Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review

West, Gillian, Melby-Lervåg, Monica and Hulme, Charles

Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review

Gillian West[1]

Monica Melby-Lervåg[2]

Charles Hulme[1]

[1] Department of Education, University of Oxford (gillian.west@education.ox.ac.uk; charles.hulme@education.ox.ac.uk)

[2]Department of Special Needs Education, University of Oslo, Norway (monica.melby-lervag@isp.uio.no)

Running head: Procedural learning in language disorder

Word Count (excluding references and tables): 10,644 words

Author note:

This review was preregistered on the PROSPERO database (registration number: CRD42016048759). The datasets for all the meta-analyses in this paper are available in the project depository on the Open Science Framework (https://osf.io/cf6tn/).

Abstract

Impaired procedural learning has been suggested as a possible cause of developmental dyslexia (DD) and Developmental Language Disorder (DLD). We evaluate this theory by performing a series of meta-analyses on evidence from the six procedural learning tasks that have most commonly been used to test this theory: the serial reaction time, Hebb learning, artificial grammar and statistical learning, weather prediction and contextual cueing tasks. Studies using serial reaction time and Hebb learning tasks yielded small group deficits in comparisons between language impaired and typically developing controls (g = -0.30 and -0.32 respectively). However, a meta-analysis of correlational studies showed that the serial reaction time task was not a reliable correlate of language-related ability in unselected samples (r = 0.03). Larger group deficits were, however, found in studies using artificial grammar and statistical learning tasks (g = -0.48) and the weather prediction task (g = -0.63). Possible reasons for the discrepancy in results from different tasks that all purportedly measure procedural learning are highlighted. We conclude that current data do not provide an adequate test of the theory that a generalized procedural learning deficit is a causal risk factor for developmental dyslexia or developmental language disorder.

Key words: Meta-analysis; procedural learning; developmental dyslexia; developmental language disorder; specific language impairment **Abbreviations:** DLD

Is A Procedural Learning Deficit A Causal Risk Factor For Developmental Language Disorder Or Dyslexia? A Meta-Analytic Review

Proficient reading and oral language skills are critically important for school performance, employment prospects and psychosocial wellbeing. Unfortunately, many children struggle to develop such skills. Understanding the cognitive causes of such difficulties is a crucial step towards developing the best assessments and treatments for those at risk. According to the procedural deficit hypothesis (Nicolson & Fawcett, 2007; 2011; Ullman, 2004; Ullman & Pierpont, 2005) a key risk factor for language learning disorders, such as developmental dyslexia (DD) and developmental language disorder (DLD), is impaired procedural learning (a deficit in an unconscious learning system that is critical for abstracting the rule-based structures of language). This theory has generated much scientific debate. A search in Google Scholar of the terms "dyslexia + procedural + memory + deficit" returns over 9,000 results, while a similar search substituting "specific language impairment" for "dyslexia" returns over 6,000 results. However, results from studies examining the theory are highly inconsistent. Here we present a series of meta-analyses to evaluate the evidence relevant to this theory. We focus both on the nature of language learning impairments and on the mechanisms of language learning. Evaluating the evidence for a procedural memory impairment in language learning disorders is also relevant to a central theoretical debate within cognitive psychology: the putative distinction between procedural and declarative learning (or implicit (unconscious) versus explicit (conscious) learning).

Developmental disorders of language learning involve difficulty in processing linguistic information, which can affect production and understanding of both spoken and written language (Hulme & Snowling, 2009). Two of these disorders are the focus of much research: developmental language disorder and developmental dyslexia. Children with developmental language disorder (previously referred to as specific language impairment or SLI) have difficulty in learning language, and sometimes impaired processing and production of the sounds of speech, in spite of normal non-linguistic cognitive development (Bishop, 2006; Bishop, Snowling, Thompson, & Greenhalgh, 2016b). The patterns of impairment seen in developmental language disorder are highly variable (Bishop et al., 2016b; Hulme & Snowling, 2009) and these difficulties can persist into adulthood. Developmental dyslexia is characterized by problems in learning to decode printed words, which adversely affect reading fluency and spelling, as well as reading accuracy (Hulme & Snowling, 2016; Snowling, 2013; Vellutino, Fletcher, Snowling, & Scanlon, 2004).

Developmental language disorder and developmental dyslexia are distinct, but frequently comorbid disorders (Bishop & Snowling, 2004), not least because both reading and writing are scaffolded on oral language (Fletcher, 2009). They are also frequently comorbid with other developmental disorders, including attention deficit hyperactivity disorder (McGrath et al., 2011; Pennington, 2006) and developmental coordination disorder (Hill, 2001; Hulme & Snowling, 2009). Both disorders occur in around 3% - 7% of the population (Snowling, 1998; Tomblin et al., 1997), although prevalence estimates vary depending on the diagnostic criteria used (Snowling, 2013). The impairments seen in both disorders are dimensional in nature (Fletcher, 2009; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). People with developmental language disorder or developmental dyslexia represent the lower end of a normal distribution in spoken language or reading ability, rather than a distinct category. As such, research into these disorders and into the normal development of language and reading can be seen as two sides of the same coin.

The procedural deficit hypothesis

Much research into developmental language disorder and developmental dyslexia has focused on understanding their cognitive bases, not least because establishing the exact nature of the cognitive impairments that underpin these disorders is crucial for guiding effective interventions. A recent cognitive level explanation for developmental language disorder and developmental dyslexia claims that both disorders may arise from a deficit in a procedural memory system (Nicholson & Fawcett, 2007; Ullman & Pierpont, 2005).

The procedural deficit hypothesis claims that an implicit (procedural) memory system is involved in the learning, storage and retrieval of the statistically regular, rule-based, features of grammar and phonology (Ullman, 2004). It has been suggested that sequence-based implicit learning is particularly implicated in language disorders (Nicolson & Fawcett, 2010), while a declarative memory system functions normally and is responsible for the associative binding of phonological or orthographical representations and meanings (Ullman, 2004). The procedural deficit hypothesis proposes that lexical knowledge in developmental language disorder may often be less impaired than grammatical skills, because acquiring lexical knowledge depends on a relatively intact declarative memory system that may compensate for weaknesses in the procedural memory system (Ullman & Pierpont, 2005). We should note that this theory rests on making a distinction between stored and computed linguistic forms (lexicon vs. grammar) which depend for their acquisition on distinct memory systems (procedural vs. declarative). However, this claim is controversial and there is growing evidence against the existence of a clear distinction between words and larger sequences, and the mechanisms used to process them (e.g. McClelland et al., 2010; Snider & Arnon, 2012; Christiansen & Arnon, 2017).

It may be useful to clarify terminology. Procedural learning as it relates to the procedural deficit hypothesis of language learning disorders is also sometimes referred to as implicit or statistical learning. The terms procedural learning, implicit learning and statistical learning refer to the same incidental learning processes and are largely synonymous (Berry & Dienes, 1993; Perruchet & Pacton, 2006; Shanks, 2005) or at least over-lapping (Seger, 1994). A task is allegedly learned implicitly if procedural knowledge develops without, or at least before, any declarative knowledge: that is to say, a person has developed the ability to perform a task without, or before, being able to give an explicit account of what they are doing. In this review the terms implicit learning and procedural learning will be used interchangeably, in line with the terminology used in different groups of studies, but this should not be taken to imply that different learning mechanisms are involved in the different tasks used.

Measuring procedural memory

There are three important issues that should be noted before we proceed to consider the empirical evidence that has been used to test the theory. First, the procedural deficit hypothesis rests on the assumption that there are separable, but interconnected, procedural and declarative memory systems in the brain (Ullman, 2004). Such a claim remains controversial and is far from universally accepted (e.g. Perruchet & Pacteau, 1990; Henke, 2010).

Second, from a methodological perspective, it is far from clear which tasks actually measure implicit learning (e.g. Arnon, 2019; Erickson, Kaschak, Theissen, & Berry, 2016; Krishnan & Watkins, 2019; Siegelman, Bogaerts, & Frost, 2017; West, Vadillo, Shanks, & Hulme, 2018; 2019) and to what extent the different tasks used measure a common (procedural or implicit) learning mechanism. In this review we therefore take a broad perspective, including all tasks that have typically been used as measures of procedural learning in the literature (summarized in Table 1). In spite of the large variation in tasks, Ullman and Pierpont (2005) claim that one of the strengths of the procedural deficit hypothesis is that it makes “testable predictions” (p.

401) and proponents of this view assert that the tasks used are valid measures of procedural learning.

There are now many studies assessing procedural learning, language and language disorders using several different tasks. However, the results across studies are highly inconsistent (e.g. Conti-Ramsden, Ullman & Lum, 2015; Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014). These studies typically use extreme group designs (whereby a group of participants with severe reading or language difficulties is compared to a control group with normal reading or language skills). In addition, research using correlational designs that explore the relationship between implicit learning and language ability in samples unselected for ability are less plentiful, but findings are also mixed. A major aim of our review is to clarify what empirical claims are justified by results to date. Clarifying the pattern of empirical findings is critical for then moving on to reach conclusions about the theoretical status of the procedural deficit hypothesis.

Finally, a major logical problem for the procedural deficit hypothesis is that it claims to provide the same cognitive explanation for two disorders with markedly different cognitive profiles (developmental dyslexia and developmental language disorder). It is not clear how a common deficit in procedural learning can account for why some children develop a language disorder (a heterogeneous condition involving multiple problems in acquiring both receptive and expressive language skills that cut across the domains of syntax, semantics and phonology), while other children develop dyslexia (a specific deficit in learning to decode printed words accurately and fluently).

Reviews in this area also present inconsistent findings (e.g. Lum, Ullman, & Conti-Ramsden, 2013; Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Schmalz, Altoe, & Mulatti, 2016; van Witteloostuijn, Boorsma, Wijnen, & Rispens, 2017). Notably, all previous meta-analyses limit their coverage to a single language disorder and exclude measures that are relevant to the debate. It should also be noted that some have reported effect sizes that are almost certainly inflated (Lum et al., 2013; 2014; Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016). These previous meta-analyses calculated effect sizes using a single standard deviation for each group for the difference between sequenced and random trials (see Siegert et al., 2006). This method, however, will result in an overestimated effect size, since the standard deviation in the raw scores for each condition will be larger than the standard deviation of the difference scores (Lund, 1988; Morris & DeShon, 2002; Ray & Shadish, 1996). See online supplemental material 1 (Elaborated methodological information) for a more detailed discussion of this issue.

Scope and aims of the current review

The current paper takes a far wider view of implicit learning and language learning than any previous meta-analyses and examines some moderating factors that may explain why studies have reached different conclusions. In particular, each meta-analysis reported here includes studies of participants with developmental language disorder, as well as those with dyslexia. While previous meta-analyses have confined themselves to a single disorder, this exclusivity is questionable, given the heterogeneity of symptoms in language disorders (Petersen & Pennington, 2015; Webster & Shevell, 2004), and their frequent comorbidity (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005; Krishnan, Watkins, & Bishop, 2016). For example, in one study (Hedenius et al., 2013) participants categorized as dyslexic, displayed scores on a test of receptive grammar (TROG: Bishop, 1982) that were on average 17 points lower than the typically developing group. The TROG test is frequently used diagnostically for developmental language disorder. Additionally, the diagnostic criteria involved in categorizing participants to groups can also vary greatly from study to study.

Language-disordered groups from different studies may not, therefore, reflect the same behavioural symptoms or underlying cognitive impairments. Importantly, the procedural deficit hypothesis claims that deficits in the procedural memory system are the basis of the impairments seen in both disorders. Examining the moderating influence of disorder type may help to clarify the extent to which the symptoms of dyslexia and developmental language disorder relate to a common procedural memory impairment.

We also examine whether the severity of language disorder in extreme groups explains variation between studies. Since there is no agreed cut-off for dyslexia or developmental language disorder, studies use samples with differing severity of symptoms. The procedural deficit hypothesis predicts that the more severe the language learning disorder, the larger the deficit in procedural learning compared with controls should be. We, therefore, examine the extent to which severity of disorder is able to explain variation in procedural learning between studies.

Another important factor to consider is the type of task used. Previous meta-analyses have typically examined only one or two different tasks. The series of meta-analyses reported here includes all of the implicit learning tasks most commonly used in this area: serial reaction time tasks, artificial grammar and statistical learning tasks, Hebb learning tasks and probabilistic category learning tasks. Including a range of different tasks, each within its own separate meta-analysis, will allow us to examine differences between the tasks in one consistently conducted review and highlight whether some tasks demonstrate larger group differences than others. A meta-analysis of studies including contextual cueing tasks (n = 4) was not conducted, as there were insufficient data in the studies to calculate effect sizes.

We also examine possible moderators operating within each type of task (i.e. serial reaction time, Hebb, artificial grammar or statistical learning and weather prediction). Task modality, i.e. whether task stimuli are verbal or nonverbal, may influence study results. A key claim of the procedural deficit hypothesis is that any procedural deficit in language disorders is domain general in nature (Ullman & Pierpont, 2005). Impaired learning on predominantly nonverbal SRT tasks in language disordered groups has been interpreted as evidence of domain generality, but any group differences may be task specific in nature and therefore, misleading. We, therefore, investigate the moderating effect of task modality (verbal vs nonverbal) within the same experimental tasks where possible.

Participant age is another potential moderator. Previous research points mainly to one of two possibilities: that procedural learning is age invariant (e.g. Meulemans, Van der Linden, & Perruchet, 1998; Reber, 1993) or declines in adulthood (Janacsek, Fiser, & Nemeth, 2012; Zwart, Vissers, Kessels, Maes, 2019). The meta-analyses here include studies in both children and adults, enabling us to investigate this question.

The majority of extreme group design studies, but not all, reported no significant differences between groups in nonverbal IQ. However, even where group differences are not significant, this may be due to limited sample sizes, and group differences in nonverbal IQ are still potentially important. For this reason, the extent to which variation in the difference in nonverbal IQ between groups reflects variation in the effect size for procedural learning is also investigated, as is the relationship between nonverbal IQ and implicit learning in correlational studies.

Research questions in the current review

The research questions addressed in the current study are outlined below. Since the results from previous reviews are inconsistent, we choose not to provide hypotheses.

1) Is there evidence for a deficit on procedural learning tasks in groups with developmental

language disorder and groups with developmental dyslexia? If so, are equivalent relationships found in correlational studies that examine the full variation of language/reading skills and procedural learning?

2) Are participants with developmental language disorder or dyslexia differentially impaired on measures of procedural learning?

3) Is there variability in the severity of the impairment in language disordered groups across studies and, if so, is severity a moderator of the size of procedural memory deficits across studies?

4) To what extent is any procedural learning deficit associated with these disorders domain general (i.e. is it present with both verbal and nonverbal tasks) and is the size of the deficit related to whether a task measures implicit sequence learning, rather than, for example, probabilistic categorization or visual search efficiency?

5) Can other moderating factors, such as participant age or nonverbal IQ explain any of the inconsistency in results across studies?

Method

Inclusion criteria

The review was preregistered on the PROSPERO database (registration number: CRD42016048759) and the meta-analyses were conducted according to the recommendations of the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009). Also, additional material and the full data set is available in the project depository in the open science framework (www.osf.io).

To be included, a study had to report either 1) a group design that compared a measure of implicit learning in children or adults with developmental language disorder or dyslexia with performance of a control group(s) or 2) a correlational design that measured the relationship between performance on an implicit learning task and language measures in individual participants. In line with methodological recommendations for meta-analyses, group and correlational designs were entered into separate analyses (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Eligible studies needed to report data on at least one of the implicit learning tasks considered here (serial reaction time (SRT); Hebb learning; artificial grammar or statistical learning; probabilistic categorization; contextual cueing). See Table 1 for an explanation of these tasks. For the purposes of our meta-analyses we combined studies of artificial grammar learning (AGL: Reber, 1967) and statistical learning (Arciuli & Simpson, 2011), since these are highly similar tasks (although the extent to which they rely on the same mechanisms has been debated, see e.g. Misyak and Christiansen, 2012). The statistical learning task has been most frequently used with children, as it involves less complex rules than the artificial grammar learning task.

Group studies needed to include means and standard deviations for performance on the tasks to enable an effect size for the difference between groups to be calculated. Correlational studies had to include a measure of effect size (*r*) for the relationship between implicit learning and language skills. However, in practice many of the eligible studies did not include sufficient information. For serial reaction time tasks, for example, means and standard deviations for the task were usually reported only graphically. Whenever the relevant information was implied but not reported, the study was considered eligible in the first instance and the authors were contacted and additional data requested.

< Insert Table 1 approximately here >

Search Strategy

Searches were conducted using the following electronic databases: Medline, PsychInfo, Web of Science, ERIC ProQuest, Google Scholar and ProQuest Theses and Dissertations. Figure 1 shows the search and flow of studies through the review. The search strategy combined terms relating to implicit learning with terms relating to language and language disorder and was developed in collaboration with subject specialist librarians at University College London (see online supplementary materials 2, Table S1, for the search syntax).

Studies that shared authors, had equal number of participants, reported the same results or use the identical task measures were further investigated to limit the risk of coding the same data twice. Duplicate reports of the same study were treated as one collective report. Data reported in theses were excluded where they were possibly the same, or partially the same, as data reported in subsequent peer-reviewed papers included in the meta-analyses.

< Insert Figure 1 here >

Coding and meta-analytic procedure

Six separate meta-analyses were conducted, one for each of the main outcomes (see Table 2). Data were analysed using the Comprehensive Meta-Analysis programme (CMA: Borenstein, Hedge, Higgins, & Rothstein, 2005).

< Insert Table 2 approximately here >

Effect size calculation for meta-analyses of group design studies. The standardised mean difference in procedural learning between groups was coded for group designs, using Hedges' g to correct for small sample sizes (Hedges & Olkin, 1985). A negative g indicates that the language disordered group is performing more poorly than the control group. Hedges' g is interpreted in terms of standard deviation units, i.e. g = -1 indicates that the group with a learning disorder are one standard deviation below the scores of the control group.

Results from group design studies for serial reaction time and Hebb learning tasks represent a group difference in the difference between conditions. For the meta-analyses of studies using these tasks, effect sizes for the group difference in procedural learning on the tasks were, therefore, calculated from the means and standard deviations of the component conditions on the tasks (serial reaction time task: means and standard deviations in milliseconds for sequenced trials and for random trials; Hebb task: means and standard deviations as a percentage score for repeated and for non-repeated trials). This method of calculating an effect size is directly analogous to the way effect sizes are calculated for randomised control trials, otherwise known as pre-test post-test control group designs (Lund, 1988; Morris, 2008; Ray & Shadish, 1996). This is the experimental design that these extreme groups studies most closely resemble structurally (see Figure 2).

< Insert Figure 2 here >

Effect size calculation for meta-analyses of correlational studies. The correlations between procedural learning and language and /or decoding measures (Pearson's *r*) were coded directly into CMA where the calculations are done with Fishers *z* and then transferred back to Pearson's *r* to ease interpretation.

Mean effect size and heterogeneity. For all meta-analyses, random effects models in CMA were used to calculate weighted averages of individual effect sizes, in order to estimate an overall effect size for each meta-analysis. 95% confidence intervals are given for each pooled effect size. In studies with more than one outcome of the same construct, the composite score of the mean of the measures was calculated. The impact of any potential outliers was examined using sensitivity analyses, which give an adjusted overall effect size after removing studies one at a time. The variation in effects sizes between studies was examined, using the *Q*-test of homogeneity (Hedges & Olkin, 1985) and I^2 was used to examine the degree of any true heterogeneity that was not attributable to sampling error (Borenstein et al., 2009).

Moderator analyses. The potential role of categorical moderators was investigated using the grouping function in CMA, while meta-regression in CMA was used to examine potential continuous moderators. The following moderators were coded:

Age. Studies were categorized as including adult or child participants.

Type of learning disorder. the language disorder (dyslexia or developmental language disorder) was coded for all group design studies.

Severity of the disorder. Hedges' g was calculated for the differences in language or reading scores between the learning disorder group and controls. Thus, the larger the group difference, the more severe the disorder.

Nonverbal IQ. Measures of nonverbal IQ for experimental and control groups were coded. These were predominantly matrix reasoning tests, e. g. Raven's standard progressive matrices (Raven, 1986) or nonverbal IQ composites from standardized batteries (e.g. WISC-III: Wechsler, 1991).

Task specific moderators. For the serial reaction time task, the following moderators were coded: task type (deterministic, alternating or probabilistic); sequence type (first or second order conditional structure); sequence length, as well as the number of sequence repetitions prior to the introduction of the random sequence (deterministic tasks) or repetitions of the sequence across the task (alternating and probabilistic versions). For the Hebb learning tasks, the following moderators were coded: modality (verbal or non-verbal) and number of repetitions of the Hebb sequence. In addition, verbal tasks were further subdivided into auditory-verbal or visual-verbal tasks, although insufficient numbers of studies meant this could not be examined as a moderator. For artificial grammar and statistical learning tasks, modality (verbal or non-verbal, including a further distinction between visual-verbal and auditory-verbal tasks) and complexity (finite artificial grammar or simple triplet structure) were coded. For probabilistic classification tasks, modality (verbal or non-verbal), number of trials and variations in cue probabilities were coded.

Publication bias. Publication bias was assessed using funnel plots in meta-analyses with significant overall effect sizes, provided they contained sufficient numbers of studies. Lau, Ionnidis, Terrin, Schmid, and Olkin (2006) do not recommend the use of funnel plots in meta-analyses with less than ten comparisons. However, funnel plots for random effects models can be difficult to interpret visually (Lau et al., 2006). For this reason, a trim and fill analysis (Duval & Tweedie, 2000) was used to impute the missing values needed to make the funnel plot symmetrical and to calculate an adjusted overall effect size based on inclusion of these imputed studies. The strength of funnel plots is that they use only the data in the meta-analysis in order to investigate bias and are, therefore, entirely representative of the result of the meta-analysis. However, we also estimated *p*-curves for each meta-analysis, including all eligible studies to investigate the potential for bias in the published literature as a whole. These are included as online supplementary material 3, accompanied by disclosure tables (Tables S2 – S5) that detail the results for the principal measures of implicit learning in every eligible study.

Inclusion and coding reliability

To establish reliability of the inclusion of studies 25% of the abstracts were coded by two independent raters. The results showed a high interrater reliability, the agreement rate was 92%. For coding accuracy assurance, all the studies included in the meta-analyses were double coded by two independent raters. Overall, across all outcomes and moderators the agreement rate was

78% while the Pearson correlation was 0.86. For details about this for the different measures, see online supplemental material 4. Any disagreements were resolved by reviewing the original article and discussion.

Results

Characteristics of the studies in each meta-analysis are in online supplementary material 5 (Tables S6 – S12). The tables include all eligible studies, indicating those for which sufficient data for an effect size was forthcoming.

Serial reaction time task

Comparisons of language-disordered groups and age-matched controls. Fifty-two eligible studies were found for the meta-analysis of group design studies including deterministic, alternating and probabilistic serial reaction time tasks (see Table S6), but only 20 of these reported or were able to supply data as means and standard deviations by group for each of the sequence types separately. Nine further studies contained fully labelled figures of sufficient quality to enable data extraction using WebPlotDgitizer (Rohatgi, 2017). This meant that the final meta-analysis contained 29 independent comparisons of procedural learning with language-disordered groups and age-matched controls on serial reaction time tasks (see Figure 3).

< Insert Figure 3 approximately here >

In total these studies included 610 participants with language disorder (mean sample size = 21.03, *SD* = 9.69, range 7 to 46) and 698 control participants (mean sample size = 24.07, *SD* = 15.11, range 10 to 87). The overall mean effect size was significant, g = -0.30, 95% CI [-0.43, -0.16]. The variation in effect sizes was not significant, although approximately 39.93% of the variance was due to true variation between the studies rather than sampling error (Q (28) = 39.93, p = 0.07, I^2 = 29.87, Tau^2 = 0.04). One effect size in the analysis was much larger than the others, so a sensitivity analysis was conducted, which showed that the overall effect size was in the range of (g = -0.26, 95% CI [-0.38, -0.14] to g = -0.32, 95% CI [-0.45, -0.18]

The effect size was slightly lower when excluding the seven studies for which data had been digitally extracted from figures (k = 20, g = -0.26, 95% CI [-0.39, -0.13]).

Moderator analyses. Analysis of categorical moderator variables showed that the difference between samples with dyslexia and samples with developmental language disorder was not significant, Q (1) = 0.17, p = 0.68, g (dyslexia) = -0.28, k = 18, 95% CI [-0.46, -0.09], $p < .01$, g (language disorder) = -0.33, k = 11, 95% CI [-0.54, -0.13], $p < .01$. Studies testing children showed a larger effect size than studies with adults, although this moderator was not significant either, Q (1) = 0.48, p = .49, g (Adults) = -0.23, k = 10, 95% CI [-0.45, 0.01], p = .04, g (Children) = -0.34, k = 19, 95% CI [-0.50, -0.16], $p < .01$.

We also considered whether the type of serial reaction time task might have a moderating effect on the overall effect size (deterministic, adapted deterministic or alternating). The difference in effect size between the studies using different types of task was significant, Q (2) = 6.35, p = .04. The effect size for tasks with a standard deterministic structure was small to moderate and significant, g = -0.28, 95% CI [-0.42, -0.14], k = 22, $p < .01$. Two studies used an adapted deterministic structure (Sengottuvel & Rao, 2013; 2014), taking the measure for random trials towards the beginning of the task (rather than at the end) when task acclimatization and effects of motor learning may have differentially contributed to speed of response across groups. These studies had by far the largest effect size, g = -0.91, 95% CI [-1.54, -0.28], p = .01. The four comparisons using tasks with an alternating structure had a very small and nonsignificant mean effect size (but note that the power here is low), g = -0.04, 95% CI [-0.33, 0.24]. The potential moderating effect on the overall effect size of sequence length and number of repetitions of the sequence was considered for the deterministic tasks. Meta regressions showed that the effect of sequence length was not significant, *ß* = 0.00, $p = .54$, $R^2 = 0.00$, $k = 22$, neither was the number of sequence repetitions *ß* = 0.00, $p = .42$, $R^2 = 0.00$, $k = 22$.

We also examined whether the severity of language and decoding problems in the language-disordered samples related to group differences in procedural learning on the serial reaction time task. For language skills, there was a large amount of variation in the degree of difference between the language-disordered groups and the comparison groups. The mean difference was $g = -1.81$, 95% CI [-2.16, -1.45], ranging from $g = -0.27$ to $g = -4.29$ and there was significant heterogeneity between the studies Q (23) = 139.56, $p < .001$, $I^2 = 83.52\%$, $k = 24$, $\text{Tau}^2 = 0.63$. However, a meta-regression showed that the degree of severity of disorder in the disordered group did not explain any significant variation in the relationship between language ability and implicit learning on the task, *ß* = 0.43, $p = .52$, $R^2 = 0.00$, $k = 12$.

For decoding, the variation between the degree of difference between the disordered and comparison group was even more marked. The mean difference was $d = -2.20$ 95% CI [-2.59, -1.81], ranging from $d = -0.63$ to $d = -7.48$ and there was significant true heterogeneity between the studies Q (20) = 122.91, $p < .01$, $I^2 = 82.92\%$, $k = 22$, $\text{Tau}^2 = 0.67$. In this case, the meta-regression did show that the degree of severity of disorder in the disordered group did explain a very small amount of variation in the relationship between decoding ability and implicit learning on the task, *ß* = 1.45, $p = 0.02$, $R^2 = 0.05$. However, one study (Jiménez-Fernández et al., 2011) contained an extremely large effect size for the difference between groups on measures of word and non-word reading accuracy ($g = -7.48$). Excluding this study changed the results of the meta-regression, *ß* = 0.41, $p = .40$, $R^2 = 0.00$, such that severity of disorder, once again, no longer explained any of the variation in the relationship between decoding ability and serial reaction time performance.

Finally, there was variation between the language disordered and control groups on the measures of nonverbal IQ (NVIQ) used in the studies, mean difference was g = -0.35, 95% CI [-0.49, -0.20], ranging from g = 0.51 to g = -1.84. There was significant true variation in effect sizes, Q (31) = 61.78, p < .01, I^2 = 49.83%, k = 32, Tau^2 = 0.09. However, once again the meta-regression showed that the degree of disparity in NVIQ between groups did not explain significant variation in the relationship between NVIQ and implicit learning on the task, *ß* = 0.01, p = .96, R^2 = 0.0.

Publication bias. A funnel plot with a trim and fill analysis (Duval & Tweedie, 2000) showed no evidence of publication bias across the 29 studies entered into the meta-analysis (see Figure 7).

Correlational studies. Six studies examined the relationship between serial reaction time performance and language ability using correlational designs (see Table S7). Five of these studies, including 441 participants (mean sample size = 88.2, SD = 25.97, range = 58 to 120), included sufficient information to be entered into a meta-analysis that calculated the effect size *(r)* for the relationship between implicit learning on the serial reaction time task and measures of language and decoding. The pooled effect size was negligible and nonsignificant (r = 0.03, 95% CI [-0.06, 0.13]), with nonsignificant variability between samples (Q (4) = 3.12, p = 0.51, I^2 = 0.00%, Tau^2 = 0.00).

Three out of five of the studies also contained sufficient information to calculate an effect size for the relationship between NVIQ and serial reaction time task implicit learning performance. The overall effect size was not significant (r = 0.10, 95% CI [0.01, 0.21], p = .08), although with only 3 studies included in this analysis, power was low.

Summary. For the serial reaction time task we found evidence of a moderate group deficit in participants with language disorders (g = -0.30). However, concurrent correlational studies showed essentially no relationship between language abilities and performance on this task (*r* = 0.03). The discrepancy between these different forms of evidence is considered in the discussion.

Hebb serial order learning

Comparisons of language-disordered groups and age-matched controls. The meta-analysis included eight studies of Hebb learning tasks with language-disordered groups and age-matched controls, which contained ten independent comparisons in total (see Table S8). All 10 comparisons calculated effect sizes using the standard deviations of the non-repeating filler and Hebb sequences themselves, not the standard deviation for the difference between them. The studies included 200 participants with a diagnosis of language disorder (mean sample size = 22.22, *SD* = 5.74, range = 12 to 29) and 201 control participants (mean sample size = 22.33, *SD* = 6.58, range = 12 to 32).

< Insert Figure 4 approximately here >

Effect sizes with confidence intervals for the different studies are shown in Figure 4. The pooled effect size was significant, g = -0.32, 95% CI [-0.52, -0.12], with language disordered groups in these studies showing less facilitation on repeated lists compared to age-matched controls without language difficulties. The variation in effect sizes between studies was not significant Q (9) = 10.36, p = .32, I^2 = 13.10%, Tau^2 = 0.01. Once again, we also calculated the effect size for these studies using the standard deviations for the non-repeated condition for both groups, which increased the effect size slightly, g = -0.35, 95% CI [-0.56, -0.13], $p < .01$.

Overall, for the moderator analyses the true heterogeneity was 13% and not significant indicating that there is little true variation between studies that might be explained by moderators. The analyses showed that the difference between samples with dyslexia and samples with developmental language disorder was not significant, Q (1) = 0.03, p = 0.86, g

(dyslexia) = -0.33, *k* = 7, 95% CI [-0.58, -0.09], *g* (language disorder) = -0.29, *k* = 3, 95% CI [-0.69, 0.11], however power was extremely low in this analysis, as there were so few studies of samples with developmental language disorder. There was also no significant difference between samples in children or adults *Q* (1) = 0.15, *p* = .70, although the effect size in samples in adults was higher than in children *g* (Adults) = -0.36, *k* = 5, 95% CI [-0.65, -0.07], *g* (Children) = -0.28, *k* = 5, 95% CI [-0.58, 0. 02].

The majority of Hebb tasks were verbal tasks (either visual or auditory). Since several comparisons included both verbal and non-verbal tasks, the question of whether modality influenced effect size was examined by calculating the effect size for verbal and non-verbal tasks separately. Given the language-related difficulties of the experimental groups, it was expected that the difference between groups for verbal tasks would be greater than for the overall analysis, but it was actually smaller *g* = -0.23, *k* = 9, 95% CI [-0.43, -0.04]. There were only four comparisons for the nonverbal tasks, so power in this analysis was low, but the effect size was moderate, *g* = -0.66, *k* = 4, 95% CI [-1.09, -0.22].

The extent to which group differences in language skills and decoding, respectively, related to the group difference in Hebb serial order learning was also analysed. The mean difference in language skills between the groups was *g* = -1.57, 95% CI [-2.55, -0.60], ranging from *g* = -0.53 to *g* = -2.99. There was significant heterogeneity between the studies (*Q* (3) = 25.94, $p < .01$, I^2 = 88.44%, *k* = 4, Tau^2 = 0.87), however, a meta-regression showed that the severity of language disorder in the disordered group did not explain any significant variation in the relationship between language ability and implicit serial learning *ß* = -0.24, *p* = .88, R^2 = 0.00.

For decoding the mean difference was *g* = -1.90, 95% CI [-2.21, -1.59], ranging from *g* = -1.39 to *g* = -2.53, but the variation between the degree of difference between the disordered and comparison group in the studies was nonsignificant in spite of a high value for I^2 (*Q* (6) = 8.58,

p = .20, I^2 = 30.04%, k = 14, Tau^2 = 0.05). A meta-regression showed that the degree of severity of disorder in the disordered group did not explain any significant variation in the relationship between decoding ability and implicit learning on the task, $ß$ = 0.70, p = 0.37, R^2 = 0.00.

Finally, there was variation between the disordered and comparison groups for the measures of NVIQ used in the studies, mean difference was g = -0.29, 95% CI [-0.57, -0.01], ranging from g = 0.21 to g = -1.39. Even with so little overall true variance to explain, this variation in effect sizes between studies was significant, Q (9) = 20.74, p = 0.01, I^2 = 56.59%, k = 10, Tau^2 = 0.11. However, a meta-regression showed that the degree of disparity in NVIQ between groups did not explain any of the variation in the relationship between language-related ability and implicit learning on the task, $ß$ = -0.26, p = .59, R^2 = 0.00.

Publication bias. A funnel plot with a trim and fill analysis (Duval & Tweedie, 2000) showed no evidence of publication bias across the 8 studies entered into the meta-analysis (see Figure 7).

Summary. For the Hebb serial learning task we found evidence of a moderate deficit in groups with language disorders (g = -0.32).

Artificial grammar and statistical learning tasks

Comparisons of language-disordered groups and age-matched controls. Twenty-three of 31 eligible studies were entered into the meta-analysis (see Table S9), which included 30 independent comparisons of artificial grammar learning and statistical learning tasks with language-disordered groups and age-matched controls. The studies included 660 participants with language disorder (mean sample size 22.00, SD = 12.40, range = 12 to 77) and 907 control participants (mean sample size 30.23, SD = 28.71 range = 12 to 146).

Effect sizes with confidence intervals for the different studies are shown in Figure 5. The pooled effect size was moderate and significant, g = -0.48, 95% CI [-0.67, -0.29] confirming that overall language disordered groups performed more poorly on artificial grammar learning and statistical learning tasks than age-matched controls without difficulties. The variation in effect sizes between studies was large and significant, Q (29) = 91.10, p < .001, I^2 = 68.17%, k = 30, Tau^2 = 0.18. A sensitivity analysis showed that the overall effect size was in the range of g = -0.45, 95% CI [-0.64, -0.27] to g = -0.52, 95% CI [-.64, -0.40].

< Insert Figure 5 approximately here >

Moderator analyses. The analysis of categorical moderator variables showed that the difference between samples with dyslexia and samples with developmental language disorder was not significant, Q (1) = 0.29, p = 0.59, g (dyslexia) = -0.44, k = 16, 95% CI [-0.75, -0.13], g (language disorder) = -0.53, k = 14, 95% CI [-0.70, -0.37]. Although studies with adults showed a smaller effect size than studies with children, this moderator was not significant either, Q (1) = 0.99, p = .32, g (Adults) = -0.39, k = 15, 95% CI [-0.69, -0.08], g (Children) = -0.57, k = 15, 95% CI [-0.75, -0. 39]. The difference between studies using artificial grammar learning or statistical learning tasks was also not significant, Q (1) = 0.01, p = .93, g (AGL tasks) = -0.48, k = 16, 95% CI [-0.79, -0.16], g (SL tasks) = -0.49, k = 14, 95% CI [-0.66, -0. 32]. Finally, the difference between studies using verbal or non-verbal stimuli was examined. This required the exclusion of two comparisons that had administered tasks of more than one modality to the same participants (Evans, Saffran, & Robe-Torre, 2009: Experiment 2; Gabay, Thiessen, & Holt, 2015). Once again, the difference was not significant Q (1) = 0.24, p = .62, although the mean effect size for verbal tasks was significant and much larger than for nonverbal tasks, g (verbal) = -0.51, k = 18, 95% CI [-0.66, -0.35], p < 0.001, g (non-verbal) = -0.39, k = 10, 95% CI [-0.83, 0. 04], p = .08.

The extent to which severity of difficulties in language and decoding, respectively, in language disordered groups, related to the group difference in artificial grammar or statistical learning task performance was also analysed. For language skills, there was variation between the degree of difference between the language-disordered groups and the comparison group. The mean difference was g = -1.70, ranging from g = -2.12 to g = -1.27 and there was significant heterogeneity between the studies Q (15) = 76.16, $p < .001$, I^2 = 80.31%, k = 16, Tau^2 = 0.59. However, a meta-regression showed that the degree of severity of disorder in the disordered group did not explain any significant variation in the relationship between language ability and implicit learning on the task, *ß* = -0.20, p = .69, R^2 = 0.00.

For decoding there was a large variation between the degree of difference between the disordered and comparison group (these were predominantly studies investigating dyslexia), mean difference was g = -1.94, ranging from g = -2.37 to g = -1.51. For decoding ability there was significant heterogeneity between the studies Q (19) = 118.01, $p < .01$, I^2 = 83.90%, k = 20, Tau^2 = 0.76. Once again a meta-regression showed that the degree of severity of disorder in the disordered group did not explain significant variation in the relationship between decoding ability and implicit learning on the task, *ß* = -0.48, p = 0.45, R^2 = 0.02. One comparison (Nigro, Jiménez-Fernández, Simpson, & Defior, 2016: Experiment 2) contained extremely large effect sizes for the difference between groups on measures of word and non-word reading accuracy (g = -7.9). However, excluding this study did not significantly change the results of the meta-regression, *ß* = -0.61, p = .24, R^2 = 0.08.

Finally, there was variation between the disordered and control groups for the measures of NVIQ used in the studies, mean difference was g = -0.31, ranging from g = -0.48 to g = -0.13. This variation in effect sizes between studies was significant, Q (25) = 51.36, $p < 0.01$, I^2 = 51.32%, k = 26, Tau^2 = 0.10. On this occasion the meta-regression showed that the degree of disparity in NVIQ between groups explained 41% of the variation in the relationship between language-related ability and implicit learning on the task, *ß* = 0.61, p = .01, R^2 = 0.41.

Publication bias. A funnel plot with a trim and fill analysis (Duval & Tweedie, 2000) indicated the presence of publication bias in the effect size for artificial grammar and statistical learning group design studies (see Figure 7), suggesting that the true effect size in the meta-analysis should be much lower, adjusted point estimate g = -0.30, 95% CI [-0.47, -0.12].

Correlational studies. Five studies examined the relationship between performance on artificial grammar (k = 1) and statistical learning (k = 4) tasks and language ability using correlational designs, with a total of 6 independent samples (see Table S10).

The five studies, including 289 participants (mean sample size = 48.17, SD = 17.44, range = 30 to 72) were entered into a meta-analysis that calculated the effect size *(r)* for the relationship between implicit learning on the statistical learning task and measures of language and decoding. The pooled effect size in this meta-analysis was moderate and significant (r = 0.30, 95% CI [0.19, 0.41]). The variability across samples was not significant (Q (5) = 1.71, p = .88, I^2 = 0.00%, Tau^2 = 0.00). However, one of the studies (Qi, Sanchez Araujo, Georgan, Gabrieli, & Arciuli, 2019), with a sample of 72 adults and children, included more in-depth reading measures for children only. When adding these word and nonword reading measures and, therefore, necessarily only including the subset of children (n = 36) the pooled effect size reduced to (r = 0.25, 95% CI [0.12, 0.36]), with nonsignificant variability across samples (Q (5) = 2.82, p = .73, I^2 = 0.00%, Tau^2 = 0.00).

Summary. For the artificial grammar and statistical learning tasks we found evidence of a group deficit (g = -0.48) that was also supported by a modest correlation between language abilities and these tasks in samples unselected for ability (r = .30).

Weather prediction task

Comparisons of language-disordered groups and age-matched controls. Six studies were entered into the meta-analysis (see Table S11), which included 101 participants with language disorder (mean sample size = 20.02, *SD* = 5.81, range = 15 to 29) and 208 control participants (mean sample size = 41.60, *SD* = 32.58, range = 15 to 87). Effect sizes with confidence intervals for the different studies are shown in Figure 6. The pooled effect size was significant, $g = -0.63$, 95% CI [-1.07, -0.19] indicating that overall language disordered groups perform poorly on weather prediction tasks compared to age-matched controls without language difficulties. The variation in effect sizes between studies was also significant, Q (4) = 11.79, $p < .02$, $I^2 = 66.09\%$, $k = 5$, Tau^2 = 016.

< Insert Figure 6 approximately here >

Moderator analyses. Only one study (Gabay, Vakil, Schiff, & Holt, 2015) tested dyslexic participants, so participant diagnosis was not examined as a moderator. The moderating effect of participant age was examined, even though power in this analysis was low (Adults $k = 2$, Children $k = 3$). The difference between studies with adults and with children was not significant, Q (1) = 0.22, $p = .64$, g (Adults) = -0.79, 95% CI [-1.59, 0.01], g (Children) = -0.55 95% CI [-1.17, 0.07].

Only two studies reported data for language tests to accompany measures of effect size for the weather prediction task (with only one of these including decoding measures), so it was not possible to examine whether the severity of language disorder was related to performance on the weather prediction task. Three studies reported data for NVIQ measures, which showed that there was a large variation between the disordered and control groups for the measures of NVIQ used in the studies, the mean difference was $g = -1.16$, 95% CI [-1.99, -.33], ranging from $g = -0.36$, 95% CI [-1.06, 0.35] to $g = -1.27$, CI [-1.93, -0.61]. This variation in effect sizes between studies was significant, Q (2) = 9.07, p = .01, I^2 = 77.94%, k = 3, Tau^2 = 0.42. However, there were not enough studies in the analysis to be able to examine the effect of this difference in NVIQ group disparity on weather prediction task performance in a meta-regression.

Publication bias. There were too few studies to investigate the moderating influence of any task related variables or to examine publication bias using a funnel plot[1].

Summary. On the weather prediction task we found evidence for a large and significant (g = -0.63) group deficit in participants with developmental language disorder and with dyslexia. However, it should be stressed that there is considerable doubt about the extent to which performance on this task depends on procedural rather than declarative memory processes (Fotiadis & Protopapas, 2014; Knowlton, Squire, & Gluck, 1994)

Contextual cueing task

Comparisons of language-disordered groups and age-matched controls. Four studies have investigated whether there is a difference in implicit learning performance on the contextual cueing task between groups with dyslexia and controls (see Table S12 for study characteristics), two with children (Jiménez-Fernández et al., 2011; Staels & Van Den Broeck, 2017) and two with adults (Howard et al., 2006; Bennett et al., 2008). The studies reported insufficient data for the contextual cueing tasks to perform a meta-analysis, but all reported comparable levels of implicit learning for both groups on the task. Indeed, Staels and Van den Broeck (2017) reported greater levels of facilitation for their group with dyslexia compared to typically developing children, but this was not significant. However, this result was found alongside

[1] Funnel plots containing fewer than 10 studies are considered unreliable indicators of publication bias (Lau et al., 2006).

better than chance performance on a post-test generation task for both groups, suggesting the facilitation effect over the course of the task was at least partially due to explicit learning.

Summary. There are only 4 studies that have examined differences between participants with language learning disorders and controls on the contextual cueing task. These studies did not provide sufficient information to allow a meta-analysis to be performed. However, there is no evidence for a deficit on the contextual cueing task in participants with language learning disorders.

Procedural learning deficits across multiple tasks

Although impaired procedural learning in language disorders is sometimes found on both verbal and non-verbal tasks, a question still remains over the domain-generality of any implicit learning impairment. The procedural deficit hypothesis is embedded in the classic multiple systems model of memory which postulates the existence of two separable memory subsytems (procedural versus declarative memory). This theory suggests that any procedural memory deficit should be domain-general, reflecting the role of a common procedural memory system that could be tapped by diverse memory tasks, involving both verbal and nonverbal stimuli. Impairments in procedural learning in language disorders would, therefore, be expected on both verbal and nonverbal procedural memory tasks.

Most studies examining the procedural deficit hypothesis have used a single implicit learning task and do not address the extent to which impairments in a language-disordered group generalize to other, similar tasks. However, fourteen group design studies (see Table 3) have used multiple tasks. Unfortunately, domain generality cannot be examined with meta-analytic techniques because most studies do not report correlations between tasks. Nevertheless, it is useful to examine the extent to which any implicit learning impairment extends across tasks. Table 3 suggests considerable task specificity, with impaired learning at a group level almost always being confined to a single task.

< Insert Table 3 approximately here >

The idea that implicit learning is task specific is also supported by two correlational studies (Schmalz, Moll, Mulatti, & Schulte-Körne, 2019; West et al., 2018). Both studies found no correlation between implicit learning measures from a serial reaction time and an artificial grammar learning task and from the serial reaction time, Hebb learning and contextual cueing tasks, respectively.

Discussion

The procedural deficit hypothesis of language learning disorders embodies two inter-related claims. First, that the human memory system is organized into separable procedural and declarative memory systems (Squire, 2004; Squire & Dede, 2015). Second, that a deficit in the procedural system is causally related to the language learning difficulties seen in children with developmental language disorder (Ullman, 2004; Ullman & Pierpont, 2005) and the difficulties in learning to decode print seen in dyslexia (Nicolson & Fawcett, 2007; 2011). Our review provides little support for this theory.

Major findings

We reported separate meta-analyses of extreme group studies of four different tasks used to assess group deficits in procedural learning (the serial reaction time task, the Hebb task, artificial grammar or statistical learning tasks, and the weather prediction (probabilistic category learning) task). Participants with language learning disorders tend to perform poorly on all four tasks compared to age-matched controls. We should stress, however, that even if such group deficits could be shown to be genuine and specific they may just as easily be a consequence of a learning disorder (language learning disorder → procedural learning deficit) as a cause (procedural learning deficit → language learning disorder).

Power in a meta-analysis will increase as the number of studies included increases, and if individual studies have larger sample sizes. The meta-analyses of extreme group studies reported here differ markedly in the number of studies included and the total sample sizes. The meta-analyses of the SRT task (29 comparisons; 1308 participants) and artificial grammar and statistical learning tasks (30 comparisons; 1567 participants), therefore, most likely give a more accurate indication of true effect sizes than those from the much smaller samples available for the Hebb (10 comparisons; 410 participants) or the weather prediction tasks (5 comparisons; 309 participants).

We found evidence of a group deficit in children with language disorders on the serial reaction time (g = -0.30) and Hebb (g = -0.32) tasks. However, the size of these group deficits are small if they reflect an important causal risk factor. We might contrast these findings with the case of the phoneme awareness deficit seen in dyslexia (which is postulated to be a major causal risk factor). The group difference in phoneme awareness between children with dyslexia and controls was reported to be d = -1.37 (Melby-Lervåg, Lyster & Hulme, 2012). Thus, if poor procedural learning is an important causal risk factor for dyslexia and DLD, we would expect larger group deficits. Moreover, studies of samples unselected for language ability showed essentially no correlation between language abilities and performance on the serial reaction time task (r = 0.03). As we discuss below, this casts serious doubt on the reliability of the findings from extreme group designs using this task.

There was a moderate deficit in extreme group designs using artificial grammar or statistical learning tasks (g = -0.48). However, the accompanying funnel plot supports the suggestion from a recent meta-analysis of artificial grammar learning in dyslexia (Schmalz et al., 2016), that the true effect size in these studies is likely to be small. Nevertheless, our meta-analysis of correlational studies yielded a moderate effect size (r = .30); which does suggest a genuine relationship between artificial grammar or statistical learning and language and reading problems. However, this may simply reflect a deficit in verbal learning amongst children with language learning disorders, since the effect size in studies using verbal stimuli is far larger than in studies using nonverbal stimuli.

Our final meta-analysis assessed evidence for impaired probabilistic category learning in the weather prediction task. The overall effect size was large and significant (g = -0.63). However, there is considerable doubt about the extent to which performance on this task reflects implicit rather than declarative memory processes (Fotiadis & Protopapas, 2014; Knowlton et al., 1994).

Methodological Issues

Limitations of extreme group designs. The majority of studies reviewed here involve extreme group designs in which groups with language learning disorders are compared to control groups. A second, less common, approach is to examine the correlation between measures of procedural learning and reading or language skills in samples unselected for ability. These two approaches should yield converging evidence. Reading and language skills are continuously distributed traits (like height and weight). Causal influences on reading or language development should therefore correlate with reading or language ability in representative (unselected) samples, as well as showing mean differences between groups at the extremes of the distribution.

The results for the serial reaction time task from extreme group designs and samples unselected for ability yield contradictory results: a group deficit in participants with language disorders, but no equivalent correlation in representative samples. In our view, this inconsistency likely reflects methodological problems inherent in extreme group studies. Extreme group designs suffer from regression to the mean and typically over-estimate the true linear relationship between variables (Preacher, Rucker, MacCallum, & Nicewander, 2005) and are frequently under-powered, leading to publication bias because only positive (not null) results get published (Button et al., 2013). In addition, both the serial reaction time and Hebb learning tasks, have poor reliability (Bogaerts, Siegelman, Ben-Porat & Frost, 2017; West et al., 2018) which may reflect the fact that difference scores (Lord, 1958; Overall & Woodward, 1975) and scores derived from short tasks tend to be unreliable (Brysbaert & Stevens, 2018; Nunally & Bernstein, 1994). Unfortunately, none of the studies reviewed here reported the reliability of their tasks.

We also need to consider comorbidities amongst participants in extreme group designs. Reading and language disorder are both highly comorbid with a range of other disorders, including problems of attention control (ADHD; Hulme & Snowling, 2009). Deficits on procedural learning tasks in children with language learning disorders may therefore reflect problems of attending during the task, rather than being a correlate of reading difficulties per se. These issues are probably better addressed in studies of unselected groups where reliable measures of reading, language and attention can be related to measures of procedural learning (see West, Shanks & Hulme, 2021).

One possible explanation for a discrepancy in results from studies using extreme groups versus samples unselected for ability might appeal to the existence of non-linear effects. Children or adults at the bottom the distribution of reading (dyslexia) or language (DLD) might have deficits on the serial reaction time task, while in the rest of the population there is no linear relationship between the two measures. Only one study to date has investigated this possibility and found no support for it (West, Shanks, & Hulme, 2021). Furthermore, even if such nonlinear effects existed (with associations in the bottom end of the distribution of reading/language skills only) we would nevertheless expect to find correlations (albeit perhaps weak ones) depending upon details of the distributions involved.

Theoretical Issues

Task "purity" and psychometric issues in identifying a procedural memory system. The procedural deficit hypothesis is founded on the claim that there are separable procedural and declarative memory systems (also known as implicit and explicit memory). It follows from this theory that diverse measures of procedural memory should tap into a common underlying procedural memory system (just as diverse measures on an IQ test should load onto a common psychometric "g" factor). Although, there are only 14 relevant studies (see Table 3) we found no evidence that any procedural learning deficit is domain-general; where more than one measure of procedural learning is used, there are typically inconsistent effects on the different tasks between groups with and without language learning difficulties.

It is important to emphasize that procedural (implicit) learning tasks are not process pure (e.g., Cleeremans, Destrebecqz, & Boyer, 1998; Reber, 1989; Shanks & St. John, 1994). For example, performance on artificial grammar and serial reaction time tasks may reflect conscious learning of fragmentary chunks of information (Buchner, Steffens, Erdfelder, & Rothkegel, 1997; Perruchet & Pacteau, 1990). Although few studies of procedural learning in dyslexia, or developmental language disorder, include separate measures of declarative learning (such as serial or free recall (see Table S6 in the online supplemental materials)), those that do, typically find poorer declarative learning in language disordered groups (Conti-Ramsden, Ullman, & Lum, 2015; Deroost et al., 2010; Henderson & Warmington, 2017; Lum & Bleses, 2012; Lum, Conti-Ramsden, Page, & Ullman, 2012; Lum, Gelgic, & Conti-Ramsden, 2010; Stoodley et al., 2006). Furthermore, a previous meta-analysis (Melby-Lervåg, Lyster & Hulme, 2012) demonstrated that children with dyslexia do poorly on one measure of verbal declarative memory (verbal memory span (d = 0.71), as do children with developmental language disorder (e.g. Gathercole & Baddeley, 1990).

Finally, many studies reviewed here simply assess a single measure of procedural learning (for example, serial reaction time) in a group of children with developmental language disorder (e.g. Perlant & Largy, 2011). Such a design cannot establish a selective deficit in procedural learning. Establishing a differential deficit requires that groups are compared on at least two tasks that purportedly measure different abilities; though making direct inferences about differences in ability from differences in task performance is fraught with difficulties (Chapman & Chapman, 1973). If children with language learning disorders do poorly on a serial reaction time task this could have many explanations: they might perform more poorly than controls on any cognitive task. Furthermore, even if children with language learning disorders really do perform more poorly on the serial reaction time task than on a single measure of declarative memory, this might reflect specific features of the two tasks, rather than differences in the underlying constructs of procedural versus declarative memory. It will be essential for future studies in this area to use multiple measures of multiple constructs if they wish to make claims about specific deficits in a procedural memory system. The measures used must also have adequate reliability, since differences in reliability can lead to spurious conclusions about differences in performance between groups (more reliable tasks are more likely to detect group differences).

Our finding that there is considerable task specificity in implicit learning in language disorders (see Table 3) is consistent with evidence from several large-scale correlational studies. Gebauer and Mackintosh (2007) found no relationship at all between performance on a serial reaction time task, an artificial grammar learning task, and a complex systems process control task in teenagers. Similarly, Pretz, Totz, and Kaufman (2010) failed to find any signficant relationship between serial reaction time and artificial grammar learning in adults. There is also mounting evidence that the same implicit learning tasks correlate poorly with one another across visual and auditory modalities (Erikson et al., 2016), as well as across verbal and nonverbal modalities (West et al., 2018). These findings all cast doubt on the psychometric validity of a procedural memory construct as assessed by current, widely used, measures.

Summary and Conclusions

The procedural deficit hypothesis of language learning disorders embodies two inter-related claims, both of which are controversial. First, that the human memory system is organized into separable procedural and declarative systems (Squire, 2004; Squire & Dede, 2015) and second, that a deficit in procedural memory is a causal risk factor for developmental language disorder (Ullman, 2004; Ullman & Pierpont, 2005) and dyslexia (Nicolson & Fawcett, 2007; 2011). In our view, it is impossible to reach clear conclusions about this hypothesis based on current evidence.

Theoretically, the claim that the human memory system is organized into separable procedural and declarative systems remains highly controversial (e.g. Perruchet & Pacteau, 1990; Henke, 2010). One major obstacle to progress is that existing measures of procedural learning are highly impure and, possibly as a consequence of this task impurity, tend to have very poor reliability. Many have noted that it is far from clear which tasks actually measure implicit learning (e.g. Arnon, 2019; Erickson, Kaschak, Theissen, & Berry, 2016; Krishnan & Watkins, 2019; Siegelman, Bogaerts, & Frost, 2017; West et al., 2018) and to what extent the different tasks used actually measure a common (procedural or implicit) learning mechanism. These problems of measurement need to be resolved before the procedural deficit hypothesis of language learning disorders can be adequately tested.

We have outlined at length a range of methodological limitations that plague this area. Future studies will need to use multiple measures of multiple constructs (procedural learning, declarative learning, and attention at a minimum) in order to demonstrate specific deficits. Our conclusion is that current evidence simply does not provide an adequate test of the procedural deficit hypothesis of language learning disorders.

References

*Aguilar, J. M., & Plante, E. (2014). Learning of grammar-like visual sequences by adults with and without language-learning disabilities. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1394-1404. https://doi.org/10.1044/2014_JSLHR-L-13-0124

*Archibald, L. M., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory & Cognition*, *41*(2), 268-280. https://doi.org/10.3758/s13421-012-0259-4

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464-473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

*Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286-304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behaviour Research*, 52, 68–81. https://doi.org/10.3758/s13428-019-01205-5

Bahl, M., Plante, E., & Gerken, L. (2009). Processing prosodic structure by adults with language-based learning disability. *Journal of Communication Disorders*, *42*(5), 313-323. https://doi.org/10.1016/j.jcomdis.2009.02.001

*Bennett, I. J., Romano, J. C., Howard Jr, J. H., & Howard, D. V. (2008). Two forms of implicit learning in young adults with dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 184-198. https://doi.org/10.1196/annals.1416.006

Berry, D., & Dienes, Z. P. (1993). *Implicit learning: Theoretical and empirical issues*. Hove. UK: Erlbaum.

Bishop, D. V. M. (1982). Test for Reception of Grammar (Medical Research Council, Oxford).

Bishop, D. V. (2006). What causes specific language impairment in children?. *Current Directions in Psychological Science*, *15*(5), 217-221. https://doi.org/10.1111/j.1467-8721.2006.00439.x

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858. https://doi.org/10.1037/0033-2909.130.6.858

Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016b). CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. Terminology. *PeerJ Preprints*, *4*, e2484v1. https://doi.org/10.7287/peerj.preprints.2484v2

*Bogaerts, L., Siegelman, N., Ben-Porat, T., & Frost, R. (2017). Is the Hebb repetition task a reliable measure of individual differences in sequence learning?. *The Quarterly Journal of Experimental Psychology*, 1-35. https://doi.org/10.1080/17470218.2017.1307432

*Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, *43*, 106-122. https://doi.org/10.1016/j.ridd.2015.06.012

Bogaerts, L., Szmalec, A., De Maeyer, M., Page, M. P., & Duyck, W. (2016). The involvement of long-term serial-order memory in reading development: A longitudinal study. *Journal of Experimental Child Psychology*, *145*, 139-156. https://doi.org/10.1016/j.jecp.2015.12.008

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis version 2. *Englewood, NJ: Biostat*, *104*.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons, Ltd.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1): 9, 1-20 http://doi.org/10.5334/joc.10

Buchner, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 631-663. https://doi.org/10.1080/713755723

Bussy, G., Krifi-Papoz, S., Vieville, L., Frenay, C., Curie, A., Rousselle, C., ... & Herbillon, V. (2011). Apprentissage procédural implicite dans la dyslexie de surface et la dyslexie phonologique. *Revue de Neuropsychologie*, *3*(3), 141-146. https://doi.org/10.3917/rne.033.0141

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), 365-376. https://doi.org/10.1038/nrn3475

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. Advances in Methods and Practices in Psychological Science, 2(2), 115-144. https://doi.org/10.1177/2515245919847196

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378-1396. https://doi.org/10.1044/1092-4388(2005/096)

Chapman, L. J., & Chapman, J. P. (1973) Problems in the measurement of cognitive deficit. *Psychological Bulletin*. 79(6), 380-395. https://doi.org/10.1037/h0034541

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, 9(3), 542-551. https://doi.org/10.1111/tops.12274

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28-71. https://doi.org/10.1006/cogp.1998.0681

Clark, G. M., & Lum, J. A. (2017a). First-order and higher order sequence learning in specific language impairment. *Neuropsychology*, *31*(2), 149. https://doi.org/10.1037/neu0000316

Clark, G. M., & Lum, J. A. (2017b). Procedural memory and speed of grammatical processing: Comparison between typically developing children and language impaired children. *Research in Developmental Disabilities*, *71*, 237-247. https://doi.org/10.1016/j.ridd.2017.10.015

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406-416. https://doi.org/10.1016/S1364-6613(98)01232-7

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01090

*Deroost, N., Zeischka, P., Coomans, D., Bouazza, S., Depessemier, P., & Soetens, E. (2010). Intact first-and second-order implicit sequence learning in secondary-school-aged children with developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *32*(6), 561-572. https://doi.org/10.1080/13803390903313556

Desmottes, L., Maillart, C., & Meulemans, T. (2017). Memory consolidation in children with specific language impairment: Delayed gains and susceptibility to interference in implicit sequence learning. *Journal of Clinical and Experimental Neuropsychology*, *39*(3), 265-285. https://doi.org/10.1080/13803395.2016.1223279

Desmottes, L., Meulemans, T., & Maillart, C. (2016b). Implicit spoken words and motor sequences learning are impaired in children with specific language impairment. *Journal of the International Neuropsychological Society*, *22*(5), 520-529. https://doi.org/10.1017/S135561771600028X

*Desmottes, L., Meulemans, T., & Maillart, C. (2016a). Later learning stages in procedural memory are impaired in children with specific language impairment. *Research in Developmental Disabilities*, *48*, 53-68. https://doi.org/10.1016/j.ridd.2015.10.010

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

*Du, W. (2013). *Associative implicit learning in adult dyslexic readers* (Doctoral dissertation, University of Strathclyde). https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.581926

Du, W., & Kelly, S. W. (2013). Implicit sequence learning in dyslexia: a within-sequence comparison of first-and higher-order information. *Annals of Dyslexia*, *63*(2), 154-170. https://doi.org/10.1007/s11881-012-0077-1

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Erickson, L., Kaschak, M., Thiessen, E., & Berry, C. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra: Psychology*, *2*(1). https://doi.org/10.1525/collabra.41

*Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-335. https://doi.org/10.1044/1092-4388(2009/07-0189)

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society*, *15*(4), 501-508. https://doi.org/10.1017/S1355617709090900

Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory & Cognition*, *42*(1), 112-125. https://doi.org/10.3758/s13421-013-0350-5

*Gabay, Y., Schiff, R., & Vakil, E. (2012a). Dissociation between online and offline learning in developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *34*(3), 279-288. https://doi.org/10.1080/13803395.2011.633499

*Gabay, Y., Schiff, R., & Vakil, E. (2012b). Dissociation between the procedural learning of letter names and motor sequences in developmental dyslexia. *Neuropsychologia*, *50*(10), 2435-2441. https://doi.org/10.1016/j.neuropsychologia.2012.06.014

*Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934-945. https://doi.org/10.1044/2015_JSLHR-L-14-0324

*Gabay, Y., Vakil, E., Schiff, R., & Holt, L. L. (2015). Probabilistic category learning in developmental dyslexia: Evidence from feedback and paired-associate weather prediction tasks. *Neuropsychology*, *29*(6), 844. https://doi.org/10.1037/neu0000194

Gabriel, A., Maillart, C., Guillaume, M., Stefaniak, N., & Meulemans, T. (2011). Exploration of serial structure procedural learning in children with language impairment. *Journal of the International Neuropsychological Society*, *17*(02), 336-343. https://doi.org/10.1017/S1355617710001724

*Gabriel, A., Maillart, C., Stefaniak, N., Lejeune, C., Desmottes, L., & Meulemans, T. (2013). Procedural learning in specific language impairment: effects of sequence complexity.

Journal of the International Neuropsychological Society, *19*(3), 264-271. https://doi.org/10.1017/S1355617712001270

*Gabriel, A., Meulemans, T., Parisse, C., & Maillart, C. (2015). Procedural learning across modalities in French-speaking children with specific language impairment. *Applied Psycholinguistics*, *36*(3), 747-769. https://doi.org/10.1017/S0142716413000490

*Gabriel, A., Stefaniak, N., Maillart, C., Schmitz, X., & Meulemans, T. (2012). Procedural visual learning in children with specific language impairment. *American Journal of Speech-Language Pathology*, *21*(4), 329-341.https://doi.org/10.1044/1058-0360(2012/11-0044)

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language*, *29*(3), 336-360. https://doi.org/10.1016/0749-596X(90)90004-J

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 34. https://doi.org/10.1037/0278-7393.33.1.34

Gould, J. H., & Glencross, D. J. (1990). Do children with a specific reading disability have a general serial-ordering deficit?. *Neuropsychologia*, *28*(3), 271-278. https://doi.org/10.1016/0028-3932(90)90020-O

*Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, *39*(2), 158-170. https://doi.org/10.1016/j.jcomdis.2005.11.004

Hall, J., Owen Van Horne, A., McGregor, K. K., & Farmer, T. (2017). Distributional learning in college students with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3270-3283. https://doi.org/10.1044/2017_JSLHR-L-17-0013

Hall, J., Owen Van Horne, A. J., McGregor, K. K., & Farmer, T. A. (2018). Individual and developmental differences in distributional learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 694-709. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0134

*Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.12734

*He, X., & Tong, S. X. (2017). Quantity matters: Children with dyslexia are impaired in a small, but not large, number of exposures during implicit repeated sequence learning. *American Journal of Speech-Language Pathology*, *26*(4), 1080-1091. https://doi.org/10.1044/2017_AJSLP-15-0190

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F Delafresnaye (Ed.).*Brain Mechanisms and Learning*, 37-46. Oxford, UK: Blackwell.

*Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities*, *34*(11), 3924-3935. https://doi.org/10.1016/j.ridd.2013.08.014

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., ... & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities*, *32*(6), 2362-2375. https://doi.org/10.1016/j.ridd.2011.07.026

Hedges, L. V. &. Olkin, I. (1985). Statistical methods for meta-analysis. *Orlando, FL: Academic Press*.

*Henderson, L. M., & Warmington, M. (2017). A sequence learning impairment in dyslexia? It depends on the task. *Research in Developmental Disabilities*, *60*, 198-210. https://doi.org/10.1016/j.ridd.2016.11.002

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*(7), 523. https://doi.org/10.1038/nrn2850

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149-171. https://doi.org/10.1080/13682820010019874

Holl, A. K., Wilkinson, L., Tabrizi, S. J., Painold, A., & Jahanshahi, M. (2012). Probabilistic classification learning with corrective feedback is selectively impaired in early Huntington's disease—Evidence for the role of the striatum in learning with feedback. *Neuropsychologia*, *50*(9), 2176-2186. https://doi.org/10.1016/j.neuropsychologia.2012.05.021

Howard, J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia*, *44*(7), 1131-1144. https://doi.org/10.1016/j.neuropsychologia.2005.10.015

*Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352-365. https://doi.org/10.1111/desc.12125

*Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, *5*, 175. https://doi.org/10.3389/fpsyg.2014.00175

Hulme, C., & Snowling, M. (2016) Reading disorders and dyslexia. Current Opinion in Paediatrics, 28, 731–735. DOI:10.1097/MOP.0000000000000411

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and cognition*. Oxford: Wiley-Blackwell.

*Iao, L. S., Ng, L. Y., Wong, A. M. Y., & Lee, O. T. (2017). Nonadjacent Dependency Learning in Cantonese-Speaking Children With and Without a History of Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *60*(3), 694-700. https://doi.org/10.1044/2016_JSLHR-L-15-0232

Inácio, F., Faísca, L., Forkstam, C., Araújo, S., Bramão, I., Reis, A., & Petersson, K. M. (2018). Implicit sequence learning is preserved in dyslexic children. *Annals of dyslexia*, *68*(1), 1-14. https://doi.org/10.1007/s11881-018-0158-x

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496-505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Jiménez, L., & Vázquez, G. A. (2008). Implicit sequence learning in a search task. *Quarterly Journal of Experimental Psychology*, *61*(11), 1650-1657. https://doi.org/10.1080/17470210701695801

*Jiménez-Fernández, G., Vaquero, J. M., Jiménez, L., & Defior, S. (2011). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of Dyslexia*, *61*(1), 85-110. https://doi.org/10.1007/s11881-010-0048-3

*Kahta, S., & Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia*, *66*(2), 235-250. https://doi.org/10.1007/s11881-016-0121-7

Katan, P., Kahta, S., Sasson, A., & Schiff, R. (2017). Performance of children with developmental dyslexia on high and low topological entropy artificial grammar learning task. *Annals of Dyslexia*, *67*(2), 163-179. https://doi.org/10.1007/s11881-016-0135-1

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, *8*(1), 43-52. https://doi.org/10.1002/dys.208

*Kemény, F., & Lukács, Á. (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 249-258. https://doi.org/10.1080/13803390902971131

*Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171. https://doi.org/10.1037/a0025405

*Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184-193. https://doi.org/10.1111/cdev.12461

*Kidd, E., & Kirjavainen, M. (2011). Investigating the contribution of procedural and declarative memory to the acquisition of past tense morphology: Evidence from Finnish. *Language and Cognitive Processes*, *26*(4-6), 794-829. https://doi.org/10.1080/01690965.2010.493735

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106-120. http://www.learnmem.org/cgi/doi/10.1101/lm.45903.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2016). Neurobiological basis of language learning difficulties. *Trends in Cognitive Sciences*, *20*(9), 701-714. https://doi.org/10.1016/j.tics.2016.06.012

Krishnan, S., & Watkins, K. E. (2019). A challenge for the procedural deficit hypothesis: how should we measure sequential learning in childhood?. *Developmental Science*, e12815. https://doi.org/10.1111/desc.12815

*Laasonen, M., Väre, J., Oksanen-Hennah, H., Leppämäki, S., Tani, P., Harno, H., ... & Cleeremans, A. (2014). Project DyAdd: Implicit learning in adult dyslexia and ADHD. *Annals of Dyslexia*, *64*(1), 1-33. https://doi.org/10.1007/s11881-013-0083-y

Iao, L. S., Ng, L. Y., Wong, A. M. Y., & Lee, O. T. (2017). Nonadjacent dependency learning in Cantonese-speaking children with and without a history of specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(3), 694-700. https://doi.org/10.1044/2016_JSLHR-L-15-0232

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. *BMJ: British Medical Journal*, *333*(7568), 597. https://doi.org/10.1136/bmj.333.7568.597

Lee, J. C. (2012). Are individual differences in language associated with differences in the corticostriatal system? A behavioral and imaging study (Doctoral dissertation, The University of Iowa). Retrieved from https://doi.org/10.17077/etd.quz11znx

Lee, J. C., Mueller, K. L., & Tomblin, J. B. (2016). Examining procedural learning and corticostriatal pathways for individual differences in language: testing endophenotypes of DRD2/ANKK1. *Language, Cognition and Neuroscience*, *31*(9), 1098-1114. https://doi.org/10.1080/23273798.2015.1089359

*Lee, J. C., & Tomblin, J. B. (2015). Procedural learning and individual differences in language. *Language Learning and Development*, *11*(3), 215-236. https://doi.org/10.1080/15475441.2014.904168

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology*, *49*(3), 150. https://doi.org/10.1037/h0045456

*Lukács, Á., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, *28*(3), 472. https://doi.org/10.1037/neu0000052

Lum, J. A., & Bleses, D. (2012). Declarative and procedural memory in Danish speaking children with specific language impairment. *Journal of Communication Disorders, 45*(1), 46-58. https://doi.org/10.1016/j.jcomdis.2011.09.001

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

Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, *51*, 1-10. https://doi.org/10.1016/j.cortex.2013.10.011

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138-1154. https://doi.org/10.1016/j.cortex.2011.06.001

*Lum, J. A., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and declarative memory in children with and without specific language impairment. *International Journal of Language & Communication Disorders*, *45*(1), 96-107. https://doi.org/10.3109/13682820902752285

*Lum, J. A., & Kidd, E. (2012). An examination of the associations among multiple memory systems, past tense, and vocabulary in typically developing 5-year-old children. *Journal of Speech, Language, and Hearing Research*, *55*(4), 989-1006. https://doi.org/10.1044/1092-4388(2011/10-0137)

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in*

Developmental Disabilities, *34*(10), 3460-3476. https://doi.org/10.1016/j.ridd.2013.07.017

Lund, T. (1988). Some metrical issues with meta-analysis of therapy effects. *Scandinavian Journal of Psychology*, *29*(1), 1-8. https://doi.org/10.1111/j.1467-9450.1988.tb00770.x

*Mainela-Arnold, E., & Evans, J. L. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*(4), 897-909. https://doi.org/10.1044/1092-4388(2005/062)

*Majerus, S., Leclercq, A. L., Grossmann, A., Billard, C., Touzin, M., Van der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: No evidence for a causal association. *Cortex*, *45*(6), 708-720. https://doi.org/10.1016/j.cortex.2008.10.006

*Mayor-Dubois, C., Zesiger, P., Van der Linden, M., & Roulet-Perez, E. (2014). Nondeclarative learning in children with specific language impairment: predicting regularities in the visuomotor, phonological, and cognitive domains. *Child Neuropsychology*, *20*(1), 14-22. https://doi.org/10.1080/09297049.2012.734293

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348-356. https://doi.org/10.1016/j.tics.2010.06.002

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., ... & Olson, R. K. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: Searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry*, *52*(5), 547-557. https://doi.org/10.1111/j.1469-7610.2010.02346.x

Melby-Lervåg, M., Lyster, S., & Hulme, C. (2012) Phonological Skills and their Role in Learning to Read: A Meta-Analytic Review. *Psychological Bulletin*, *138*, 322-352. https://doi.org/10.1037/a0026744

Menghini, D., Carlesimo, G. A., Marotta, L., Finzi, A., & Vicari, S. (2010). Developmental dyslexia and explicit long-term memory. *Dyslexia*, *16*(3), 213-225. https://doi.org/10.1002/dys.410

*Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., & Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. *NeuroImage*, *33*(4), 1218-1226. https://doi.org/10.1016/j.neuroimage.2006.08.024

*Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., & Vicari, S. (2008). Structural correlates of implicit learning deficits in subjects with developmental dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 212-221. https://doi.org/10.1196/annals.1416.010

Merrill, E. C., Conners, F. A., Roskos, B., Klinger, M. R., & Klinger, L. G. (2013). Contextual cueing effects across the lifespan. *The Journal of Genetic Psychology*, *174*(4), 387-402. https://doi.org/10.1080/00221325.2012.694919

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*(3), 199-221. https://doi.org/10.1006/jecp.1998.2442

*Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386. https://doi.org/10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105. https://doi.org/10.1037/1082-989X.7.1.105

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders?. *TRENDS in Neurosciences*, *30*(4), 135-141. https://doi.org/10.1016/j.tins.2007.02.003

Nicolson, R. I., & Fawcett, A. J. (2010). *Dyslexia, learning, and the brain*. Cambridge, Mass.; London: MIT press.

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, *47*(1), 117-127. https://doi.org/10.1016/j.cortex.2009.08.016

*Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of Dyslexia*, *66*(2), 202-218. https://doi.org/10.1007/s11881-015-0116-9

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd edition). MacGraw-Hill, New York.

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01245

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin*, *82*(1), 85. https://doi.org/10.1037/h0076158

*Pavlidou, E. V., & Williams, J. M. (2010). Developmental dyslexia and implicit learning: evidence from an AGL transfer study. *Procedia-Social and Behavioral Sciences*, *2*(2), 3289-3296. https://doi.org/10.1016/j.sbspro.2010.03.503

Pavlidou, E. V., & Williams, J. M. (2014). Implicit learning and reading: Insights from typical children and children with developmental dyslexia using the artificial grammar learning (AGL) paradigm. *Research in Developmental Disabilities*, *35*(7), 1457-1472. https://doi.org/10.1016/j.ridd.2014.03.040

*Pavlidou, E. V., Kelly, L. M., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning?. *Dyslexia, 16*(2), 143-161. https://doi.org/10.1002/dys.400

*Pavlidou, E. V., Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of Dyslexia*, *59*(1), 55-77. https://doi.org/10.1007/s11881-009-0023-z

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413. https://doi.org/10.1016/j.cognition.2006.04.008

Perlant, A. S., & Largy, P. (2011). Are implicit learning abilities sensitive to the type of material to be processed? Study on typical readers and children with dyslexia. *Journal of Research in Reading*, *34*(3), 298-314. https://doi.org/10.1111/j.1467-9817.2010.01464.x

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General*, *119*(3), 264. https://doi.org/10.1037/0096-3445.119.3.264

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233-238. https://doi.org/10.1016/j.tics.2006.03.006

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Plante, E., Bahl, M., Vance, R., & Gerken, L. (2010). Children with specific language impairment show rapid, implicit learning of stress assignment rules. *Journal of*

Communication Disorders, *43*(5), 397-406. https://doi.org/10.1016/j.jcomdis.2010.04.012

*Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, *35*(5), 453-462. https://doi.org/10.1016/S0021-9924(02)00094-1

*Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia*, *10*(1), 61-76. https://doi.org/10.1002/dys.266

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178. https://doi.org/10.1037/1082-989X.10.2.178

Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2010). The effects of mood, cognitive style, and cognitive ability on implicit learning. *Learning and Individual Differences*, *20*(3), 215-219. https://doi.org/10.1016/j.lindif.2009.12.003

*Przekoracka-Krawczyk, A., Brenk-Krakowska, A., Nawrot, P., Rusiak, P., & Naskręcki, R. (2017). Unstable Binocular Fixation Affects Reaction Times But Not Implicit Motor Learning in Dyslexia. *Investigative Ophthalmology & Visual Science*, *58*(14), 6470-6480. https://doi.org/10.1167/iovs.16-21305

*Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, *23*(1), 101-115. https://doi.org/10.1080/10888438.2018.1485680

Raven, J. C., Court, JH, & Raven, J. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis & Co.

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size?. *Journal of Consulting and Clinical Psychology*, *64*(6), 1316. https://doi.org/10.1037/0022-006X.64.6.1316

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855-863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219. https://doi.org/10.1037/0096-3445.118.3.219

Reber, A. S. (1993). *Implicit learning and tacit knowledge: an essay on the cognitive unconscious*. Oxford: Oxford University Press.

Rohatgi, A. (2017). WebPlotDigitizer, version 4.0. Austin, TX. Retrieved from http://automeris.io/WebPlotDigitizer/.

*Rüsseler, J., Gerth, I., & Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology*, *28*(5), 808-827. https://doi.org/10.1080/13803390591001007

*Samara, A., & Caravolas, M. (2017). Artificial grammar learning in dyslexic and nondyslexic adults: Implications for orthographic learning. *Scientific Studies of Reading*, *21*(1), 76-97. https://doi.org/10.1080/10888438.2016.1262865

*Schiff, R., Sasson, A., Star, G., & Kahta, S. (2017). The role of feedback in implicit and explicit artificial grammar learning: A comparison between dyslexic and non-dyslexic adults. *Annals of dyslexia*, *67*(3), 333-355. https://doi.org/10.1007/s11881-017-0147-5

Schmalz, X., Altoè, G., & Mulatti, C. (2016). Statistical learning and dyslexia: a systematic review. *Annals of Dyslexia*, 1-16. https://doi.org/10.1007/s11881-016-0136-0

*Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why?. *Scientific Studies of Reading*, *23*(1), 64-76. https://doi.org/10.1080/10888438.2018.1482304

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163. https://doi.org/10.1037/0033-2909.115.2.163

*Sengottuvel, K., & Rao, P. K. (2013). Aspects of grammar sensitive to procedural memory deficits in children with specific language impairment. *Research in Developmental Disabilities*, *34*(10), 3317-3331. https://doi.org/10.1016/j.ridd.2013.06.036

*Sengottuvel, K., & Rao, P. K. (2014). Sequence learning pattern in children with specific language impairment. *International Journal on Disability and Human Development*, *13*(1), 55-62. https://doi.org/10.1515/ijdhd-2013-0003

*Sengottuvel, K., Rao, P., & Bishop, D. V. (2016). Declarative capacity does not trade-off with procedural capacity in children with specific language impairment. Autism & Developmental Language Impairments, 1, 1-17. https://doi.org/10.1177/2396941516674416

Shanks, D. R., 2005. Implicit learning. In Lamberts, K., & Goldstone, R. L. *Handbook of cognition*, 202-220. London: Sage. http://dx.doi.org/10.4135/9781848608177.n8

Shanks, D. R., & St John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367-395. https://doi.org/10.1017/S0140525X00035032

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145-150. https://doi.org/10.1056/NEJM199201163260301

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: how do patients with Parkinson's disease learn?. *Behavioral Neuroscience*, *118*(4), 676. https://doi.org/10.1037/0735-7044.118.4.676

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418-432. https://doi.org/10.3758/s13428-016-0719-z

Siegert, R. J., Taylor, K. D., Weatherall, M., & Abernethy, D. A. (2006). Is implicit sequence learning impaired in Parkinson's disease? A meta-analysis. *Neuropsychology*, *20*(4), 490. https://doi.org/10.1037/0894-4105.20.4.490

*Sigurdardottir, H. M., Danielsdottir, H. B., Gudmundsdottir, M., Hjartarson, K. H., Thorarinsdottir, E. A., & Kristjánsson, Á. (2017). Problems with visual statistical learning in developmental dyslexia. *Scientific Reports*, *7*(1), 606. https://doi.org/10.1038/s41598-017-00554-5

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666-681. https://doi.org/10.1177/1745691614553988

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In S. Gries & D. Divjak (Eds.), *Frequency effects in language* (pp. 127–163). Berlin, Germany: Mouton de Gruyter.

Snowling, M. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review*, *3*(1), 4-11. https://doi.org/10.1111/1475-3588.00201

Snowling, M. J. (2013). Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs*, *13*(1), 7-14. https://doi.org/10.1111/j.1471-3802.2012.01262.x

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171-177.

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*(3), a021667. https://doi.org/10.1101/cshperspect.a021667

*Staels, E., & Van den Broeck, W. (2015). No solid empirical evidence for the solid (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 650. https://doi.org/10.1037/xlm0000054

*Staels, E., & Van den Broeck, W. (2017). A specific implicit sequence learning deficit as an underlying cause of dyslexia? Investigating the role of attention in implicit learning tasks. *Neuropsychology*, *31*(4), 371. https://doi.org/10.1037/neu0000348

*Stoodley, C. J., Harrison, E. P., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, *44*(5), 795-798. https://doi.org/10.1016/j.neuropsychologia.2005.07.009

*Stoodley, C. J., Ray, N. J., Jack, A., & Stein, J. F. (2008). Implicit learning in control, dyslexic, and garden-variety poor readers. *Annals of the New York Academy of Sciences*, *1145*(1), 173-183. https://doi.org/10.1196/annals.1416.003

*Szmalec, A., Loncke, M., Page, M., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1270. https://doi.org/10.1037/a0023820

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*(4), 269-293. https://doi.org/10.1080/15475440701377477

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260. https://doi.org/10.1044/jslhr.4006.1245

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231-270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433. https://doi.org/10.1016/s0010-9452(08)70276-4

Vakil, E., Lowe, M., & Goldfus, C. (2015). Performance of Children With Developmental Dyslexia on Two Skill Learning Tasks—Serial Reaction Time and Tower of Hanoi Puzzle: A Test of the Specific Procedural Learning Difficulties Theory. *Journal of Learning Disabilities*, *48*(5), 471-481. https://doi.org/10.1177/0022219413508981

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2-40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(10), 1392-1397. http://dx.doi.org/10.1136/jnnp.2004.061093

*Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia*, *41*(1), 108-114. https://doi.org/10.1016/S0028-3932(02)00082-9

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., & Curran, T. (2003). Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits?. *Journal of Experimental Child Psychology*, *84*(4), 338-354. https://doi.org/10.1016/S0022-0965(03)00030-4

Webster, R. I., & Shevell, M. I. (2004). Topical review: Neurobiology of specific language impairment. *Journal of Child Neurology*, *19*(7), 471-481. https://doi.org/10.1177/08830738040190070101

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children – Third Edition (WISC-III)*. San Antonio,TX: The Psychological Corporation.

West, G., Melby-Lervåg, M, & Hulme, C. (2021, February 10). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. Retrieved from osf.io/cf6tn

West, G., Hulme, C., Melby-Lervåg, M. (2016). Procedural learning impairment in language disorder: a systematic review and meta-analysis. PROSPERO Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016048759

West, G., Shanks, D. R., & Hulme, C. (2021). Sustained attention, not procedural learning, is a predictor of reading, language and arithmetic skills in children. Scientific Studies of Reading, 25(1), 47-63. https://doi.org/10.1080/10888438.2020.1750618

*West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science, 21(2), e12552*. https://doi.org/10.1111/desc.12552

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2019). The procedural deficit hypothesis of language learning disorders: We still see some serious problems. *Developmental Science*, *22*(4). https://doi.org/10.1111/desc.12813

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2017). Visual artificial grammar learning in dyslexia: A meta-analysis. *Research in Developmental Disabilities*, *70*, 126-137. https://doi.org/10.1016/j.ridd.2017.09.006

*Yang, Y., & Hong-Yan, B. (2011). Unilateral implicit motor learning deficit in developmental dyslexia. *International Journal of Psychology*, *46*(1), 1-8. https://doi.org/10.1080/00207594.2010.509800

*Yang, J., & Li, P. (2013). Brain networks of explicit and implicit learning. *PloS One*, *7*(8), e42993 https://doi.org/10.1371/journal.pone.0042993

Zwart, F. S., Vissers, C. T. W., Kessels, R. P., & Maes, J. H. (2018). Implicit learning seems to come naturally for children with autism, but not for children with specific language impairment: Evidence from behavioral and ERP data. *Autism Research*, *11*(7), 1050-1061. https://doi.org/10.1002/aur.1954

Zwart, F. S., Vissers, C. T. W., Kessels, R. P., & Maes, J. H. (2019). Procedural learning across the lifespan: A systematic review with implications for atypical development. *Journal of Neuropsychology*., 13(2), 149-182. https://doi.org/10.1111/jnp.12139

Table 1

Implicit learning tasks

Task	Task description	Implicit learning measure
The serial reaction time task (SRT: Nissen & Bullemer, 1987)	Participants press buttons to indicate the which of four possible locations a stimulus occupies on a screen. Successive stimuli may be random or conform to a covert sequence.	Shorter response times to trials conforming to a covert sequence, compared to random trials, provide evidence of implicit learning.
The Hebb serial order learning task (Hebb, 1961)	Participants recall sequences of items in order. A covert repeated sequence is included at regular intervals across trials.	Better recall for the repeated as opposed to non-repeating sequences provides evidence of implicit learning.
Artificial grammar (AGL: Reber, 1967) **and statistical learning tasks** (SL: Arciuli & Simpson, 2011)	Participants are presented with strings of stimuli that conform to an undisclosed set of combinatory rules (grammar) and judge whether new strings conform to, or violate, these rules. AGL grammars are usually complex, while stimulus strings in SL tasks conform to a simpler (base triplet) structure.	Above chance performance is taken to reflect implicit learning. (NB: this measure reflects declarative knowledge for implicitly learned information).
The weather prediction task (Knowlton, Squire, & Gluck, 1994)	Participants classify combinations of four possible stimuli into one of two possible outcomes. The stimuli each have a fixed probability of a certain outcome.	A trial is scored correct if it accords with the conditional probabilities of the stimuli shown. Above chance performance is taken as evidence of implicit learning.
Contextual cueing task (Chun & Jiang, 1998)	Participants identify the location of a target stimulus within matrices of distractor stimuli as quickly as possible. The target position in some matrices is predictable.	Shorter response times to predictable compared to unpredictable matrices provides evidence of implicit learning.

Table 2

Series of meta-analyses with numbers of eligible studies and final inclusion numbers

Meta-analysis	Experimental Design	Eligible studies*
Serial reaction time task	Group	52 (29)
Serial reaction time task	Correlational	6 (5)
Hebb serial order learning task	Group	9 (8)
Artificial grammar learning and statistical learning tasks	Group	31 (23)
Artificial grammar learning and statistical learning tasks	Correlational	5 (5)
Probabilistic category learning tasks	Group	6 (5)
Contextual cueing task	Group	4 (Insufficient[1])

*Number of studies considered eligible. The final number included in the meta-analysis is noted in brackets. [1]Four studies were eligible, but there was insufficient data for meta-analysis

Table 3

Extreme groups studies testing participants on more than one procedural task, highlighting the tasks that report significant implicit learning differences between groups.

Study	**Age**	**DLD / DD**	**Correlation between implicit tasks**	**SRT**	**AGL / SL**	**Hebb**	**WPT**	**CC**	**Other**
Bennett et al., 2008	Adult	DD	Not reported	✓				x	
Du, 2013[a]		DD	Not reported	x	✓				
Henderson & Warmington, 2017	Adult	DD	Nonsignificant	x		✓			
Howard et al., 2006	Adult	DD	Not reported	✓				x	
Hsu & Bishop, 2014	Child	DLD	SRT and Hebb (r = .23, p = .09)	✓		✓			x [b]
Jiménez-Fernández et al., 2011	Child	DD	Different children	✓				x	
Laasonen et al., 2014	Adult	DD	Not reported	x	x				

Lee, 2012[a]	Adult	DLD	Not reported	x		✓	
Lee & Tomblin, 2015	Adult	DLD	Nonsignificant, except WPT and repetition priming (r = .35, p = .01)	x		✓	✓ [b, c]
Lukacs & Kemeny, 2014	Child	DLD	Nonsignificant	✓	✓	x	
Mayor-Dubois et al., 2014	Child	DLD	Not reported	x		x	
Rüsseler et al., 2006	Adult	DD	Not reported	x	x		
Vakil et al., 2015	Child	DD	Not reported	x			x [d]
Vicari et al., 2005	Child	DD	Not reported	✓			✓ [e]

✓ = significant implicit learning-related differences between groups; x = no significant difference between groups; [e] = unpublished thesis; [b] = Pursuit Rotor; [c] = Repetition priming; [d] = Tower of Hanoi; [e] = Mirror Drawing;

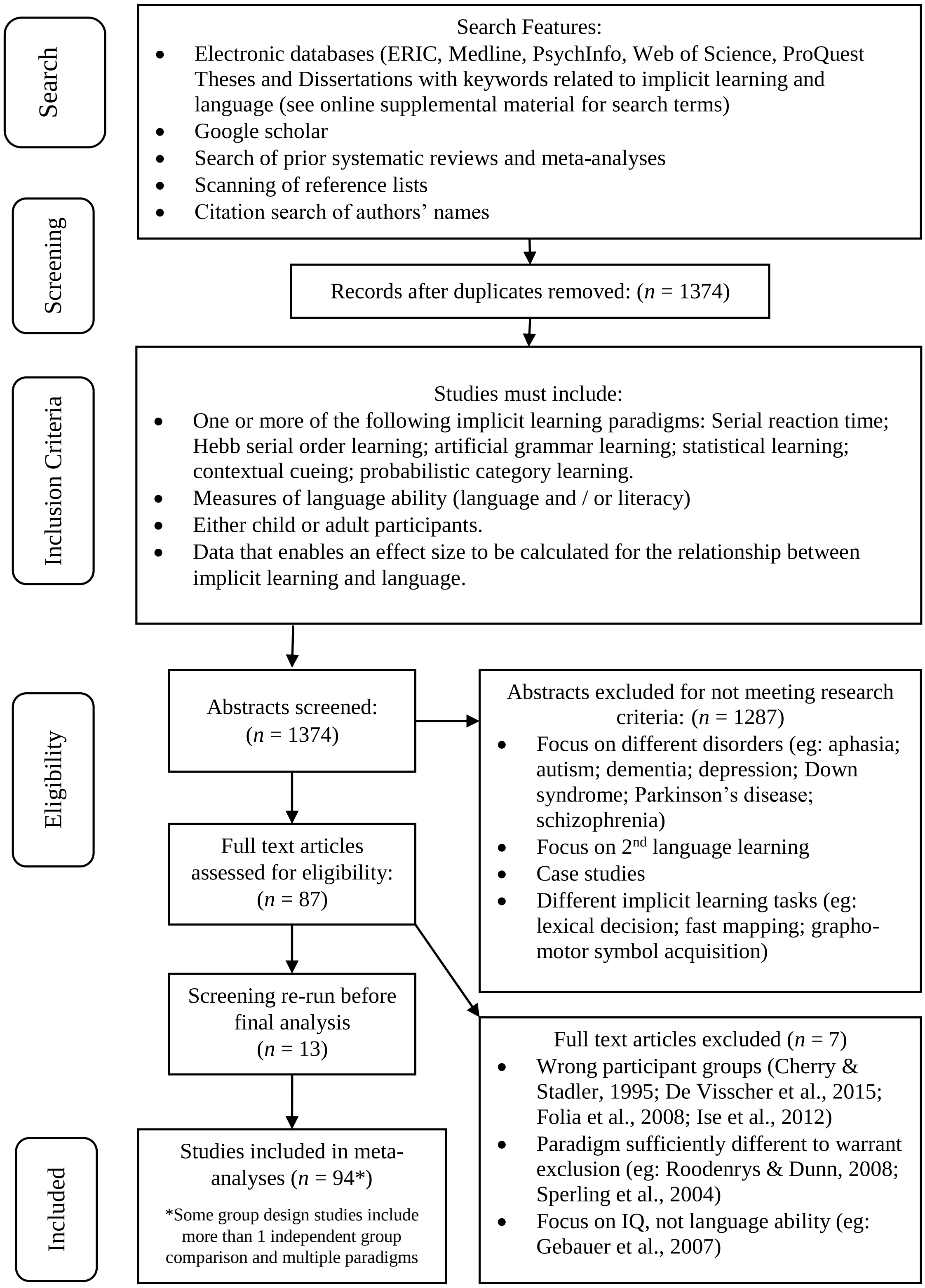


Figure 1. PRISMA flow chart for the search and inclusion criteria for studies in this series of meta-analyses.

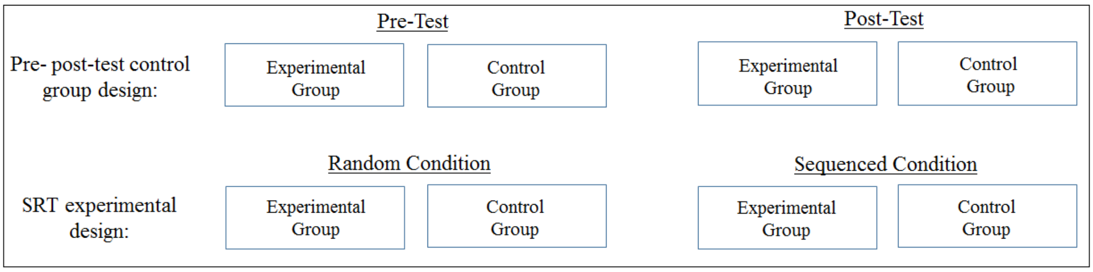


Figure 2. Analogous design structure for randomised control trials and implicit learning tasks based on difference scores.

Study name	Hedges's g	Standard error
Vicari et al., 2003	-1.378	0.364
Sengottuvel & Rao, 2013	-1.262	0.346
Jimenez-Fernandez et al., 2011	-0.942	0.388
Lum et al., 2010	-0.678	0.372
Gabay et al., 2012b	-0.643	0.387
Sengottuvel & Rao, 2014	-0.615	0.276
Przekoracka-Krawczyk, 2017	-0.573	0.363
Stoodley et al., 2006	-0.572	0.317
Gabay et al., 2012a	-0.557	0.403
Menghini et al., 2006	-0.453	0.372
Hsu & Bishop, 2014	-0.421	0.267
Yang et al., 2013	-0.406	0.459
Gabriel et al., 2013	-0.360	0.293
Menghini et al., 2008	-0.314	0.431
Gabriel et al., 2012	-0.234	0.358
He & Tong, 2017	-0.233	0.267
Desmottes et al., 2016a	-0.210	0.243
Yang & Hong-Yan, 2011	-0.194	0.269
Sengottuvel et al., 2016	-0.180	0.255
Lukacs & Kemeny, 2014	-0.124	0.216
Stoodley et al., 2008	-0.116	0.210
Lee & Tomblin, 2015	-0.105	0.284
Deroost et al., 2010	-0.046	0.264
Hedenius et al., 2013	-0.019	0.367
Henderson & Warmington, 2017	0.020	0.257
Staels & Van den Broeck, 2017	0.034	0.241
Bennett et al., 2008	0.144	0.336
Gabriel et al., 2015	0.277	0.369
Russeler et al., 2006	0.484	0.400
	-0.298	0.068

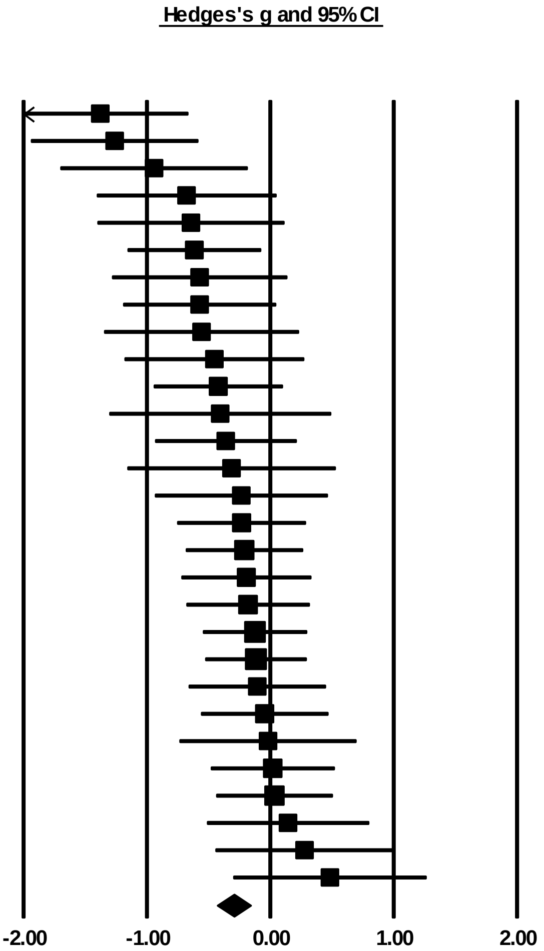


Figure 3. Forest plot showing effect sizes for the group difference in performance on the SRT task in 29 studies using means and standard deviations for both sequenced and random trials per group. Although Lukács & Kemény (2012) reported *z*-transformed scores in their paper, means and SDs in milliseconds supplied by the authors were used in the meta-analysis.

Study name	Hedges's g	Standard error
Hsu & Bishop, 2014	-0.865	0.301
Szmalec et al., 2011	-0.715	0.357
Bogaerts et al., 2015 (Expt 2)	-0.568	0.337
Bogaerts et al., 2016	-0.524	0.299
Staels, & Van den Broeck, 2015 (Expt 1)	-0.327	0.262
Bogaerts et al., 2015 (Expt 1)	-0.262	0.285
Archibald & Joanisse, 2013	-0.135	0.280
Henderson & Warmington, 2017	-0.129	0.259
Staels, & Van den Broeck, 2015 (Expt 2)	-0.048	0.263
Majerus et al., 2009	0.341	0.397
	-0.319	0.101

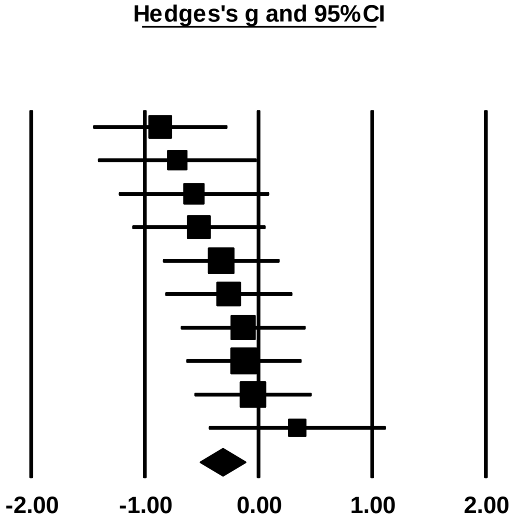


Figure 4. Overall average effect size for the group difference in performance on Hebb tasks (displayed by ♦) with 95% confidence interval for each study.

Study name	Hedges's g	Standard error
Pavlidou & Williams, 2010	-1.338	0.383
Kahta & Schiff, 2016	-1.246	0.397
Mainela-Arnold & Evans, 2014	-1.126	0.335
Evans et al., 2009 (Expt 2)	-0.979	0.380
Plante et al., 2002	-0.929	0.364
Pavlidou et al., 2009	-0.855	0.361
Mayor-Dubois et al., 2014	-0.828	0.268
Gabay et al., 2015	-0.809	0.359
Du, 2013 - Expt 2a	-0.691	0.417
Pavlidou et al., 2010	-0.681	0.355
Haebig, Saffran, & Weismer, 2017	-0.676	0.290
Aguilar & Plante, 2014 (Expt 1)	-0.601	0.405
Schiff et al. 2017	-0.569	0.297
Sigurdardottir et al., 2017	-0.563	0.235
Lao et al., 2017	-0.561	0.353
Aguilar & Plante, 2014 (Expt 2)	-0.558	0.269
Lukacs & Kemeny, 2014	-0.556	0.219
Evans et al., 2009 (Expt 1)	-0.480	0.205
Nigro et al., 2016 (Expt 2)	-0.439	0.307
Laasonen et al., 2014	-0.425	0.236
Russeler et al., 2006	-0.237	0.396
Nigro et al., 2016 (Expt 1)	-0.204	0.304
Grunow et al., 2006	-0.203	0.288
Hsu et al., 2014 - LV task	-0.196	0.311
Samara & Caravolas, 2017 (Expt 2)	-0.175	0.272
Hsu et al., 2014 - HV task	-0.093	0.315
Samara & Caravolas, 2017 (Expt1)	-0.066	0.280
Hsu et al., 2014 - MV task	0.044	0.310
Du, 2013 - Expt 2b	0.137	0.397
Pothos & Kirk, 2004	0.658	0.144
	-0.479	0.096

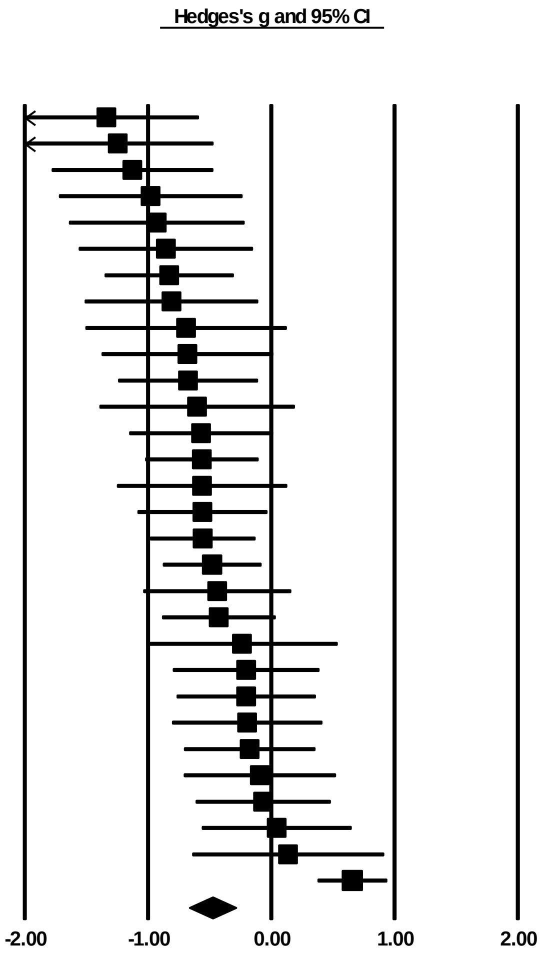


Figure 5. Forest plot showing effect sizes for group difference in performance on artificial grammar learning and statistical learning tasks (displayed by ♦) with 95% confidence interval for each study.

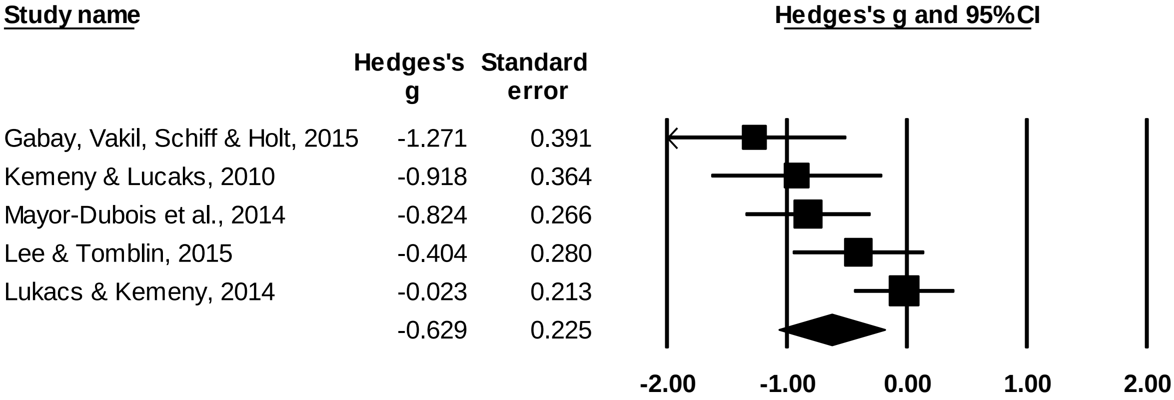
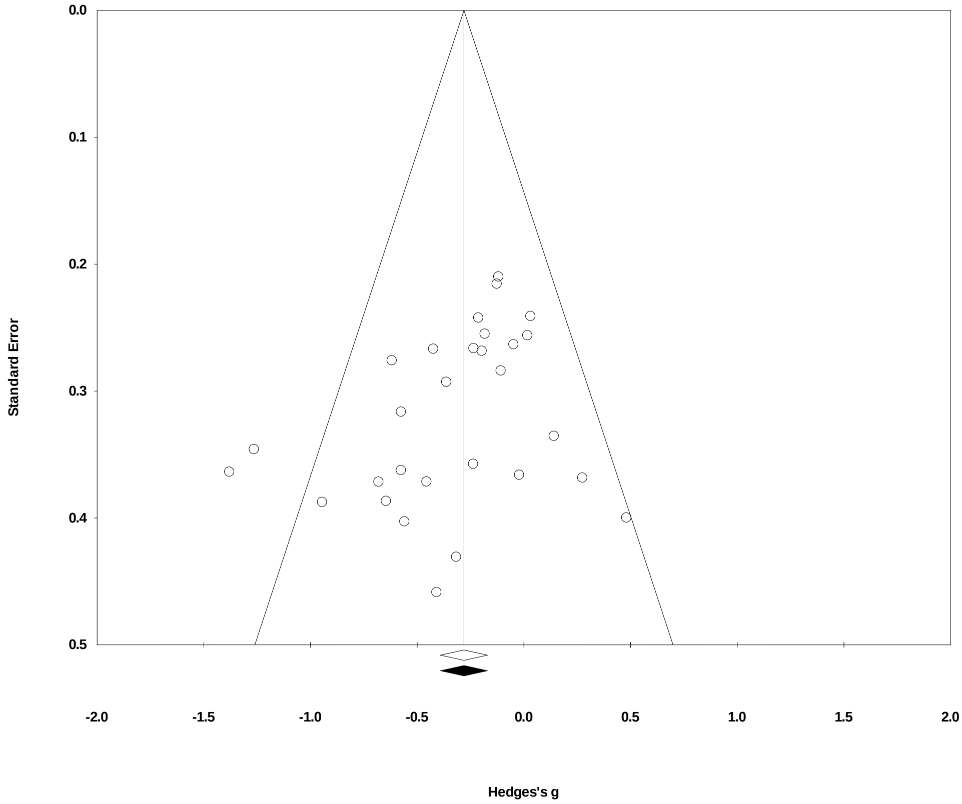
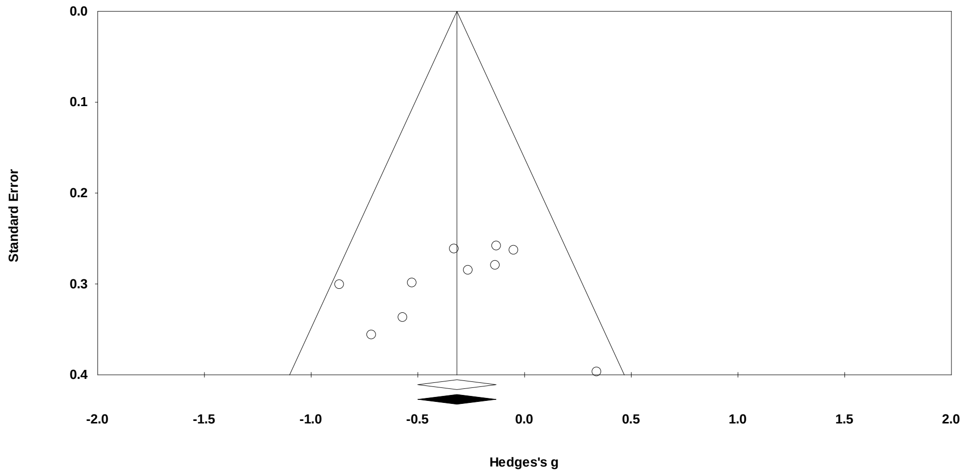


Figure 6. Overall effect size for the group difference in performance on weather prediction tasks (displayed by ♦) with 95% confidence interval for each study.

Funnel plot for group design studies using SRT tasks (point estimate of *d* = -0.30, 95% CI [-0.43, -0.16]):



Funnel plot for group design studies using Hebb learning tasks (point estimate of *d* = -0.32, 95% CI [-0.52, -0.12]):



Funnel plot for group design studies using artificial grammar or statistical learning tasks (adjusted point estimate of *d* = -0.30, 95% CI [-0.47, 0.12]):

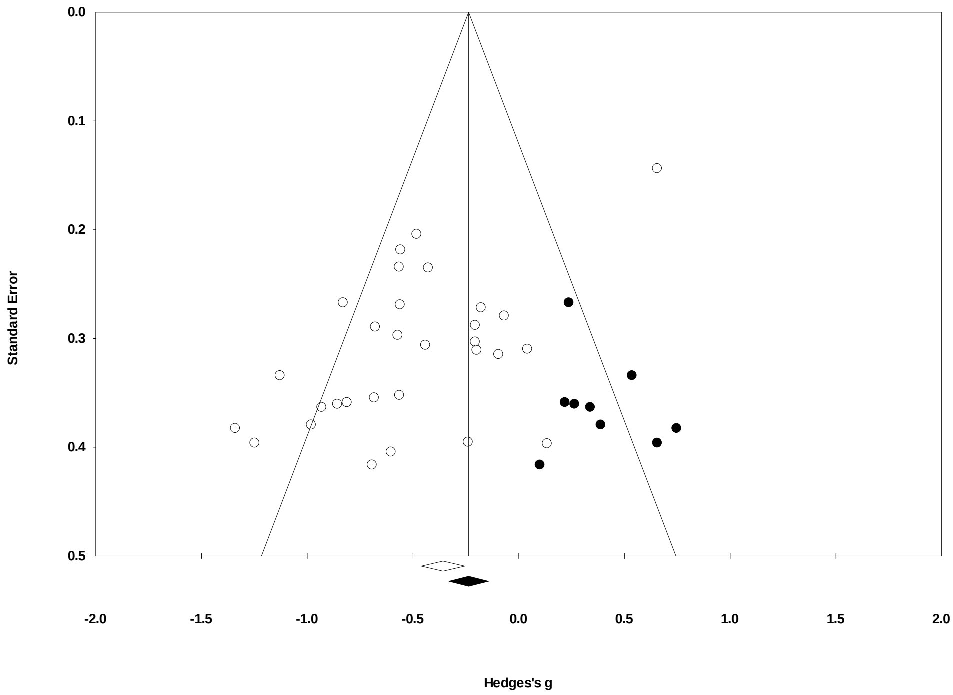


Figure 7. Funnel plots examining evidence of publishing bias for the four meta-analyses of group design studies (top to bottom: SRT; Hebb learning and AGL and statistical learning tasks). Open circles and diamond correspond to observed studies and point estimate. Filled circles and diamond correspond to imputed missing studies and adjusted point estimate, following Duval and Tweedie's Trim and Fill procedure.