controlling for order effects. As in their 2017 study, children's learning was measured through a word definition task, but in contrast to their previous study, outcome measures were administered at two time points posttreatment (an average of 5 days and 21 days post) and were also tracked during treatment. This was an important aspect of the study, as it revealed that children learned more words during treatment (an average of 10) than they retained after treatment was withdrawn (an average of four words). Only 40% of the words that were correctly defined at the end of treatment were retained 5 or 6 days later, and only 30% of words were retained at the 21-day time point. Word learning was, however, consistent with their previous study, in that children defined an average of four to five words correctly immediately postintervention at this exposure level. The drop in word learning calls into question our previous suggestion that perhaps the word learning advantage shown in the study by Aguilar et al. (2018) was due to the timing of the postintervention outcome measure (6 weeks post) and that this potentially served as a consolidation period. Results from Storkel et al. suggest that the delayed outcome measure revealed decay rather than consolidation.

With respect to treatment scheduling, the manipulation of dose and dose frequency while maintaining 36 exposures in both conditions did not result in differences in word learning outcomes. This finding suggests that it is the overall dose (number of exposures) that has greater impact on children's word learning than the frequency of the treatment schedule. It is also in keeping with that reported by Bellon-Harn (2012), Meyers-Denman and Plante (2016), and Balthazar and Scott (2018; presented later in this review article) with respect to morphosyntax, all of whom reported no learning advantage for a spaced rather than a more concentrated treatment schedule, when overall dose is controlled.

Morphosyntax

Dose: number of properly administered teaching episodes during a single session

Only two studies with morphosyntax outcomes included in the review manipulated dose. Proctor-Williams and Fey (2007) [7] investigated the effect of three recast densities of novel irregular past tense verbs on spontaneous conversational productions in two groups of children. Recasts were provided in the context of a child-led, play-based activity and were defined as "immediate adult responses to child utterances, that repeat some of the child's words and correct or modify the morphologic or syntactic form of the child's prior utterance, while maintaining the central meaning of the child's production" (Proctor-Williams & Fey, 2007, p. 1029). Children with DLD (between 7 and 8 years) and language-matched typically developing children (5–6 years) were exposed to recasts of six novel verbs at a conversation-level density (0.19/min) or at an intervention-level density (0.47/min), or no recasts over a period of five

sessions. The recast exemplars were distributed equally across the six verbs, that is, three in the low-density recast condition and three in the high-density recast condition. Low-density recasting translated as two per verb in each of the five sessions (30 recasts), and high-density recasting translated as five per verb in Sessions 4 and 5 only (30 recasts). Therefore, while dose per session was manipulated, total dose was equal across high- and low-density conditions. Cumulative learning was measured as the number of correct elicited irregular past tense verb productions, directly postintervention.

Contrary to the authors' expectations, the children with DLD did not improve their production accuracy at higher intervention-like recast densities; however, the sample size was small (n = 13). It may also be that difference in dose density was not sufficient to yield a difference in children's verb learning across only five sessions or that high-density recasting was not high enough to effect change. We note that the effective density of recasting reported in the Meyers-Denman and Plante (2016) study (see below) is higher than that reported here (1.25/min vs. 0.47/min). It might also be the case that an equal total dose over the course of the intervention reduces the likelihood of significant differences emerging when manipulating dose per session, particularly over such a short intervention duration. It is also noteworthy that, for both high- and low-density conditions, the total dose is only 30 recasts. This is in stark contrast to the Meyers-Denman and Plante study in which the treatment duration was equivalent (150 min), but the total dose was considerably higher, at 125 recasts. It is also unfortunate that the distribution of the five intervention sessions was not controlled, which resulted in a substantial range in total intervention duration (4–44 days). Interestingly, when the authors tested the relationship between the length of time (in days) that it took to complete the five sessions and the accuracy of past tense productions in both the low- and high-density conditions, the results indicated that the longer that children were in the experiment, the less accurately they produced the verbs. Following on from this, they investigated whether a gap of 5 days or more between any of the sessions affected the children's accuracy of spontaneous productions and found that it did not. The impact of recasting is further complicated by the fact that children were given at least five opportunities to produce each of the irregular past tense verbs in each session, regardless of density condition. Children's production levels were therefore similar across conditions and may have gone some way toward reducing the effect of recasting input on their production outcomes.

Dose: the distribution/ density of episodes within the session

Building on the work by Proctor-Williams and Fey (2007), a more recent study carried out by Plante et al. (2019) reported on within-session manipulation of the dose density of enhanced conversational recasting. An additional study distinction was that Plante et al. kept overall intervention duration constant. Twenty children with DLD (ages 4;0–5;11) were exposed to 24 unique recasts of different morphological forms per session. Recasts were given in the context of dialogic book reading and free-play activities. Treatment took place 5 days a week for 5 weeks, and targets included *–ed, third person –s, Aux is,* and *possessive*. Half of the group heard the recasts over a 30-min period (one recast every 1.25 min), and the other half heard them over a 15-min period (one recast every 38 s) while maintaining session length at 30 min. The study was designed to ascertain which of the two treatments was more effective and efficient, and how many children generalized their targets in that time frame. Children's learning was measured through the use of generalization and retention morpheme probes. The former were administered before each Monday, Wednesday, and Friday session, and the latter were given 6 weeks postintervention. Results indicated that the majority of children showed a strong treatment effect. However, there were no significant differences between the two treatment conditions on any of the outcome measures (probe or spontaneous performance, number of treatment responders, follow-up performance). In addition, there was a significant relationship between children's performance at the end of treatment and at follow-up. The authors conclude that within-session, high-density dose delivery does not offer any advantage over a lower density delivery, if dose and overall intervention duration are constant. However, the sample size was again small (n = 10). They also note that, although children retained the gains that they made in treatment, they did not show any independent improvement in target morpheme use following treatment. Findings from this study differed from the Proctor-Williams and Fey study in that the treatment itself was effective, but given the overall dose differences (30 recasts vs. between 528 and 600), this is not surprising. An important difference between the two studies was how the dose density manipulation was implemented. In the Proctor-Williams and Fey study, the low-density condition was distributed across the five sessions, but the high-density condition was implemented in Sessions 4 and 5 only. Therefore, the density manipulation was achieved by altering the number of sessions in which the recasts were given (2 × 5 sessions, 5 × 2 sessions), and as result, dose frequency was a confound. In contrast, Plante et al. altered the session length in which an equal number of recasts were given (24 recasts in 15 min vs. 24 recasts in 30 min), and this was constant across all sessions. Despite these differences, both studies showed no differences between the high- and low-density groups when dose was constant. As previously stated, it may have been the case that the dose was too low in the Proctor-Williams and Fey study to have an effect and to reveal any differences. In contrast, Plante et al. implemented a high dose, which resulted in a strong treatment effect, but even then, no differences emerged. These findings support the premise that the within-session dose maybe more important in treatment effectiveness than the session length, during which the doses are given in the context of an equivalent overall intervention duration. However, further research with larger samples is needed to validate this finding.

Dose frequency: number of sessions over a given time frame

Dose frequency was manipulated in five of the eight studies within the morphosyntax domain. Bellon-Harn (2012)

[2] reported on a study in which they examined the effect of different dose frequencies on the morphosyntactic abilities of preschool children with DLD (M = 4.61 years). Children were enrolled into either a concentrated (4 times a week for 6 weeks) or spaced treatment schedule (twice a week for 12 weeks), in which the dose, dose form, total number of intervention sessions, and therefore total number of treatment hours (8 hr) were kept constant. However, total intervention duration was not controlled. Using books as the stimuli, the therapy was described as a "scaffolded language intervention" in which techniques such as expansions, cloze procedures, and models were integrated, with an implicit method of instruction. Baseline and immediate posttreatment measures were taken using language sample analysis and probes designed to elicit targets (such as the use of auxiliary, copula, third-person singular). While the authors report positive outcomes following both treatment schedules, there were no differences in how children performed in either the concentrated or spaced treatments. This result is not consistent with previous literature in relation to typical language learners (Ambridge et al., 2006) or children with DLD (Desmottes et al., 2017); however, the sample size is particularly small (six per group), and consequently, these results should be interpreted with some caution. It is also worth noting that there is considerable variation across studies as to what is considered spaced or concentrated in treatment delivery and how this interacts with the total duration of the intervention. Indeed, even the more concentrated treatment in this study is delivered over a 6-week period. In addition, it is noteworthy that, although the authors suggest that dose is kept constant in this study, they acknowledge that, in a scaffolded language therapy, there is no predetermined script or target. As a result, dose was not closely controlled, that is, the frequency of linguistic forms within each cloze procedure, expansion, and model. The authors suggest that dose for both treatment schedules was high and may therefore mask any dose frequency effect. It may also be the case that a total of 8 hr of intervention, which was constant across conditions, was not so lengthy as to reach the point of diminishing returns, which would potentially result in a smaller effect for the more frequent schedule.

The second study in which dose frequency was manipulated with respect to morphosyntax was carried out by Smith-Lock, Leitao, Lambert, Prior, et al. (2013) [11]. The study (which included a larger sample than that by Bellon-Harn, 2012) compared the effectiveness of two different dose frequencies in relation to a school-based treatment of expressive grammar. Five-year-old children with DLD were assigned to either eight 1-hr sessions of treatment given over an 8-week period (a spaced treatment) or eight 1-hr sessions given over an 8-day period (a concentrated treatment). Once again, total intervention duration was not controlled. Therapeutic techniques were integrated into naturalistic play sessions and included explicit instruction, focused stimulation, recasting, and imitation. Treatment targets were individualized and included accurate use of past and present tense, pronouns, and possessives. Learning was measured immediately on the Grammar Elicitation Test (Smith-Lock, Leitao, Lambert, & Nickels, 2013), as well as 8 weeks postintervention. While results showed significant improvement in the group that received the spaced treatment (relative to the same time period prior to treatment), this was not the case for the concentrated treatment group. Single-subject analyses indicated that 46% of children who received the spaced schedule and 17% of those who received the concentrated schedule showed a significant treatment effect. This result is in keeping with previous findings indicating advantages for spaced learning but is contrary to results by Bellon-Harn (2012). Of interest is the fact that the number of therapy hours is equivalent for both studies; however, in addition to the sample size, a notable difference between the two studies is the total intervention duration. In Smith-Lock, Leitao, Lambert, Prior, et al (2013), the concentrated intervention takes place over a relatively short period (8 days). The spaced intervention duration (8 weeks) is, however, quite similar to the concentrated intervention duration in the Bellon-Harn (2012) study (6 weeks). We might suppose that, given a total number of therapy hours that is effective and equal in both conditions, differences only emerge between spaced and concentrated treatment schedules for children with DLD, when the time frame between the beginning and end of the treatment is significantly shorter for one condition than the other (e.g., 1 week vs. 8 weeks).

It is also the case that, while Smith-Lock, Leitao, Lambert, Prior, et al. (2013) provided teachers with scripts and detailed activity plans, dose was not controlled for in this study. Research suggests that dose frequency effects (i.e., number of sessions) can be mitigated if dose per session is high (Fey et al., 2013), but the authors do not give us any sense of dose in this study. Additionally, there are a number of treatment techniques used in both aforementioned studies, such that dose in relation to each technique is likely to be somewhat diluted and to vary between each treatment session.

Meyers-Denman and Plante (2016) [5] is the third included study to examine the effects of treatment dose frequency on grammatical morpheme remediation in young children with DLD. Again, the sample size was small at eight per group. Using enhanced conversational recasts, treatment was given in both concentrated (3 × 10-min sessions within a 4-hr period, 5 days a week) and spaced conditions (1 × 30-min session 5 times a week). The concentrated condition resulted in fifteen 10-min sessions, while the spaced condition resulted in five 30-min sessions. Specifically, with respect to dosage, a significant difference between this study and that by Bellon-Harn (2012) was, regardless of whether treatment was administered in the concentrated or spaced condition, the treatment dose (24 conversational recasts per day), rate of delivery (one recast every 1.25 min), total intervention hours (2.5 hr), and total intervention duration were controlled. Children's learning was measured through a play-based generalization probe, in which they were required to use the target morphemes with untreated lexical items. Pre- and postassessments revealed a significant improvement in morpheme production in both dose frequencies, with no change in untreated morpheme use. There were, however, no differences in the effect of treatment for the

concentrated or spaced conditions. The authors conclude that enhanced conversational recast treatment can produce positive results in a short time for children with DLD. This study appears to lend further support to the idea that if the dose itself is high (in this case, one recast every 1.25 min), it facilitates more effective encoding, and dose frequency can be reduced. One could argue that both treatment frequencies were relatively high, as treatment was given daily in both conditions. On the other hand, given the small sample size, it may be that there was no sufficient statistical power to detect differences between the two conditions. In any case, optimal dose frequency, relative to dose, has yet to be established.

In a more recent study, Balthazar and Scott (2018) [1] manipulated dose frequency, with respect to the treatment of complex sentences, in older children with DLD (10–14 years of age). Adverbial, object complement, and relative clauses were taught following a once or twice weekly treatment protocol. Total intervention duration was 9 weeks, and session length ranged between 40 and 60 min, resulting in total intervention time of 6–9 hr for the once weekly condition and 12–18 hr for the twice weekly condition. Importantly, dose was kept constant at a planned rate of 30 stimuli per session and an actual rate of 26 items per session (236 in total) in the once weekly condition, and 28 items per session (502 items in total) in the twice weekly condition. Stimuli presentation was through modeling, repetition, and manipulation of a complex sentence, with scaffolding and clinician feedback. Primary outcome measures were sentence probes administered before, during, and after treatment as well as standardized language tasks reflecting a broad range of oral and written language. Interestingly, while treatment was effective, as measured by the sentence production probes, there was no advantage for the higher dosage group on any oral language measure. This finding was contrary to the authors' hypothesis, and they suggest a number of possible explanations for this result: Given that three sessions were devoted to each sentence type, even in the once weekly group, there may have been no advantage to the additional sessions; they acknowledge that treatment maintenance was not examined, and they question whether the difference in the two dose frequencies was sufficient to yield a difference. We suggest that the findings of this study are in keeping with previous studies and support the notion that high dose reduces the need for high intervention frequency. However, it is important to consider maintenance effects.

An additional study, in which dose frequency was statistically analyzed in the treatment of complex syntax, was carried out by Siegmüller et al. (2017) [10]. Intervention outcomes were children's ability to use subordinate clause structures. Intervention dose form was implicit and carried out in six steps, which included (1) intensive modeling of (a) verbs and their associated arguments and (b) different grammatical subcomponents of the sentence, (2) questions eliciting the production of the main clause, and (3) modeling expansions of the main clause to subordinate clause structures. Children were assigned to different steps depending on their pretest performance, and treatment was discontinued when the child reached Step 5 (showed the ability to use subordinate clauses). To analyze the effects of dose frequency on the outcome, the children were divided into two groups: those who had therapy once and those who had therapy twice weekly. The aim was to establish the effect of dose frequency on how many sessions the children needed to reach the intervention goal. The maximum number of intervention sessions was 22. In support of a spacing effect advantage, the results showed that the children who received less intensive treatment (once weekly) needed fewer sessions to achieve the therapy goals than the children who received more intensive treatment. When analyzing the effect of age on achieving the intervention outcome, a significant moderate correlation was found between age and number of sessions. To study this further, the children were divided into two groups: young and old. There was a significant difference between the groups in the number of sessions needed, with younger children requiring fewer sessions. The authors suggest that younger children might react faster and more easily to intervention than older children. However, given the fact that we have no information on dose (of each dose form), it is difficult to draw strong conclusions from this study. As was the case with work already described (Bellon-Harn, 2012; Meyers-Denman & Plante, 2016), if the dose of each aspect of the treatment protocol was high, then this may have negated any benefits of a more frequent intervention. On the contrary, the participant numbers are greater in this study, therefore revealing a spacing advantage that perhaps could not be detected with smaller sample sizes.

Total intervention duration

Only one study was identified for inclusion in the review in which total intervention duration was manipulated. Bellon-Harn et al. (2014) [3] examined the effects of interactive storybook reading on children's use of microstructure elements within language samples. The study included 12 preschool children with DLD (M = 4.63 and 4.78 years), randomly assigned to two intervention durations. In one intervention, children received 42 sessions across 14 weeks, and in the other, they received 24 sessions across 6 weeks. As a result, dose frequency (although not identical) was minimally different (3 vs. 4 times a week), while there was a considerable difference in total intervention duration. However, keeping dose frequency fairly similar while manipulating the total intervention duration necessitates a considerable difference in the *total* number of intervention sessions per group (almost double), which is also likely to translate into dose differences (unless intentionally controlled for). The authors do not provide specific dose information, and we can therefore assume dose differences. The outcomes of interest were the frequency with which children used coordinate and subordinate clauses as well as the number of words within clauses. Although results indicated positive outcomes, there were no group differences between those who received 24 sessions versus those who received double this amount of treatment. The authors suggest that gains in narrative microstructure elements are obtained with less total treatment time, although it is worth noting that the treatment

frequency for both groups was intense at 3–4 times weekly. It would also be interesting to replicate this finding, while controlling for dose and with a larger sample size. It may have been the case that the dose per session was sufficiently high, that the longer intervention duration served no advantage. This would support the premise that if dose is high, not only frequency but total intervention duration can be reduced. Again, it is important to temper our interpretation based on the very small sample size included in the study. Previous findings by Fey et al. (2013) in relation to toddlers with intellectual disabilities suggest that increases in treatment frequency are only advantageous when dose is decreased, perhaps this is also the case in relation to total intervention duration and children with DLD. As previously discussed with respect to diminishing returns in word learning, a lack of advantage for the longer morphosyntax intervention is also in keeping with deficient processing theories of learning (Cepeda et al., 2006), with a suggested reduction in children's attention levels, when what is being learned becomes overly familiar, in a very lengthy intervention.

Omnibus Outcomes

Two further articles investigated how dosage characteristics interact to contribute to children's global language outcomes [4, 9]. These articles are based on a unique study that used data from a large clinically identified sample of children with DLD (n = 233), who were receiving language treatment within the U.S. public school system, over an academic year. Natural variations in treatment intensity data allowed the authors to examine the impact of different aspects of dosage on children's language outcomes, as well as the extent to which treatment outcomes vary as a function of one or more dosage parameters. Treatment centered on one of nine language focused targets, and outcomes were the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003) Core Language scores and the Picture Vocabulary subtest from the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2001). It is important to note that, in both articles, the term *dose* is defined and operationalized differently to Warren et al. (2007). Here, it is defined as the total amount of time spent addressing any one of nine language-focused targets, in contrast with the now more usual definition of the number of administered teaching episodes in a given intervention session. Hence, it is a proxy measure with less specificity and accuracy than a measure of dose, and it precludes a clear definition of dose form. On the other hand, this approach allows an examination of dosage effects in a much larger sample than found in other intervention studies and scrutinizes dosage schedules used in real-world clinical contexts. Using this approach, Schmitt et al. (2016) [9] examined the extent to which dose, dose frequency, and the interplay between the two were associated with language gains over the school year. Using structured equation modeling, the results showed that children receiving low dose/high frequency (intervention sessions of approximately 2 min, at a rate of 2–3 times per week), or high dose/low frequency (intervention sessions of approximately 20 min, at a rate of 1 per week or fortnight) had better outcomes than those receiving high frequency/high dose (20 min, 2/3 times weekly), high frequency/average dose (12 min, 2/3 times weekly), or low frequency/low dose treatment (2 min, 1 per week or fortnight). It must be noted when considering clinical application that the total intervention duration here was a school year and not discrete "blocks" of therapy found in many health care systems (McKean et al., 2019). Therefore both "optimal" conditions have relatively high total intervention hours (low dose/high frequency: 2 min × 3 sessions × 28 weeks = 168 min (2 hr 48 min); high dose/low frequency: 20 min × 1 session × 28 weeks = 560 min (9 hr 20 min per year).

Justice et al. (2017) [4] aimed to make recommendations about the quantity of treatment required to achieve the optimal amount of language gain for children with DLD using this same data set. Outcomes were retrospectively analyzed with respect to dose, dose frequency (intensity), and cumulative intensity of therapy. Multilevel modeling allowed the authors to predict language gains from each dosage parameter, and regression weighting guided a recommended amount of treatment. The process allowed the authors to develop an empirically derived equation/algorithm for use by SLPs to calculate optimal language outcomes (defined as an increase of 0.6 *SD* units). Therefore, if a clinician knows the session frequency (e.g., once weekly) and number of weeks they can work with a child over the course of the school year (e.g., 25), using baseline language scores and 0.6 *SD* as the desired amount of change, the algorithm can identify the amount of time that should be spent working on language skills, within each of those 25 sessions. Because baseline language scores are used, the algorithm that is highly innovative takes account of the severity of the disorder and provides therapists with a scientific alternative to making decisions about treatment, rather than those based on caseload size or common practice. Additionally, by manipulating the session frequency and the amount of time spent on a given language goal, therapists can also determine the degree of spacing both within and between sessions, in relation to what is being learned. With respect to limitations, the authors acknowledge that the algorithms are based on correlational data and cannot therefore be interpreted causally. We also do not know how dosage interacts with SLP decision making and whether the schedule and its relative success was influenced by therapy goals, which may be more suited to one schedule than another (e.g., past tense *–ed* vs. narrative macrostructure). In addition, although the diversity of goals and speech-language pathology practice in the schools does suggest that a range of dose forms can be effective, the ways in which targets were addressed by clinicians is likely to have varied considerably. Finally, there is a need to better understand interactions between child-level factors such as language severity and treatment intensity. The literature is unclear regarding whether children with more severe DLD might benefit from higher frequency interventions or from those in which learning opportunities are more spaced, thereby facilitating consolidation and enhanced attention.

Discussion

In this study we aimed to ascertain to what degree the quantitative aspects of dosage have been specifically manipulated in intervention studies with children with DLD, in which there were phonology, vocabulary, or morphosyntax outcomes. In addition, we aimed to identify optimal quantitative dosage characteristics in each of these domains and to highlight gaps in the literature and difficulties in interpreting the evidence. The dominant finding of the review is the lack of intervention studies across domains, in which quantitative aspects of dosage have been experimentally or statistically manipulated for children with DLD. In addition, a number of studies included in the review have been carried out with particularly small sample sizes, causing us to call into question the validity of these findings. Consequently, there is a significant need for further research to inform clinical practice. Significantly, there were no studies with phonological outcomes in this population of children in which quantitative aspects of dosage were manipulated. It is possible that the literature relating to children with speech sound disorder can be directly applied to DLD. However, this has not been tested, and given the metalinguistic skills and abstract concepts invoked in many phonological interventions, it would seem likely that modifications in dosage and/or other aspects of the interventions would be required and should be tested in empirical studies. Given high comorbidity between DLD and speech sound disorder, such work would likely have significant clinical impact (Eadie et al., 2014).

Vocabulary

The finding that there were only three studies in the vocabulary domain, in which quantitative aspects of dosage were manipulated, again highlights the dearth of research in this area. Hence, there is limited evidence on which practitioners can draw to inform the implementation of interventions and advise managers and policy makers regarding optimal dosage. The work by Storkel et al. (2017) has been both pioneering in its use of an escalation design and unique in showing diminishing benefits following a specific number of word exposures. However, in many ways, this important work represents a starting point from which to grow research in this domain. Given that frequency and total intervention duration were not constant for each dose, further work is required to determine whether this finding is replicated under constant frequency or duration conditions. It is also significant to note the differences in children's responses with respect to outcome measures (43%, word definition vs. 86%, word naming). Within word learning studies alone, outcomes can include forced choice comprehension probes, naming, word definition, and synonym comprehension or production tasks—all of which may use experimental or unfamiliar referents and which can occur during intervention, immediately after or following a consolidation period. If we are to build the necessary evidence upon which to base clinical decisions, the use of consistent outcome measures will be required to make meaningful cross-study comparisons. In addition, the timing of outcome measures is central to how we interpret study findings. This is highlighted in the work reported by Storkel et al. (2019), in which there was a 40% drop in word learning a mere 5–6 days postintervention.

Based on the findings of their earlier study (Storkel et al., 2017) and reinforced by this most recent study (Storkel et al., 2019), when measured with a naming or word definition task, 36-word exposures appear to be the optimal dose for word learning in 5- to 6-year-old children with DLD. However, this age range is narrow, and it would be interesting to investigate a potential interaction between age and number of exposures: an interaction revealed by Siegmüller et al. (2017) in relation to morphosyntax outcomes. Finally, it is important to consider the interaction between dose and dose form. There is some evidence suggesting that increasing object variability (how a referent is presented) may result in word learning at a lower dose (see Aguilar et al., 2018), and this would seem a fruitful avenue for further research.

Research examining dose frequency effects in word learning interventions in children with DLD is also scarce. This is despite the number of articles in the general verbal learning literature suggesting an advantage for distributed over massed learning (see meta-analysis completed by Janiszewski et al., 2003). Although Riches and colleagues addressed this in their 2005 article (Riches et al., 2005), as we have already noted, there was no blinding of outcomes; only four verbs were included in the study, and there were only two intervals of learning. In addition, both the massed and the spaced learning intervals were relatively concentrated, that is, the spaced condition was over 4 days, rather than a period of weeks, as in the Storkel et al. (2017) article and in much clinical practice. Recent work by Storkel et al. (2019) manipulating dose and dose frequency sheds further light on this topic, in that a much larger set of words were taught; there was some blinding of outcomes, and outcome measures were taken 21 and 5 days postintervention. Interestingly, when overall dose was controlled, the spacing of the treatment schedule did not impact children's word-learning outcomes, and the authors concluded that when treatment is given over a period of weeks, overall dose is more important than the frequency of the treatment schedule. In this study, the massed condition was over a period of 4 weeks (× 9 doses), and the spaced condition was for 9 weeks (× 4 doses). However, how spaced and massed learning conditions are defined is problematic throughout the language learning literature. One study's "spaced" presentation is another study's "massed," and there is significant variation in the total intervention duration and the total intervention hours implemented. Future work is clearly required to ascertain what is optimal dosage for children with DLD. We recommend the systematic examination of a broad range of learning intervals across a range of ages, together with a consideration of how those learning intervals interact with number of exposures.

Morphosyntax

Although quantitative aspects of dosage have been more extensively studied in morphosyntax, it is revealing

that only two studies investigated the effect of dose in interventions for children with DLD. Each study investigated a different dose subcomponent (the average rate of teaching episodes per unit of time and the distribution of episodes within the session). Examination of dose frequency would suggest that where dose is high, then dose frequency can be reduced (e.g., Balthazar & Scott, 2018). However the optimal dose per session has not yet been identified. Following dose manipulation through the presentation of recasts in low (0.19/min) and high (0.47/min) density conditions, Proctor-Williams and Fey (2007) reported no improvement in irregular past tense production accuracy in the high-density condition. This article is a telling example of the complex interactive nature of dosage and shows the difficulty involved in manipulating one aspect at a time. While cumulative intervention intensity was equivalent across groups and children's expressive dose was equal in both density conditions (such that the manipulation was only with respect to the number of recasts children heard), the authors operationalized the manipulation of dose by significantly impacting dose frequency. In addition, total intervention duration was uncontrolled and very variable (4–44 days). There is an important gap in the evidence, with studies needed taking a systematic approach to the examination of dose with respect to morphosyntax interventions. One such study was carried out by Plante et al. (2019). High-dose recast density was manipulated within sessions, while at the same time controlling for dose, dose frequency, and overall intervention duration. The high dose resulted in a treatment effect, but no differences emerged as a result of the density with which the dose was given. Because other aspects of dosage were controlled, we can be clearer about conclusions drawn from this study. However, the number of participants per group was small ($n = 10$). The findings suggest that within-session dose may be more important than the session length in which the doses are given; however, to increase confidence in this result, replication is required with a larger sample. Potentially, this has important implications for therapists, many of whom have large caseloads, who may be able to deliver high-dose effective morphosyntax interventions while allocating shorter time periods per session.

In addition, an escalation design, as implemented by Storkel et al. (2017) for vocabulary, has the potential to be informative for morphosyntax, while controlling for dose frequency. In clinical practice, dose is rarely operationalized and measured. When considering dosage characteristics, clinicians use proxy measures such as the number of intervention hours given over a specific time, the ratio of clinicians to children in an intervention service, and the degree of parent or child participation in a service over time. Without measurement of dose, these can only ever yield rough approximations of dosage characteristics.

Bellon-Harn et al. (2014) found tentative evidence that gains in morphosyntax in a narrative context can be achieved in a much shorter total intervention duration, when dose frequency is relatively intense. Unfortunately, due to the small sample size and no information on dose, the study sheds little light on why almost double the number of sessions over a longer intervention duration offered no further advantage. We suggest that, in keeping with deficient processing theories of learning (Cepeda et al., 2006), children's focus may decrease when cumulative intervention intensity becomes too high.

Lastly, dose frequency is the aspect of dosage most commonly examined in the morphosyntax domain, and much of the discussion with respect to dose frequency centers around the concepts of concentrated versus distributed learning. Study findings are mixed, and in keeping with our conclusions in relation to vocabulary, cross-study comparisons are difficult due to variation in many study characteristics. In particular, the inconsistency with which the terms *distributed* and *concentrated* are defined is problematic. While findings by Smith-Lock, Leitao, Lambert, Prior, et al. (2013) and Siegmüller et al. (2017) support a distributed learning advantage, Bellon-Harn (2012), Meyers-Denman and Plante (2016), and Balthazar and Scott (2018) found no differences in the effect of treatment for concentrated versus distributed conditions. However, sample sizes were particularly small in two of the three studies in which no differences were detected and therefore may obfuscate the true result. It is also worth noting that, in both articles that report a distributed learning advantage, we are given no information on dose. In contrast, two of the three studies reporting no differences between conditions (Balthazar & Scott, 2018; Meyers-Denman & Plante, 2016) control carefully for the effect of dose. Treatment dose was also very high in each study (24 recasts per day at a rate of one every 1.25 min; 26 or 28 sentence stimuli per session, respectively). Interestingly, both studies also controlled for total intervention duration. In summary, the research to date suggests no difference in a morphosyntax treatment effect between concentrated and distributed conditions if the treatment dose, rate of delivery, total treatment hours, and total intervention duration are controlled. In addition, one study has shown that if the rate of delivery within session is manipulated (massed vs. distributed), no learning advantage emerges (Plante et al., 2019) However, significantly more research is required with respect to concentrated and distributed intervention schedules, and optimal dose frequency relative to dose has yet to be established.

Omnibus Outcomes

Insights regarding the interaction between dose and dose frequency have been gained from the two included articles, which measure global language outcomes, where dose was defined as the amount of time spent on a given language target. Findings suggest that the best outcomes are achieved when children receive either "little and often" (frequent sessions [approximately 3 times per week] in which the focus on a specific language target is very short [2 min]) or "more and less often" (less frequent sessions [approximately weekly] in which specific goals are targeted for longer periods [20 min]). The evidence for "little and often," if embedded within longer sessions with mixed goals, may be confounded by an increase in variability and intervention

context. By changing the target after 2 min, both variability and context change, both of which are thought to be advantageous to children's language learning (Haebig et al., 2019; Plante et al., 2014). What is unclear is how many times (or if at all) the target was revisited within a single session, that is, whether there was a within-session spacing effect. In the "more, less often" scenario, there is greater spacing between sessions, which may have been a facilitating factor in increasing learning. More work is required to illuminate what is driving these effects.

Summary and Recommendations for the Future

This review highlights the limited research base available from which to identify optimal quantitative dosage characteristics in the domains of phonology, vocabulary, and morphosyntax. The need for future research to inform clinical practice is significant. Dosage characteristics and their interactions in SLT are complex. To summarize what has been reported to date, more is not always better, and studies show a point of diminishing returns for both vocabulary (number of exposures) and morphosyntax (frequency/total number of intervention sessions). There is some evidence suggesting that younger children may require fewer sessions to achieve the same results (in relation to morphosyntax), but dose frequency and total intervention duration have not been systematically examined in relation to age, and dose form techniques were not accounted for in this finding. Study findings also suggest that if dose is high (the number of learning opportunities within a session), then frequency can be reduced, particularly in relation to morphosyntax. Although results suggest no spaced learning advantage between sessions (for morphosyntax) if all other dosage characteristics are controlled, inconsistencies in the definitions of spaced/distributed and massed/concentrated have been problematic, making cross-study comparison and clinical application difficult. Within-session spacing has been underresearched, and while Plante et al. (2019) report no differences in treatment effects based on the within-session density with which the dose was given, changes in dose form context which inadvertently create within-session spacing have been found to be advantageous (Haebig et al., 2019). Finally, frequent interventions (2/3 times per week) that target language goals for short periods or less frequent interventions (1 per week or fortnight) targeting language goals for longer have been found to yield the best outcomes in relation to composite language measures. However, more nuanced research is required to examine the facilitators of these effects.

Although there are clear gaps in the evidence, some implications for practice arise from this review. Findings from Schmitt et al. (2016) and Justice et al. (2017) support the current practice of weekly or fortnightly sessions as an efficient model, but only if dose is high. Ensuring intervention sessions contain high levels of the "active ingredients" of interventions is therefore vital. Furthermore, "little and often" practice would also seem to be supported as being a potentially effective approach. Such intervention schedules are often more accessible to parents and educational practitioners, working in partnership with SLPs. However, efficacy would depend on appropriate treatment fidelity, such that the dose form delivers the necessary active ingredients of the intervention. This review also demonstrates that there are minimum cumulative interventions dosages required for children's performance to improve on intervention goals and also that too many may bring diminishing returns. While the review has not been able to identify a "magic number" for success, it does suggest that simply delivering the number of intervention hours that are part of local custom and practice is not defensible. Rather, to ensure dosage is sufficient to have an effect, children's progress should be monitored over the course of therapy and delivered until a child reaches a predetermined criterion of success, and to ensure resources aren't wasted, the focus of an intervention should be changed when progress plateaus. The implications for research are clear. A systematic program of studies is required, which manipulate individual dosage characteristics while keeping others constant. Specifically, in relation to vocabulary, it would be beneficial to explore any potential interaction (a) between age and number of exposures/dose, (b) between object referent variability and dose, and (c) between a range of learning intervals and dose. The potential to leverage spacing effects to maximize efficiency appears promising (for both morphosyntax and vocabulary), but more work is needed, particularly in the area of within-session spacing. Finally, we suggest the application of a dose escalation design to morphosyntax research (used by Storkel et al., 2017, for vocabulary), while controlling for dose frequency. To facilitate this research, we recommend the development of a minimum data set of agreed outcome measures across the discipline, together with the more widespread adoption of open data practices. This would allow data pooling and meta-analyses to be conducted, enabling the consideration of the relative contribution of different dosage characteristics on intervention effects, and to identify the optimal dosage characteristics with which to efficiently, effectively, and ethically intervene to make a difference to the lives of individuals with DLD.

Acknowledgments

The systematic review was part of the work of the Cost Action IS1406 network entitled *Enhancing Children's Oral Language Skills Across Europe and Beyond—A Collaboration Focusing on Interventions for Children With Difficulties Learning Their First Language.* The network was supported by COST (European Cooperation in Science and Technology), funded by the European Union (Grant COST 106/14). This work was also supported in part by an National Health and Medical Research Council fellowship to C. M. (Centre of Research Excellence in Child Language 1023493). The authors wish to thank Virve Vihman and Marta Valdes for their assistance with screening and Anna Ceroni and Lorna Bateman for their assistance in updating the final search.

References

Aguilar, J., Plante, E., & Sandoval, M. (2018). Exemplar variability facilitates retention of word learning by children with specific

language impairment. *Language, Speech, and Hearing Services in Schools, 49*(1), 72–84. https://doi.org/10.1044/2017_LSHSS-17-0031

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/10.1044/1092-4388(2006/068)

Ambridge, B., Theakston, A., Lieven, E., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21*(2), 174–193. https://doi.org/10.1016/j.cogdev.2005.09.003

Balthazar, C. H., & Scott, C. M. (2018). Targeting complex sentences in older school children with specific language impairment: Results from an early-phase treatment study. *Journal of Speech, Language, and Hearing Research, 61*(3), 713–728. https://doi.org/10.1044/2017_JSLHR-L-17-0105

Bellon-Harn, M. L. (2012). Dose frequency: Comparison of language outcomes in preschool children with language impairment. *Child Language Teaching and Therapy, 28*(2), 225–240. https://doi.org/10.1177/0265659012445936

Bellon-Harn, M. L., Byers, B. A., & Lappi, J. (2014). Treatment intensity. *Communication Disorders Quarterly, 35*(4), 226–236. https://doi.org/10.1177/1525740114524051

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Botting, N., & Marshall, C. (2017). Domain-specific and domain-general approaches to developmental disorders. In L. Centifanti & D. Williams (Eds.), *The Wiley handbook of developmental psychopathology* (pp. 139–159). Wiley. https://doi.org/10.1002/9781118554470.ch6

Brandel, J., & Frome Loeb, D. (2011). Program intensity and service delivery models in the schools: SLP survey results. *Language, Speech, and Hearing Services in Schools, 42*(4), 461–490. https://doi.org/10.1044/0161-1461(2011/10-0019)

Broomfield, J., & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomized control trial. *International Journal of Language & Communication Disorders*, *46*(6), 628–640. https://doi.org/10.1111/j.1460-6984.2011.00039.x

Cepeda, N., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology, 24*(2), 237–255. https://doi.org/10.1044/2015_AJSLP-14-0105

Desmottes, L., Meulemans, T., Patinec, M. A., & Maillart, C. (2017). Distributed training enhances implicit sequence acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 60*(9), 2636–2647. https://doi.org/10.1044/2017_JSLHR-L-16-0146

Eadie, P., Morgan, A., Ukoumunne, O., Ttofari Eecen, K., Wake, M., & Reilly, S. (2014). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology, 57*(6), 578–584. https://doi.org/10.1111/dmcn.12635

Fey, M., & Finestack, L. (2009). Research and development in children's language intervention: A 5-phase model. In R. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513–531). Psychology Press.

Fey, M., Yoder, P., Warren, S., & Bredin-Oja, S. (2013). Is more better? Milieu communication teaching in toddlers with intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 56*(2), 679–693. https://doi.org/10.1044/1092-4388(2012/12-0061)

Glogowska, M., Roulstone, S., Enderby, P., & Peters, T. (2000). Randomised controlled trial of community based speech and language therapy in preschool children. *BMJ, 321*(7266), 923–923. https://doi.org/10.1136/bmj.321.7266.923

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)

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weberb, 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

Higgins, J., Altman, D., Gotzsche, P., Juni, P., Moher, D., Oxman, A., Savović, J., Schulz, K. F., Weeks, L., Sterne, J. A. C., & Cochrane Statistical Methods Group, Cochrane Bias Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343,* d5928–d5928. https://doi.org/10.1136/bmj.d5928

Jacoby, G., Lee, L., Kummer, A., Levin, L., & Creaghead, N. (2002). The number of individual treatment units necessary to facilitate functional communication improvements in the speech and language of young children. *American Journal of Speech-Language Pathology, 11*(4), 370–380. https://doi.org/10.1044/1058-0360(2002/041)

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*(1), 138–149. https://doi.org/10.1086/374692

Justice, L. M., Logan, J., Jiang, H., & Schmitt, M. B. (2017). Algorithm-Driven Dosage Decisions (AD 3): Optimizing treatment for children with language impairment. *American Journal of Speech-Language Pathology, 26*(1), 57–68. https://doi.org/10.1044/2016_AJSLP-15-0058

Law, J., & Conti-Ramsden, G. (2000). Treating children with speech and language impairments. *BMJ, 321*(7266), 908–909. https://doi.org/10.1136/bmj.321.7266.908

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder. *Journal of Speech, Language, and Hearing Research, 47*(4), 924–943. https://doi.org/10.1044/1092-4388(2004/069)

Law, J., Garrett, Z., & Nye, C. (2005). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Campbell Systematic Reviews, 1*(1), 1–85. https://doi.org/10.4073/csr.2005.5

Law, J., McKean, C., Murphy, C. A., & Thordardottir, E. (Eds.) (2019). *Managing children with developmental language disorder: Theory and practice across Europe and beyond.* Routledge. https://doi.org/10.4324/9780429455308

McKean, C., Frizelle, P., Tolonen, A. K., Kunnari, S., Murphy, C. A., Saldana, D., Law, J., & Tulip, J. (2017). Interventsions for children with primary language impairment: A COST IS1406 systematic review of the influence of dosage in interventions addressing vocabulary, phonology and morphosyntax outcomes.

McKean, C., Gerrits, E., Tulip, J., & Tolonen, A. (2019). Service delivery for children with language disorders across Europe and beyond. In J. Law, C. McKean, C. A. Murphy, & E. Thordardottir (Eds.), *Managing children with developmental language disorder: Theory and practice across Europe and beyond* (pp. 84–109). Routledge. https://doi.org/10.4324/9780429455308-5

Meyers-Denman, C. N., & Plante, E. (2016). Dose schedule and enhanced conversational recast treatment for children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 47*(4), 334–346. https://doi.org/10.1044/2016_LSHSS-15-0064

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1–9. https://doi.org/10.1186/2046-4053-4-1

Morgan, A., & Vogel, A. (2008). Intervention for childhood apraxia of speech. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD006278.pub2

Nye, C., Foster, S., & Seaman, D. (1987). Effectiveness of language intervention with the language/learning disabled. *Journal of Speech and Hearing Disorders, 52*(4), 348–357. https://doi.org/10.1044/jshd.5204.348

OCEBM Levels of Evidence Working Group. (2011). *The Oxford levels of evidence 2*. A. K. T. https://www.cebm.net/index.aspx?o=5653

Plante, E., Mettler, H. M., Tucci, A., & Vance, R. (2019). Maximizing treatment efficiency in developmental language disorder: Positive effects in half the time. *American Journal of Speech-Language Pathology, 28*(3), 1233–1247. https://doi.org/10.1044/2019_AJSLP-18-0285

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., Lieser, A. M., & Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23*(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038

Proctor-Williams, K., & Fey, M. E. (2007). Recast density and acquisition of novel irregular past tense verbs. *Journal of Speech, Language, and Hearing Research, 50*(4), 1029–1047. https://doi.org/10.1044/1092-4388(2007/072)

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37*(1), 106–121. https://doi.org/10.1044/jshr.3701.106

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

Roberts, M., & Kaiser, A. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*(3), 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Ruggero, L., McCabe, P., Ballard, K. J., & Munro, N. (2012). Paediatric speech-language pathology service delivery: An exploratory survey of Australian parents. *International Journal of Speech-Language Pathology, 14*(4), 338–350. https://doi.org/10.3109/17549507.2011.650213

Schmitt, M. B., Justice, L. M., & Logan, J. A. R. (2016). Intensity of language treatment: Contribution to children's language outcomes. *International Journal of Language & Communication Disorders, 52*(2), 155–167. https://doi.org/10.1111/1460-6984.12254

Sciberras, E., Westrupp, E., Wake, M., Nicholson, J., Lucas, N., Mensah, F., Gold, L., & Reilly, S. (2014). Healthcare costs associated with language difficulties up to 9 years of age: Australian population-based study. *International Journal of Speech-Language Pathology, 17*(1), 41–52. https://doi.org/10.3109/17549507.2014.898095

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. The Psychological Corporation.

Siegmüller, J., Baumann, J., & Höppe, L. (2017). Language intervention in developmental language disorder using modelling—A controlled study based on experimental data. L.O.G.O.S. *Interdisziplinair, 25*(4), 253–263. http://doi.org/10.7345/prolog-1704253

Smith-Lock, K., Leitao, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal Of Language & Communication Disorders, 48*(3), 265–282. https://doi.org/10.1111/1460-6984.12003

Smith-Lock, K., Leitao, S., Lambert, L., Prior, P., Dunn, A., Cronje, J., Newhouse, S., & Nickels, L. (2013). Daily or weekly? The role of treatment frequency in the effectiveness of grammar treatment for children with specific language impairment. *International Journal of Speech-Language Pathology, 15*(3), 255–267. https://doi.org/10.3109/17549507.2013.777851

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

Sugden, E., Baker, E., Munro, N., Williams, A. L., & Trivette, C. M. (2018). Service delivery and intervention intensity for phonology-based speech sound disorders. *International Journal of Language & Communication Disorders, 53*(4), 718–734. https://doi.org/10.1111/14606984.12399

Warren, S., Fey, M., & Yoder, P. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Riverside.

Zeng, B., Law, J., & Lindsay, G. (2012). Characterizing optimal intervention intensity: The relationship between dosage and effect size in interventions for children with developmental speech and language difficulties. *International Journal of Speech-Language Pathology, 14*(5), 471–477. https://doi.org/10.3109/17549507.2012.720281