Children with Down Syndrome can Benefit from Language Interventions; Results from a Systematic Review and Meta-analysis

Smith, Elizabeth[1] [2], Hokstad, Silje[3], & Næss, Kari-Anne B.[3]

[1]= Cardiff University, UK

[2]= University of Bristol, UK

[3]= University of Oslo, Norway

Correspondence concerning this article should be addressed to

Liz Smith, PhD, School of Psychology, Cardiff University

Tower Building, 70 Park Place, Cardiff, CF10 3AT

Mail: SmithL57@Cardiff.ac.uk

Acknowledgements

This work was supported by grant number 238030 from the Finnut program from the Research Council of Norway

Abstract

Language disorder is a cardinal challenge for children with Down syndrome, and their learning capacity has been debated. The aim of the current study was to conduct a systematic review and meta-analysis of existing language interventions for children with Down syndrome to reveal knowledge about the effects of these interventions and identify any common characteristics specific to effective or ineffective interventions. A systematic search was conducted in databases relevant for education, speech and language therapy, and psychology. Based on a set of predefined inclusion criteria, the hits were screened and coded. Eight studies were synthesised in a systematic review and four in a meta-analysis. The overall effect of the interventions was large (g = 1.01), but significant transfer effects to untrained aspects of language were rarely found. Interventions showing significant effects varied with regards to numerous characteristics including the age of the target group, the intervention approach, the dosage, and the implementer. The common characteristic across the effective interventions was simply *the aim* of improving language skills in children with Down syndrome. Overall, there was a moderate to high risk of bias across all studies. To conclude, children with Down syndrome have the potential to respond to language intervention. However, more interventions that reach transfer effects are needed to maximise children's language outcomes. Based on the limited number of studies and a moderate to high risk of bias across the studies, there is a great need for more robust intervention studies to ensure that future interventions are informed by high-quality research.

Keywords: Down syndrome, trisomy 21, language, intervention, evidence-based practice

Children with Down Syndrome can Benefit from Language Interventions; Results from a Systematic Review and Meta-analysis

Language is key to communication, social functioning, development and learning (Boudreau, 2002; Hartup, 1983; Hulme et al., 2012; Næss, Nygård, Ostad, Dolva, & Lyster, 2016; Næss & Smith, 2018). Due to substantial language disorders (Martin, Klusek, Estigarribia, & Roberts, 2009; Næss, Lyster, Hulme, & Melby-Lervåg, 2011), individuals with Down syndrome will typically spend large amounts of time and energy undertaking interventions. However, their potential for progress is sometimes underestimated due to their intellectual disabilities (see e.g., Rynders et al., 1997; Turner & Alborz, 2003). Therefore, there are both ethical and economic reasons for gaining knowledge about long-term effective language intervention adapted for their language profile. However, a national study of Norwegian third graders with Down syndrome reported that one third of the children did not receive systematic language intervention and the interventions provided are rarely indicated to be effective through research (Næss, Engevik, & Hokstad, 2017).

The aim of this study is to systematically summarize the findings from existing group-design language interventions for children with Down syndrome to determine the effect of interventions and to reveal knowledge about what makes an effective language intervention for this group of children. This knowledge is essential to inform decision-making in the field of practice and to inform the development of future intervention programmes.

Language profile in children with Down syndrome

Across the three core components of language: content (e.g., vocabulary), form (e.g., phonology, morphology, and syntax), and use (e.g., pragmatics) (Bloom & Lahey, 1978), children with Down syndrome usually show a specific profile of strengths and weaknesses.

For language content, the number of words the children understand (receptive vocabulary) is often found to be a strength, while expressive vocabulary tends to be poorer relative to typically developing children matched on non-verbal mental age (Næss, Lyster, Hulme, & Melby-Lervåg, 2011), and relative to their own receptive vocabulary (Chapman, 1997). However, more recent research highlights that children with Down syndrome also tend to have a weakness in receptive vocabulary depth (how well they know the words) relative to typically developing peers matched on receptive vocabulary breadth (Laws et al., 2015). Furthermore, language form also tends to be an area of difficulty for individuals with Down syndrome. Particular weaknesses are seen in phonology (e.g., phonological awareness; Laws, 1998; Næss, Lyster, Hulme, & Melby-Lervåg, 2011; Næss, 2016), morphology (e.g., the use of morphological units; Eadie, Fey, Douglas, & Parsons, 2002), and syntax (e.g., structuring sentences correctly; Fowler, 1990). Finally, language use (e.g., understanding the meaning of language across different contexts) has been found to be significantly less impaired relative to structural language (content and form) among individuals with Down syndrome (Laws & Bishop, 2004; Smith, Næss, & Jarrold, 2017). However, language use is still found to be weaker relative to norms of typically developing children (Smith et al., 2017).

Previous reviews of language interventions for children with Down syndrome

The number of studies on language intervention for children with Down syndrome has been increasing over the last decade (Hokstad, Smith, & Næss, 2019), and some reviews have been carried out to summarize the existing research base. The most recent review focussed mainly on expressive communication, including speech (Neil & Jones, 2018). However, pragmatic language was not included in the search, and the summary analysis and discussions focussed on case series designs. Further, no risk of bias analysis was included. The remaining

systematic reviews in this field are restricted to specific approaches to intervention or specific language subcomponents, such as parent-mediated interventions (O`Toole, Lee, Gibbon, van Bysterveldt, & Hart, 2018; Te Kaat-van den Os, Jongmans, Volman, & Lauteslager, 2017), speech and language therapy interventions (Smith & van der Gaag, 2012), and training of phonological awareness (Lemons & Fuchs, 2010).

No meta-analysis summarizing the overall effect from group comparison studies exists. Therefore, it remains unknown whether language intervention for children with Down syndrome is more effective than no intervention and what characterizes an effective language intervention for this group of children. More knowledge about non-responsive children is also needed, as this question has not been raised in previous reviews. The value of adding a meta-analysis lies in the ability to increase power and thereby increase the chance of detecting a real effect from interventions (Cohn & Becker, 2003). It is valuable to aggregate studies for a greater level of power to investigate overall effects when individual studies are small scale (Button et al., 2013), as is commonly the case in the Down syndrome research field.

Current review and meta-analysis

Due to the gap in existing literature and the language profile of children with Down syndrome, it is essential to investigate the effect of language interventions, including all key aspects of language (content, form, and use). In this systematic review and meta-analysis, we will investigate: 1. The overall mean effect on language outcomes for children with Down syndrome receiving a language intervention, compared to business as usual or another comparison intervention, and 2. Whether there are any common characteristics specific to effective or ineffective interventions.

Method

The current systematic review and meta-analysis was conducted based on a structured literature search, following the guidelines in the Cochrane Handbook (Higgins & Green, 2011) and following the PRISMA statement (Moher et al., 2015).

Search strategy

The following electronic databases were searched based on topic: PsycINFO, ERIC, and Web of Science Core Collection. The key topics for this search were *Down syndrome*, *language*, and *intervention*. Additionally, we attempted to include all relevant associated terms. An example of the full electronic search strategy for PsycINFO is presented in Table 1.

[Insert Table 1 here]

The search was conducted on 15/03/18. No limits were set regarding the publication date. In December 2018, we searched for additional records via the following sources: reference lists in the included studies, an ongoing literature review on implementation quality of language interventions carried out by Hokstad, Smith and Næss (2019), Google Scholar, and conference proceedings for recent international conferences in the Down syndrome field. We also contacted the authors of the included articles to enquire about any ongoing or unpublished research.

Inclusion criteria for systematic review and meta-analysis

Articles were selected on the basis of meeting a set of predefined inclusion criteria. The studies had to include:

(1a) Language interventions targeting content, form and/or use, in any potential setting (i.e., at home with parents, in clinical practice with speech and language pathologists (SLPs), or in school settings with teachers). Interventions targeting pre-linguistic skills and articulation were excluded.

(1b) Participants with Down syndrome. Studies with mixed aetiology groups were included if participants with Down syndrome were treated as a separate independent group in the analysis. This criterion was to ensure the findings from the study were directly relevant to children with Down syndrome and effects were not driven by individuals with other aetiologies.

(2) Children between 0-18 years old. To ensure that any prevention studies for young children targeting language (e.g., vocabulary) were included, no cut-off for the early years was applied.

(3) Control group that did not receive the targeted intervention, such as a group receiving business as usual or an active control group.

(4) Quantitative outcome data (pre and post or growth) for both groups in at least one language component: content (vocabulary, semantics), form (phonology, grammar (syntax, morphology)), and/or use (pragmatics) to allow for comparison(s).

(5) Description of the intervention methods, including materials and procedures used; this was to allow interpretation of study findings.

(6) Original data to ensure the data were independent.

In addition to these criteria, studies had to fulfil the following criteria to be included in the meta-analysis: (7) Effect sizes or sufficient information to allow the calculation of

effect sizes (means, standard deviation, and sample size for training and control group pre- and post-intervention).

Screening process

All abstracts identified in our systematic search were individually assessed according to the inclusion criteria to determine eligibility. For abstracts in which inclusion criteria appeared to be met, or if it was unclear if inclusion criteria were met, full text articles were read by both the first and the second author and discussed with the third author. Details regarding the number of included studies at each stage of the screening process, as well as reasons for exclusion, are shown in Figure 1.

Coding

All included articles were coded in duplicate by the first and second author and discussed in detail among all authors according to the following information:

(1) Name of author and publication year.

(2) Source.

(3) The country of origin and the language used.

(4) Intervention aim, method, and approach. The aim was defined in terms of what aspect of language or specific skill(s) the intervention aimed to improve. The method was coded with a brief description of the training tasks, and approaches were dichotomised as either systematic or naturalistic intervention programmes. Studies were defined as having a systematic approach when they included a predefined programme with specified procedures, frequency, and content to be followed for all children. Studies were defined as naturalistic if they involved the delivery of ‘strategies that identify and use opportunities for learning that occur throughout the child`s natural activities, routines, and interactions; follow the child`s lead; and use natural consequences’ (Shelden & Rush, 2001, p.2).

(5) The aspect(s) of language trained and assessed was coded according to content (vocabulary), form (phonology, morphology, or syntax), and use (pragmatics), or a combination of two or more of the subcategories.

(6) The study design and the type(s) of assessment(s) used (e.g., experimental test; observation).

(7) Intensity of training (a. the length of the training, b. the frequency of the training, and c. the session duration).

(8) Implementer(s)/Setting.

(9) Whether the intervention was designed or adapted specifically for children with Down syndrome was binary coded; yes/no.

(10) Sample size.

(11) The mean age and age range of participants (in months).

(12) The statistical tests used to assess the effects of the intervention, along with what type of effect sizes were reported and for which measures. In instances where effect sizes were not reported, we note which other type of statistical significance values are reported (e.g., *p* values) if any.

(13) The effect sizes for all language outcomes measured, as reported in the paper for each study. If the effect sizes were not reported in the paper, then we instead coded the *p* values. If no effect sizes and no *p* values were reported, only the descriptive information was coded.

(14) Whether or not the study included a follow up, the time frame of the follow up (number of months post-intervention), and the results of the follow up were coded.

(15) The number of non-responsive children in the intervention group (defined as those reported to have an absence of any gains), reported for each targeted outcome measure separately.

Risk of bias

A quality appraisal was carried out based on the Cochrane Handbook guidelines for the assessment of risk of bias (Higgins & Green, 2011; Higgins, Altmam, & Sterne, 2017). Each study was assessed in five key areas:

1. Selection bias (the method used to generate allocation sequence and the method used to conceal allocation sequence).
2. Performance bias (were the personnel involved in implementing the intervention and the participants themselves blinded to the conditions?).
3. Attrition bias (were children excluded or withdrawn from the study?).
4. Detection bias (were the individuals assessing the outcomes blinded to the conditions?).
5. Reporting bias (was reporting done selectively, i.e., are results reported for some but not other measures?).

Two reviewers independently assessed each study based on these five areas, judging the risk of bias for each study as low risk, high risk, or unclear. After reviewing each study independently, the two reviewers met to discuss their risk of bias judgments for each study to ensure agreement.

Analysis for systematic review

The analysis of the data from the systematic review includes a narrative synthesis of the content of all the interventions (the aim(s), intervention method(s), and approach(es)) and procedures (implementers, time frame, and intensity) included in the review. Furthermore, we assessed the frequency of different study characteristics, identifying commonalities and differences (e.g., in research design, intervention content, and intervention outcomes) across all included studies.

Meta-analysis procedure

Studies that included measures with enough reported information to calculate effect sizes were included in the meta-analysis. The effect size (Hedge's g) for each language outcome was calculated by the research team wherever possible, and these are presented in supplementary Table 1b (also available online via the Open Science Framework (OSF): https://osf.io/scxam/)). The Hedge's g was calculated from the mean difference pre- to post-test, the standard deviation at pre-test, and the sample size, for the intervention group relative to the control group. We used this method to calculate effect sizes for each study to ensure a fair comparison, as some studies used an Analysis of covariance (ANCOVA) model to calculate effect size whereas other studies used a pre-post t-test. These approaches can lead to slightly different effect sizes. To ensure independence, a summary effect size of all language measures was calculated for each study (see Borenstein, Hedges, Higgins, & Rothstein, 2011; Higgins & Green, 2011) via averaging the effect sizes for each of the individual language outcomes within the study.

The meta-analysis was carried out using Jamovi programming software (Jamovi project, 2017). The effect size measure used was Hedge's g. The overall effect size across the studies was calculated based on a weighted average and was computed based on a random effects model. Weighted effect sizes account for differences in statistical power in the different studies, thereby allowing for fair comparisons of the different studies and making it easier for practitioners and researchers to interpret the effectiveness of different interventions. The random effects model was used to account for the possibility of systematic variation across studies. Ninety-five percent confidence intervals (CI) were also estimated for each measure, and exact values for these are shown in Appendix A. A forest plot was carried out to illustrate the summary effect size of each study.

Results

As shown in Figure 1, there were a total of eight studies included in the systematic review and four of these studies met criteria to be included in the meta-analysis. The main reason for excluding language interventions in the systematic review was the lack of a comparison group.

The information coded for each of the studies included in the systematic review can be found in supplementary Table 1a (also available online via OSF: https://osf.io/scxam/), while the statistics reported in each of the studies and statistics included in the current meta-analysis can be found in supplementary Table 1b (also available online via OSF: https://osf.io/scxam/).

[Insert Figure 1 here]

Research design variables

Number of participants and demographics

Sample sizes in the studies included in the systematic review varied between three and 28 children in the training group (mean: 13.50, with a SD of 9.50) and between three and 26 children in the control group (mean: 12.90, with a SD of 8.90). A sensitivity power analysis carried out with G*Power, assuming standard power (.80) (Cohen, 1988), shows that a sample of 13 children in each group would allow for the detection of effect sizes at d = 1.15 and above. Therefore, only very large effect sizes can be detected with these small samples, given that a Cohen's d value of .80 or more is considered large (Cohen, 1988). It should be noted that there were three studies with larger sample sizes positively skewing the mean (Baxter, Hulme, Rees, & Perovic, 2018; Burgoyne et al., 2012; Yoder, Woynaroski, Fey, & Warren, 2014); as a result, the median sample size was nine participants in the training group

and 9.50 participants in the control group. These small studies have less power to detect true effects and a lower likelihood of reproducibility.

The training group had a mean age of 8.30 years (mean based on the data provided in six of the eight studies. The difference in age between the youngest and oldest child *within* samples was often wide (mean: 5 years), with Cleave, Bird and Bourassa (2011) representing the largest range in age (5-16 years). On average, the youngest children in the studies were age 5 and the oldest children were age 10 (based on data reported in seven of the eight studies). No intervention was specifically designed for adolescents.

Description of interventions and aspects of language trained and assessed

Content and form were trained within the included studies. However, language use (pragmatics) was not trained. Six out of eight studies focussed on a single language component. Language form was the most common aspect of language trained, with five out of eight studies targeting this aspect, and an additional study trained language form in combination with language content (Burgoyne et al., 2012). The remaining two studies in the review trained language content with a focus on expanding children's expressive vocabulary (Girolametto, Weitzman, & Clements-Baartman, 1998; Yoder et al., 2014). Of the studies training language form, either phonology or grammar were targeted. Both interventions training phonology focussed on phonological awareness, targeting rhyming, initial and final phoneme awareness (Cleave et al., 2011), phoneme segmentation, and phoneme blending (Goetz et al., 2008). Both phonological awareness interventions aimed to improve children's language as well as reading abilities. In the current review, we consider only the language outcomes. In the Burgoyne et al. (2012) study training both form and content, the training focussed on phoneme blending and vocabulary. In terms of grammatical form, the interventions aimed to improve understanding of past tense (Baxter et al., 2018), increase

utterance length and complexity (MacDonald et al., 1974), or to improve various features of morpho-syntax (Sepulveda, López-Villaseñor, & Heinze, 2013). Overall, four out of eight of the intervention approaches were designed or adapted *specifically* for children with Down syndrome.

In addition to assessing the aspects of language trained, some studies included near transfer measures (untrained measures within the same language aspect) (Baxter et al., 2018; Burgoyne et al., 2012; Goetz et al., 2008). Three studies additionally assessed far transfer to other aspects of language; Girolametto et al. (1998) assessed whether vocabulary training improved syntax (as well as vocabulary), Sepulveda (2013) assessed whether syntax and morphology training improved semantics, and pragmatics (alongside syntax/morphology), and Burgoyne et al. (2012) assessed whether training phonology and vocabulary led to improvements in grammar.

The most common assessment tools were experimental tests. The only other measures used for assessment were parental reports (Yoder et al., 2014) or a combination of behaviour observations and parental reports (Girolametto et al., 1998).

Intervention approaches

A naturalistic approach was used to train targeted vocabulary in one study (Girolametto et al.,1998) involving parents interacting with children in naturalistic settings during play activities, following the child's lead while using techniques such as modelling the target words. Two further studies (MacDonald et al., 1974; Yoder et al., 2014) used a combination of systematic and naturalistic intervention, with MacDonald et al. (1974) involving lab-based systematic sessions with professional language trainers and more naturalistic sessions at home with parents. The Yoder et al. (2014) study can be characterised as systematic in terms of the frequency of sessions and naturalistic in terms of the

intervention content within the sessions. All remaining studies used a systematic approach, such as Sepulveda et al. (2013), involving pre-defined structured tasks with a speech therapist, with the same pre-defined set procedure for each child.

Intervention procedures

Time and Intensity

The included studies ranged in session duration from 20 minutes (Baxter et. al., 2018) up to 150 minutes (Girolametto et. al., 1998). The mean session duration was 53 minutes; this was positively skewed by one study with a session duration over 60 minutes (Girolametto et al., 1998), hence the median was somewhat lower: 40 minutes.

The regularity of sessions ranged from five days a week (Baxter et al., 2018; Burgoyne et al., 2012; Goetz et al., 2008; MacDonald et al., 1974; Yoder et al., 2014) to once per week (Girolametto et al., 1998). Most studies had multiple sessions per week, and five out of eight had daily sessions (5 days a week).

The length of the intervention periods ranged from two months (Goetz et al., 2008) to nine months (Yoder et al., 2014), and the mean length of interventions was 4.5 months.

Implementer/Setting

The most frequent implementers in the interventions were teaching assistants. It was less common for researchers, speech language pathologists, or parents to implement the intervention, and no study reported using teachers as implementers. Four out of eight studies were implemented in schools. The remaining four studies were each implemented in different settings: home, speech and language pathologists' rooms, or a combination (clinic and home/childcare centre and home).

Characteristics of the included interventions are shown graphically in Appendix B (also available online via OSF: https://osf.io/6gpu4/).

Systematic review of the effect of language intervention for children with Down syndrome

Seven out of eight studies reported tests of statistical significance on the language assessments, and one of the included studies did not (MacDonald et al., 1974). MacDonald et al. (1974) describes substantial improvements in the targeted language area (syntax). Of the studies reporting statistics, two used nonparametric tests (Cleave et al., 2011; Girolametto et al., 1998), reported as a means to overcome the small sample sizes, or due to data not being normally distributed.

Of the studies reporting statistics, three were excluded from the meta-analysis (Cleave et al., 2011; Girolametto et al., 1998; Yoder et al., 2014) because a) information reported was not sufficient to allow for the calculation of effect sizes (Girolametto et al., 1998; Yoder et al., 2014), or b) the comparison groups were active control groups and were therefore inappropriate to include in the same meta-analysis as studies with no treatment control groups (Cleave et al., 2011). Based on *p* values and effect sizes calculated by the study authors (N=7), five studies showed significant effects. Based on Hedge's g calculated by the review authors (N=5; Baxter et al., 2018; Burgoyne et al., 2012; Cleave et al., 2011; Goetz et al., 2008; Sepulveda et al., 2013), two studies showed a significant effect overall (Baxter et al., 2018; Sepulveda et al., 2013); both studies were Randomised Controlled Trials (RCTs), and both targeted grammar (morphology/syntax/past tense). One study fell just short of significance on language outcomes overall (Burgoyne et al., 2012); this study targeted both reading and language. Two studies did not show a significant effect on language outcomes overall (Cleave et al., 2011; Goetz et al., 2008). One of these involved an active control group (Cleave et al., 2011), and the other (Goetz et al., 2008) primarily targeted reading rather than

language. Effect sizes of the single outcomes in each study are shown in supplementary Table 1b (also available online via OSF: https://osf.io/scxam/).

Follow up

Two out of eight studies reported long-term effects by follow up assessments a period of time after the intervention period and after post-intervention assessments (Cleave et al., 2011; Goetz et al., 2008). Both studies trained phonological awareness and found an increase in gains for the training relative to the control group at the follow up on one of the outcome measures. The gains at the follow up for both studies were on outcome measures for which the gains fell short of significance at the post-test assessment. The gain was for a directly trained measure in the Cleave et al. (2011) study and for a transfer measure in the Goetz et al. (2008) study.

Number of non-responsive children

Only two out of eight studies reported the number of children that were non-responsive to training (Cleave et al., 2011; MacDonald et al., 1974). In Cleave et al. (2011), six out of 16 children were non-responsive on the initial phoneme measure, and 12 out of 16 children were non-responsive on the rhyme measure. In MacDonald et al. (1974), none of the three participants were non-responsive.

Meta-analysis results

Across the four studies in the meta-analysis, the total number of participants in the training group was 72 children (sample size range = 8-28; mean age = 107 months, *SD*= 19 months), and the total in the control group was 69 children (sample size range = 7-26; mean

age = 111 months, *SD* = 24 months). The overall average effect size across the four studies was *g* = 1.01 (CI = -0.54, 2.57), indicating that children with Down syndrome can be responsive to language interventions and have the potential to experience large gains (Cohen, 1988). Summary effect sizes for each study and the overall effect size are displayed in the forest plot (see Figure 2).

[Insert Figure 2 here]

Heterogeneity

According to Higgins and Thompson (2002), the studies showed high heterogeneity (Q(3) = 39,241, $p < .01$) and large true variability (I2= 92,355).

Risk of bias results

Detailed information about the risk of bias results, based on the guidelines of the Cochrane Handbook (Higgins & Green, 2011; Higgins et al., 2017) can be found in Appendix C.

Selection bias

Two components contribute to the overall risk of selection bias:

1. *Random sequence allocation:* For five out of eight studies, the method of random sequence generation was unclear due to insufficient information about the allocation process. Two studies (Goetz et al., 2008; MacDonald et al., 1974) were deemed to be at a higher risk, as allocation was not stated to be randomised, while one study (Yoder et al., 2014) was deemed to be at a lower risk due to reporting the use of a random number generating programme.

2. *Allocation concealment:* For six out of eight studies, the method of allocation concealment was unclear due to insufficient information about the allocation process, while

the remaining two studies (Baxter et al., 2018; Burgoyne et al., 2012) were deemed to be at a lower risk, as participants were randomly allocated to one arm of the intervention after initial screening for eligibility.

Performance bias

Six out of eight studies were deemed to be at a higher risk of performance bias due to the likelihood that the implementer would know whether they were implementing the study or continuing business as usual. Of the remaining two studies, one (Cleave et al., 2011) was deemed to be at a lower risk, as personnel delivering the interventions were blinded, and one (Yoder et al., 2014) was deemed to have an unclear risk of performance bias, as the active control allows for potential blinding of the hypothesis, but blinding was not mentioned in the study.

Attrition bias

Seven out of eight studies were deemed to be at a lower risk of attrition bias, as there were no or very few instances of attrition for the outcome measures, and these instances were well justified (Burgoyne et al., 2012). The remaining study (Yoder et al., 2014) was deemed to be at a higher risk, as there was withdrawal as well as exclusions made (due to attending less than two-thirds of the intervention) and a lack of explanations for the withdrawal or low attendance.

Detection bias

Five out of eight studies were deemed to have an unclear risk of detection bias, as the blinding of outcome assessors was not stated in the study. One study (Cleave et al., 2011) was deemed to be at a lower risk, as the assessors were blind to the child's group assignment. Two studies (Girolametto et al., 1998; MacDonald et al., 1974) were deemed to be at a higher

risk, as the assessors for either all or some measures were the same individuals that implemented the intervention, meaning they would not be blind.

Reporting bias

All eight studies were deemed to be at a lower risk of reporting bias, as they provided results for all reported measures.

Overall risk of bias

In summary, all studies had two or more areas reviewed as a high risk or unclear risk due to missing information in the reporting of the study, with selection bias, performance bias, and detection bias representing the highest proportion of high or unclear risks of bias. For an overview of the overall risk of bias across studies see Appendix D.

Discussion

The results from the systematic review and meta-analysis showed that children with Down syndrome can benefit from receiving language intervention. The effect varied across studies and there were few common characteristics across effective interventions. Limited transfer effects were found, and the studies were reviewed to have a moderate to high risk of bias.

Do children with Down syndrome benefit from language interventions?

A positive effect in language was found for all but two of the included studies in the systematic review (Cleave et al., 2011; Goetz et al., 2008). The overall effect size in our meta-analysis is positive and large, showing children with Down syndrome can benefit from language interventions. However, large variation in effect sizes was found across the individual studies with a high degree of heterogeneity (Higgins and Thompson, 2002). Some

possible explanations for the variation in the results across the studies should be highlighted. While the reading intervention by Goetz et al. (2008) resulted in significant reading gains, no significant improvements were found on their language outcomes (phonological awareness). In contrast, the combined reading and language intervention by Burgoyne et al. (2012) did lead to significant gains in some language outcomes in addition to the significant reading gains. Therefore, results may be influenced by the main aim of the interventions. Furthermore, whether the language skills trained are rule-based or not might also be decisive, as all intervention studies targeting grammar showed significant effects, with especially large effect sizes in the larger scale intervention study by Baxter et al. (2018). Also, variance can be related to methodological aspects of the studies, such as controlling for age (Burgoyne et al., 2012) versus not controlling for age (Cleave et al, 2011), as well as measures not being sensitive enough to detect change in this group of children (e.g., floor effect at post-test (Goetz et al., 2008)). The type of control group may also play a role, where studies using an active control group, such as Cleave et al. (2011), are potentially less vulnerable to the Hawthorne effect, as the control group receives equal attention to the intervention group.

Few studies reported the near transfer effects of the intervention on untrained items and far transfer effects to other aspects of language. Overall, there was limited evidence of transfer effects from the trained materials to untrained items/materials or to other language components not focussed on in the intervention. Two exceptions were found; Baxter et al. (2018) found significant gains in near transfer measures of morpho-syntax as a result of training past tense. This highlights that in some instances, children with Down syndrome can apply language rules to other closely related abilities. Sepulveda et al. (2013) found that a better understanding of syntax/morphology may also improve children's understanding of language content. However, the results from Sepulveda may also reflect a degree of overlap between the semantic and syntax/morphology assessment, specifically the Bateria de

Lenguaje Objetiva Criterial (BLOC-C; Puyuelo et al., 1998). The semantic assessment in the BLOC-C included syntactic/morphological elements, which may explain why the training in morpho-syntax led to improvement on the semantic measure.

Few studies reported the number of non-responsive children, and the details concerning non-responsiveness were limited. Cleave et. al. (2011) reported that some of the children were responsive and others were not, while MacDonald et al. (1974) reported that all children were responsive. Cleave et al. (2011) reported particularly low levels of responsivity on the rhyme measure, which may reflect the pronounced difficulties in rhyme awareness that have previously been reported in the Down syndrome population (Snowling, Hulme, & Mercer, 2002). The overall findings across the studies revealed no conclusion on intervention responsiveness on an individual level (characteristics of the non-responsive individuals), which is essential for the applicability of the research, potentially enabling researchers to develop differentiated and more broadly effective interventions in the future, as well as providing more precise descriptions of the target group for the intervention.

The lack of transfer effects is in line with what is often seen in intervention studies for typically developing children (Barnett & Ceci, 2002; Melby-Lervåg, Redick, & Hulme, 2016). These findings indicate that intervention programmes should be focussed on directly training children to learn items and skills deemed to be the most relevant and useful. Additionally, these findings highlight the need for continuous intervention in order to maximise language progress. However, carrying out interventions across different contexts (e.g., in the classroom and various natural settings) may improve children's possibilities to transfer the trained language knowledge to new contexts.

To summarize, the results of the systematic review and meta-analysis indicate that targeted language interventions tend to be more effective than business as usual or another

comparison intervention for children with Down syndrome. In line with research on typically developing children, transfer effects of language interventions are limited. However, our results also indicate the effects for children with Down syndrome can accumulate over time (e.g., Cleave et al., 2011). Therefore, if time and effort are invested in the delivery of language intervention for children with Down syndrome, long-term effects may potentially be achieved.

What characterizes the language intervention studies?

There was large variation across the effective studies in terms of the implementer, targeted age group, intervention content, and intervention dosage.

Overall, interventions were most frequently implemented by teaching assistants; this caregiver group is likely to spend the most time with the child with Down syndrome in an educational setting (Webster & Blatchford, 2015). Schools were the most common setting for the interventions in this review, with half of the studies implemented in school. All but one of the remaining studies were carried out at home, but these tended to be in combination with other settings (childcare centre and clinic) with other professionals. Teachers and SLPs also typically spend a substantial amount of time with children with Down syndrome in language learning contexts (Næss et al., 2017). However, there were no interventions involving teachers and only two interventions involving SLPs. One of these interventions was solely implemented by SLPs, and one involved SLPs supporting the research staff in intervention implementation. The results from the systematic review and the meta-analysis indicate that all these groups of implementers can be successful in implementing language interventions with children with Down syndrome. However, based on the existing studies, it is not possible to conclude whether trained specialists such as SLPs, implementing the same interventions,

or in combination with parents, may promote more successful intervention outcomes than non-specialist trainers, such as parents alone or teaching assistants.

All the interventions targeted children across the ages of the early school years. However, based on non-significant results in the wide age-ranged sample of Cleave et al. (2011) and the impact of age on effect in Burgoyne et al. (2012), it could be suggested that more targeted interventions for specific age groups could be beneficial. For example, Cleave et al. (2011) included children aged 5 years to 16 years. As such, the children would most likely have very different needs and abilities, and tasks motivating for a pre-schooler would not necessarily be expected to motivate a teenager and vice versa. Sepulveda et al. (2013) was the only study other than Cleave et al. (2011) to extend to teenagers. Sepulveda et al. (2013) also included a relatively broad age range (6 years to 14 years), highlighting that teenagers were only targeted in studies that also involved younger children, and there were no studies focussing specifically on teenagers with Down syndrome. Younger children may be targeted more regularly due to the greater potential for knock-on effects for later development. However, some evidence from uncontrolled studies suggests language interventions are also beneficial at later ages (Buckley, 1995), and continued support is needed at later ages if skills are to be maintained (Chapman & Bird, 2011). This again reveals the need for more age-specific interventions for individuals with Down syndrome.

The content focussed on in the language interventions seems to be driven by descriptions of a specific language profile in children with Down syndrome in previous research literature. Language form was the most frequent language component targeted, which is an area found to be a specific weakness among children with Down syndrome (e.g., Næss, 2016), while vocabulary, which is often suggested to be a relative strength (e.g., Laws & Bishop, 2003; Roberts, Price, & Malkin, 2007), has not been focussed on to the same degree. However, a relative gap between receptive and expressive skills tends to be found

(e.g., Laws & Bishop, 2003), with many children with Down syndrome particularly struggling to express themselves in comparison to relatively greater receptive vocabulary abilities. Furthermore, the studies often showing receptive vocabulary as a strength have mainly investigated vocabulary breadth, such as how many words the child correctly selects in multiple choice picture vocabulary tests (e.g., Glenn & Cunningham, 2005), with no in-depth measure, such as an understanding of semantic relations (see Laws et al., 2015). Conversely, studies investigating vocabulary depth show a significant deficit in semantic knowledge (Laws et al., 2015; Smith & Jarrold, 2014), indicating superficial word knowledge limiting the functionality of their vocabulary skills and a dissociation between breadth and depth of vocabulary. Since vocabulary is often seen as a bottleneck for language development in general (e.g., Lee, 2011), these results highlight the need for more intervention studies focussing on both expressive and receptive vocabulary breadth and depth for children with Down syndrome.

No intervention targeted language use, which may reflect the findings of pragmatics being a *relative* strength in this population (Martin, Klusek, Estigarribia, & Roberts, 2009). However, pragmatic communication is nonetheless a weakness relative to mental age among children with Down syndrome (Smith et al., 2017) and should therefore be given more attention in future language interventions for individuals with Down syndrome.

There is large variability within the interventions regarding the intervention dosage, as such there are no common characteristics with regards to time and intensity. However, it should be noted that Yoder et al. (2014) found children with Down syndrome receiving a higher frequency of sessions (5 days a week) had greater vocabulary outcomes than those in an active control group receiving a lower frequency of intervention sessions (once per week). In line with Yoder et al. (2014), it may be suggested that high frequency intervention leads to

better outcomes. Therefore, it may be argued that lower frequency interventions could show increased gains if they were to increase intervention intensity.

Risk of bias

Overall, there was a high risk of bias across the studies in this review and many instances in which information was unclear. In particular, selection bias, performance bias, and detection bias were judged to be relatively unclear or high risk across all studies. To ensure replicability and high-quality research, it is essential that future language intervention studies involving children with Down syndrome aim to address these components of bias where possible and increase the level of reporting for these design aspects. The journals publishing these intervention studies therefore have an important role in insisting on high and consistent standards for reporting (Craig et. al., 2008).

Limitations, strengths, and future directions

The number of included studies was limited. A relatively large number of studies were excluded because they did not fulfil inclusion criteria related to study design. However, it is not possible to carry out a meaningful comparison across different designs since they are asking different research questions. For example, case series designs can answer research questions related to trends, but they are descriptive rather than able to determine cause and effect (Carey & Boden, 2003). In contrast, studies using control comparison groups ask questions about the gains in the intervention group relative to the gain in a control group, drawing inferences about cause and effect. As such, a separate review of studies using case series designs or baseline designs is appropriate.

Randomized Controlled Trial (RCT) studies are considered to be the gold standard for assessing treatment effects (see e.g., Hariton & Locascio, 2018). Of the included studies, the

majority reported using an RCT design. However, the relatively high risk of bias across the studies indicates that studies referred to as RCT do not necessarily meet the RCT standard according to the Cochrane guidelines (Higgins & Green, 2011; Higgins et al., 2017). Randomization only functions to control for differences between groups when the number of participants is very large. When conducting studies with smaller samples, other methods, such as a Single Case Research Design (SCRD), might therefore be a more appropriate alternative to the RCT (Gast & Ledford, 2014). However, it is important to note that designs like SCRD do not answer the exact same research questions, and the findings cannot be generalized in the same way as a well-designed RCT. This review also highlights the crucial need for well-powered, controlled language intervention studies of high quality, as this is vital in order to inform and deliver the most effective language interventions for children with Down syndrome in the future.

None of the studies in this review used technological approaches. Recent research indicates that technology (e.g., an iPad) can be used to successfully support learning for children with developmental disabilities (Kagohara et al., 2013). This highlights another future avenue to explore in the field of language interventions for children with developmental disabilities such as Down syndrome.

Because children with Down syndrome can be responsive to interventions that aim to improve language, they should receive research-based language intervention. However, more knowledge about the characteristics of the interventions that contribute to the effects, and what is effective for whom at what age and stage of development, is needed. In addition, more interventions that reach long-term and transfer effects are needed in future research to maximise children's language outcomes. Based on the limited number of studies and a moderate to high risk of bias across the studies, there is a great need for more robust

intervention studies, reducing the risk of bias, to ensure future language intervention practices are informed by high-quality research.

Conclusion

The findings from the present study suggest that providing interventions aiming to improve language for children with Down syndrome is more effective than business as usual or another comparison intervention. The Hawthorne effect (see Gast, 2010) could contribute to this outcome to some extent, where additional attention and new energy invested toward the child's language activities may motivate and encourage both the implementer and the child. The only common component across all the effective studies is the aim to improve language skills, highlighting the need for prioritisation of language interventions in the children's curriculum and in practice. Additionally, little is known about individual variation in response to interventions. At this time, interventions should therefore be provided to all children with Down syndrome, as responsiveness cannot be predicted.

References

Studies marked * are included in the systematic review, and studies marked ** are included in the meta-analysis.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612-637

**Baxter, R., Hulme, C., Rees, R. & Perovic, A. (2018, July). Evaluating a language intervention for children with Down syndrome. Poster presented at World Down syndrome Congress, Glasgow, UK.

Bloom, L., & Lahey, M. (1978). Language development and language disorders. New York: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). Introduction to meta-analysis. Chichester, West Sussex, UK: John Wiley & Sons.

Boudreau, D. (2002). Literacy skills in children and adolescents with Down syndrome. *Reading and Writing, 15,* 497–252.

Buckley, S. (1995) Improving the expressive language skills of teenagers with Down syndrome. *Down Syndrome Research and Practice*, *3*(3), 110-115. doi:10.3104/reports.57

**Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *53*(10), 1044-1053.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365.

Carey, T. S., & Boden, S. D. (2003). A critical guide to case series reports. *Spine*, *28*(15), 1631-1634.

Chapman, R. S. (1997). Language development in children and adolescents with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, *3*(4), 307-312.

Chapman, R. S., & Kay-Raining Bird, E. (2011). Language development in childhood, adolescence, and young adulthood in persons with Down syndrome. In J. Burack, R. Hodapp, G. Iarocci, & E. Zigler (Eds.), *The Oxford Handbook of Intellectual Disability and Development* (2nd edition, pp. 167 - 183). Toronto, Canada: Oxford University Press.

*Cleave, P. L., Bird, E. K. R., & Bourassa, D. C. (2011). Developing Phonological Awareness Skills in Children with Down Syndrome. *Canadian Journal of Speech-Language Pathology & Audiology*, *35*(4), 332-343.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological methods*, *8*(3), 243.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, *337*, a1655.

Eadie, P. A., Fey, M. E., Douglas, J. M., & Parsons, C. L. (2002). Profiles of grammatical morphology and sentence imitation in children with specific language impairment and Down syndrome. *Journal of Speech, Language, and Hearing Research*, *45*(4), 720-732.

Fowler, A. E. (1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. *Children with Down syndrome: A developmental perspective*, *9*, 302-328.

Gast, D. L. (2010). General factors in measurement and evaluation. In J. Ledford (Ed.), *Single subject research methodology in behavioral sciences* (pp. 91-109). New York, NY: Routledge.

Gast, D. L., & Ledford, J. R. (2014). Applied research in education and behavioral sciences. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology: Applications in special education and behavioral sciences,* (pp. 1-18). New York, NY: Routledge.

*Girolametto, L., Weitzman, E., & Clements-Baartman, J. (1998). Vocabulary intervention for children with Down syndrome: Parent training using focused stimulation. *Infant-Toddler Intervention: The Transdisciplinary Journal*, *8*(2), 109-25.

Glenn, S., & Cunningham, C. (2005). Performance of young people with Down syndrome on the Leiter-R and British picture vocabulary scales. *Journal of Intellectual Disability Research*, *49*(4), 239-244.

**Goetz, K., Hulme, C., Brigstocke, S., Carroll, J. M., Nasir, L., & Snowling, M. (2008). Training reading and phoneme awareness skills in children with Down syndrome. *Reading and Writing*, *21*(4), 395-412.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology*, *125*(13), 1716.

Hartup, W. W. (1983). Peer Relations. In E. M. Hetherington (Ed.), P. H. Mussen (Series Ed.), *Handbook of Child Psychology: Vol. 4. Socialization, personality and social development* (pp. 103-196). New York: Wiley.

Higgins, J. P. T., Altman, D.G., & Sterne, J. A. C. (Editors). Chapter 8: Assessing risk of bias in included studies. In Higgins, J. P. T., Churchill, R., Chandler, J. & Cumpston, M. S. (Eds), Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017), Cochrane, 2017. Retrieved from www.training.cochrane.org/handbook

Higgins, J. P. T., & Green, S. (Editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Retrieved from www.handbook.cochrane.org

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a metaanalysis. *Statistics in Medicine*, *21*, 1539-1558.

Hokstad, S., Smith. E., & Næss, K. B. (2019). *Implementation Quality in Language Intervention Research: A systematic review of previous studies of children with Down syndrome.* Manuscript in preparation.

Hulme, C., Goetz, K., Bringstocke, S., Nash, H.M., Lervåg, A. & Snowling, M. (2012). The growth of reading skills in children with Down syndrome. *Developmental Science*, *15*(3), 320-329.

Jamovi project (2017). jamovi (Version 0.8) [Computer Software]. Retrieved from https://www.jamovi.org

Kagohara, D. M., van der Meer, L., Ramdoss, S., O'Reilly, M. F., Lancioni, G. E., Davis, T. N., ... & Green, V. A. (2013). Using iPods® and iPads® in teaching programs for individuals with developmental disabilities: A systematic review. *Research in developmental disabilities*, *34*(1), 147-156.

Te Kaat-van den Os, , D. J. A., Jongmans, M. J., Volman, M. C., & Lauteslager, P. E. (2017). Parent-Implemented Language Interventions for Children with a Developmental Delay: A Systematic Review. *Journal of Policy and Practice in Intellectual Disabilities*, *14*(2), 129-137.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(1), 69-92.

Laws, G. (1998). The use of nonword repetition as a test of phonological memory in children with Down syndrome. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(8), 1119-1130.

Laws, G., & Bishop, D. V. (2003). A comparison of language abilities in adolescents with Down syndrome and children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1324-1339.

Laws, G., & Bishop, D. V. (2004). Pragmatic language impairment and social deficits in Williams syndrome: a comparison with Down's syndrome and specific language impairment. *International Journal of Language & Communication Disorders*, *39*(1), 45-64.

Laws, G., Briscoe, J., Ang, S. Y., Brown, H., Hermena, E., & Kapikian, A. (2015). Receptive vocabulary and semantic knowledge in children with SLI and children with Down syndrome. *Child Neuropsychology*, *21*(4), 490-508.

Lemons, C. J., & Fuchs, D. (2010). Phonological awareness of children with Down syndrome: Its role in learning to read and the effectiveness of related interventions. *Research in Developmental Disabilities*, *31*(2), 316-330.

*MacDonald, J. D., Blott, J. P., Gordon, K., Spiegel, B., & Hartmann, M. (1974). An experimental parent-assisted treatment program for preschool language-delayed children. *Journal of Speech and Hearing Disorders*, *39*(4), 395-415.

Martin, G. E., Klusek, J., Estigarribia, B., & Roberts, J. E. (2009). Language characteristics of individuals with Down syndrome. *Topics in Language Disorders*, *29*(2), 112-132.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512-534.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, *4*(1), 1.

Neil, N., & Jones, E. A. (2018). Communication intervention for individuals with Down syndrome: Systematic review and meta-analysis. *Developmental neurorehabilitation, 21*(1), 1-12.

Næss, K. A. B., Engevik, L. I., & Hokstad, S. (2017). Prioriteres språkstimulering i opplæringstilbudet til 3. klasse-elever med Down syndrom? *Norsk tidsskrift for logopedi, 3.* 6-13.

Næss, K. A. B., Nygård, E., Dolva, A. S., Ostad, J., & Lyster, S. (2016). The profile of Social Functioning in Children with Down Syndrome, *Disability and Rehabilitation*, 1-12.

Næss, K. A. B. (2016). Development of phonological awareness in Down syndrome: A meta-analysis and empirical study, *Developmental Psychology, 52(2),* 177-190.

Næss, K. A. B., Lyster, S. A. H., Hulme, C., & Melby-Lervåg, M. (2011). Language and verbal short-term memory skills in children with Down syndrome: A meta-analytic review. *Research in developmental disabilities*, *32*(6), 2225-2234.

Næss, K. A. B, & Smith, E. (2018). *The prevalence of reading skills in Down syndrome. What predicts early reading success?* Manuscript in preparation.

O'Toole, C., Lee, A. S., Gibbon, F. E., van Bysterveldt, A. K., & Hart, N. J. (2018). Parent-mediated interventions for promoting communication and language development in young children with Down syndrome. *Cochrane Database of Systematic Reviews (10).*

Puyuelo, M., Wiig, E. H., Renom, J. & Solanas, A. (1998). Bater´ıa de Lenguaje Objetiva y Criterial (BLOC). Barcelona: Masson.

Roberts, J. E., Price, J., & Malkin, C. (2007). Language and communication development in Down syndrome. *Mental retardation and developmental disabilities research reviews, 13(*1), 26-35.

Rynders, J., Abery, B. H., Spiker, D., Olive, M. L., Sheran, C. P., & Zajac, R. J. (1997). Improving educational programming for individuals with Down syndrome: Engaging the fullercompetence. *Down Syndrome Quarterly, 2*, 1–11.

**Sepúlveda, E. M., López-Villaseñor, M. L., & Heinze, E. G. (2013). Can individuals with Down syndrome improve their grammar?. *International journal of language & communication disorders*, *48*(3), 343-349.

Shelden, M. L., & Rush, D. D. (2001). The Ten Myths about Providing Early Intervention Services in Natural Environments. *Infants & Young Children, 14*(1), 1-13.

Smith, E., & Jarrold, C. (2014). Grouping, semantic relation and imagery effects in individuals with Down syndrome. *Research in developmental disabilities*, *35*(11), 3162-3174.

Smith, E., Næss, K. A. B., & Jarrold, C. (2017). Assessing Pragmatic Communication in Children with Down syndrome. *Journal of Communication Disorders, 68,* 10-23.

Smith, L., & van der Gaag, A. (2012). Speech and language therapy for children with Down syndrome: systematic literature review. Report prepared for Down syndrome Association.

Snowling, M. J., Hulme, C., & Mercer, R. C. (2002). A deficit in rime awareness in children with Down syndrome. *Reading and Writing, 15*(5), 471-495.

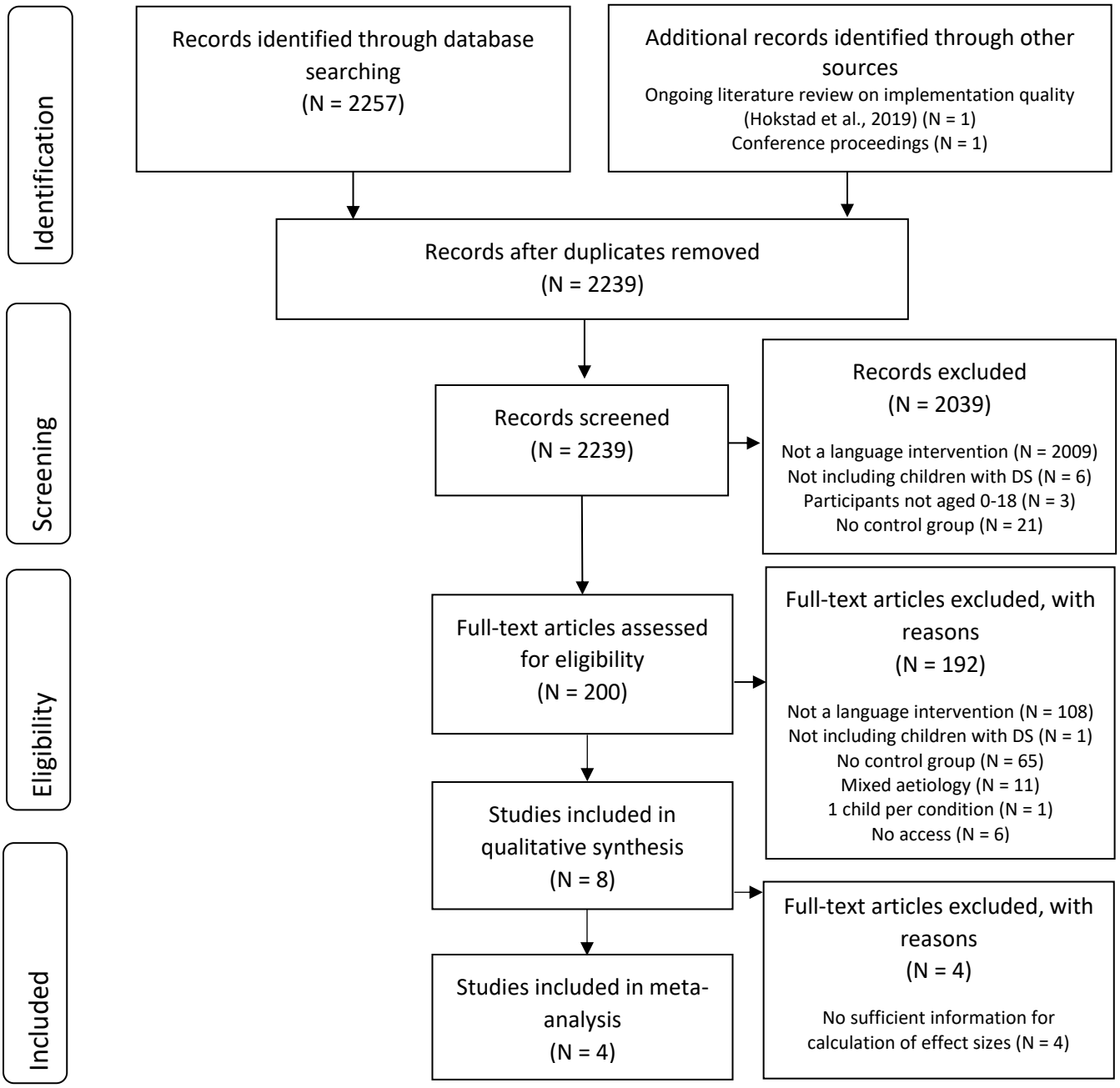
Te Kaat-van den Os, , D. J. A., Jongmans, M. J., Volman, M. C., & Lauteslager, P. E. (2017). Parent-Implemented Language Interventions for Children with a Developmental Delay: A Systematic Review. *Journal of Policy and Practice in Intellectual Disabilities*, *14*(2), 129-137.

Turner, S., & Alborz, A. (2003). Academic attainments of children with Down's syndrome: A longitudinal study. *British Journal of Educational Psychology*, *73*(4), 563-583.

Webster, R., & Blatchford, P. (2015). Worlds apart? The nature and quality of the educational experiences of pupils with a statement for special educational needs in mainstream primary schools. *British Educational Research Journal, 41(2),* 324-342.

*Yoder, P., Woynaroski, T., Fey, M., & Warren, S. (2014). Effects of dose frequency of early communication intervention in young children with and without Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *119*(1), 17-32.

Figures



Note. DS = Down syndrome.

Figure 1. Identification and screening process (based on Flow chart template from Prisma statement)

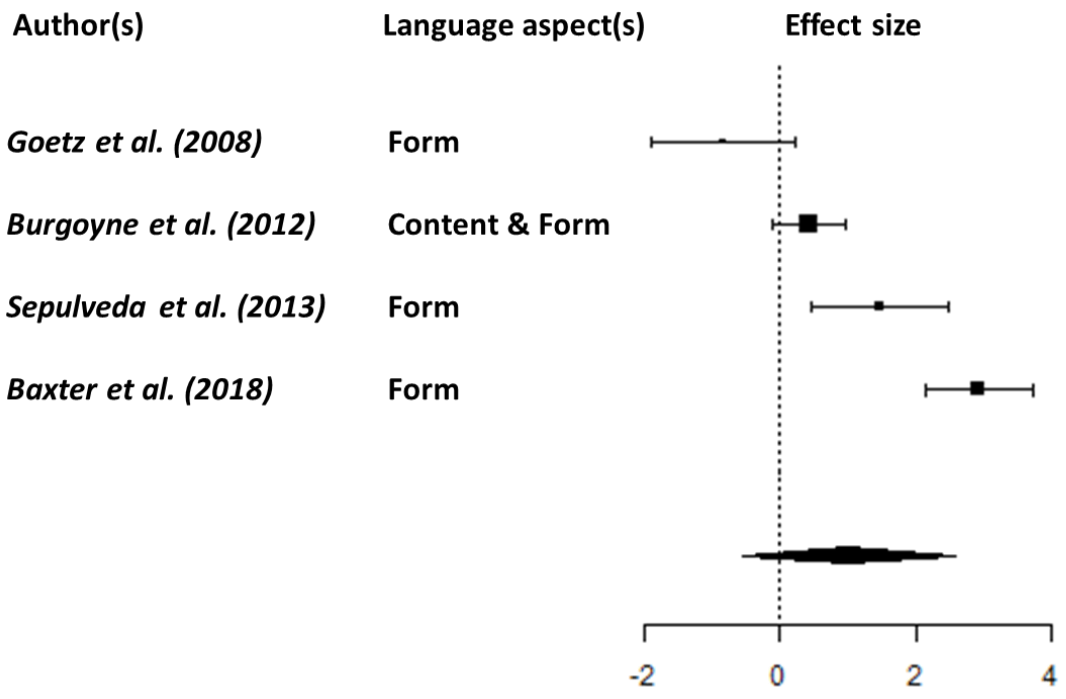


Figure 2. Forest plot showing relative weights, and overall average effect size (displayed by ◆); standardized mean difference (Hedge's g) with confidence intervals.

Table 1

Search strategy for PsychINFO

Database	Filters	Search strategy
PsychINFO via Ovid	Language (English Norwegian Swedish Danish)	1 (Language or Word* or Linguistic or Lexicon or Vocabulary or Semantic* or Phonology* or Phonemic or Expressive or Receptive or Grammar* or Morphology* or Syntax* or Pragmatic* or Comprehension or Production or Concept).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] 2 (Trial* or Outcome* or Effect* or Evaluate* or Improve or Intervention or Measure or Compare or Comparison or Randomised or Training or Enhance* or Efficacy or Report or Program or Teaching or Influence or Evidence or Control).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] 3 (Down* syndrome* or Trisomy 21).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] 4. 1 and 2 and 3

Note: Due to language constraints, the search was limited to studies published in English, and Scandinavian languages.

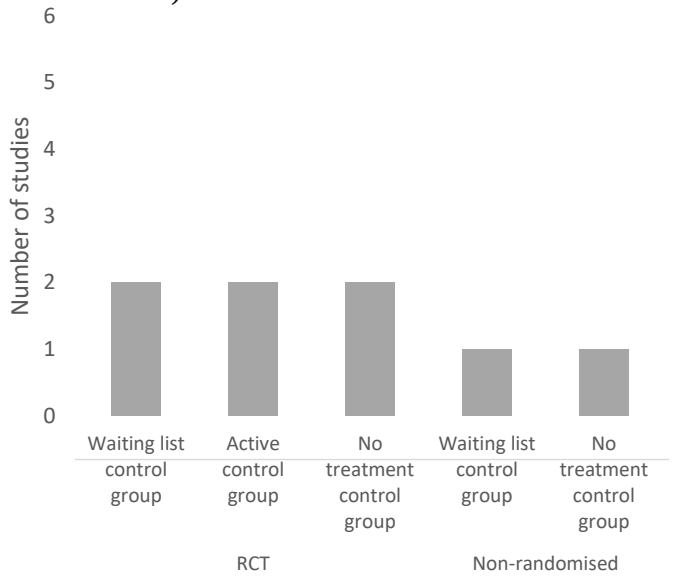
Appendices

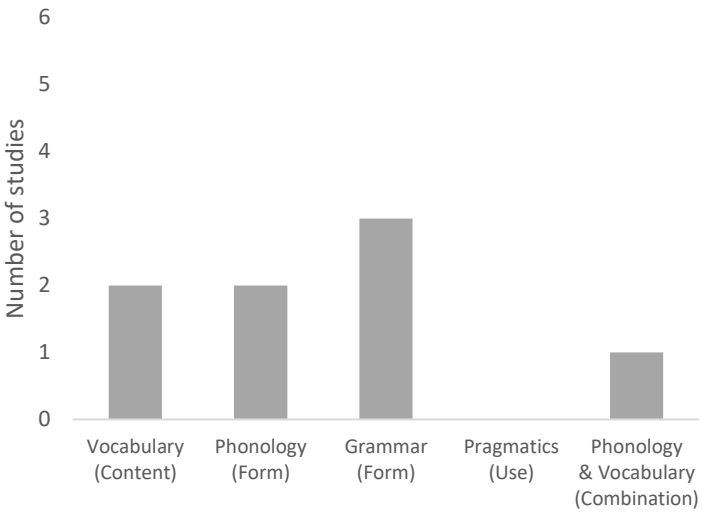
Appendix A. Summary effect size for each study in the meta-analysis

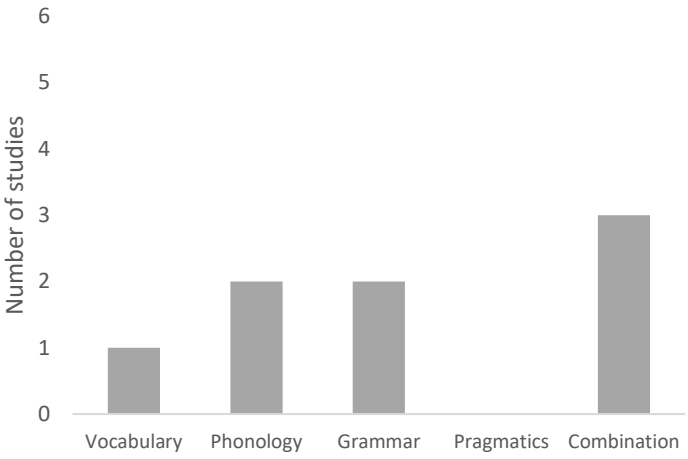
Table A1. *Summary effect size with 95% confidence intervals for each study*.

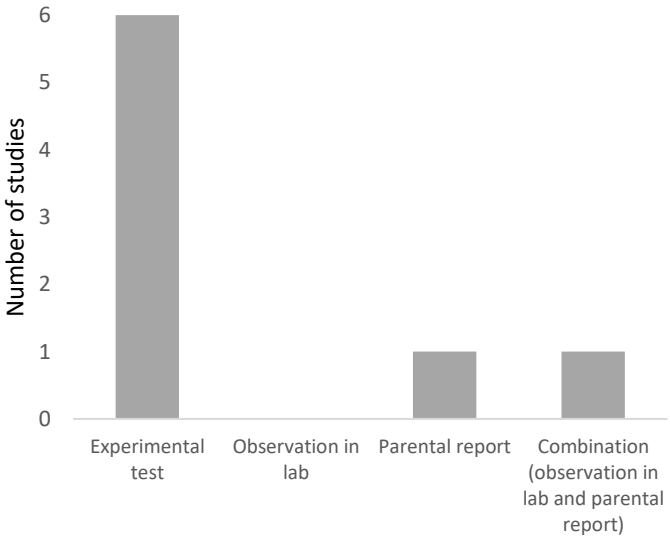
Authors:	*Effect size:*	*Lower CI:*	*Upper CI:*
Goetz et al. (2008)	-0.83	-1.89	0.23
Burgoyne et al. (2012)	0.44	-0.10	0.98
Sepulveda et al. (2013)	1.47	0.47	2.47
Baxter et al. (2018)	2.92	2.13	3.72

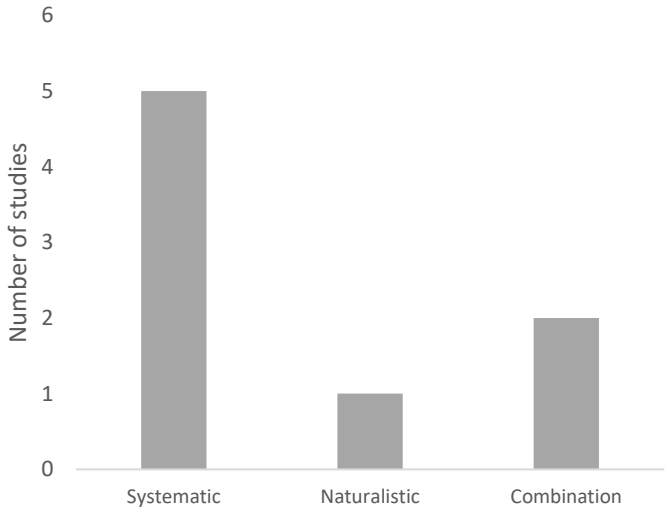
Appendix B. Characteristics of interventions included in the systematic review (Figures A1 and A2)



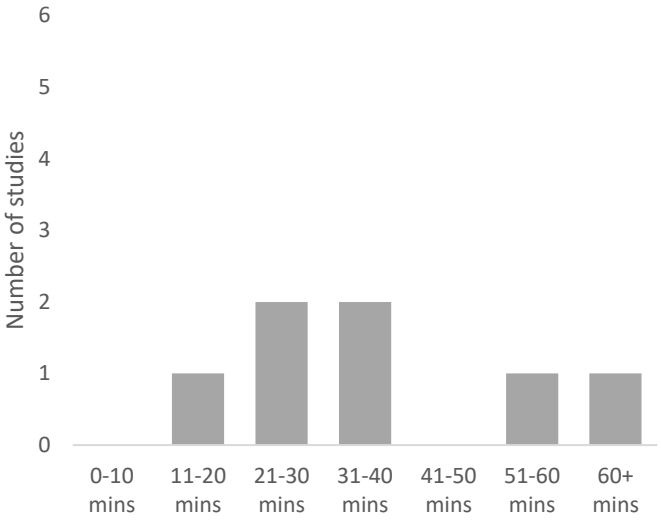


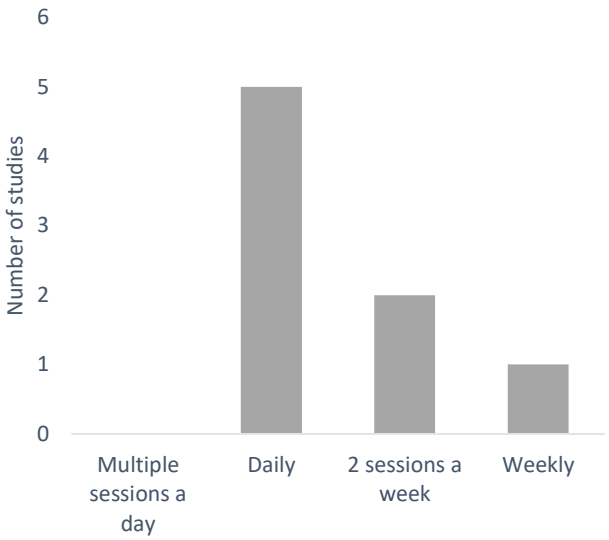


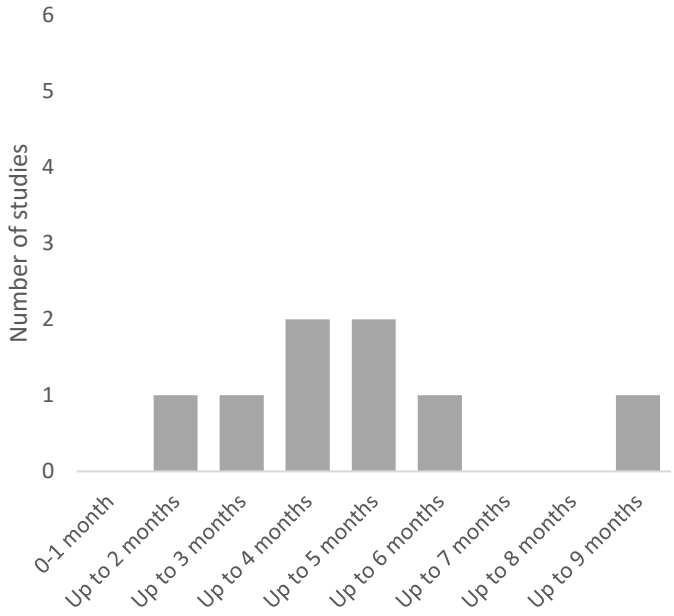


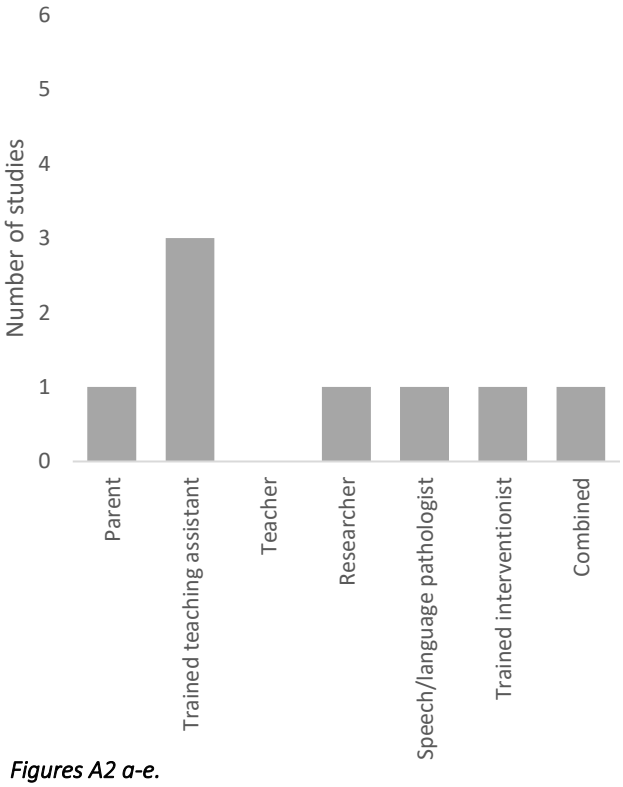


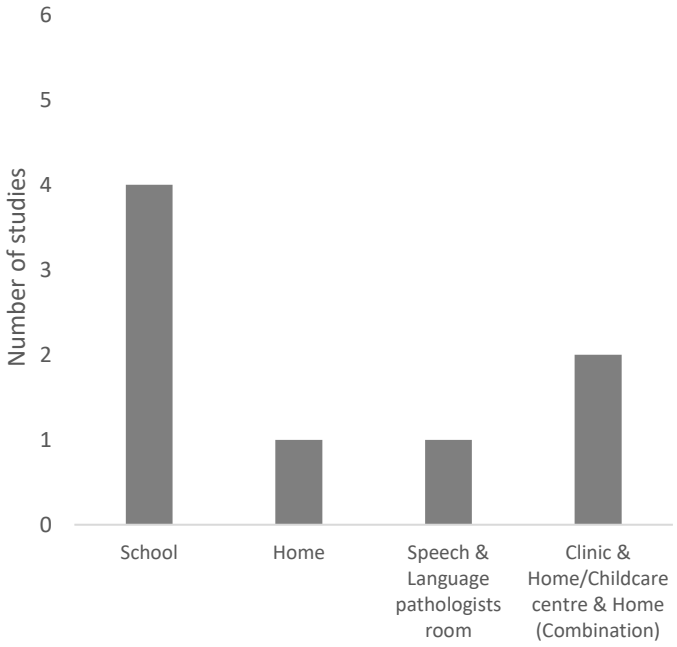
Figures A1 a-e. Frequency bar charts showing study characteristics across the 8 studies. Figure a (top left panel) shows the study designs used. Figure b (top right panel) shows the aspects of language trained. Figure c (middle left panel) shows the aspects of language assessed; note: combination consists of phonology and vocabulary (1 study), vocabulary and syntax (1 study), morphology, syntax, semantics/vocabulary and pragmatics (1 study)). Figure d (middle right panel) shows the different types of assessment tools that were used. Figure e (bottom left panel) shows the use of systematic versus a naturalistic intervention approach.











Figures A2 a-e.
Frequency bar charts showing study characteristics. Figure a (top left panel) shows number of studies for each session duration; note: one study (MacDonald et al., 1974) is not included as session duration was not reported in the study. Figure b (top right panel) shows number of studies within each range of intensity. Figure c (middle left panel) shows number of studies at each intervention length. Figure d (middle right panel) shows number of studies with each type of implementer; note: 'trained interventionist' refers to individuals with post-secondary education and experience working with young children with special needs, provided with training to deliver the programs, and combined here refers to trained professionals and parents. Figure e (bottom left panel) shows number of studies in each setting.

Appendix C. Risk of bias justifications for judgements of low risk (green boxes), unclear risk (yellow boxes), and high risk (pink/red boxes).

	Baxter et al. 2018	**Burgoyne et al. 2012**	**Cleave et al. 2011**	**Goetz et al. 2008**	**Girolametto et al. 1998**	**MacDonald et al. 1974**	**Sepulveda et al. 2013**	**Yoder et al. 2014**
Selection bias (Random sequence generation)	Insufficient information about the generation process	Insufficient information about the generation process	Insufficient information about the generation process	Probably not done, allocation not stated to be randomised	Insufficient information about the generation process	Probably not done, allocation not stated to be randomised	Insufficient information about the generation process	Generation process described. A computer programme using a random number generator was utilized
Selection bias (Allocation concealment)	Participants randomly allocated to one arm of the intervention after initial screening for eligibility	Participants randomly allocated to one arm of the intervention after initial screening for eligibility	Insufficient information about allocation concealment	Insufficient information about allocation concealment	Insufficient information about allocation concealment	Insufficient information about allocation concealment	Insufficient information about allocation concealment	Insufficient information about allocation concealment
Performance bias	Probably not blind as implementer will know whether he/she is implementing the study or continuing practise as usual	Probably not blind as implementer will know whether he/she is implementing the study or continuing practise as usual	Blinding of personnel: 'the interventionists were not told of the study`s hypothesis, but they were aware of the goals of the intervention programme they were implementing. They had no involvement in the testing and did not know the measures that were used to assess PA-skills.'	Probably not blind as implementer will know whether he/she is implementing the study or continuing practise as usual	Not blind because the mothers delivering the intervention were aware that it was a training programme	Not blind because trainers were involved in designing the programme and carried out assessments, and the mothers delivering the intervention were aware that it was a training programme	Probably not blind as implementer will know whether he/she is implementing the study or continuing practise as usual	Not stated in current paper but refers to Fey et al., 2013. Fey et al. do not specifically state that parents implementing training are blinded. Parents will know the dose they are giving, but they may not have been aware that this was a manipulation (i.e., being compared to the low dose active control group).

	Baxter et al. 2018	Burgoyne et al. 2012	Cleave et al. 2011	Goetz et al. 2008	Girolametto et al. 1998	MacDonald et al. 1974	Sepulveda et al. 2013	Yoder et al. 2014
Attrition bias	No attrition for outcome measure	Only 1/29 excluded from analyses in experimental group (participant moved schools) and 2/28 excluded from analyses in control group (1 participant moved schools, and 1 refused to participate in testing)	No attrition for outcome measure	No attrition for outcome measure	No attrition for outcome measure	No attrition for outcome measure	No attrition for outcome measure	Unclear how many participants have withdrawn from the study vs how many have been excluded from analysis. '12 children were excluded from the analysis due to failure to attend at least 2/3 of the intervention sessions or due to early withdrawal from the study.' It is unknown how many of these children had Down syndrome.
Detection bias	Not stated in paper	Not stated in paper	Blind outcome assessors: 'all testing was completed by speech-language pathologists who were blind to the child`s group assignment.'	Not stated in paper	Parents providing parent report are not blind as they are involved in implementing the intervention. Unclear if blinding of assessors for the remaining outcomes.	Experimenters carried out training as well as assessments, so blinding of assessors not possible	Not stated in paper	MB-CDI vocabulary checklist filled out by the parents. Unclear if they were present during intervention sessions. Method may be described more clearly in previous paper(s).
Reporting bias	Results for all measures provided	Results for all measures provided	Results for all measures provided	Results for all measures provided	Results for all measures provided	Results for all measures provided	Results for all measures provided	Results for all measures provided

Appendix D. Risk of bias summary