Title	The influence of quantitative intervention dosage on oral language outcomes for children with developmental language disorder: a systematic review and narrative synthesis
Authors	Frizelle, Pauline;Tolonen, Anna-Kaisa;Tulip, Josie;Murphy, Carol-Anne;Saldana, David;McKean, Cristina
Publication date	2021-01-19
Original Citation	Frizelle, P., Tolonen, A.-K., Tulip, J., Murphy, C.-A., Saldana, D. and McKean, C. (2021) 'The Influence of Quantitative Intervention Dosage on Oral Language Outcomes for Children With Developmental Language Disorder: A Systematic Review and Narrative Synthesis', Language, Speech, and Hearing Services in Schools, 52(2), pp. 738-754. doi: 10.1044/2020_LSHSS-20-00058
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://pubs.asha.org/doi/10.1044/2020_LSHSS-20-00058 - 10.1044/2020_LSHSS-20-00058
Rights	© 2021 American Speech-Language-Hearing Association. This is the authors' accepted manuscript of an article that has been published in its final definitive form in Language, Speech, And Hearing Services In Schools, https://doi.org/10.1044/2020_LSHSS-20-00058
Download date	2024-04-26 05:48:07
Item downloaded from	https://hdl.handle.net/10468/11375

The influence of quantitative intervention dosage on oral language outcomes for children with Developmental Language Disorder: a systematic review and narrative synthesis.

Pauline Frizelle, [1, 6] Anna-Kaisa Tolonen [2, 6] Josie Tulip [3, 6], Carol-Anne Murphy [4, 6], David Saldana [5, 6] and Cristina McKean [3, 6]

[1] (corresponding author) Department of Speech and Hearing Sciences, University College Cork, Republic of Ireland, 00353 21 4901649, p.frizelle@ucc.ie

[2] Department of Logopedics, University of Oulu, Oulu, Finland

[3] School of Education, Communication & Language Sciences, Newcastle University, Newcastle, UK

[4] School of Allied Health, University of Limerick, Limerick, Ireland

[5] Departamento de Psicología Evolutiva y de la Educación, University of Seville, Seville, Spain

[6] European COST Action 1406 Management Committee

Conflict of Interest: The authors declare that no competing interests existed at the time of publication.

Funding Statement: 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 NHMRC fellowship to CMK. (Centre of Research Excellence in Child Language 1023493).

RUNNING HEAD: Quantitative dosage manipulation in language interventions for DLD

Key words: Dose, Dosage, Intervention, Intensity, DLD,

Abstract

Purpose: 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 morpho-syntax 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, RCT or cohort analytic studies, published in any language between January 2006 to May 2020. Included papers reported on participants with DLD (M= 3-18 years); oral language interventions with phonology, vocabulary or morpho-syntax outcomes; and experimental manipulation or statistical analysis of any quantitative aspect of dosage. Studies were appraised using the Cochrane risk-of-bias tool.

Results: 244 papers 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 papers reported phonological outcomes, three vocabulary and eight morpho-syntax. Dose frequency was the most common characteristic manipulated.

Conclusion: Research is in its infancy and significant further research is required to inform SLPs 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 (2/3x per week; ~2mins) and less frequent, long sessions (1x per week; ~ 20mins) have yielded the best outcomes when

composite language measures have been used, however replication and further research is required before clinicians can confidently integrate these findings into clinical practice.

Introduction

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 Randomised Controlled Trials (RCTs), that interventions can and do effect meaningful change for children and young people with speech, language and communication disorders (Law et al., 2005; 2004; Broomfield & Dodd, 2011; 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 is required for positive outcomes to be achieved. Whilst 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 synthesises 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 neuro-developmental 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 speech and language therapy 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 stigmatisation 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 are 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 behavioural 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 i.e. *dose x dose frequency x 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 is 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 non-linear 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 hours 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 6.2 hours, 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 versus 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 hours 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. one to 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. (1987) reported that session lengths shorter than 90 minutes yielded higher effect sizes than longer sessions. Jacoby et al. (2002) studied the number of individual 'treatment units' (i.e. 15-minute 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 hours 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 morpho-syntax 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 are 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 (post intervention and at follow up) is tapping the level of consolidation or decay that has occurred.

Current study

Since the publication of the Zeng et al. (2012) review, a number of studies which directly manipulate aspects of intervention dosage have been published. In order to inform evidence-based service delivery, commissioning and policy, this paper 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 analysed. 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, Non-randomized 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 morpho-syntax, and reports findings separately to determine whether optimal dosage characteristics differ across domains.

We addressed the following questions

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? Do they differ across domains?
3) What gaps remain in the evidence?

Method

This systematic review was registered with PROSPERO (ID=CRD42017076663): and is one of a series completed as part of European COST Action 1406. Action 1406 focussed 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 Procedures

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 terminology used in each study were extracted and are presented in Table 1. 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 to December 2015. Three updated searches were then completed; the first to include studies published between January 2016 and October 2017; the second between November 2017 and May 2019 and the third between June 2019 and May 2020. Search terms were developed through discussion between authors and consultation with a research librarian. The search string is published in our pre-registration (McKean et al. 2017). Reference lists of all papers included on full text and relevant systematic reviews were also hand searched for any additional papers.

Inclusion / Exclusion criteria

Included papers met the following criteria:

- Research design – either 1) RCTs; 2) Quasi-experimental designs (non-random assignment) with an element of control; 3) Cohort analytic 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 - May 2020.
- Participants with a mean age ≥3 years and ≤18 years,
- Participants identified as having a) developmental language disorder or an equivalent term such as primary language impairment or specific language impairment and b) difficulties on at least one oral language assessment (vocabulary, morpho-syntax 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. ADHD, dyslexia) were included. Children with childhood apraxia of speech (CAS) 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 morpho-syntax

- Experimentally manipulated or statistically analysed an aspect of either dose, dose frequency or cumulative intervention intensity, whilst keeping other variables constant.

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

Paper 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 papers evaluating interventions for children with DLD across all language domains (vocabulary, phonology, morpho-syntax and pragmatics). These papers 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, (CAM & DS for the initial search and CAM & PF 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 1198 papers. All non-English papers at this and subsequent stages were considered by either author AKT (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 PF after translation.

Stage 2: To identify those specifically relevant to vocabulary, phonology or morpho-syntax outcomes considered in this review, two independent reviewers (PF and AKT) screened 100% of the papers included after stage 1 on title and abstract. Agreement rate of 93%. This yielded 698 papers.

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 papers emerging from stage 3 to identify those with a specific focus on dosage characteristics, which were experimentally manipulated or statistically analysed, 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 papers 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 Figure 1 for PRISMA flowchart.

----Insert Figure 1 about here---

Data Extraction

The first author extracted the following data from the papers 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 author (PF and CMK) 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 paper is shown in Figure 2. ----Insert Figure 2 about here---

Results

Thirty-nine papers reported on studies in which dosage was experimentally or statistically manipulated and 13 of these focused on the quantitative aspects of dosage. These 13 papers came from 8 journals, 9 of which were in the English language and 1 which was in German. Of the 13 papers, 3 (23%)[1] were from the Journal of Speech, language and Hearing Research; 3 (23%) from Language Speech and Hearing Services in schools; 2 (15%) from the American Journal of Speech-Language pathology; and 1 (8%) from each of the following journals; the International Journal of Language and communication disorders; the 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%); 1 in the United Kingdom (8%); and 1 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 (Med = 25) and children with DLD had an average age range from 3;11 to 12;01 years. See Table 1 in supplemental materials.

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 1) a composite score of below 1 standard deviation on a standardized language measure such as the Clinical Evaluation of Language Fundamentals (CELF 4) or the Structured Photographic Expressive Language Test (SPELT-3) 2) non-verbal IQ scores within 1 standard deviation of

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

the norm on a test of cognitive functioning 3) hearing within the normal range (shown by passing a pure tone hearing screening) and 4) no known neurological, social-emotional or psychiatric disorders. With respect to the two studies that took place in the community [4,8] 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 an MLU of 3 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 i.e. a standard score of 80 [1] and 75 [6, 7]. No evidence of speech impairment was specified in 3 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?

No studies manipulating quantitative dosage characteristics were identified which focussed 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 3 studies (23%) specific to vocabulary, and 8 (62%) specific to morphosyntax. Lastly, there were two studies with omnibus outcomes (15%) in which dosage was statistically manipulated. Figure 2 summarises 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 3 of the 5 focussed on morphology; only 2 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, 2 recruited from a single setting, 2 recruited from multiple settings and one was unclear. Randomization was always with respect to the treatment condition. Each of the trials were 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 papers, with 4 of the 5 focussed on morphology. In broad terms these studies were non-equivalent group designs although in some studies there was an attempt to match across variables, such as non-verbal 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 4 of the 5 studies and similarly there was no reported power calculation for 4 of the 5 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? Do they differ across domains? And RQ 3. What gaps remain in the evidence?

The following provides a narrative summary of the findings of the papers identified, organised by outcome (Vocabulary, Morpho-syntax, Phonology, Omnibus Measure). In each section we

report on each of Warren and colleagues quantitative dosage components in turn (dose, dose frequency, total intervention duration and cumulative treatment intensity), identifying whether evidence exists, summarising the findings and describing the level of confidence in those findings. Table 1 also summarises the data extracted from the papers.

Vocabulary

For this domain studies manipulating dose (n = 1) [Study 12 – Table 1][2], 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 and colleagues in their 2017 paper [12], in the context of interactive book reading using a novel escalation design methodology. Twenty-seven children with DLD (M = 5;08 years) 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, 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 post intervention), 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

[2] Numbers in square brackets indicate the study number in Summary Table 1

399 the outcome measure, for each treatment intensity. Children showed the ability to define the
400 most words (n=5) following 36 exposures and word learning began to diminish at 48 exposures.
401 In addition, results from the naming task indicated 36 exposures to be the optimal dose (with
402 86% of children responding). A decrease in treatment response was again evident as the
403 number of exposures increased to 48. The finding that children's optimal performance was
404 following 36 exposures supports the theory that there is a critical minimum number of
405 exposures required to allow adequate encoding of words to occur. On the other hand,
406 diminishing returns at 48 exposures may be in keeping with deficient-processing theories of
407 learning, which suggest that learning effectiveness is dependent on the degree of attention
408 directed towards what is being learned. A reduction in attention is thought to occur as what is
409 being learned becomes overly familiar, and while this has previously been discussed in relation
410 to massed practice (Cepeda et al., 2006), it could also occur in the context of too many word
411 exposures within a given time period.
412 While this study is highly innovative, in the application of an escalation design to the field of
413 language learning, there are a number of points to note with respect to dosage. The number of
414 treatment sessions ranged between 10 and 20 and were given 2 to 3 times a week. Therefore,
415 the total intervention duration is a confound as it was not constant for each dose. It is also
416 noteworthy that children's response to treatment was very low at all exposures, when using the
417 definition task as a measurement of learning. Only 43% of children responded at optimal
418 dosage and only 5 treatment words were correctly defined. A more optimistic result was
419 evident using the naming task as the outcome measure, with 86% of children responding at
420 optimal dose and 60% responding at a minimum of 12 exposures. The authors posit that
421 semantic knowledge is measured by the definition task and that the naming task is a measure
422 of phonology. We suggest this may be an overly conservative approach to the measurement of
423 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 and colleagues found that with high variability in the referent presented, preschool children with DLD had the ability to learn 3 new words having been exposed to them 18 times over 3 sessions and asked to name the items once per session. However, learning was measured through a comprehension probe in the Aguilar 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 (2017) where the outcomes were administered immediately post intervention, the Aguilar retention outcome measure was administered at follow up (6 weeks post intervention), 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;06 years) and 24 language matched control children were taught four novel verbs, using a dual morphological frame (*Look its dacking, see it dacks*) modelled through play activities. The manipulation of the number of exposures along with the spacing of the treatment sessions resulted in four experimental conditions 1) massed 12, with 12 exposures on a single day; 2) massed 18 with 18 exposures on a single day; 3) spaced 12, with 12 exposures spread over 4 days (3 each day); 4) spaced 18, with 18 exposures spread over 4 days (either 4 or 5 each day). Outcome measures were carried out directly following, and one week post intervention, 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 3 objects) which one was *verb*ing?. Post-test 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 i.e. 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. Firstly, 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, whilst 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 5;0 and 6;02 years of age) were give 36 exposures to two word sets, 60 words in total consisting of nouns, verbs and adjectives. For the first word set a 6 dose x 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 x 9 dose frequency or 9 dose x 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 post treatment (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 4 words). Only 40% of the words that were correctly defined at the end of treatment were retained 5 /6 days later and only 30% of words were retained at the 21 day timepoint. Word learning was however consistent with their previous study, in that children defined an average of 4 -5 words correctly immediately post intervention 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 and colleagues (2018) was due to the timing of the post intervention outcome measure (6 weeks post), and that this potentially served as a consolidation period. Results from Storkel et al., (2019) 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 et al. (2016) and Balthazar and Scott (2018) (presented later in this review) 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 .

Morpho-syntax.

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 (.19 per minute), at an intervention level density (.47 per minute) or no recasts, over a period of 5 sessions. The recast exemplars were distributed equally across the 6 verbs i.e. three in the low-density recast condition and three in the high-density recast condition. Low density recasting translated as 2 per verb in each of the five sessions (30 recasts) and high density as 5 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 post intervention.

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 (2016) study (see below), is higher than that reported here (1.25 per minute v's .47 per minute). 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 study in which the treatment duration was equivalent (150 minutes) 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 to 44 days). Interestingly, when the authors tested the relationship between length of time, in days, that it took to complete the five sessions and 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 5 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 towards 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 a more recent study carried out by Plante and colleagues (2019) [13] reported on within-session manipulation of the dose density of enhanced conversational recasting. An additional study distinction was that Plante and colleagues kept overall intervention duration constant. Twenty children with DLD (4 – 5;11 years) 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, 3rd person – s, Aux is and possessive. Half of the group heard the recasts over a 30 minute period (1 recast every 1.25 minutes) and

548 the other half heard them over a 15 minute period (1 recast every 38s) while maintaining
549 session length at 30 minutes. The study was designed to ascertain which of the two treatments
550 was more effective and efficient and how many children generalised their targets in that
551 timeframe. Children's learning was measured through the use of generalisation and retention
552 morpheme probes. The former were administered before each Monday, Wednesday and Friday
553 session and the latter were given 6 weeks post intervention. Results indicated that the majority
554 of children showed a strong treatment effect. However there were no significant differences
555 between the two treatment conditions on any of the outcome measures (probe or spontaneous
556 performance, number of treatment responders, follow up performance). In addition there was
557 a significant relationship between children's performance at the end of treatment and at follow
558 up. The authors conclude that within-session high density dose delivery does not offer any
559 advantage over a lower density delivery, if dose and overall intervention duration are constant.
560 However, the sample size was again small (n = 10). They also note that although children
561 retained the gains that they made in treatment they did not show any independent improvement
562 in target morpheme use following treatment. Findings from this study differed from Proctor-
563 Williams and Fey (2007) in that the treatment itself was effective but given the overall dose
564 differences (30 recasts v's between 528 – 600) this is not surprising. An important difference
565 between the two studies was how the dose density manipulation was implemented. In Proctor-
566 Williams and Fey (2007) the low density condition was distributed across the 5 sessions but
567 the high density condition was implemented in sessions 4 and 5 only. Therefore the density
568 manipulation was achieved by altering the number of sessions in which the recasts were given
569 (2x5 sessions, 5x2 sessions) and as result dose frequency was a confound. In contrast, Plante
570 and colleagues (2019) altered the session length in which an equal number of recasts were
571 given (24 recasts in 15 minutes v's 24 recasts in 30 minutes) and this was constant across all
572 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 study to have an effect and to reveal any differences. In contrast Plante and colleagues (2019) 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 5 of the 8 studies within the morphosyntax domain. Bellon-Harn and colleagues (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 so total number of treatment hours (8 hours) 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 immediately post-treatment measures were taken using language sample analysis and probes designed to elicit targets (such as the use of auxiliary, copula, third person singular *3s*). 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 (6 per group) and consequently these results

598 should be interpreted with some caution. It is also worth noting that there is considerable
599 variation across studies as to what is considered spaced or concentrated in treatment delivery
600 and how this interacts with the total duration of the intervention. Indeed, even the more
601 concentrated treatment in this study is delivered over a six-week period. In addition it is
602 noteworthy that although the authors suggest that dose is kept constant in this study, they
603 acknowledge that in a scaffolded-language therapy, there is no predetermined script or target.
604 As a result, dose was not closely controlled i.e. the frequency of linguistic forms within each
605 cloze procedure, expansion, and model. The authors suggest that dose for both treatment
606 schedules was high and may therefore mask any dose frequency effect. It may also be the case
607 that a total of 8 hours of intervention, which was constant across conditions, was not so lengthy
608 as to reach the point of diminishing returns, which would potentially result in a smaller effect
609 for the more frequent schedule.
610 The second study in which dose frequency was manipulated with respect to morphosyntax, was
611 carried out by Smith-Lock and colleagues (2013) [11]. The study (which included a larger
612 sample than that by Bellon-Harn (2012)) compared the effectiveness of two different dose
613 frequencies in relation to a school-based treatment of expressive grammar. Five-year old
614 children with DLD were assigned to either 8 one-hour sessions of treatment given over an 8-
615 week period (a spaced treatment), or 8 one-hour sessions given over an 8-day period (a
616 concentrated treatment). Once again total intervention duration was not controlled. Therapeutic
617 techniques were integrated into naturalistic play sessions and included explicit instruction,
618 focused stimulation, recasting and imitation. Treatment targets were individualised and
619 included accurate use of past and present tense, pronouns and possessives. Learning was
620 measured on The Grammar Elicitation Test (Smith-Lock et al., 2013) immediately and 8 weeks
621 post intervention. While results showed significant improvement in the group that received the
622 spaced treatment, (relative to the same time period prior to treatment), this was not the case for

623 the concentrated treatment group. Single-subject analyses indicated that 46% of children who
624 received the spaced schedule and 17% of those who received the concentrated schedule showed
625 a significant treatment effect. This result is in keeping with previous findings indicating
626 advantages for spaced learning but is contrary to results by Bellon-Harn and colleagues (2012).
627 Of interest is the fact that the number of therapy hours is equivalent for both studies, however,
628 in addition to the sample size, a notable difference between the two studies is the total
629 intervention duration. In Smith-Lock et al. (2013) the concentrated intervention takes place
630 over a relatively short period (8 days). The spaced intervention duration (8 weeks) is however
631 quite similar to the concentrated intervention duration in the Bellon-Harn (2012) study (6
632 weeks). We might suppose that, given a total number of therapy hours that is effective and
633 equal in both conditions, differences only emerge between spaced and concentrated treatment
634 schedules for children with DLD, when the time frame between the beginning and end of the
635 treatment is significantly shorter for one condition than the other (e.g. one-week v's 8 weeks).
636 It is also the case that while Smith-Locke and colleagues (2013) provided teachers with scripts
637 and detailed activity plans, dose was not controlled for in this study. Research suggests that
638 dose frequency effects (i.e. number of sessions) can be mitigated if dose per session is high
639 (Fey et al. 2013) but the authors do not give us any sense of dose in this study. Additionally,
640 there are a number of treatment techniques used in both aforementioned studies, such that dose
641 in relation to each technique is likely to be somewhat diluted and to vary between each
642 treatment session.
643 Meyers-Denman and colleagues (2016) [5] is the third included study to examine the effects
644 of treatment dose frequency on grammatical morpheme remediation in young children with
645 DLD. Again the sample size was small at eight per group. Using enhanced conversational
646 recasts, treatment was given in both concentrated (3 x 10 minute sessions within a 4 hour
647 period, 5 days a week) and spaced conditions (1 x 30 min session 5 times a week). The

concentrated condition resulted in 15 ten-minute sessions, while the spaced condition resulted in 5 thirty-minute 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 ½ hours) and total intervention duration were controlled. Children's learning was measured through a play-based generalisation probe, in which they were required to use the target morphemes with untreated lexical items. Pre- post- assessments 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 period of 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 minutes, 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 not 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). Adverbial, object complement and relative clauses were taught following a once or twice weekly treatment protocol. Total intervention duration was nine weeks and session length ranged between 40 and 60 minutes, resulting in total intervention time of 6-9 hours for the once weekly condition and 12-18 hours 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)

673 in the once weekly condition and 28 items per session (502 items in total) in the twice weekly
674 condition. Stimuli presentation was through modelling, repetition and manipulation of a
675 complex sentence with scaffolding and clinician feedback. Primary outcome measures were
676 sentence probes administered before, during and after treatment as well as standardized
677 language tasks reflecting a broad range of oral and written language. Interestingly while
678 treatment was effective as measured by the sentence production probes, there was no advantage
679 for the higher dosage group on any oral language measure. This finding was contrary to the
680 authors' hypothesis and they suggest a number of possible explanations for this result: given
681 that 3 sessions were devoted to each sentence type, even in the once weekly group, there may
682 have been no advantage to the additional sessions; they acknowledge that treatment
683 maintenance was not examined; and they question whether the difference in the two dose
684 frequencies was sufficient to yield a difference. We suggest that the findings of this study are
685 in keeping with previous studies and support the notion that high dose reduces the need for
686 high intervention frequency. However, it is important to consider maintenance effects.
687 An additional study in which dose frequency was statistically analysed in the treatment of
688 complex syntax was carried out by Siegmüller and colleagues (2017) [10]. Intervention
689 outcomes were children's ability to use subordinate clause structures. Intervention dose form
690 was implicit and carried out in 6 steps which included 1) intensive modelling of a) verbs and
691 their associated arguments and b) different grammatical subcomponents of the sentence; 2)
692 questions eliciting the production of the main clause 3) modelling expansions of the main
693 clause to subordinate clause structures. Children were assigned to different steps depending on
694 their pre-test performance and treatment was discontinued when the child reached step 5
695 (showed the ability to use subordinate clauses). To analyse the effects of dose frequency on the
696 outcome, the children were divided into two groups: those who had therapy once weekly and
697 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 analysing 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 et al., 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 which 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 and colleagues (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 v's 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 co-ordinate and sub-ordinate 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 v's 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 to 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 papers investigated how dosage characteristics interact to contribute to children's global language outcomes [4, 9]. These papers are based on a unique study that used data from a large clinically identified sample of children with DLD (n = 233), who were receiving

748 language treatment within the U.S. public schools system, over an academic year. Natural
749 variations in treatment intensity data, allowed the authors to examine the impact of different
750 aspects of dosage on children's language outcomes, as well as the extent to which treatment
751 outcomes vary as a function of one or more dosage parameters. Treatment centred on one of 9
752 language focused-targets and outcomes were the CELF-4 (Semel et al., 2003) core language
753 scores and the picture vocabulary subtest from the Woodcock Johnson III Tests of
754 Achievement (Woodcock et al., 2001). It is important to note that in both papers the term dose
755 is defined and operationalised differently to Warren et al., (2007). Here it is defined as the total
756 amount of time spent addressing any one of nine language-focused targets, in contrast with the
757 now more usual definition of the number of administered teaching episodes in a given
758 intervention session. Hence it is a proxy measure with less specificity and accuracy than a
759 measure of dose and it precludes a clear definition of dose form. On the other hand, this
760 approach allows an examination of dosage effects in a much larger sample than found in other
761 intervention studies and scrutinises dosage schedules used in real-world clinical contexts.
762 Using this approach Schmitt and colleagues (2016) [9] examined the extent to which dose, dose
763 frequency, and the interplay between the two were associated with language gains over the
764 school year. Using structured equation modelling the results showed that children receiving
765 low dose /high frequency (intervention sessions of approximately 2 minutes, at a rate of 2 to
766 3 times per week), or high dose / low frequency (intervention sessions of approximately 20
767 minutes, at a rate of 1 per week or fortnight) had better outcomes than those receiving high
768 frequency/ high dose (20 minutes, 2/3 times weekly), high frequency/ average dose (12
769 minutes, 2/3 times weekly) or low frequency/ low dose treatment (2 minutes, 1 per week or
770 fortnight). It must be noted when considering clinical application that the total intervention
771 duration here was a school year and not discrete 'blocks' of therapy found in many healthcare
772 systems (McKean et al., 2019). Therefore both 'optimal' conditions have relatively high total

intervention hours (low dose/high frequency: 2mins x 3 sessions x 28 weeks = 168 minutes (2 hours 48 minutes); high dose/low frequency: 20 minutes x 1 session x 28 weeks = 560 minutes (9 hours 20 minutes per year).

Justice and colleagues (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 dataset. Outcomes were retrospectively analysed with respect to dose, dose frequency (intensity) and cumulative intensity of therapy. Multi-level modelling 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 SLP's to calculate optimal language outcomes (defined as an increase of .6*SD* units). Therefore if a clinician knows the session frequency (e.g. once weekly) and number of weeks s/he can work with a child over the course of the school year (e.g. 25), using baseline language scores and .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 which 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 versus narrative macro-structure). In addition, although the diversity of goals

798 and SLP practice in the schools does suggest that a range of dose forms can be effective, the
799 ways in which targets were addressed by clinicians is likely to have varied considerably.
800 Finally, there is a need to better understand interactions between child-level factors such as
801 language severity and treatment intensity. The literature is unclear regarding whether children
802 with more severe DLD might benefit from higher frequency interventions or from those in
803 which learning opportunities are more spaced, thereby facilitating consolidation and enhanced
804 attention.

805 Discussion

806 In this study we aimed to ascertain to what degree the quantitative aspects of dosage have been
807 specifically manipulated in intervention studies with children with DLD, in which there were,
808 phonology, vocabulary or morphosyntax outcomes. In addition, we aimed to identify optimal
809 quantitative dosage characteristics in each of these domains; to highlight gaps in the literature;
810 and difficulties in interpreting the evidence. The dominant finding of the review is the lack of
811 intervention studies across domains, in which quantitative aspects of dosage have been
812 experimentally or statistically manipulated for children with DLD. In addition, a number of
813 studies included in the review have been carried out with particularly small sample sizes,
814 causing us to call into question the validity of these findings. Consequently, there is a
815 significant need for further research to inform clinical practice. Significantly, there were no
816 studies with phonological outcomes in this population of children in which quantitative aspects
817 of dosage were manipulated. It is possible that the literature relating to children with Speech
818 Sound Disorder (SSD) can be directly applied to DLD. However, this has not been tested and,
819 given the meta-linguistic skills and abstract concepts invoked in many phonological
820 interventions it would seem likely that modifications in dosage and/or other aspects of the
821 interventions would be required and should be tested in empirical studies. Given high

822 comorbidity between DLD and SSD such work would likely have significant clinical impact
823 (Eadie et al., 2014).
824 *Vocabulary*: The finding that there were only three studies in the vocabulary domain, in which
825 quantitative aspects of dosage were manipulated, again highlights the dearth of research in this
826 area. Hence there is limited evidence on which practitioners can draw, to inform the
827 implementation of interventions and advise managers and policy makers regarding optimal
828 dosage. The work by Storkel and colleagues (2017), has been both pioneering in its use of an
829 escalation design, and unique in showing diminishing benefits following a specific number of
830 word exposures. However, in many ways this important work represents a starting point from
831 which to grow research in this domain. Given that frequency and total intervention duration
832 were not constant for each dose, further work is required to determine whether this finding is
833 replicated under constant frequency or duration conditions. It is also significant to note the
834 differences in children's responses with respect to outcome measures (43%, word definition
835 versus 86%, word naming). Within word learning studies alone, outcomes can include forced
836 choice comprehension probes; naming; word definition; and synonym comprehension or
837 production tasks; all of which may use experimental or unfamiliar referents and which can
838 occur during intervention, immediately after or following a consolidation period. If we are to
839 build the necessary evidence upon which to base clinical decisions the use of consistent
840 outcome measures will be required to make meaningful cross study comparisons. In addition
841 the timing of outcome measures is central to how we interpret study findings. This is
842 highlighted in the work reported by Storkel and colleagues (2019), in which there was a 40%
843 drop in word learning a mere 5 to 6 days post intervention.
844 Based on the findings of their earlier study (2017) and reinforced by this most recent study
845 (2019), when measured with a naming or word definition task, 36-word exposures appear to
846 be the optimal dose for word learning in 5-6-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 and colleagues (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 papers 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 paper, 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 i.e. the spaced condition was over 4 days, rather than a period of weeks as in the Storkel et al., (2017) paper and in much clinical practice. Recent work by Storkel and colleagues (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 as well as 5 days post intervention. 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 (x9 doses) and the spaced condition was for 9 weeks (x4 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 (.19 per minute) and high (.47 per minute) density conditions, Proctor-Williams and Fey (2007) reported no improvement in irregular past tense production accuracy in the high density condition. This paper 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 operationalised 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 and colleagues (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 and colleagues (2017) for vocabulary has the potential to be informative for morphosyntax, while controlling for dose frequency. In clinical practice, dose is rarely operationalised and measured. When considering dosage characteristics clinicians use proxy measures such as the number of intervention hours given over a specific period of 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 and colleagues (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 centres 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 et al. (2013) and Siegmüller et al. (2017) support a distributed learning advantage, Bellon-Harn (2012), Meyers-Denman et al. (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 papers 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 (Meyers-Denman et al., 2016; Balthazar and Scott (2018) 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 minutes; 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 versus 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 papers 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 (~ 3 times per week) in which the focus on a specific language target is very short (2 minutes) ; or 'more and less often': less frequent sessions (~ weekly) in which specific goals are targeted

for longer periods (20 minutes). 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 two minutes, 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, i.e. 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 speech and language therapy are complex. To summarise 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 under-researched and while Plante and colleagues (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. Whilst the review has not been able to identify a 'magic number' for success it does suggest that simply delivering the number of intervention hours which 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 pre-determined 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 whilst keeping others constant. The potential to leverage spacing effects to maximise efficiency appears promising, but more work is needed. 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 so identify the optimal dosage characteristics with which to efficiently, effectively and ethically intervene to make a difference to the lives of individuals with DLD.

Acknowledgements

The authors wish to thank Virve Vihman and Marta Valdes for 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, 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

Bathazar, 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, Vol. 61, 713–728 http://doi.org/10.23641/asha.5923318

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

Baker-Ward, M. L., Byers, B. A., & Lappi, J. (2014). Treatment Intensity. *Communication Disorders Quarterly*, *35*(4), 226–236. http://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

Bishop, 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). John Wiley & Sons Ltd. 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

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 inter- vention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology*, *24*(2), 237–255.

1047 Cleave, P., Becker, S., Curran, M., Van Horne, A., & Fey, M. (2015). The Efficacy of Recasts in
1048 Language Intervention: A Systematic Review and Meta-Analysis. *American Journal Of*
1049 *Speech-Language Pathology*, *24*(2), 237-255. https://doi.org/10.1044/2015_ajslp-14-0105
1050 Desmottes, L., Meulemans, T., Patinec, M.A., & Maillart, C. (2017). Distributed training enhances
1051 implicit sequence acquisition in children with specific language impairment. *Journal of*
1052 *Speech, Language, and Hearing Research,* 60, 2636–2647.
1053 https://doi.org/10.1044/2017_JSLHR-L-16-0146
1054 Eadie, P., Morgan, A., Ukoumunne, O., Ttofari Eecen, K., Wake, M., & Reilly, S. (2014). Speech
1055 sound disorder at 4 years: prevalence, comorbidities, and predictors in a community cohort of
1056 children. *Developmental Medicine & Child Neurology*, *57*(6), 578-584.
1057 https://doi.org/10.1111/dmcn.12635
1058 Hess, M., Yoder, P., Warren, S., & Bredin-Oja, S. (2013). Is More Better? Milieu Communication
1059 Teaching in Toddlers With Intellectual Disabilities. *Journal Of Speech, Language, And*
1060 *Hearing Research*, *56*(2), 679-693. https://doi.org/10.1044/1092-4388(2012/12-0061)
1061 Kaiser, M., & Finestack, L. (2009). Research and development in children's language intervention: A 5-
1062 phase model. In R. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513-531).
1063 Psychology Press.
1064 Kelly, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts
1065 success? *Journal of Speech, Language, and Hearing Research*, 46, 56–67.
1066 https://doi.org/10.1044/1092- 4388(2003/005)
1067 Higgins, J., Altman, D., Gotzsche, P., Juni, P., Moher, D., & Oxman, A. et al. (2011). The Cochrane
1068 Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(oct18 2), d5928-
1069 d5928. https://doi.org/10.1136/bmj.d5928
1070 Mädebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S.L., Usler, E., Kueser, J.B., Souto, S.,
1071 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*, Vol. 62, 944–964. https://doi.org/10.23641/asha.7927112

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

Hadley, 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. http://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. http://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

McKean, C., Frizelle, P., Tolonen, AK., Kunnari, S., Murphy, CA., Saldana, D., Law, J., & Josie Tulip, (2017). Interventions for children with primary language impairment: A COST IS1406 systematic review of the influence of dosage in interventions addressing vocabulary, phonology and morphosyntax outcomes. PROSPERO CRD42017076663 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017076663

McKean, C., Gerrits, E., Tulip, J., Tolonen, A. (2019. Service delivery for children with language disorders across Europe and beyond. pp 84 – 109 In J. Law, C. McKean, CA. Murphy, E. Thordardottir (Eds.), *Managing Children with Developmental Language Disorder: Theory and Practice Across Europe and Beyond*. Routledge

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

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. http://doi.org/10.1044/2016_LSHSS-15-0064

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*. Oxford Centre for Evidence-Based Medicine. Retrieved from https://www.cebm.net/index.aspx?o=5653

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

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, 1233–1247. https://doi.org/10.1044/2019_AJSLP-18-0285

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. http://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, 106–121. https://doi.org/10.1044/jshr.3701.106

Rispoli, M., Hadley, P. A., & Holt, J. K. (2009). Stalled Sentence Growth in Children With ASD. *Journal of Autism and Developmental Disorders*, *39*(1), 22–32. https://doi.org/10.1007/s10803-008-0597-7

Rispoli, M., Hadley, P. A., & Holt, J. K. (2012). Sequence Analysis of Normal and Late Syntax Development in Children with Autism Spectrum Disorder. *Journal of Speech, Language, and Hearing Research*, *55*(1), 204–216. https://doi.org/10.1044/1092-4388(2011/10-0237)

Roberts, M. Y., & Kaiser, A. P. (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. http://doi.org/10.1111/1460-6984.12254

Law, E., Westrupp, E., Wake, M., Nicholson, J., Lucas, N., & Mensah, F. et al. (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.* 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., Prior, P., Dunn, A., Cronje, J., et al. (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. http://doi.org/10.3109/17549507.2013.777851

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

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

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. http://doi.org/10.1044/2019_LSHSS-VOIA-18-0131

Wren, E., Baker, E., Munro, N., Williams, A.L. and Trivette, C.M. (2018), Service delivery and intervention intensity for phonology-based speech sound disorders. *International Journal of Language & Communication Disorders*, 53: 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.

Zhang, 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

Table 1. Summary of included intervention studies, with vocabulary, phonology or morphosyntax outcomes, in which aspects of *dose frequency* were manipulated.

Study		Participants (Intervention and Comparison)		Treatment					Outcomes		
Study	**Study design**	**Number**	**Age (M, SD)**	**Dose Form / Intervention context**	**Treatment /control targets**	**Dosage manipulation**	**Planned/ Received dose**	**Planned/ received Dose frequency**	**Intervention duration and session length**	**Nature and timing of measures**	**Main findings**
1. Balthazar, C. H., & Scott, C.M. (2018)	Quasi Exper-imental	30 children with SLI Once weekly n= 14 Twice weekly n= 16	Once weekly group 11;06 years Twice weekly group 12;01 years	Modelling, repetition, manipulation of a complex sentence	Production of adverbial clauses, object complement clauses and relative clauses	Cumulative intervention intensity	Planned - 30 stimulus presentations per session (15 or modelling and repetition and 15 of sentence manipulation). Received for once weekly group 26 items per session. Twice weekly 28 items per session	Once weekly or twice weekly	9 weeks (40 to 60 minutes per session)	Complex sentence probes (before, during & after treatment). Standardized language tests and criterion referenced tasks. Pre and post intervention.	Treatment effective as measured by the sentence production probes. No advantage for the higher dosage group on any oral language measure.

2. Bellon-Harn, M. L. (2012)	RCT	12 children with SLI Concentrat-ed n = 6 Distributed n= 6	Both groups (concentrated and distributed) 4.61 years	Wh questions, expansions, cloze procedures (at varying levels of semantic complexity), models. Scaffolded language intervention in the context of book reading.	Expressive semantic and morphologic abilities	Dose Frequency	Average number of cloze procedure, expansion, or model used per minute ranged between 7 and 13 during each sampled session for all children. Authors note that - in a scaffolded-language therapy, there is no predetermined script or target. As such, questions remain about the frequency	Concentrated group - 4 times per week. Distributed group - twice a week. Received dose frequency was as planned	6 weeks (4 times a week) 12 weeks (twice a week) 20 minutes per session	Language sample analysis and expressive language probes. Pre and post intervention	Positive outcomes following both treatment schedules, no differences in how children performed in either the concentrated or spaced treatments

							of linguistic forms within each cloze procedure, expansion, and model.				
3. Bellon-Harn, M. L., Byers, B. A., & Lappi, J. (2014)	RCT	12 children with SLI (from low SES area) 24 sessions n= 6 42 sessions n = 6	4 times weekly group - 4.63 years 3 times weekly group - 4.78 years	Cloze procedures, expansions, models - Interactive Book Reading	Micro-structure components of narrative	Total intervention duration	Based on 20% sample, Average number of cloze procedures, expansions, or models used per minute ranged between 7 and 13. Average number of coordinating clauses between 0.2 and 3 times per minute Average	4 times per week (yielding 24 sessions in total) or 3 times per week (yielding 42 sessions in total)	6 weeks or 14 weeks - 20 minute per session	Language sample analysis. Week 1, midpoint and final week of intervention. Measure-ments from samples: Frequency with which children used co-ordinate and sub-ordinate clauses as well as the number of words within clauses.	Results indicated positive outcomes, but no group differences between those who received 24 sessions v's those who received 42 sessions of treatment.

							number of subordinating clauses between 0.6 and 3 times per minute				
4. Justice, L. M., Logan, J., Jiang, H., & Schmitt, M. B. (2017).	Cohort Analytical	233 children with language impairment.	76 months (ranged from 59 to 96 months)	Not specified as it depended on the target. Business as usual treatment carried out within the public schools system in the USA	One of 9 language focussed targets (grammar, communi-cative functions, discourse, narrative, listening comprehension Abstract language, meta-linguistics, literacy) 3 speech focused (articulation, phonology,	Treatment intensity (dose and frequency)	Children received language-focused treatment for about 12 min per session (mean = 11.8, SD = 4.73; range = 0.94–22.69), correspond-ing to about 49% of children's time in treatment Children received an average cumulative	1.3 sessions per week (range .5 to 4.1) Planned - 40.8 minutes per week Received - an average of 36.11 min/week an average of 46.4 treatment sessions in total across the academic year (SD =	Estimates of each child's treatment intensity was based on an average of 28 weeks of the current academic year (range = 7–39 weeks)	The four core subtests of the CELF-4 (con-cepts and following directions, word structure, recalling sentences, and formulating sentences) were administered in the fall and spring of the academic year. The vocabulary subtest from the Woodcock	children receiving high frequency/ low dose, or low frequency/ high dose treatment had better outcomes than those receiving high frequency/ high dose, high frequency/ average dose or low frequency/ low dose treatment.

					fluency, voice) 2 other (behaviour management, transitioning activities and null: no discernible target		intervention intensity of 1092.3 min (SD = 609.10, range = 66.45–3505.86 min) over an academic year of services.	16.56; range = 16–154), corresponding to about one session per week		Johnson III Tests of Achievement	
5. Meyers-Denman, C. N., & Plante, E. (2016)	RCT	16 children with SLI Massed n = 8 Spaced n = 8	Group A (massed) 5;03 years Group B (spaced condition) 5;04 years	Focused conversational recasts targeting a single morpheme. Focused recasts that used vocabulary that was unique to that recast and was administered to a child who was attending (i.e., looking	One expressive morpheme different for each child. In Group A (massed condition) past -ed x3, is - ing x2, 3psx1, she x1, hasx1. In Group B spaced condition 3x is-ing, 1x3ps,1xpast, 1xdoesn't, 1x she,1xhas	Dose Frequency (intervention given in massed or spaced conditions)	Planned: 24 conversational recasts per day targeting a specific grammatical morpheme, regardless of whether these were administered in the massed or spaced condition. Overall rate of delivery	Group A: 5 times per week Group B: 15 times a week	5 weeks (21 - 26 days, mean 25 days) Group A: One session of 30 minutes Group B: 3 X 10 minute sessions within a 4hr period	Baseline, end treatment, follow up Generalisation probes administered post treatment and at follow up – measuring child's use of the target / or control morphemes during a play based activity	Results indicated a significant improvement in morpheme production in both dose frequencies with no change in untreated morpheme use. No differences in the effect of treatment for the

				at the clinician) during the recast.)			controlled across spaced and massed conditions at eight recasts per each 10-min block of time (one recast every 1.25 min.) Received: cumulative intervention intensity of approximately 600 conversational recasts containing the target morpheme - range 504- 624 recasts			that obligated the use of the morphemes with untreated lexical items.	concentrated or spaced conditions.

6. Plante, Mettler, Tucci, & Vance, (2019).	Quasi-Exper-imental	20 children with DLD High density n = 10 Low density n = 10	M = 5;0 years Range 4;01 to 5;11	Enhanced conversational recast treatment – in the context of free play and dialogic book reading	Morphological forms used (spontaneously or elicited) less than 30% of the time were assigned as treatment or control Including -ed, 3rd person – s, Aux. is, possessive	High density or low density of Enhanced conversation-al recast treatment	High density – 24 recasts administered in the first 15 min of a 30 minute session (1.6 per min) Low density- 24 recasts over 30 minutes (.8 per minute) Cumulative intervention intensity of 528 – 600 across children	5 days per week	5 weeks (22 – 25 days) 30 minute sessions	Generalisation probes – administered immediately before the days treatment session on Monday, Wednesday and Friday Retention probes administered 6 weeks post treatment (m= 42, range 35 – 49 days) Number of treatment responders	No significant differences between treatment conditions on any outcome measure Strong relationship between performance at end of treatment and follow up 7/10 treatment responders low density condition 8/10 treatment responders high density condition
7. Proctor-Williams, K., & Fey,	Cohort Analytical (treatment	26 children	SLI group - 7;10 years	Recasts - in the context of a	Novel verb learning (6 verbs).	Dose Frequency (described by	Planned – Total dose of 30 recasts in	Distribution of sessions was not	Duration in weeks not specified. 2	During the intervention - Correct	Children with DLD did not improve their

M. E. (2007).	words randomly assigned)	children with SLI n = 13 (all children assigned verbs in two conditions) younger TD participants n = 13	TD group - 5;6 years	play based activity	Syntactically all verbs were transitive, causative and telic. Phonologically all were single-syllable verbs that marked tense with a vowel shift.	authors as high or low density)	both density conditions. Low density = .2 per min High density condition = .5 per minute (no recasts in the first three sessions, last 2 sessions included 5 irregular past tense recasts for each of the three high-density verbs Received - low density .19 per min; High density .47.	tightly controlled - substantial range in the number of days from the first to the fifth experimental sessions (4 to 44, respectively; M = 14 days; SD = 8.95).	Training sessions of 31 minutes - 5 experimental sessions - average 31 minutes.	spontaneous productions of irregular past tense novel verbs in obligatory contexts in Sessions 4 and 5. Post intervention - the number of correct irregular past tense verb productions (maximum 12 per condition)	production accuracy at higher intervention recast densities
8. Riches, N. G., Tomasello, M., & Conti-	Quasi Experimental	45 children. 23 children with SLI; 22 younger	SLI group mean age 5;6 years	Novel verb modelling using an intransitive frame and a	Comprehension and production of four novel verb forms -	Dose (number of exposures) and Dose Frequency	Four planned doses - Massed 12, 12 exposures on a single day;	4 days in one week Or 1 day	1 week (between 2 and 10 minutes per sessions)	3 probes; an action probe (what does it do, can you show me), a	In relation to comprehension children with SLI benefitted from a greater

Ramsden, G. (2005)		typically developing All children assigned one of 4 verbs in 4 conditions	TD group mean age 3;5 years	dual morphological frame, alternating between the -ing form and the third person form e.g. dacking, dacks using both was counted as one presentation of the target verb. Dose form was given while playing with a series of objects.	dack, tam, meek, gorp	(spaced or concentrated)	Massed 18, 18 exposures on a single day; Spaced 12, 12 exposures spread over 4 days (3 per day), and (c) Spaced 18, 18 exposures spread over 4 days (4,5,4,5).			production probe (what's it doing, can you tell me?) and a comprehension probe (from a choice of three objects - which one was verbing?) were carried out immediately post and one week post intervention.	number of exposures to novel verbs. For production the spacing effect was greater than the effect of the number of exposures i.e. children had better learning after 12 presentations when the exposures were spaced, than after 18 presentations when the exposures were massed.
9. Schmitt, M. B., Justice, L. M., & Logan, J. A. (2016)	Cohort Analytical	233 children with language impairment.	76 months (ranged from 59 to 96 months)	Not specified as it depended on the target.	One of 9 language focussed targets (grammar,	Treatment intensity (dose and frequency)	Children received language-focused treatment for	1.3 sessions per week (range .5 to 4.1)	Estimates of each child's treatment intensity was based on an	The four core subtests of the CELF-4 (con-cepts and following	children receiving high frequency/ low dose, or low frequency/

				Business as usual treatment carried out within the public schools system in the USA	communi-cative functions, discourse, narrative, listening comprehension Abstract language, meta-linguistics, literacy) 3 speech focused (articulation, phonology, fluency, voice) 2 other (behaviour management, transitioning activities and null: no discernible target		about 12 min per session (mean = 11.8, SD = 4.73; range = 0.94–22.69), correspond-ing to about 49% of children's time in treatment Children received an average cumulative intervention intensity of 1092.3 min (SD = 609.10, range = 66.45–3505.86 min) over an academic year of services.	Planned - 40.8 minutes per week Received - an average of 36.11 min/week an average of 46.4 treatment sessions in total across the academic year (SD = 16.56; range = 16–154), corresponding to about one session per week	av- erage of 28 weeks of the current academic year (range = 7–39 weeks)	directions, word structure, recalling sentences, and formulating sentences) were administered in the fall and spring of the academic year. The vocabulary subtest from the Woodcock Johnson III Tests of Achievement	high dose treatment had better outcomes than those receiving high frequency/ high dose, high frequency/ average dose or low frequency/ low dose treatment.
10. Siegmüller, J., Baumann,	Quasi Experimental	30 children with DLD /SLI both	Mean age given for 48 children (30	Intensive modelling of a) verbs and their	Use of subordinate	Dose Frequency	Not specified	Once or twice per week	Maximum number of sessions 22.	Post testing completed with all	Results showed that children who received

J., & Höppe, L. (2017)	(Dosage factors studied retro-spectively	terms are used in the paper Once weekly n = 15 Twice weekly n = 15	of whom were studied in relation to dosage) 3;11years (SD 14;77 months)	associated arguments and b) different grammatical subcomponents of the sentence; Questioning to elicit main clause production; Modelling expansions of the main clause to subordinate clause structures.	clause structures				Average number 17. 98. Intervention discontinued before 22 sessions when child showed the ability to expand a given structure to a subordinate clause. Each session an estimated 45 minutes.	children after 16 sessions	once weekly treatment needed fewer sessions to achieve therapy goals than the children who received twice weekly sessions. Younger children required fewer sessions to achieve goals.
11. Smith-Lock, K. M., Leitao, S., Lambert, L., Prior, P., Dunn, A., Cronje, S., Newhouse,	Quasi Experimental	36 children with SLI (31 in the analyses) Daily treatment n = 18	Group A 63.61 months Group B 62.08 months	Repeated modelling of grammatical targets, opportunities for the child to produce the targets,	pronouns, possessives, past tense, present tense.	Dose frequency	Not specified	4 times a week for two weeks, once a week for 8 weeks	2 weeks or 8 weeks (1 hour sessions)	Grammar elicitation test (administered 4 times). Each child completed the section of the test relevant to	Results showed significant improvement in the spaced treatment group (but not the concentrated

S., & Nickels, L. (2013).		Weekly treatment n = 13		feedback to the child, opportunities for child to correct him/herself. Detailed activity plans provided for use in a natural play context						their grammatical target. Gain between Tests 1 and 2 (pre-treatment gain) compared with gains made between Tests 2 and 3 (post treatment gain). Pre-treatment gain compared with gain between test 2 and follow-up	group), relative to the same time period prior to treatment. Single-subject analyses indicated a significant treatment effect in 46% of children in the spaced group and 17% in the concentrated group.
12. Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017).	RCT - recruitment not random but children were randomly assigned to one of four	27 children with language impairment All children assigned words in	M = 5;08 years Range 5;0 - 6;05	children heard: the target word in a book, a definition of the target word, a synonym of the target word,	Word learning	Dose (using an escalation design 12,24,36 or 48 exposures)	Planned: Depending on treatment intensity the no. of exposures per session were 3,4,6 and 6.	2 to 3 sessions per week	Dependant on treatment intensity 4-5 weeks (10 sessions); 5-	Ability to give word definitions was measured pre and immediately post treatment.	Results from the word definition and naming tasks indicated 36 exposures to be the optimal dose (43% of

	treatment intensity conditions.	one of four treatment intensities		the target word used in a supported context sentence.			For 12 cumulative exposures (target word exposure 3 times in each book and book read 4 times over the course of the intervention); 24 exposures (4X6); 36 exposures (6x6) and 48 exposures (6x8) Received: Based on 20% of sessions - dividing the total number of exposures administered by the intended		8 weeks (15 sessions); and 7-10 weeks (20 sessions) 20 to 30 minutes per session.	% of children that responded to treatment on the basis or word definitions % of children that responded to treatment on the basis of naming	children responded based on definitions and 86% responded based on naming).

							number of exposures, was 99.92%.				
13. Storkel, Komesidou, Pezold, Pitt,., Fleming, & Romine, (2019)	RCT - recruitment not random but children were randomly assigned to treatment arms	34 children with DLD All children exposed to words with dose 6 X frequency 6 protocol In addition half children received dose 9 X frequency 4 protocol and half received dose 4 X frequency 9 protocol	M = 5;6 years SD = 0;4	Pre-book reading activity (showing 6 target words in colourful pictures with orthographic label) Reading of book in which target words are highlighted by a box Post-book reading activity reviewing 6 target words with different colourful pictures and orthographic label.	Word learning of two word sets (60 words in total – 16 nouns, 25 verbs, 19 adjectives) 6 words targeted in a given book	Dose and dose frequency	4, 6 or 9 exposures	9, 6 or 4 book reading sessions Typically, two treatment sessions per week (2 books per session)	12 weeks (4x9 condition) 8 weeks (6x6 condition) 5 weeks (9x4 condition) Average session length was 13 min (4x9 condition) 14 min (6x6 condition 16 min (9x4 condition)	Primary outcome - Definition task - administered pre, 5 / 6 days post each treatment session and approx. 21 days post each treatment session Secondary outcome – Interim definition and naming task (at 4 points during each treatment, the final test following the	36 exposures supports significant word learning in children with DLD There was a significant drop in children's performance once treatment was withdrawn (60% drop 5/6 days post, 70% drop 21 days post) Manipulation of dose x dose frequency did not result in

				In each activity children heard the target word, a definition of the target word, a synonym of the target word, the target word used in a supported context sentence.						last treatment session)	significant differences in word learning outcomes Naming data not reported as they showed the same pattern as the definition data, which were more complete.

NOTE: The term dose is used differently in the Justice et al. (2017) and Schmitt et al. (2016) studies and is defined as the total amount of time spent addressing any one of nine language-focused targets (it was based on 3 videos but they correlated very highly so was considered representative of each child's dose).

Figure 1. PRISMA flowchart showing literature search process.

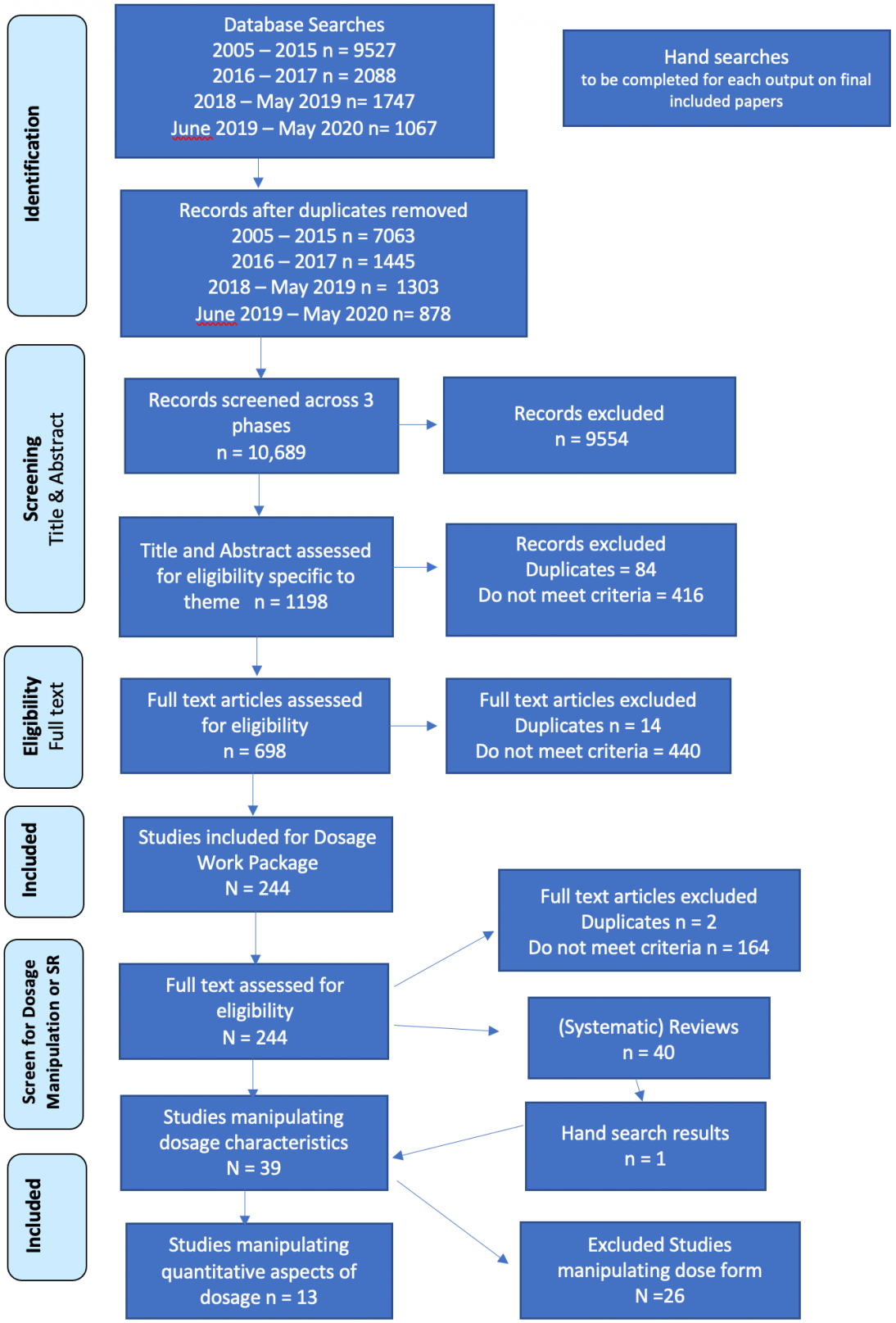
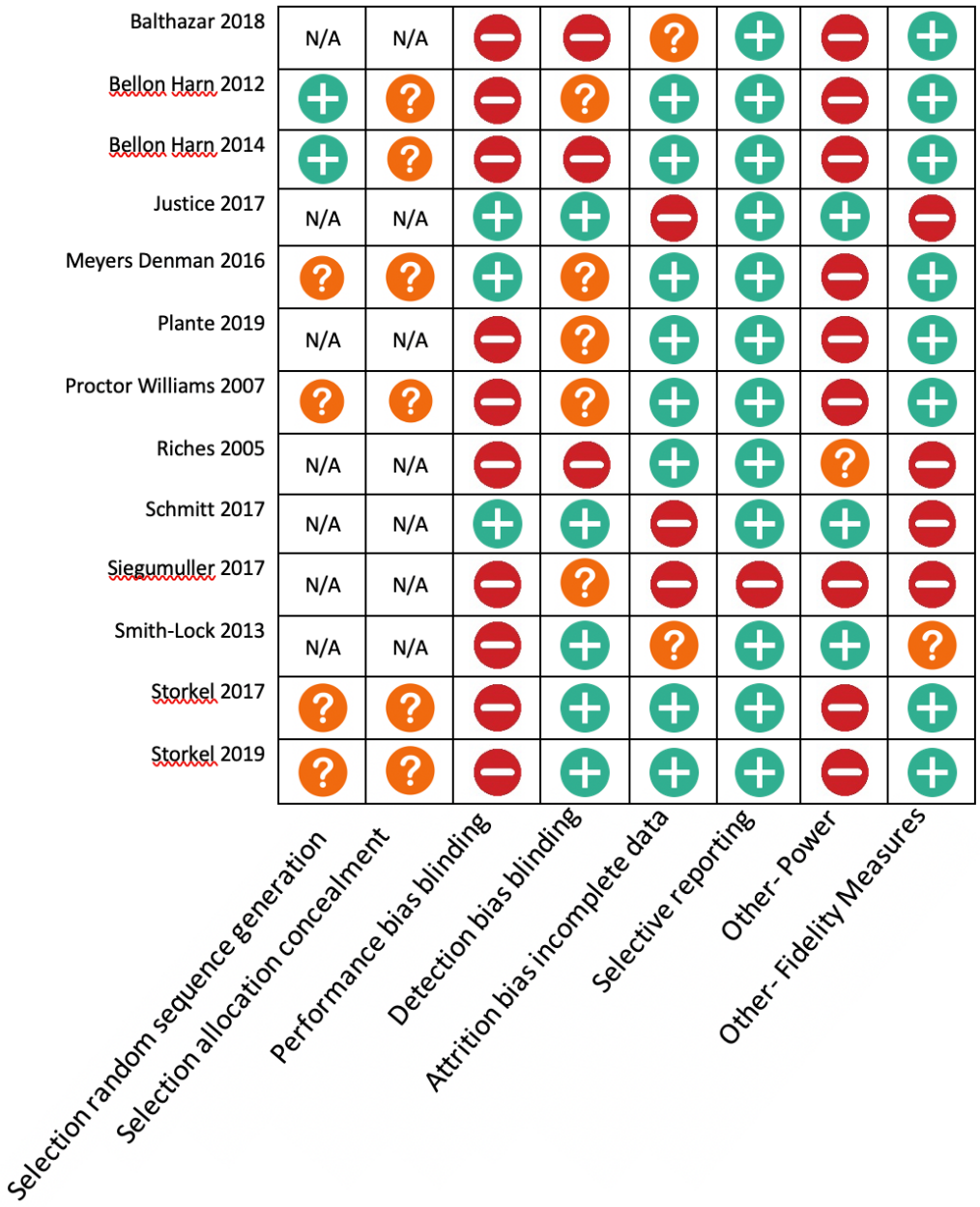


Figure 2. Critical Appraisal of each included study.



Appendix A

Search strategy for “Systematic reviews of interventions aimed at the vocabulary, morpho-syntax and phonology of children with Developmental Language Disorder within the COST IS1406 network”

PubMeD

MH AND Teaching OR training OR treatment OR "clinical trial" OR intervention OR therapy OR rehabilitation OR remediation OR “special education” OR "dynamic assessment" OR "response to intervention") AND ("language impairment" OR "language delay" OR "language disorder" OR "language disability" OR "language development disorders" OR dysphasi* OR aphasi* OR "developmental communication disorder") AND (child* OR preschool* OR adolescen* OR teenage* OR youth) NOT (adult OR deaf OR autis* OR "hearing impairment" OR "Down syndrome" OR "intellectual disability" OR "traumatic brain injury" OR "acquired brain injury" OR "physical disability" OR "learning disability" OR "severe learning difficulties" OR "severe learning difficulty" OR disease) AND (("2006/01/01"[PDat] : "2015/12/31"[PDat]))

Web of Science

TOPIC: ((teaching OR training OR treatment OR "clinical trial" OR intervention OR therapy OR rehabilitation OR remediation OR "special education" OR "dynamic assessment" OR "response to intervention") AND (language AND (impairment OR delay OR disorder OR disability) OR "language development disorder" OR dysphasi* OR aphasi* OR "developmental communication disorder") AND (child* OR preschool* OR adolescen* OR teenage* or youth) NOT (adult OR deaf OR autis* OR "hearing impairment" OR "Down syndrome" OR "intellectual disability" OR "traumatic brain injury" OR "acquired brain injury" OR "physical disability" OR "learning disability" OR "severe learning difficulties" OR disease)) 2006-2015 excluding Chemical abstracts

ERIC

((teaching OR training OR treatment OR "clinical trial" OR intervention OR therapy OR rehabilitation OR remediation OR "special education" OR "dynamic assessment" OR "response to intervention") AND (“language impairment” OR “language delay” OR “language disorder” OR “language disability” OR "language development disorder" OR dysphasi* OR aphasi* OR "developmental communication disorder") AND (child* OR preschool* OR adolescen* OR teenage* or youth) NOT (adult OR deaf OR autis* OR "hearing impairment" OR "Down syndrome" OR "intellectual disability" OR "traumatic brain injury" OR "acquired brain injury" OR "physical disability" OR "learning disability" OR "severe learning difficulties" OR disease))

Setting the year limit on (2006-2015).

PsychInfo

((teaching OR training OR treatment OR "clinical trial" OR intervention OR therapy OR rehabilitation OR remediation OR "special education" OR "dynamic assessment" OR "response to intervention") AND (language AND (impairment OR delay OR disorder OR disability) OR "language development disorder" OR dysphasi* OR aphasi* OR "developmental communication disorder") AND (child* OR preschool* OR adolescen* OR teenage* or youth) NOT (adult OR deaf OR autis* OR "hearing impairment" OR "Down syndrome" OR "intellectual

disability" OR "traumatic brain injury" OR "acquired brain injury" OR "physical disability" OR "learning disability" OR "severe learning difficulties" OR disease))

SCOPUS

(TITLE-ABS-KEY (teaching OR training OR treatment OR "clinical trial" OR intervention OR therapy OR rehabilitation OR remediation OR "special education" OR "dynamic assessment" OR "response to intervention") AND TITLE-ABS-KEY ("language impairment" OR "language delay" OR "language disorder" OR "language disability" OR "language development disorder" OR dysphasi* OR aphasi* OR "developmental communication disorder") AND TITLE-ABS-KEY (child* OR preschool* OR adolescen* OR teenage* OR youth) AND NOT TITLE-ABS-KEY (adult OR deaf OR autis* OR "hearing impairment" OR "Down syndrome" OR "intellectual disability" OR "traumatic brain injury" OR "acquired brain injury" OR "physical disability" OR "learning disability" OR "severe learning difficulties" OR disease)) AND SUBJAREA (mult OR medi OR nurs OR vete OR dent OR heal OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 2005 AND PUBYEAR < 2016 AND (EXCLUDE (DOCTYPE , "le"))

APPENDIX B

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item
ADMINISTRATIVE INFORMATION		
Title:		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:		
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:		
Sources	5a	Indicate sources of financial or other support for the review
Sponsor	5b	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known

Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS		
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:		
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)

	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.D