Statistical learning and dyslexia: A systematic review

Xenia Schmalz[1], Gianmarco Altoè, & Claudio Mulatti

Dipartimento di Psicologia dello Sviluppo e della Socializzazione

Università degli Studi di Padova

Padova, Italy.

[1] Corresponding author. Address: DPSS, Via Venezia 15, Padova, 35131, Italy.

Email: xenia.schmalz@gmail.com, phone: +39 320 639 2523, fax: +39 049 827 6547.

Gianmarco Altoè: Address: DPSS, Via Venezia 8, Padova, 35131, Italy. Email:

gianmarco.altoe@unipd.it

Claudio Mulatti: Address: DPSS, Via Venezia 8, Padova, 35131, Italy. Email:

Claudio.mulatti@gmail.com

Abstract

The existing literature on developmental dyslexia (hereafter: dyslexia) often focuses on isolating cognitive skills which differ across dyslexic and control participants. Among potential correlates, previous research has studied group differences between dyslexic and control participants in performance on statistical learning tasks. A statistical learning deficit has been proposed to be a potential cause and/or a marker effect for early detection of dyslexia. It is therefore of practical importance to evaluate the evidence for a group difference. From a theoretical perspective, such a group difference would provide information about the causal chain from statistical learning to reading acquisition. We provide a systematic review of the literature on such a group difference. We conclude that there is insufficient high-quality data to draw conclusions about the presence or absence of an effect.

Keywords: Theory of dyslexia, deficit, publication bias, implicit learning.

Statistical learning and dyslexia: A systematic review

Learning to read requires the involvement of multiple cognitive skills: the child needs to learn to identify letters, map the orthographic information onto phonology, recognise whole words, and directly access the words' pronunciation and semantics. Research on developmental dyslexia (hereafter: dyslexia) often aims to understand the cognitive components that may be deficient in dyslexia. These could be candidates for an underlying cause of reading problems or markers that could be used for early detection of potential reading problems.

Here, we focus on the proposal that there is a group difference between participants with dyslexia and controls on tasks which measure statistical learning ability. Statistical learning is arguably a domain-general computational process (e.g., Kirkham, Slemmer, & Johnson, 2002; Pacton, Fayol, & Perruchet, 2005; Perruchet & Pacton, 2006; but see also Frost, Armstrong, Siegelman & Christiansen, 2015; Siegelman & Frost, 2015), which extracts regularities from a quasi-regular system and applies this statistical knowledge to predict future events. Recent studies have shown a relationship between reading and performance on non-linguistic statistical learning tasks. This has been found for adults and children (Arciuli & Simpson, 2012), as well as for adult L2 learners of a new orthography (Frost, Siegelman, Narkiss, & Afek, 2013).

Furthermore, it has been proposed that a deficit in statistical learning may underlie dyslexia. Studies have reported group differences in performance on statistical learning tasks between dyslexic and control children and adults (e.g., Pavlidou, Kelly, & Williams, 2010; Pavlidou & Williams, 2010; Pavlidou, Williams, & Kelly, 2009; Stoodley, Harrison, & Stein, 2006; Vicari et al., 2005; Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003). However, there have also been failures to

find a group difference (Kelly, Griffiths, & Frith, 2002; Laasonen et al., 2014; Menghini et al., 2010; Nigro, Jiménez-Fernández, Simpson, & Defior, 2015; Pothos & Kirk, 2004; Roodenrys & Dunn, 2008; Rüsseler, Gerth, & Münte, 2006; Waber et al., 2003), which raises questions about the reality of this effect.

Here, we aim to provide a systematic review of the evidence for a statistical learning deficit in dyslexia. This involves, first and foremost, a theoretical discussion of what a statistical learning deficit in dyslexia would mean for theories of statistical learning, reading, and dyslexia. We provide a brief discussion of these issues, followed by a description of methodological and statistical issues. Prior to conducting a meta-analysis, it is also important to assess the possibility that the current literature is biased towards positive effects (Egger, Smith, Schneider, & Minder, 1997; Simonsohn, Nelson, & Simmons, 2014a, 2014b). Finding evidence for publication bias in the available literature, or being unable to determine whether publication bias is present, limits the utility of a meta-analysis and suggests that more data is needed – especially from large-scale pre-registered studies – before conclusions can be drawn (Van Elk et al., 2015). With these issues in mind, we critically examine the existing studies on a statistical learning deficit in dyslexia.

What would a statistical learning deficit in dyslexia mean?

If there is a statistical learning mechanism in dyslexia, there should be consistent evidence for lower average performance on statistical learning tasks of dyslexic participants compared to matched controls. This should hold true for any statistical learning task, given the assumption that the statistical learning mechanism is domain-general (Kirkham et al., 2002; Pacton et al., 2005; Perruchet & Pacton, 2006). Further discussion of what constitutes statistical learning is outside the scope of the current paper. In short, there is little consensus about what statistical learning

is, and what specific aspects may relate to reading (see Perruchet & Pacton, 2006; Siegelman & Frost, 2015). However, we can meaningfully address the question of whether there are group differences in performance on statistical learning tasks, provided we assume that these tasks capture something about this statistical learning construct. Also, a full overview of potential causal pathways that would lead from a domain-general statistical learning deficit to dyslexia is outside the scope of the current paper. Possible intermediate steps would involve the learning of print-to-speech correspondences (Apfelbaum, Hazeltine, & McMurray, 2013; Arciuli & Simpson, 2012) and/or knowledge of the statistical distributions of letter patterns (Cassar & Treiman, 1997; Pacton et al., 2005; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Pollo, Kessler, & Treiman, 2009; Rothe, Schulte-Körne, & Ise, 2014).

Finding a consistent group difference in performance on statistical learning tasks would allow us to evaluate several propositions: (1) There is a domain-general statistical learning mechanism, (2) this mechanism is possibly related to reading acquisition, and (3) all or most children with dyslexia have a deficit in statistical learning. A finding of a group difference across tasks would indicate that different statistical learning tasks are likely to tap the same underlying cognitive construct. This would be an important insight, because recent studies have argued that statistical learning mechanism is not domain-general, as the correlation between various statistical learning tasks is rather low (Frost, Armstrong, Siegelman, & Christiansen, 2015; Siegelman & Frost, 2015). A consistent group difference would also strengthen the view that statistical learning is linked to reading acquisition (Arciuli & Simpson, 2012; Frost et al., 2013; Rothe et al., 2014). A positive finding would justify further

exploring this relationship, which should lead towards a well-specified theory of a causal chain between statistical learning and reading acquisition.

Similarly, failing to find a group difference would raise important questions for future research. First, it would be necessary to establish whether it is reasonable to assume that statistical learning is a domain-general construct. Equivalent performance on statistical learning tasks is compatible with the view that a statistical learning deficit may be present in dyslexia, but it affects only (or predominantly) learning processes that are specifically linked to the reading system. An alternative explanation would be that statistical learning is intact in dyslexia.

Second, it is necessary to understand the role of statistical learning in reading acquisition, regardless of whether there is a group difference on statistical learning tasks between dyslexic and control participants. Specific aspects of statistical learning could influence specific aspects of learning to read, and the causal chain that connects them may or may not be different in children with dyslexia compared to controls. Creating a comprehensive theoretical framework of this hypothetical relationship would allow for more targeted experimental studies in the future, and possibly more effective interventions and reading instruction programs, if statistical learning is found to have a causal role in reading acquisition. However, we first need to evaluate the existing evidence in order to constrain the number of plausible causal chains.

Finally, future research could focus on individual differences in dyslexia. From a theoretical perspective, it is unlikely that in all children with dyslexia the reading deficit is caused by the same underlying cognitive deficit. The reading process requires the collaboration of multiple cognitive skills, and a deficit in any of these could make reading acquisition more difficult (Castles & Coltheart, 1993; Coltheart, 2015; Pennington, 2006; Ziegler et al., 2008). Thus, it would be somewhat

surprising to find a consistent group difference in statistical learning ability. However, future research could aim to determine whether particular subgroups show a statistical learning deficit. To this end, large-scale and/or longitudinal studies could be particularly useful. Specifically, one could test a strong prediction of a causal theory, that all children who have a statistical learning deficit prior to the onset of reading instruction will go on to become poor readers.

Methodological and statistical issues

The Artificial Grammar Learning (AGL) and Serial Reaction Time (SRT) tasks

The AGL (Reber, 1967) and SRT (Nissen & Bullemer, 1987) tasks are most commonly used to study statistical learning. We focus on these two tasks, as there is an increasing number of experiments where these tasks were used to compare dyslexic participants to controls. Our overall aim is to assess whether there are group differences in a hypothetical domain-general construct of statistical learning. If this is the case, we should find group differences in both tasks, even though they rely on different input domains. If we were to find a group difference on one task, but not the other, this would provide evidence against a domain-general statistical learning deficit, and instead suggest that task-specific aspects of the particular task are linked to reading ability.

For the AGL task, participants are shown strings of letters or symbols. In the first phase, the participants are asked to perform a cover task, such as memorising the letter strings for immediate subsequent recall. Unbeknown to the participants, the strings are created according to a set of rules (a grammar), where some letter or symbol combinations are legal in certain positions, whereas others are not. Before the second phase, the participants are generally informed that the strings were created

according to a set of complex rules, but they are not told what these rules are. In the second phase, the participants are shown both legal and illegal strings, and are asked to indicate the perceived legality.

If dyslexic participants have a statistical learning deficit, they should be less likely than control participants to correctly classify strings as legal or illegal. This should be evident from a main effect of dyslexia, either on the number of correctly classified strings, or on the sensitivity index d', which takes into account individual differences in response bias. Previous research has reported inferior performance of dyslexic participants compared to controls (Kahta & Schiff, 2016; Pavlidou et al., 2010; Pavlidou & Williams, 2010, 2014; Pavlidou et al., 2009), equal performance across the two groups (Laasonen et al., 2014; Nigro et al., 2015; Rüsseler et al., 2006), and superior performance of dyslexic compared to control participants (Pothos & Kirk, 2004).

An SRT task measures motor learning. Participants are typically shown stimuli in varying positions on a screen. The task is to press a button which corresponds to the location of the stimulus (e.g., top, bottom, left, right). Unbeknown to the participants, the sequence of the stimuli's locations repeats throughout the experiment. Performance of participants improves throughout the experiments, but – importantly – decreases for a control block which is inserted near the end of the experiment. In the control block, the order of the locations is random, thus disrupting the repeated sequence. The performance on this random block is generally compared to both the preceding and succeeding sequential blocks. If statistical learning has occurred, participants make more errors and take more time to determine the location throughout the random compared to the sequential blocks.

If there is a group difference on this task, we would expect it to show both in the slope of the learning curve, and in the comparison of the sequential versus random blocks. The former, by itself, cannot give evidence for implicit learning, because it could reflect practice effects. However, in a between-subject comparison, a difference in statistical learning ability would lead to a steeper slope among the participants with better statistical learning ability. Thus, if dyslexic participants are worse at statistical learning than controls, we would expect both a flatter learning curve for dyslexic than control participants, and a larger difference between the random and sequential blocks.

Dependent variables that can be used for this task include the difference between the random and sequential blocks, using reaction times (RTs) or accuracy rates, or the increase in accuracy or RTs across the sequential blocks. Previous studies have used slight variations of this SRT task, and analysed different dependent variables. This opens up a garden of forking paths, where a significant group difference in either dependent variables can be used to support the alternative hypothesis, at the risk of an increased Type-I error (false alarm) rate (Cramer et al., 2015; Gelman & Loken, 2014). Evaluating the consistency of the literature is further complicated by different tasks which are used by different authors. For example, different experimenters have used different stimuli: some, as in the original task (Nissen & Bullemer, 1987), use simple stimuli which appear in different location on the screen (e.g., Waber et al., 2003), while others use colourful stimuli, and rather than determining the location of the stimulus, the participants need to press a button corresponding to the stimulus' colour (e.g., Stoodley, Ray, Jack, & Stein, 2008).

Taken at face value, the results from this task are mixed. Several studies reported a group difference (Howard, Howard, Japikse, & Eden, 2006; Jiménez-

Fernández, Vaquero, Jiménez, & Defior, 2011; Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006; Stoodley et al., 2006; Stoodley et al., 2008; Vicari et al., 2005; Vicari et al., 2003), while others reported equivalent performance of dyslexic and control groups (Deroost et al., 2010; Hedenius et al., 2013; Kelly et al., 2002; Roodenrys & Dunn, 2008; Rüsseler et al., 2006; Waber et al., 2003).

The difference between significant and not significant is not in itself significant

In order to conclude that there are group differences, one needs to show a significant effect of group (in the AGL task), or a significant group-by-block interaction (i.e., in the SRT task, steeper increase in controls compared to dyslexic participants across the sequential blocks, and/or larger difference between the random and sequential blocks in controls compared to dyslexic participants). In addition, it is possible to check whether both or either group performed significantly above chance. However, finding significant statistical learning in the control group and non-significant statistical learning in the dyslexic group, in the absence of a group effect or group interaction, does not constitute evidence for a group difference (Gelman & Stern, 2006). This needs to be considered when evaluating studies which report support for a group difference. Specifically, a group difference in AGL can only be concluded if it is shown that the dyslexic and control group are significantly different.

Furthermore, it is unclear what to make of studies which show a significant statistical learning effect for controls, but not for dyslexic participants (especially in the absence of a significant group effect). First, p-values greater than 0.05 cannot be used to conclude that the learning effect is not different from zero (Dienes, 2014; Rouder, Speckman, Sun, Morey, & Iverson, 2009). Second, it is highly unlikely that participants with dyslexia are completely incapable of statistical learning: given the proposed importance of statistical learning in various aspects of language learning

(Kempe & MacWhinney, 1998; Petersson, Folia, & Hagoort, 2012; Saffran, Aslin, & Newport, 1996; Seidenberg & Gonnerman, 2000), one would expect severe difficulties in language development if statistical learning ability were absent altogether. This would predict that dyslexia cannot occur independently of language problems – and it does (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005). Statistical learning is also likely to play a role in learning other complex everyday tasks, ranging from social skills to intuitive decision making and categorisation (Kahneman, 2011; Pothos, 2005; Reber, 1989). While a large amount of research to date has focused on finding dyslexic deficits, we know of no study that would suggest potential group differences on these everyday tasks.

Resolving conflicts: Issues of replicability

As seen from the overview below, there are studies which support the notion of a group difference in statistical learning, and studies which fail to find a group difference. There are, broadly speaking, three possibilities: the positive findings could be Type-I errors, the negative findings could be Type-II errors, or there is a hidden moderator that causes the effect to emerge in some situations but not others. Assessing whether or not there is a hidden moderator is not the aim of the current study. Quests for moderators should be ideally theory-driven and specified *a priori*, preferably in a pre-registration report. This is because there are many possible variables that could act as covariates, including information about participants or details about the experimental procedure that were not reported in the papers. A comprehensive *post-hoc* exploration is very likely to yield a variable that appears to explain differences across studies, but such findings are more likely to be Type-I errors and should be interpreted with caution until they are confirmed by a subsequent study. In the particular question of statistical learning and dyslexia, we have no *a*

priori reason to expect moderation effects: if there is a domain-general statistical learning mechanism underlying dyslexia, it should emerge regardless of the exact experimental procedures or participant characteristics. Thus, we do not speculate on potential moderators, but future research may well identify conditions under which a group difference is more or less likely to emerge.

Alternatively, it is also possible that the studies with conflicting results all reflect random noise around a true effect size. In our view, it is important to consider this possibility before pursuing the possibility of a hidden moderator (see Meehl, 1990; Schmidt, 1992; Schmidt, 1996; Swanson, 1996). However, even if we assume that all studies come from the same population, there are some obstacles that need to be overcome before one can conclude whether or not the true effect size is different from zero.

Generally, positive findings are more likely to be published than null-findings in psychological research (Rosenthal, 1979; Simonsohn et al., 2014a, 2014b; Sterling, Rosenbaum, & Weinkam, 1995). This is because of a mixture of file-drawer problems (where studies with non-significant effects are not written up) and publication bias (where journals prefer to publish positive findings while rejecting null-results – hereafter, we refer to file-drawer problems and journal-level publication bias collectively as publication bias). This problem is exacerbated by questionable research practices, which range from creatively removing outliers or adding *post-hoc* moderators to data fabrication (Fanelli, 2009; Simmons, Nelson, & Simonsohn, 2011).

Questionable research practices in psychology are further enabled by generally small sample sizes and low statistical power (Button et al., 2013; Christley, 2010; Gelman & Weakliem, 2009; Ioannidis, 2005; Royall, 1986). These decrease the

precision of effect size estimates (Gelman & Carlin, 2014; Gelman & Weakliem, 2009), and in combination with publication bias and questionable research practices may lead to an inflated Type-I error rate (Button et al., 2013; Christley, 2010; Ioannidis, 2005). By definition, low statistical power also means that a study is less likely to determine a true effect. Thus, the overall small sample sizes in statistics may also inflate the number of published Type-II errors.

In sum, general issues in replicability may skew the existing literature either towards an inflated or deflated effect size. If the literature contains a disproportionate number of false positives, the effect size of a meta-analysis will be biased towards a larger effect size than the true effect size, even if the true population mean is zero. If the literature contains a disproportionate number of false negatives, a true effect in a meta-analysis may be diminished or masked.

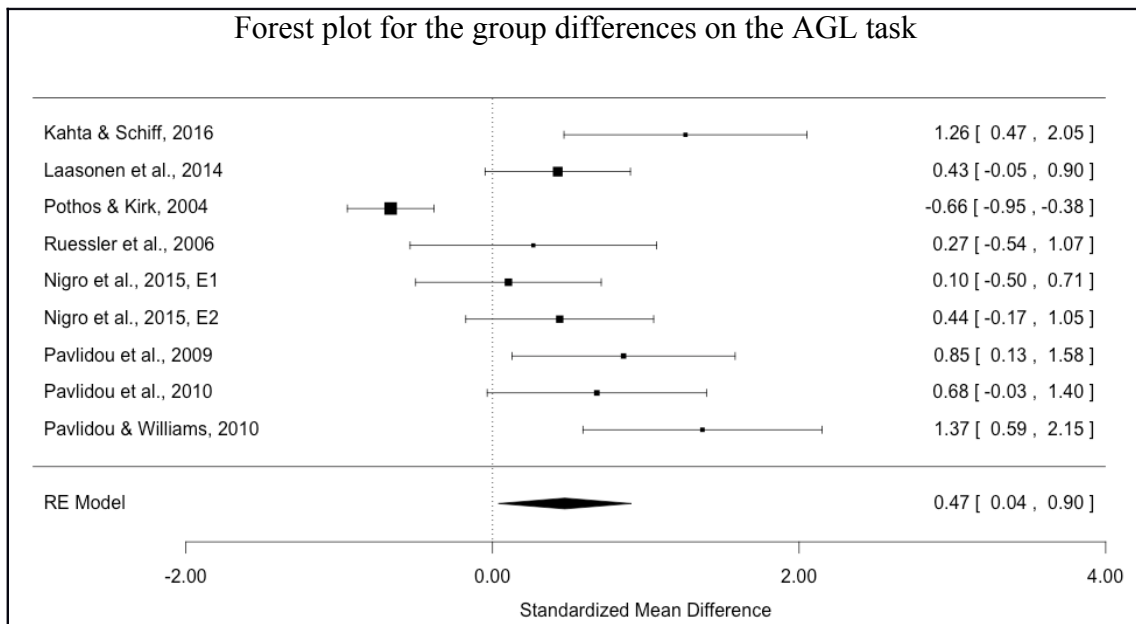## Artificial Grammar Learning: Is there a group difference?

Establishing whether or not dyslexic participants perform worse on AGL tasks than control participants can be measured by assessing a simple main effect of group: our hypothesis is that dyslexic participants make more errors than controls at classifying strings as grammatical or ungrammatical. As accuracy rates tend to be low, researchers do not typically measure or analyse RTs.

In this review we include experiments which compared accuracy rates of dyslexic versus control participants on AGL tasks. These were found through Google Scholar and by scanning the reference lists of related papers Search keywords were reading or dyslexia, and statistical learning, implicit learning, artificial grammar learning, or serial reaction time task, in all possible combinations. Further efforts were made to find unpublished studies on AGL and dyslexia. To this end, we scanned

available abstracts of relevant conferences and contacted authors who work in the field. However, we did not find any unpublished studies.

Figure 1 summarises the effect sizes of the available studies, to serve as a point of reference (a duplicate study was excluded – see details below). It shows a forest plot of the observed effect sizes and 95% CI of all studies which compared dyslexic participants to controls on AGL tasks. The standardised mean differences were calculated from the reported correct classifications in each group, using the R package metafor (Viechtbauer, 2010). When the observed mean accuracy and standard deviation (or standard error) were not reported, we contacted the corresponding author to obtain this information. This allowed us to include all of the published studies in the analysis. The meta-analysis effect size was calculated by the restricted maximum-likelihood estimator. This is a random-effect model, which assumes that the effect sizes reflect sampling from a normal distribution (Borenstein,

Hedges, Higgins, & Rothstein, 2009, 2010)[1]. The R script and data used in this

analysis can be downloaded here: https://osf.io/ruj3v/.

Forest plot for the group differences on the AGL task

Kahta & Schiff, 2016	1.26 [0.47 , 2.05]
Laasonen et al., 2014	0.43 [-0.05 , 0.90]
Pothos & Kirk, 2004	-0.66 [-0.95 , -0.38]
Ruessler et al., 2006	0.27 [-0.54 , 1.07]
Nigro et al., 2015, E1	0.10 [-0.50 , 0.71]
Nigro et al., 2015, E2	0.44 [-0.17 , 1.05]
Pavlidou et al., 2009	0.85 [0.13 , 1.58]
Pavlidou et al., 2010	0.68 [-0.03 , 1.40]
Pavlidou & Williams, 2010	1.37 [0.59 , 2.15]
RE Model	0.47 [0.04 , 0.90]

Standardized Mean Difference

        1 For the current purposes, we chose a random-effect model rather than a

fixed-effect model. Under a fixed-effect model, the assumption is that there is a single

true effect across all studies, and that the effect size estimates differ due to sampling

error. Thus, the latter uses stronger weightings as a function of sample size, as large

studies provide less noisy estimates (Borenstein et al., 2009, 2010). In our case, this

gives greater weighting to the large-sample study by Pothos and Kirk (2004),

resulting in a diminished effect size estimate, $d = 0.15$, 95% CI [-0.02, 0.33]. The

fixed effect model could be considered more appropriate in the current context,

because we specified *a priori* that we do not expect the effect to differ across studies

as a function of moderating variables. However, note that we use the meta-analysis

effect size estimate for descriptive purposes only – either way, this estimate is

unreliable due to the small number of studies, and the heterogeneity across labs

(discussed below). We continue to use the random-effect estimate throughout this

section, as it is more sensitive to the observed heterogeneity of the effect sizes.

From the studies described in Table 1, the top four tested dyslexic and control adults, and only one found a group difference in the expected direction, with a total sample size of twenty-nine participants (Kahta & Schiff, 2016). The second study, with 71 participants, found only a marginally significant difference between dyslexic and control groups, $p = 0.09$ (Laasonen et al., 2014). The third study found an effect in the opposite direction with 109 participants (Pothos & Kirk, 2004), and the fourth study with 24 participants found no effect (Rüsseler et al., 2006).

The other four papers from Table 1 are developmental studies (Nigro et al., 2015; Pavlidou et al., 2010; Pavlidou & Williams, 2010; Pavlidou et al., 2009). Note that two published studies by Pavlidou and colleagues report identical data (Experiments 1 of Pavlidou & Williams, 2010, 2014): the sample descriptives (the children's average scores on digit span, matrix reasoning, vocabulary, word reading and spelling tasks) are identical, as are the accuracy rates and standard deviations for the AGL task. This is unlikely to occur by chance under any circumstances. We therefore treat this as a duplicate publication and exclude Pavlidou and Williams (2014) from all further analyses.

This leaves us with four papers. As the three papers by Pavlidou and colleagues report a (mostly highly) significant group difference while Nigro and colleagues report no significant group difference, it seems that there is a lab effect, where the size of the effect depends on the lab in which it was conducted. For the studies of Pavlidou and colleagues, the estimated effect size (using the same meta-analytic model as above) is $d = 0.95$. For the studies of Nigro and colleagues, the estimated effect size is $d = 0.27$. Assuming the effect size of Pavlidou and colleagues,

Nigro et al. (2015) should have had a high chance to detect an effect in a single study (for a two-tailed *t*-test, effect size $d = 0.95$, alpha = 0.05, and total sample size = 42, power = 0.85). Instead, Nigro et al. (2015) obtain two non-significant *p*-values. Similarly, if we assume the observed effect size of Nigro and colleagues, the outcomes of Pavlidou et al. are highly unlikely. A single study of Pavlidou's, assuming a population effect-size of $d = 0.27$, a two-tailed *t*-test sample size of 32, and two-tailed alpha-level of 0.05, should have an 11.5% chance to detect an effect. The probability of obtaining a significant effect in three consecutive studies of the same sample size is 0.115^3, or 0.0015.

Given the results of the adult studies, we tentatively conclude that a true effect size should be small: the meta-analytic effect size estimate of the four adult studies is $d = 0.27$. In this case, the results of Pavlidou and colleagues report an inflated effect size. Unfortunately, we were unable to further explore the potential reasons for this unusually large effect size, because the raw data have been lost in a computer crash (E. Pavlidou, personal communication, 5.2.16).

Obtaining an inflated effect size in multiple consecutive studies is unlikely enough to warrant scepticism about the credibility of the data, as it suggests that questionable research practices were used to obtain them (Schimmack, 2012). If we assume that the studies of Pavlidou and colleagues, for whatever reason, report inflated effect sizes, the remaining studies leave us with an effect size estimate of $d = 0.25$, 95% CI [-0.26, 0.77]. It is worth noting that the confidence intervals are very wide, thus showing that there is a high degree of uncertainty in this effect size estimate. There is also a high correlation between the observed effect and standard error, $r(5) = 0.73$. This is because the studies reporting large effects tend to have smaller samples and consequently larger variance. This suggests publication bias,

where studies with small samples are only published when they show significant results, and large studies show a more realistic effect size (Egger et al., 1997).

Taking all of these issues into account, it is difficult to draw conclusions. At face value, the effect size estimate is larger than zero. However, we found evidence of publication bias and questionable research practices, thus we cannot draw any conclusions about the presence or absence of an effect based on the available data. Note that publication bias and questionable research practices result in an inflated effect size, but it is still unclear whether the real effect size is different from zero or not. The limited number of studies would make further meta-analyses with bias correction difficult to interpret. Furthermore, there is no agreed-upon reliable method of correcting for publication bias, and different methods can yield different results (Van Elk et al., 2015). Thus, future research – preferably a large-scale pre-registered study – is needed to establish whether or not there is a group difference between dyslexic and control participants on AGL tasks. If an effect exists, we estimate its size to be around $d = 0.25$.[2] A future – adequately powered and pre-registered – study could be conducted to assess whether there is evidence for a small effect. Assuming the optimistic estimate of $d = 0.25$, such a future study would need to test a total of almost 400 participants to achieve 80% power to find a significant group difference in a one-tailed t-test. This demonstrates that the existing studies are underpowered to find an effect, if it is real.

Serial Reaction Time Task: Is there a group difference?

Compared to the results of the AGL task, the outcomes of the SRT task are more difficult to evaluate because there are different dependent variables that

---

2 Note that this is an optimistic approximation, as a fixed-effect model provides the estimate of $d = -0.10$, 95% CI [-0.30, 0.1].

researchers have used. This precludes the calculation of a single effect size. Furthermore, it is rare for researchers to report the descriptive statistics across blocks, as it is more common to represent the data in a figure. We emailed all authors of the published studies to ask for the condition averages, but the few positive responses were not sufficient to generate a meta-analysis based on the available data. Instead, we turn to a description of the existing studies, issues that make some studies difficult to interpret, and tentatively conclude that the evidence for a group difference is somewhat underwhelming.

Chronologically, the first study which used an SRT task in dyslexic versus control participants was by Kelly et al. (2002). The authors used this task to evaluate the proposal that motor learning is impaired in dyslexia (Nicolson & Fawcett, 1990). The participants were 14 adults with dyslexia, and 14 controls. None of the interactions involving group approached significance, all $F < 1.72$, though the participants with dyslexia were marginally slower at responding to all stimuli than the controls.

In a subsequent study, Vicari et al. (2003) reported evidence for a group difference with 18 children and adolescents with dyslexia, and 18 age-matched controls. The authors found a significant interaction between the reduction in RTs across blocks, suggesting a steeper learning curve for control than dyslexic participants, $F(5,170) = 5.95$, $p < 0.0001$. However, this was due to a significant *increase* in reaction time across blocks for dyslexic participants, and an approximately symmetrical decrease in RTs for control participants. In the comparison between the last sequential and the random block, the authors do not report whether or not the interaction between block and group reached significance. The trends were in the expected direction: for controls, the RTs increased for the random block from 397.3

ms to 488.2 ms ($p < 0.05$), while for the dyslexic group, there was a non-significant decrease from 550.1 to 542.7 ms. In sum, dyslexic participants showed overall higher RTs, an increase in reaction time across blocks, and an apparent total lack of learning as shown by the comparison of the last sequential and the random block. Altogether, this is in line with the explanation that poor overall performance was caused by lack of attention (see also Ise & Schulte-Körne, 2012; Staels, 2016). If the dyslexic participants were paying less attention during the task, as evidenced by overall higher RTs, the apparent absence of statistical learning may simply reflect reduced attention to the sequences. It is possible that the pattern of results reflects worse statistical learning for the dyslexic than control group masked by random noise, however, the study does not provide strong evidence for such a statistical learning deficit.

Three further studies from the same lab addressed the possibility of a group difference in statistical learning using the SRT task. Menghini et al. (2006) tested 14 adults with dyslexia and 14 controls. The conclusion of a group difference is based mainly on the finding that the dyslexic group did not show significantly different RTs to the random compared to the last sequential block, while the control group did. As discussed in the introduction, this does not constitute evidence for a group difference (Gelman & Stern, 2006), because two p-values falling on different sides of the arbitrary significance cut-off may still come from the same population. The authors also report that the decrease in RTs across sequential blocks was not significantly different across groups, F for the interaction < 1. Vicari et al. (2005) report group differences in a sample of 32 children, but this conclusion is based, again, on a significant difference between the sequential and random blocks for the control but not for the dyslexic children. In a subsequent study, Menghini et al. (2010) used a larger sample of children and adolescents, namely 60 with dyslexia, and 65 controls.

The participants were tested on numerous cognitive tasks, including an SRT task. Here, the authors found no significant differences between dyslexic participants and controls.

Other studies provide some evidence for a group difference on the statistical learning task. Two of these are studies with adults. Stoodley et al. (2006) tested 40 participants found a significant interaction between the random and repeated blocks, with the expected larger difference between block types for control than dyslexic participants. RTs were also significantly slower for dyslexic than control groups in the final sequential blocks (though it was not reported whether the interaction between block and group was significant). Similarly, in a study with 23 adults, Howard et al. (2006) report significantly smaller differences between random and sequential trials for dyslexic and control participants, though there was no significantly different decrease in RTs as a function of trial type for the two groups.

Further two studies provide evidence for a group difference in children. Stoodley et al. (2008) report a significantly larger difference between the sequential and random block for control than dyslexic groups, with a total of 89 participants. Jiménez-Fernández et al. (2011) found both the expected increased difference in controls in the comparison of the random and final sequential blocks, in a sample of 28 children. Furthermore, they report a significant interaction between block and group, showing a flatter learning curve for dyslexic compared to control participants.

However, there are also additional failures to find group differences. Bennett, Romano, Howard Jr, and Howard (2008) report a steeper learning curve of dyslexic compared to control adults, which goes against the prediction of a statistical learning deficit in dyslexia. Further studies report no group differences on SRT tasks in adults (Rüsseler et al., 2006) and children (Hedenius et al., 2013). Waber et al. (2003) used a

large sample size of over 400 children and regression analyses to test for the effect of dyslexia on SRT task performance. They found no group differences: overall differences in speed were better accounted for by symptoms of attentional difficulties, which had also been measured.

As discussed in the introduction, the mixed findings could reflect either a disproportionate amount of Type-I errors due to publication bias, or a disproportionate amount of Type-II errors due to small samples, which may yield insufficient power to detect a true effect. In the former case, we should expect that larger studies are more likely to yield non-significant findings. In the latter case, large studies should be more likely to show an effect as they should have more power to find a hypothetical real group difference (Egger et al., 1997). In contrast to the AGL task, the varying experimental procedures used in SRT tasks – and the absence of the condition means and standard deviations for most of the studies – do not allow us to meaningfully compare effect sizes across experiments. We can, however, calculate average sample size of studies which found significant group differences versus those that did not. Five studies show group differences on at least one of the dependent variables used in the SRT task, and have an average sample size of 43.2 participants (SD = 26.5, range = 23-89). In contrast, eleven studies have found no significant group differences on any measures between dyslexic and control participants, with an average sample size of 82.6 participants (SD = 116.4, range = 24-422).

In conclusion, multiple studies have addressed the possibility that performance on SRT tasks is worse in participants with dyslexia compared to controls. However, the results are mixed. Two larger studies (N > 100) found no group differences; smaller studies report a mixture of positive and null-findings. Thus, the number of positive results is likely to be inflated due to publication bias. It is also noteworthy

that the null-findings outnumber the positive findings, further suggesting that the effect is less robust than may appear at first glance. As for the AGL results, it remains difficult to conclude whether or not the true effect size is different from zero: even in the presence of publication bias, the possibility remains that an effect exists, but is smaller and less reliable than would appear from the literature.

Due to variations in the experimental procedure and dependent variables we are unable to estimate a single effect size that could be used to guide future research. Instead, a future pre-registered study could specify the expected effect size after pilot testing, using the same procedure that would be employed in the subsequent experiment.

Directions for future research and conclusions

In the current systematic review, we aimed to determine whether there is evidence for a group difference in statistical learning between dyslexic and control participants. We found that the existing evidence does not allow us to draw strong conclusions, due to the presence of publication bias and questionable research practices.

While it would have been more satisfying to find evidence either for or against the group difference, the current conclusion is also important. It suggests that researchers may not want to assume that there is a statistical learning deficit when designing theories and planning experiments that build on the existing literature. This could lead to unfruitful spending of the researchers' resources. Conversely, researchers should also not assume the absence of a group difference. The findings of our review stress the need to adequately-powered, pre-registered studies (Van Elk et al., 2015). Such a study should be designed specifically in such a way that it would allow the researcher to obtain evidence for or against a group difference. This

involves maximising the strength of the manipulations and the reliability of the experimental measures, and ensuring that the sample size is large enough to be likely to yield informative data.

Future research would also benefit from a thorough consideration of the theories on statistical learning and dyslexia. First, it is not yet established that performance on statistical learning tasks reflects a domain-general statistical learning mechanism (Frost et al., 2015; Siegelman & Frost, 2015). If no such domain-general mechanism exists, this would substantially change the conclusions that can be drawn from studies that find group differences between dyslexic and control participants. A domain-general statistical learning mechanism would be expected to be correlated with reading ability, presumably because it is necessary for learning the print-to-speech correspondences (Apfelbaum et al., 2013; Arciuli & Simpson, 2012) and/or acquiring knowledge of the statistical distributions of letter patterns (Cassar & Treiman, 1997; Pacton et al., 2005; Pacton et al., 2001; Pollo et al., 2009). If, instead, statistical learning is domain-specific, one would not expect a strong correlation between reading ability and performance on statistical learning tasks. If such a correlation is found, it is more likely attributable to broader cognitive abilities, such as attentional capacity. As such, a relationship between performance on statistical learning tasks and a diagnosis of dyslexia would be uninformative about the causal chain that leads from statistical learning to reading ability.

To circumvent this problem, it is, first and foremost, important to understand if and to what extent statistical learning is domain-general. Even if statistical learning is constrained by the input domain, it is still possible that there is a universal computational mechanism that operates with the same efficiency across all domains in the mind of a given individual. More specifically to reading, future research could

focus on statistical learning processes that are directly linked to the cognitive reading system. Previous research has studied the relationship between print-to-speech correspondences and letter-sound distributions to reading (Apfelbaum et al., 2013; Cassar & Treiman, 1997; Deacon, Benere, & Castles, 2012; Pollo et al., 2009; Rothe, Cornell, Ise, & Schulte-Körne, 2015; Rothe et al., 2014; Treiman, Kessler, Zevin, Bick, & Davis, 2006), but to date little to nothing is known about the direction of causality of the relationships, or the role of these relationships in dyslexia.

Finally, future research should also bear in mind the heterogeneity of dyslexia. The cognitive profiles of children with dyslexia vary (Castles & Coltheart, 1993; Pennington, 2006; Ziegler et al., 2008). This is proposed to reflect the complex system of skills which children need in order to become efficient readers: as a deficit in any single skill or in a combination of skills could result in reduced reading ability, it is unlikely that there is a single cause of dyslexia (Coltheart, 2015). In turn, a skill which is consistently impaired in all children with dyslexia is unlikely to be a cause of dyslexia. If poor reading is caused by, say, a statistical learning deficit in only some of the children with dyslexia, finding a consistent group difference in all or most studies would be more likely to reflect systematic bias (Gunn et al., 2016), such as a confounding variable, a cognitive deficit that is epiphenomenal to dyslexia, or one that is caused by low exposure to reading.

In conclusion, future research is needed to determine whether there is evidence for a statistical learning deficit in dyslexia. The existing studies did not allow us to draw conclusions, as we detected evidence of publication bias and/or questionable research practices. Future research could conduct pre-registered studies, and consider in depth what a group difference would mean on a theoretical level.

References

Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology, 49*(7), 1348.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286-304.

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.

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*(6), 858.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. West Sussex, UK: John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Confidence and precision increase with high statistical power. *Nature Reviews Neuroscience, 14*(8). doi:10.1038/nrn3475-c4

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89*(4), 631.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149-180.

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.

Christley, R. (2010). Power and error: increased risk of false positive results in underpowered studies. *Open Epidemiology Journal, 3*, 16-19.

Coltheart, M. (2015). What kinds of things cause children's reading difficulties? *Australian Journal of Learning Difficulties*, 1-10.

Cramer, A. O., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P., . . . Wagenmakers, E.-J. (2015). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, 1-8.

Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg? Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition, 122*(1), 110-117.

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.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629-634.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one, 4*(5), e5738.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117-125.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243-1252.

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science, 9*(6), 641-651.

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis - a "garden of forking paths" - explains why many statistically significant comparisons don't hold up. *American Scientist, 102*, 460-465. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*(4), 328-331.

Gelman, A., & Weakliem, D. (2009). Of beauty, sex and power: Too little attention has been paid to the statistical challenges in estimating small effects. *American Scientist, 97*(4), 310-316.

Gunn, L. J., Chapeau-Blondeau, F., McDonnell, M., Davis, B., Allison, A., & Abbott, D. (2016). Too good to be true: when overwhelming evidence fails to convince. *arXiv preprint arXiv:1601.00900.*

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.

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.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine, 2*(8), 696-701. doi:10.1371/journal.pmed.0020124

Ise, E., & Schulte-Körne, G. (2012). Implizites Lernen und LRS: Spielen Defizite im impliziten Lernen eine Rolle bei der Entstehung von Schwierigkeiten im Lesen und Rechtschreiben? *Lernen und Lernstörungen, 1*(2), 79-97. doi:10.1024/2235-0977/a000011

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.

Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.

Kahta, S., & Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia*, 1-16.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia, 8*(1), 43-52.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition, 20*(04), 543-587.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-B42.

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.

Meehl, P. E. (1990). Why Summaries of Research on Psychological Theories Are Often Uninterpretable. *Psychological Reports, 66*(1), 195-244. doi:10.2466/Pr0.66.1.195-244

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., . . . Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. *Neuropsychologia, 48*(4), 863-872.

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.

Nicolson, R., & Fawcett, A. (1990). Automaticity: A new framework for dyslexia research? *Cognition, 35*(2), 159-182.

Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2015). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of Dyslexia*, 1-17.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*(1), 1-32.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development, 76*(2), 324-339.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: the case of orthographic regularities. *Journal of Experimental Psychology: General, 130*(3), 401.

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.

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.

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.

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.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238.

Petersson, K.-M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language, 120*(2), 83-95.

Pollo, T. C., Kessler, B., & Treiman, R. (2009). Statistical patterns in children's early writing. *Journal of Experimental Child Psychology, 104*(4), 410-426.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences, 28*(01), 1-14.

Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia, 10*(1), 61-76.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855-863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219.

Roodenrys, S., & Dunn, N. (2008). Unimpaired implicit learning in children with developmental dyslexia. *Dyslexia, 14*(1), 1-15.

Rosenthal, R. (1979). The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin, 86*(3), 638-641.

Rothe, J., Cornell, S., Ise, E., & Schulte-Körne, G. (2015). A comparison of orthographic processing in children with and without reading and spelling disorder in a regular orthography. *Reading and Writing, 28*(9), 1307-1332.

Rothe, J., Schulte-Körne, G., & Ise, E. (2014). Does sensitivity to orthographic regularities influence reading and spelling acquisition? A 1-year prospective study. *Reading and Writing, 27*(7), 1141-1161.

Rouder, J. N., Speckman, P. L., Sun, D. C., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237. doi:10.3758/Pbr.16.2.225

Royall, R. M. (1986). The Effect of Sample-Size on the Meaning of Significance Tests. *American Statistician, 40*(4), 313-315. doi:10.2307/2684616

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.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*(10), 1173.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*(2), 115-129. doi:10.1037//1082-989x.1.2.115

Seidenberg, M., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*(9), 353-361.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105-120.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22(11),* 1359-1366. Doi: 10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). p-Curve and Effect Size Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science, 9*(6), 666-681.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534-547. doi:10.1037/a0033242

Staels, E. (2016). *Cognitive causes of dyslexia: New hypotheses investigated using improved methodologies.* Vrije Universiteit Brussel.

Sterling, T. D., Rosenbaum, W., & Weinkam, J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*(1), 108-112.

Stoodley, C. J., Harrison, E. P., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia, 44*(5), 795-798.

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.

Swanson, H. L. (1996). Meta-analysis, replication, social skills, and learning disabilities. *The Journal of Special Education, 30*(2), 213-221.

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology, 93*(1), 1-24. doi:10.1016/J.Jecp.2005.06.008

Van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Frontiers in Psychology, 6*.

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.

Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia, 41*(1), 108-114.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. Retrieved from http://www.jstatsoft.org/v36/i03/.

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.

Ziegler, J., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition, 107*(1), 151-178.