Statistical Learning in Specific Language Impairment

A Meta-Analysis

Lammertink, I.; Boersma, P.; Wijnen, F.; Rispens, J.

Published in:
Journal of Speech, Language, and Hearing Research

DOI:
10.1044/2017_JSLHR-L-16-0439

JSLHR

Research Article

Statistical Learning in Specific Language Impairment: A Meta-Analysis

Imme Lammertink,[a] Paul Boersma,[a] Frank Wijnen,[b] and Judith Rispens[a]

Purpose: The current meta-analysis provides a quantitative overview of published and unpublished studies on statistical learning in the auditory verbal domain in people with and without specific language impairment (SLI). The database used for the meta-analysis is accessible online and open to updates (Community-Augmented Meta-Analysis), which facilitates the accumulation and evaluation of previous and future studies on statistical learning in this domain.
Method: A systematic literature search identified 10 unique experiments examining auditory verbal statistical learning in 213 participants with SLI and 363 without SLI, aged between 6 and 19 years. Data from qualifying studies were extracted and converted to Hedges' *g* effect sizes.
Results: The overall standardized mean difference between participants with SLI and participants without SLI was 0.54, which was significantly different from 0 ($p < .001$, 95% confidence interval [0.36, 0.71]).
Conclusion: Together, the results of our meta-analysis indicate a robust difference between people with SLI and people without SLI in their detection of statistical regularities in the auditory input. The detection of statistical regularities is, on average, not as effective in people with SLI compared with people without SLI. The results of this meta-analysis are congruent with a statistical learning deficit hypothesis in SLI.
Supplemental Material: https://doi.org/10.23641/asha.5558074

Natural languages are structured at the level of sound (phonology), word formation (morphology), and sentence (syntax). These structures are reflected by statistical regularities in speakers' verbal output. Children learning their native language unconsciously detect and extract these regularities (Romberg & Saffran, 2010). This process, called statistical learning, is thought to be fundamental for the earliest stages of language acquisition (Evans, Saffran, & Robe-Torres, 2009). Two types of statistical learning are generally distinguished: distributional statistical learning and sequential statistical learning. Distributional statistical learning is about the detection of frequencies with which certain linguistic elements or structures occur. Sequential statistical learning concerns the detection of the sequential ordering and co-occurrence of concrete elements (e.g., syllables) in the auditory input in time (Kerkhoff, Bree, & Wijnen, submitted). This meta-analysis focuses on sequential statistical learning, and therefore, from here onward, the term *statistical learning* refers to sequential but not distributional statistical learning.

Individual performance on statistical learning tasks has been shown to predict sentence comprehension (Misyak & Christiansen, 2012), processing of relative clause sentences with long-distance dependencies, and lexical and oral language skills in participants' native language (Evans et al., 2009; Mainela-Arnold & Evans, 2014). Because tracking statistical patterns appears crucial for language acquisition and people differ in their ability to do this, it is not surprising that deficits in the ability to detect statistical patterns and relations in the input have been put forward as an explanation for impairments of language acquisition, notably specific language impairment (SLI; Evans et al., 2009; Hsu & Bishop, 2011; Ullman & Pierpont, 2005). A considerable number of studies looked at the domain specificity of this type of learning deficit in SLI. A recent meta-analysis by Obeid, Brooks, Powers, Gillespie-Lynch, and Lum (2016) summarized these findings and concluded that people with SLI perform worse on statistical learning tasks compared with typically developed people but that this difference in performance did not vary as a function of task modality (visual; visual motor and auditory modality) or age. The current meta-analysis provides a more extensive

[a]Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands
[b]Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands
Correspondence to Imme Lammertink: i.l.lammertink@uva.nl

Editor: Sean Redmond
Associate Editor: Klara Marton
Received November 28, 2016
Revision received April 20, 2017
Accepted July 6, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0439

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

quantitative investigation of the difference in statistical learning ability between people with and without SLI[1] in the auditory domain. Different from Obeid et al. (2016), our focus is on the auditory linguistic domain. Specifically, we were interested to see whether a difference in statistical learning performance between people with and without SLI varied as a function of linguistic level (word segmentation vs. grammar) or age at which learning took place.

Statistical Learning in the Laboratory

Many experimental studies of statistical learning focus on learning dependencies. These dependencies can be learned at different linguistic levels (word segmentation vs. grammar). We first discuss examples of artificial word segmentation studies followed by examples of artificial grammar learning studies.

In experiments that simulate word segmentation, participants are exposed to a continuous stream of syllables that are organized according to a set of statistical regularities. The stimuli are designed in such a way that transitional probabilities of sequences of certain adjacent syllables are higher than transitional probabilities of other adjacent syllables (continuous relationship), reflecting word boundaries (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). After exposure, participants perform a lexical decision task (or word recognition via a preferential looking paradigm in the case of infant studies) in which they hear sequences of syllables that had high transitional probabilities in the exposure phase (reflecting words) as well as sequences of syllables that had low to zero transitional probabilities in the exposure phase. Accordingly, adult participants have to indicate whether the words they are presented with are part of the language they were familiarized with or not. In infant studies, the listening times to the sequences of syllables with high transitional probabilities versus sequences of syllables with low transitional probabilities are compared. Results show that adults and infants are able to distinguish such artificial high-probability words from artificial low-probability words on the basis of adjacent transitional probabilities.

Contrary to word segmentation studies, the stimuli in artificial grammar learning studies consist of already segmented words that have primary stress and minimal coarticulation and are separated by pauses. Artificial grammar learning studies aim to resemble grammatical phenomena present in natural language. In natural language, for example, grammatical relations are present among functional elements (e.g., *is* and *–ing*) across interleaved lexical elements (e.g., Grandma *is singing*; example taken from Sandoval & Gómez, 2013). In experimental designs that test this type of learning, participants are exposed to strings generated, unknown to the learner, by a miniature artificial grammar. The grammar follows a set of nonadjacent (discontinuous) dependency relations (Gómez, 2002), a set of predictive relations (cf. Saffran, 2002), or a set of finite rules (finite state grammar; Gómez & Gerken, 1999). The procedure of artificial grammar learning designs is similar to the procedure in word segmentation studies: After a period of exposure to the language, participants are tested with strings that either conform to the grammar (grammatical items) or violate the grammar (ungrammatical items), and participants have to indicate whether the string they hear is grammatical or ungrammatical. More important, participants are asked to judge strings with elements that they have heard during the familiarization phase of the experiment as well as strings with novel elements that they have not heard before to test for generalization of the rule (although not all artificial grammar learning studies test for generalization; see Grama, Kerkhoff, & Wijnen, 2016).

Cognitive Processes Involved in Auditory Verbal Statistical Learning

As stated in our operational definition, statistical learning requires sensitivity to regularities in the input (e.g., statistical cues like transitional probabilities in word segmentation and [non]adjacent dependencies in artificial grammar learning). However, there are also other cognitive processes involved in statistical learning in the auditory linguistic domain such as phonological awareness (the ability to analyze and manipulate incoming phonemes and syllables), verbal short-term memory, and verbal working memory. Both word segmentation and artificial grammar learning involve the temporary storage of incoming input, which is necessary to pick up the statistical regularities between elements in the input (verbal short-term memory). In addition, artificial grammar learning, compared with word segmentation, requires processing of long-distance dependencies and generalizing those dependencies to novel items. Long-distance dependencies have been argued to put more demand on working memory than adjacent dependencies (see, e.g., theoretical models on resource limitation of Gibson [1998]), and generalization is more demanding than recognition of items previously introduced (Thompson & Newport, 2007). Therefore, we hypothesize that artificial grammar learning, compared with word segmentation, is more demanding on working memory capacity. In the following section, we discuss how this difference between both levels of learning might disadvantage individuals with SLI in their auditory verbal statistical learning performance.

Statistical Learning in SLI

In natural language, SLI is characterized by problems at the grammatical level (e.g., subject–verb agreement, past-tense marking; Leonard, 2014) as well as at the word segmentation level (e.g., lexical–phonological deficits observed in gating and nonword repetition tasks; see Mainela-Arnold, Evans, & Coady, 2010; Graf Estes, Evans, &

[1]When we speak of people without SLI, we mean people who are matched in age and/or IQ to participants with SLI (see Table 1), who have no reported (history of) hearing, language, or learning problems and no reported (history of) neurological impairment or illness.

Else-Quest, 2007). In artificial language, we see a similar pattern: Most studies investigating auditory verbal statistical learning in SLI show that participants without SLI outperform participants with SLI both in word segmentation and in grammar learning tasks (word segmentation: Evans et al., 2009; grammar: Hsu, Tomblin, & Christiansen, 2014; Lukács & Kemény, 2014; Mainela-Arnold & Evans, 2014; Mayor-Dubois, Zesiger, Van der Linden, & Roulet-Perez, 2014). It is known that people with SLI exhibit deficits in verbal short-term memory and verbal working memory as well (Archibald & Gathercole, 2006; Marton, Eichorn, Campanelli, & Zakariás, 2016; Montgomery, 2003). As these processes are involved in auditory statistical learning, it might well be the case that these deficits influence the auditory verbal statistical learning abilities of people with SLI. Previous research, however, suggests that memory problems cannot solely explain auditory statistical learning problems. For example, individuals with SLI have problems with statistical learning in the nonverbal domain (Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Lum, Conti-Ramsden, Page, & Ullman, 2012; Obeid et al., 2016), which are unlikely to be caused by verbal short-term and working memory problems. In addition, Hsu and Bishop (2014) report poor verbal sequence learning in children with SLI, even after controlling for limitations of verbal short-term memory (Hebb repetition task). Taken together, results of previous studies are congruent with the hypothesis that SLI is associated with a "statistical learning disadvantage." The magnitude and moderators of this disadvantage, however, are unknown. Therefore, the primary purpose of the current meta-analysis was to assess the magnitude of this statistical learning disadvantage in the auditory verbal domain. The second goal was to explore the potential impact of linguistic level and age at which learning takes place. We wanted to explore whether the statistical learning disadvantage is more severe in artificial grammar learning than word segmentation studies, as the former type of learning is more demanding on verbal working memory capacity, which is generally affected in SLI. With the second meta-regression, we explore whether age moderates the statistical learning disadvantage. Previous studies investigating the influence of age in statistical learning have provided mixed results. Obeid and colleagues (2016) reported no effect of age on statistical learning differences between people with and without SLI across different modalities of learning. Lum and colleagues (2014), however, reported smaller differences in visuospatial statistical learning performance between people with and without SLI for older compared with younger participants. Likewise, studies investigating the developmental trajectory of statistical learning in typically developing people have reported mixed results. Some studies report that there is no evidence for a difference in statistical learning performance between adults and children (visual domain: Kirkham, Slemmer, & Johnson, 2002; auditory domain: Saffran et al., 1997), whereas others do report that statistical learning performance improves with age (visual domain: Arciuli & Simpson, 2011; auditory domain: Lukács & Kemény, 2015; visuospatial: Meulemans, Van der Linden, & Perruchet, 1998).

The Current Study

The current meta-analysis provides an estimate of the magnitude of the statistical learning disadvantage in people with SLI by means of a quantitative overview of both published and unpublished studies that investigate statistical learning in the auditory linguistic domain in people with and without SLI. In a first step, we calculated the standardized averaged mean difference (effect size measure) in performance on statistical learning tasks in people with and without SLI. In a second analysis, we explored whether the effect size measure was moderated by linguistic level (word segmentation vs. grammar) and age.

Method

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement to organize the current meta-analysis (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). Effect size calculations were done in the statistical software R (R Core Team, 2016): Formulas were implemented via the R *compute.es* package (Del Re, 2013), and statistical analyses on the effect size measures were conducted with the R *meta* (Schwarzer, 2015) and *metafor* (Viechtbauer, 2010) packages.

Literature Search

Systematic searches for empirical articles were conducted in February 2016 using a combination of prespecified key word combinations (details of all key words, Boolean operators, and syntax used for each database can be found in Supplemental Material S1). We conducted our searches in five different sources including PubMed, Education Resources Information Center, PsycINFO, Linguistics and Language Behavior Abstracts, and Open Access Theses and Dissertations. In addition, we asked experts in the field to inform us of any published or unpublished studies via two different calls (LINGUIST List and Cogdevsoc list; July 2016). These combined searches yielded 161 articles (PubMed: 26 hits, Education Resources Information Center: 25 hits, PsycINFO: 64 hits, Linguistics and Language Behavior Abstracts: 38 hits, Open Access Theses and Dissertations: five hits, and experts in the field: three hits).

Inclusion Criteria and Study Selection

To be included in the meta-analysis, studies were required to meet the following criteria: (a) A study should report on original empirical research data. Both published and unpublished studies were eligible, including articles in refereed journals, nonrefereed journals, dissertations, and conference presentations; (b) a study should have an experimental design that tests sequential statistical learning in the auditory verbal domain assessed via a word segmentation, grammaticality judgment, or related task; (c) as we aimed

to test whether participants implicitly detected the statistical regularity, participants should not receive any explicit instruction or feedback regarding the underlying structure of the artificial language to be learned or on their behavior during the training or test phase; and (d) selected studies include one group of participants with SLI and one group of age-matched controls who do not have language impairments. More important, we only included studies that identified participants with SLI on the basis of inclusionary and exclusionary criteria typical for SLI. Therefore, studies had to report scores on standardized language tests[2] or use a test battery that differentiates between participants with and without a history of SLI (e.g., Tomblin battery; Tomblin, Freese, & Records, 1992; see Table 1). In addition, a nonverbal IQ measure[3] and no history of neurological or emotional delays should be reported for both participant groups. It is important to mention that the inclusion and exclusion criteria for SLI vary across the studies in our sample (see Table 1). We only included studies, however, that based their inclusion criteria on both standardized language tests and IQ scores. If studies failed to report on one of these criteria (or if information on these criteria could not be confirmed via contact with the authors), the study was excluded from the analysis. In addition, when studies included children with an IQ below 80, the control group and the group with SLI had to be matched on nonverbal IQ to ensure that differences in statistical learning performance are not the result of lower IQ scores. Finally, to be included in the analysis for the current article, studies had to be conducted before September 2016 (but see footnote 4). However, as our database is accessible online and open to update, future studies can be added, which facilitates the accumulation and evaluation of previous and future studies on statistical learning in this domain (Tsuji, Bergmann, & Cristia, 2014). No start date for publications was set to find as many studies as possible. For an overview of the exact inclusionary and exclusionary criteria for the studies in our final sample, see Table 1.

After removing duplicates, 81 studies (78 published articles and three unpublished conference posters) remained. Two reviewers independently conducted the study selection procedure. In a first step, both reviewers performed a full-text inspection of the 19 studies (16 published articles and three nonpublished conference posters) that were selected, based on screening of the title and abstract. The reviewers independently screened these full-text articles and posters according to the inclusion criteria. There was 95% (18/19 studies) agreement on the selection of these full-text studies (eight studies included, 10 studies excluded, one study for discussion). After discussion, the reviewers decided not to include the one study they had disagreed on because participants in this study had received feedback on their behavior during the test phase (Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013). As a result, the initial final selection consisted of eight studies (five published articles and three nonpublished conference posters).[4] For a visual representation of the literature search procedure, see Figure 1.

Four of the eight studies reported multiple individual experiments or multiple outcomes per participant group (Evans et al., 2009; Grunow, Spaulding, Gómez, & Plante, 2006; Hsu & Bishop, 2014; Torkildsen, 2010). If the data necessary to compute the individual effect size were available for each experiment separately and the groups of participants tested in the experiments were independent (i.e., different participants), all of the experiments of that study were included in the meta-analysis. Only the study of Hsu et al. (2014) met these criteria. For the other three studies with multiple experiments (Evans et al., 2009; Grunow et al., 2006; Torkildsen, 2010), only one effect size measure was incorporated into the final analysis (for more details on our decisions with respect to this part, see the subsection Effect Size Calculation). This resulted in a final sample of 10 experiments.

Sample Description

The eight studies (10 experiments) we included in our analysis were published (six studies) or presented (two studies) between 2006 and 2017 (see footnote 4). The experiments collectively examined 213 participants with SLI and 363 controls, all between 6 and 19 years old. The dependent variable was slightly different across the 10 experiments. In six experiments, the outcome variable was the overall accuracy scores on a grammaticality judgment task; in three experiments, the outcome variable was the overall accuracy score on a word segmentation task; and in one experiment, the outcome variable was an event-related potential (ERP: P600).

Effect Size Calculation

For each individual experiment, we calculated the effect size (Hedges' *g*) as the standardized mean difference (SMD)[5] in performance between the participants with and without SLI. The SMD was chosen over the raw mean difference, because the dependent variables differed across studies (ERP amplitude vs. grammaticality judgments).

All formulas used to calculate the SMD and the approximation of the variance of the SMD for each individual experiment are shown in Figure 2 and were taken from the R *compute.es* package (Del Re, 2013). The effect size was calculated so that positive values indicated that

[2]Participants with SLI scored at least 1.25 standard deviations below age norms.

[3]Nonverbal IQ had to fall within the normal range (> 80), or when the lower limit of IQ was < 80, the control group and the group with SLI had to be matched on nonverbal IQ.

[4]During the review of our current meta-analysis, the poster of Haebig and colleagues got published as an article in *The Journal of Child Psychology and Psychiatry*. Therefore, the final data set consists of six published articles and two nonpublished conference posters.

[5]The standardized mean difference expresses the size of the effect in each study relative to the variability observed in that study (Higgins & Green, 2011).

Table 1. Overview of the study sample characteristics for each individual experiment in our sample.

Study	Native language	Sample size		Mean age (years)		SLI inclusionary criteria	Matching participants within study
		SLI	Control	SLI	Control		
Evans et al. (2009)	English	35	78	9.6	7.9	(a) Nonverbal IQ > 85 (LIPS-R) (b) Normal hearing (c) Normal corrected vision (d) Normal oral speech and motor abilities (e) Expressive language composite score > 1.5 *SD*s below mean (CELF-3)	Age, nonverbal IQ
Evans et al. (2010)	English	14	14	16.5	15.6	(a) Nonverbal IQ > 80 (LIPS-R, WISC-R, WPPSI) (b) Normal hearing (c) No major neurological abnormalities (d) Absence of other dev. disorders (e) Expressive language composite score > 1.5 *SD*s below mean (CELF-R) (f) Receiving speech and language services	Age
Grunow et al. (2006)	English	22	22	19.1	18.5	(a) Normal nonverbal IQ (Test of Nonverbal Intelligence III[a]) and IQ should not differ from controls (b) Normal hearing (c) No history of seizures/head trauma (d) No diagnosis of ADHD (e) Language impairment status attested via Tomblin battery[b]	Age, nonverbal IQ
Haebig et al. (2017)	English	25	30	10.4	10.4	(a) Normal nonverbal IQ (WISC-4) (b) Language assessment via PPVT-4 and CELF-4	Age, nonverbal IQ
Hsu et al. (2014)[c]	English	20	20	13.8	14.3	(a) Nonverbal IQ should not differ from controls (WISC-3) (b) Significant poorer language scores on composite scores from CELF-3, PPVT-R, CREVT, and listening QRI-2	Age, nonverbal IQ
Hsu et al. (2014)[d]	English	20	20	14.1	14.2	(a) Nonverbal IQ should not differ from controls (WISC-3) (b) Significant poorer language scores on composite scores from CELF-3, PPVT-R, CREVT, and listening QRI-2	Age, nonverbal IQ
Hsu et al. (2014)[e]	English	20	20	14.2	13.9	(a) Nonverbal IQ should not differ from controls (WISC-3) (b) Significant poorer language scores on composite scores from CELF-3, PPVT-R, CREVT, and listening QRI-2	Age, nonverbal IQ
Lukács & Kemény (2014)	Hungarian	29	87	9.1	9.1	(a) Nonverbal IQ > 85 (RAVEN) (b) Normal hearing (c) No neurological impairment (d) 1.5 *SD*s below age norms at two of the language tests assessed (Hungarian PPVT, TROG, MAMUT, NWR)	Age

(*table continues*)

Table 1. *(Continued).*

Study	Native language	Sample size		Mean age (years)		SLI inclusionary criteria	Matching participants within study
		SLI	Control	SLI	Control		
Mayor-Dubois et al. (2014)	French	18	65	10.1	10.0	(a) Nonverbal IQ within normal range (> 80; WISC-4) (b) SLI diagnosis confirmed via standardized language assessment by speech and language therapist (c) Participants with SLI were still pursuing speech and language therapy	Age
Torkildsen (2010)	Norwegian	14	14	6.1	5.9	(a) Nonverbal IQ > 70 (b) Normal hearing and (corrected) vision (c) No history of epilepsy, cerebral palsy, or brain hemorrhage (d) No structural abnormalities in speech system (assessed by speech pathologist) (e) 1.25 *SD*s below age norms at standardized language tests (assessment done by pedagogical–psychological services)	Age, nonverbal IQ

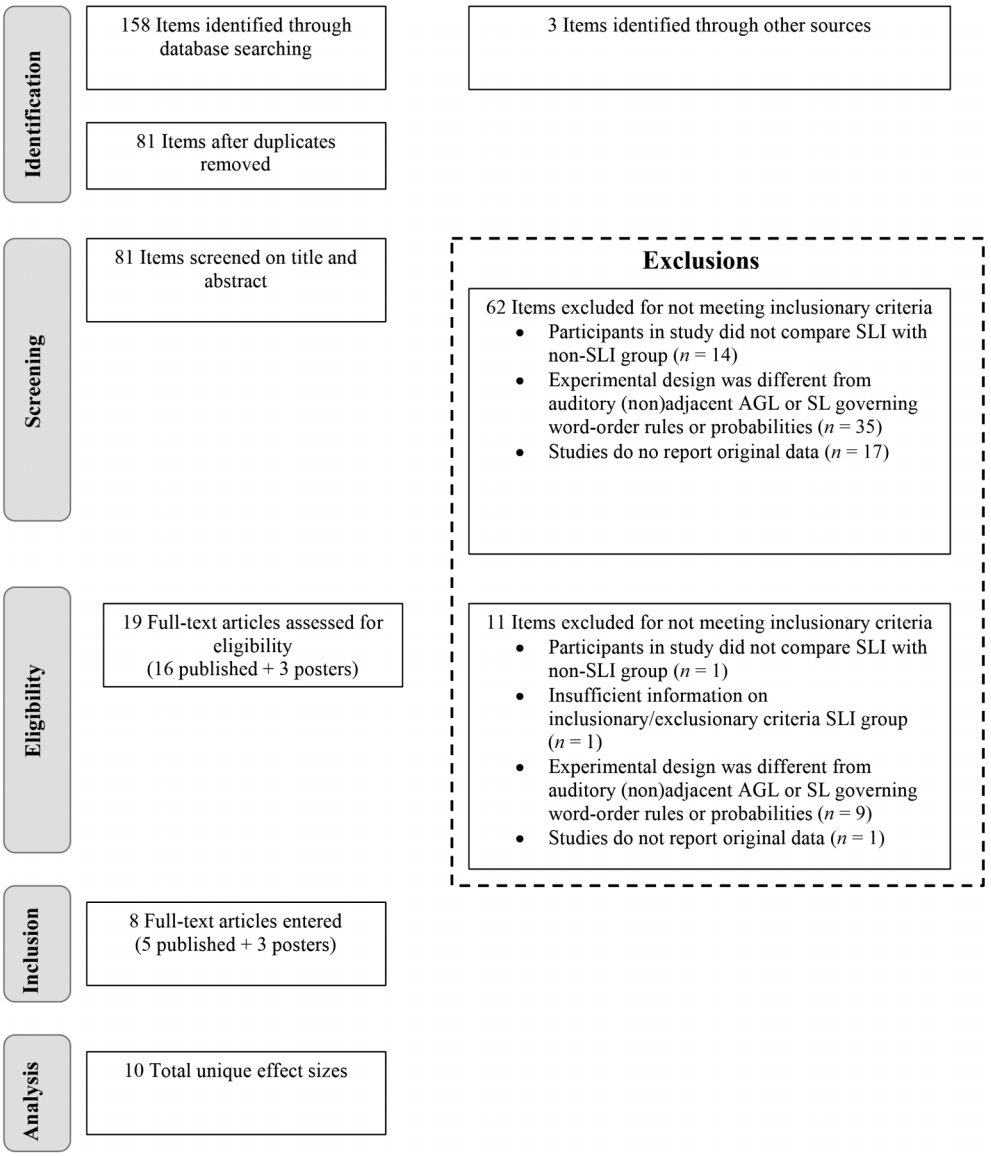
Note. SLI = specific language impairment; LIPS-R = Leiter International Performance Scale–Revised (Roid & Miller, 1997); CELF-3 = Clinical Evaluation of Language Fundamentals–Third Edition (Semel et al., 1995); WISC-R = Wechsler Intelligence Scale for Children–Revised (Wechsler, 1974); WPPSI = Wechsler Preschool and Primary Scale of Intelligence (no version reported); CELF-R = Clinical Evaluation of Language Fundamentals–Revised (Semel et al., 1987); WISC-4 = Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003); PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003); WISC-3 = Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991); PPVT-R = Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981); CREVT = Comprehensive Receptive and Expressive Vocabulary Test: Adult (Wallace & Hammill, 1997); QRI-2 = Qualitative Reading Inventory–Second Edition (Leslie & Caldwell, 1995); RAVEN = Raven's Progressive Matrices and Raven's Coloured Matrices (Raven et al., 1987); PPVT = Peabody Picture Vocabulary Test (Csányi, 1974); TROG = Test for Reception of Grammar (Lukács et al., 2012); MAMUT = Magyar Mondatutánmondási Teszt (Hungarian Sentence Repetition Test; Kas & Lukács, 2011); NWR = nonword repetition task (Racsmány et al., 2005).

[a]Brown et al. (1997). [b]Tomblin et al. (1992). [c]Participant characteristics for set size $x = 24$. [d]Participant characteristics for set size $x = 12$. [e]Participant characteristics for set size $x = 2$.

the participants without SLI outperformed the participants with SLI. For seven of the 10 experiments (Evans et al., 2009; Evans, Hughes, Hughes, Jackson, & Fink, 2010; Haebig, Saffran, & Weismer, 2017; Hsu et al., 2014; Lukács & Kemény, 2014), the SMD was calculated with the mean overall accuracy scores and the standard deviation scores for both participant groups (*mes2* function in the R *compute.es* package). For two experiments (Grunow et al., 2006; Torkildsen, 2010), the SMD was calculated from the reported *F* statistic on the main effect of group (*fes* function from the *compute.es* package), and for one experiment (Mayor-Dubois et al., 2014), the reported *t* statistic was used to calculate the SMD (*tes* function in the *compute.es* package).

As mentioned in the Inclusion Criteria and Study Selection section, it was not always possible to calculate multiple effect sizes for studies that ran multiple experiments. In the case of the Grunow et al. (2006) experiments, we calculated one effect size because the statistical information necessary to calculate a separate effect size for each of the different experimental conditions (low vs. high intervening X-category, generalization vs. nongeneralization items) was not available. Likewise, one effect size was obtained from the study by Evans at al. (2009), which reported on two different experiments. The second experiment was conducted 6 months after the first. The participants of Experiment 2, however, had all participated in Experiment 1, rendering the data of the second experiment correlated with a part of the data of the first experiment. A combined effect size, taking the correlation term between Experiments 1 and 2 into account, would have been the ideal solution because it would take into account the increased precision of within-subject measures (Borenstein, Hedges, Higgins, & Rothstein, 2009, pp. 28–30). However, it was impossible to determine the correlation term between the two experiments, because only parts of the data were correlated. Therefore, we included only the first experiment, which had twice as many participants as the second experiment. Last, Torkildsen (2010) recorded ERPs during both the exposure phase and the test phase. As we have no measures of performance during the exposure phase for the other studies in our sample, only the effect size measure of the ERPs recorded during the test phase is included.

Figure 1. Flowchart indicating data exclusion at each stage of the literature search procedure. SLI = specific language impairment; AGL = artificial grammar learning; SL = statistical learning.



Finally, we applied Hedges' *g* correction for small sample sizes to all 10 effect sizes, because most of the experiments had a sample size of less than 20 (Borenstein et al., 2009, p. 27).

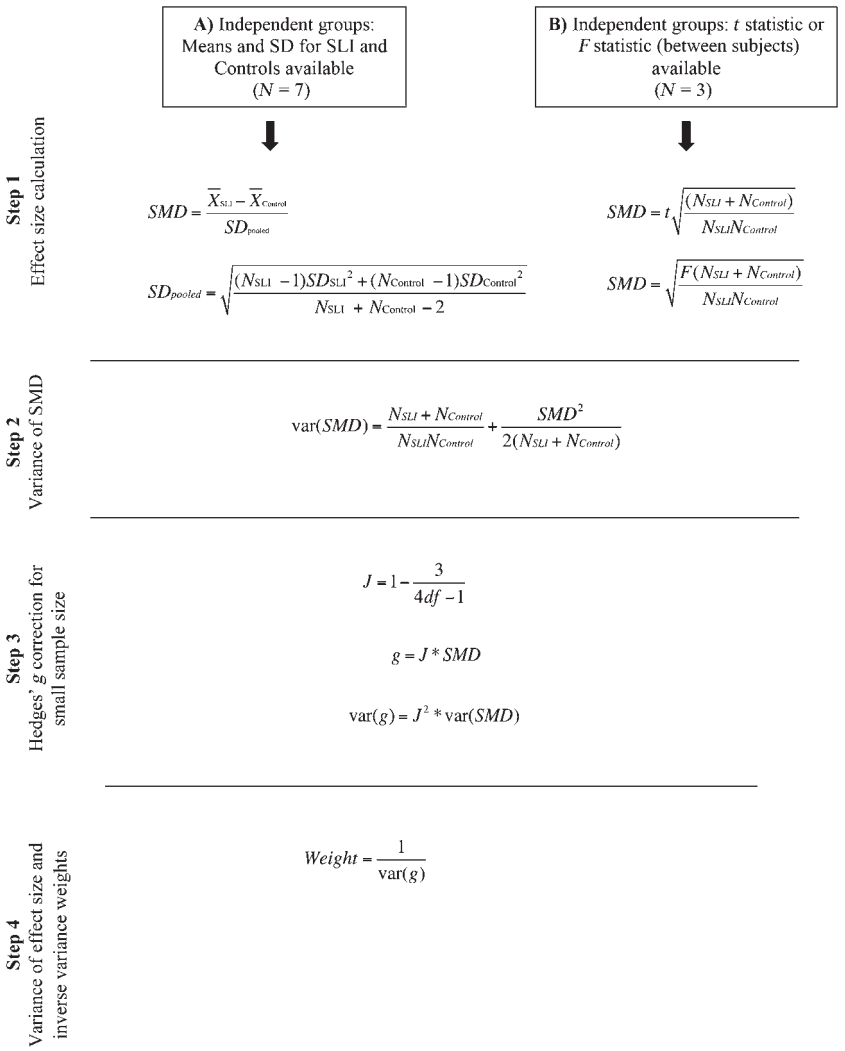
Results

Publication Bias

Meta-analyses are generally sensitive to publication bias. Publication bias reflects the tendency of a higher publication rate for studies with significant results compared with studies with nonsignificant results (Dickersin, 2005). Because it is more likely that published studies end up in a meta-analysis, the overall combined effect size might be overestimated when there is a publication bias in the sample used to compute the combined effect sizes (Borenstein et al., 2009, p. 278).

In the current meta-analysis, we analyzed funnel plot asymmetry as a potential indicator of publication bias (Egger, Smith, Schneider, & Minder, 1997). In our funnel plot (see Figure 3), the effect size of a particular experiment is plotted against the standard error of that particular experiment. The standard error can be interpreted as a measure of experiment size, as generally experiments with fewer participants have higher standard errors. In the absence of publication bias, a funnel plot is symmetric and funnel shaped; large experiments appear toward the top (low standard error) of the plot and generally cluster around

Figure 2. Formulae through which effect sizes (standardized mean difference/Hedges' *g*), variance (var), and weights for each individual study were calculated (Steps 1–4), and formula through which the average weighted standardized mean difference (SMD) was calculated (Step 5). Here, SLI refers to the values for the group with participants with specific language impairment (SLI). Control refers to the values for the control group. *N* signifies the number of participants in a given experiment. *F* is the reported *F* value for the main effect of group, and *t* is the reported *t* value for the between-groups effect.



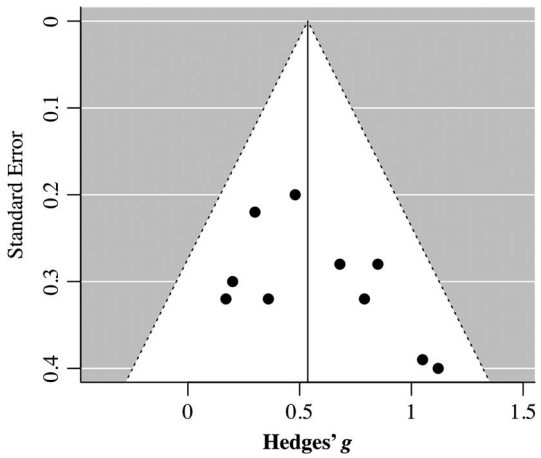
the mean effect size, whereas smaller experiments appear toward the bottom (higher standard error) of the graph and tend to be spread across a broader range of values. Visual inspection of our funnel plot (see Figure 3) seems to suggest asymmetry such that smaller experiments tend to have greater effect sizes (i.e., they appear more to the right side of the mean effect size than the left side). The latter could indicate publication bias, as small experiments are more likely to be found (or published) when the effect size is large compared with when the effect size is small. We performed a linear regression on funnel plot asymmetry (Egger et al., 1997). The test on funnel plot asymmetry was performed using the *regtest* function in the *metafor* (Viechtbauer, 2010) R package. The regression on funnel plot asymmetry was not significant ($z = 1.52$, $p = .13$). Therefore, we have no statistical evidence for a publication bias in the current sample.

Primary Analysis: Effect Size and Heterogeneity

We estimated the average weighted SMD and heterogeneity of the sample with a random-effects model with the restricted maximum-likelihood estimator for the amount of heterogeneity. All 10 observed effect sizes and their weights were included to estimate the median effect size. No further moderator variables were specified in the model. Sample heterogeneity was assessed via Cochran's *Q* test for heterogeneity.

The overall weighted mean effect size and the observed effect sizes for the individual experiments are shown in Figure 4. The average observed weighted mean effect size (intercept) under our random-effects model (random effect = study) was 0.54 ($SE = 0.09$, 95% confidence interval [CI] [0.36, 0.70]). The observed effect size was significantly

Figure 3. Funnel plot showing standard error of the effect size Hedges' *g* as a function of effect size. The vertical line indicates the overall model estimate. The triangle-shaped unshaded region represents a pseudo confidence interval region with bounds equal to ± 1.96 *SE*.



different from zero ($z = 5.98$, $p < .001$) and positive, which indicates that people without SLI, on average, outperform people with SLI on statistical learning tasks in the auditory verbal domain. In other words, the value of 0.54 can be regarded as our estimate for the statistical learning disadvantage in people with SLI. Furthermore, the CI ranges from 0.36 to 0.70, indicating that we reliably detected any effect size up to 0.36, which means that we can speak of a moderate-to-large statistical learning disadvantage in people with SLI.

As a measure of heterogeneity, the total amount of variance between the experiments was $\tau^2 = 0.0$ ($SE = 0.036$). Cochran's Q test for heterogeneity was not significant ($Q(9) = 10.11$, $p = .34$). This means that there is no statistical evidence that the true effect sizes differ between the studies in our sample. It is important to note, however, that, whereas a significant Q test provides evidence that the true effects vary, a nonsignificant Q test alone should not be taken as evidence that the true effect sizes are consistent. The low number of experiments in our design could well explain the finding of nonsignificant heterogeneity (Borenstein et al., 2009, p. 113).

Secondary Analysis: Meta-Regressions on Linguistic Level and Age

As mentioned in the introduction, we were interested in seeing whether the linguistic level (word segmentation vs. grammar) and age at which the experiments were performed influence the SMD. We do realize, however, that our sample includes only 10 studies, which renders it unlikely that we will find a significant effect. Nevertheless, we decided to continue our meta-regression, as assessing the impact of the moderator variables linguistic level and age was part of our research question. As our moderator variables are correlated, the impact of both moderators is evaluated by means of two separate meta-regression models.

To assess whether the linguistic level at which the experiments were performed (word segmentation vs. grammatical structure) influences the SMD, we added linguistic level as a between-experiments moderator variable to the random-effects model described above. When we coded experiments at the word segmentation level as −0.5 and experiments at the grammatical level as 0.5, the resulting mixed-effects model detected no significant effect of linguistic level (estimate of the SMD difference = −0.15, $SE = 0.18$, $z = -0.80$, $p = .43$, 95% CI [−0.51, 0.21]).

As can be seen in Figure 4 (and Table 1), the studies in our sample included participants between 6 and 19 years old. To test for age effects, we fit a second meta-regression model with age in years (log-transformed) as the continuous predictor variable. The mixed-effects model detected no significant effect of age (estimate of the SMD difference = −0.10, $SE = 0.11$, $z = -0.91$, $p = .36$, 95% CI [−0.32, 0.12]).

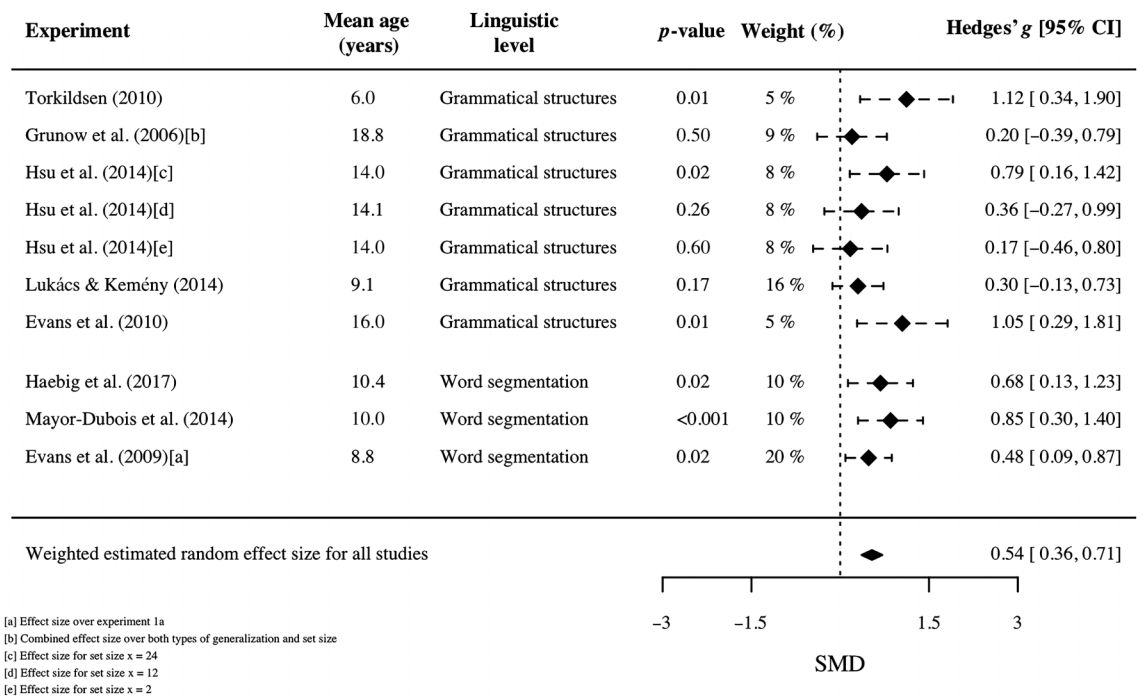
In summary, we found no evidence that linguistic level or age influences the statistical learning disadvantage in people with SLI.[6] The potential effects of these moderators might be too small to detect with meta-regression due to the relative small number of studies in our sample.

Discussion

The primary purpose of our meta-analysis was to provide a quantitative overview of published and unpublished studies on auditory verbal statistical learning in SLI to evaluate the magnitude of the auditory verbal statistical learning disadvantage in people with SLI. We found that, on average, the detection of statistical regularities in the input was not as effective in people with SLI as in people without SLI (statistical learning disadvantage) and that this difference in performance was moderate to large. The results supplement the findings of Obeid et al. (2016) on statistical learning across different modalities in people with SLI. Different from Obeid and colleagues, our focus was on statistical learning in the auditory linguistic domain, which allowed us (a) to add five additional studies on statistical learning in this domain that were not included in the Obeid et al. study and (b) to further explore whether differences in statistical learning ability between people with and without SLI arise as a function of linguistic level. Following on the latter, the second goal of our meta-analysis was to

[6]In addition, we conducted an exploratory meta-regression with the moderator variable adjacency type. This regression revealed no significant effects either. As one of our reviewers pointed out, however, a meta-regression with the moderator variable adjacency type is problematic, as adjacency type is highly correlated with linguistic level (i.e., all word segmentation studies feature an adjacent dependency learning paradigm, whereas the artificial grammar learning studies featured a mix of adjacent and nonadjacent dependency types).

Figure 4. Forest plot showing overall and individual average weighed effect sizes (Hedges' *g*) and 95% confidence interval (CI). A positive effect size indicates that the control group outperformed the group with specific language impairment. SMD = standardized mean difference.



investigate whether the magnitude of statistical learning disadvantage in people with SLI was moderated by the linguistic level (word segmentation vs. grammar) or age at which learning takes place. We did not find evidence that the difference in statistical learning performance between people with and without SLI is moderated by the linguistic level and age at which learning takes place. Although the absence of the effect of linguistic level is a null effect and therefore difficult to interpret, it is in line with previous research reporting absences of associations between verbal working memory and sequence repetition learning (Hsu & Bishop, 2014; Lum et al., 2012). Alternatively, the potential influence of both moderators might have been too small to detect with our meta-regressions due to the relative small number of studies in our sample.

In all, our results extend previous findings on a visual statistical learning disadvantage in SLI (Lum et al., 2012, 2014; Obeid et al., 2016) to the auditory verbal domain and underline the assumption of a general cognitive deficit in the implicit detection of statistical regularities and/or dependencies in people with SLI that contributes to the language problems seen in this population (see also Evans et al., 2009; Hsu & Bishop, 2011, 2014; Ullman & Pierpont, 2005).

Relevance for Clinicians Working With SLI

The current meta-analysis provides strong evidence that people with SLI have more difficulties with statistical learning than people without SLI. These findings support the use of evidence-based interventions that facilitate and stimulate the detection of (statistical) regularities in the input for people with SLI. A concrete example of such a statistical learning–based intervention is the conversation recast treatment for morpheme errors in children with SLI (Plante et al., 2014). Plante and colleagues base their training method on findings from artificial language studies. In such studies, strings have an *aXb* structure in which the *a* and *b* elements always co-occur (Gómez, 2002). It has been found that participants only learn the nonadjacent dependencies when the variability (i.e., different numbers of *X* elements) of the intervening *X* element is high enough (Gómez, 2002). Likewise, Plante et al. (2014) showed that children's use of trained morphemes improved for children who were trained on these morphemes in a high-variability context (24 verbs). They found no evidence of such a treatment effect for children in the low-variability (12 verbs) context. It thus seems that both people with and without language impairment benefit from variability and not only repetition in their language input (Plante et al., 2014). High variability facilitates rule learning rather than rote learning, as participants need to look for regularities and patterns in the input as soon as they notice that memorization is not an option in case of high variability (exceeding working memory capacity). These results suggest that clinicians working with children with SLI need to provide a great number of examples when explaining new rules.

Publication Bias

We would like to stress that, although the regression on funnel plot asymmetry did not reach significance, one should always be cautious for the possibility of publication bias in the literature on auditory statistical learning in SLI. Such a potential publication bias relates to the validity of the classical statistical learning paradigms to measure statistical learning efficiency. Recently, more and more researchers stress the importance of an online measure of statistical learning (e.g., Bogaerts, Franco, Favre, & Rey, 2016; Isbilin, McCauley, Kidd, & Christiansen, 2016; Misyak, Christiansen, & Tomblin, 2010) or a test phase that is more sensitive to individual variation. As mentioned by Siegelman, Bogaerts, and Frost (2016), a large proportion of the participants in a statistical learning study perform at chance level. On the group level, test performance is usually just above chance, and an accuracy score higher than 60% is rarely obtained. For these reasons, we consider it likely that more research groups have unpublished (pilot) data on auditory statistical learning in SLI that did not yield statistical significance. Inclusion of these unpublished data could have made our estimates more precise, and we therefore invite researchers who have such unpublished null results to contribute to our Community-Augmented Meta-Analysis via https://osf.io/4exbz/.

Recommendations for Future Studies

The results of the current meta-analysis show that there is a moderate-to-large statistical learning disadvantage in people with SLI. The moderators of this disadvantage, however, remain unknown. Therefore, we recommend that future studies test the effects of potential moderators such as linguistic level and age within a single study in which the variables are within-subject predictors. Longitudinal designs can be used to test statistical learning performance of the same participants but at different ages. Furthermore, we recommend the use of more sensitive and elaborate (e.g., online) measures of statistical learning at both the individual and group levels. For example, our meta-analysis included only one ERP study (Torkildsen, 2010). Interestingly, the difference between people with and without SLI in this particular study was relatively high (see Figure 4). Potentially, the ERP measure compared with the accuracy measure is more sensitive in picking up differences in performance between people with and without SLI. We recommend future studies to further investigate this potential difference in a within-subject design with results of both measurement types for each individual.

Conclusion

In conclusion, the result of our meta-analysis shows that there is a moderate-to-large statistical learning deficit in people with SLI. This result is congruent with the hypothesis that people with SLI are less effective in statistical learning in the auditory verbal domain than people without language impairment. These results motivate the development of statistical learning–based interventions for children with SLI. More studies are needed, however, to perform more fine-grained analyses on the determinants of statistical learning deficiencies in the auditory linguistic domain in people with SLI.

Acknowledgments

This project is funded by the project "Examining the Contribution of Procedural Memory to Grammar and Literacy" awarded by NWO to Judith Rispens (PI). We would like to thank Julia Evans, Eileen Haebig, Hsinjen Julie Hsu, Janne von Koss Torkildsen, and Agnes Lukács for providing additional data and information that enabled us to perform the meta-analysis. In addition, we would like to thank Sho Tsuji and Christina Bergmann for their feedback and advice and for sharing their thoughts on creating a Community-Augmented Meta-Analysis. Finally, we would like to thank Sacha Couvee for her help during the literature search.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal on Language & Communication Disorders, 41*(6), 675–693.

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

Bogaerts, L., Franco, A., Favre, B., & Rey, A. (2016, June). *Speech onset latencies as an online measure of regularity extraction*. Poster session at the Fifth Implicit Learning Seminar, Lancaster, United Kingdom.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: John Wiley & Sons.

Brown, L., Sherbenou, R., & Johnsen, S. A. (1997). *Test of Nonverbal Intelligence–Third Edition*. San Antonio, TX: The Psychological Corporation.

Csányi, I. (1974). *Peabody Szókincs-Teszt [Peabody Vocabulary Test]*. Budapest, Hungary: Bárczi Gusztáv Gyógypedagógiai Foiskola.

Del Re, A. C. (2013). *Compute.es: Compute effect sizes. R package version 0.2-2*. Retrieved from http://cran.r-project.org/web/packages/compute.es

Dickersin, K. (2005). Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11–34). West Sussex, United Kingdom: Wiley & Sons.

Dunn, M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Dunn, M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. Circle Pines, MN: American Guidance Service.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. https://doi.org/10.1136/bmj.315.7109.629

***Evans, J. L., Hughes, C., Hughes, D., Jackson, K., & Fink, T.** (2010, June). *SLI—A domain specific or domain general implicit*

learning deficit? Modality-constrained statistical learning of auditory and perceptual motor sequences in SLI. Poster session presented at the symposium on research in child language disorders, Madison, WI.

***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)

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-years-old leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50,* 177–195.

Grama, I. C., Kerkhoff, A., & Wijnen, F. (2016). Gleaning structure from sound: The role of prosodic contrast in learning non-adjacent dependencies. *Journal of Psycholinguistic Research, 45*(6), 1427–1449. https://doi.org/10.1007/s10936-016-9412-8

***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,* 158–170. https://doi.org/10.1016/j.jcomdis.2005.11.004

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

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions, version 5.1.0 (updated March 2011)*. Retrieved from http://www.handbook.cochrane.org

Hsu, H. J., & Bishop, D. V. M. (2011). Grammatical difficulties in children with specific language impairment: Is learning deficient? *Human Development, 53*(5), 264–277. https://doi.org/10.1159/000321289

Hsu, H. J., & Bishop, D. V. M. (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,* 1–10. https://doi.org/10.3389/fpsyg.2014.00175

Isbilin, E., McCauley, S., Kidd, E., & Christiansen, M. (2016, June). *Testing implicit-statistical learning implicitly*. Poster session at the Fifth Implicit Learning Seminar, Lancaster, United Kingdom.

Kas, B., & Lukács, Á. (2011). *Magyar Mondatutánmondási Teszt [Hungarian Sentence Repetition Test]*.

Kerkhoff, A., Bree, E. D., & Wijnen, F. (submitted). Dyslexia, language, and statistical learning. In T. Mintz (Ed.), *Current trends in statistical approaches to language acquisition*. Abingdon, United Kingdom: Taylor & Francis.

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.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Leslie, L., & Caldwell, J. (1995). *Qualitative reading inventory* (2nd ed.). New York, NY: Addison-Wesley.

Lukács, Á., Győri, M., & Rózsa, S. (2012). A TROG pszichometriai jellemzoinek magyar vizsgálata, a normák kialakítása [The psychometric analysis of Hungarian data from the TROG]. In D. V. M. Bishop (Ed.), *TROG—Test for reception of grammar handbook* (pp. 47–86). Budapest, Hungary: OS Hungary.

***Lukács, Á., & Kemény, F.** (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology, 28*(3), 472–483. https://doi.org/10.1037/neu0000052

Lukács, Á., & Kemény, F. (2015). Development of different forms of skill learning throughout the lifespan. *Cognitive Science, 39,* 383–404. https://doi.org/10.1111/cogs.12143

Lum, J. A. G., 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.

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48,* 1138–1154.

Mainela-Arnold, E., & Evans, J. L. (2014). Do statistical segmentation abilities predict lexical–phonological and lexical–semantic abilities in children with and without SLI? *Journal of Child Language, 41*(2), 327–351. https://doi.org/10.1017/S0305000912000736

Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2010). Explaining lexical–semantic deficits in specific language impairment: The role of phonological similarity, phonological working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research, 53,* 1742–1756.

Marton, K., Eichorn, N., Campanelli, L., & Zakariás, L. (2016). Working memory and interference control in children with specific language impairment. *Language and Linguistics Compass, 10*(5), 211–224. https://doi.org/10.1111/lnc3.12189

***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), 1–9. https://doi.org/10.1080/09297049.2012.734293

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

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1,* 31. https://doi.org/10.3389/fpsyg.2010.00031

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The 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.pmed1000097

Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders, 36,* 221–231.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

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,* 1245. https://doi.org/10.3389/fpsyg.2016.01245

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23,* 530–545.

Racsmány, M., Lukács, Á., Németh, D., & Pléh, C. (2005). A verbális munkamemória magyar nyelvu˝ vizsgálóeljárásai [Hungarian methods for studying verbal working memory]. *Magyar Pszichológiai Szemle, 60,* 479–505.

Raven, J., Court, J., & Raven, J. (1987). *Raven's Progressive Matrices and Raven's Coloured Matrices*. London, United Kingdom: H. K. Lewis.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roid, M., & Miller, L. (1997). *Leiter International Performance Scale–Revised*. New York, NY: Springer.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R. (2002). Constraints on statistical learning. *Journal of Memory and Language, 47,* 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Sandoval, M., & Gómez, R. L. (2013). The development of non-adjacent dependency learning in natural and artificial languages. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(5), 511–522. https://doi.org/10.1002/wcs.1244

Schwarzer, G. (2015). *Meta: General package for meta-analysis. R package version 4.3-2*. Retrieved from https://CRAN.R-project.org/package=meta

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition*. San Antonio, TX: The Psychological Corporation.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Semel, E. M., Wiig, E. H., & Secord, W. (1987). *CELF-R: Clinical Evaluation of Language Fundamentals–Revised*. San Antonio, TX: The Psychological Corporation.

Siegelman, N., Bogaerts, L., & Frost, R. (2016). 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

Thompson, S. P., & Newport, E. L. (2007). Statistical Learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1–42. https://doi.org/10.1207/s15473341lld0301_1

Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing specific language impairment in adults for the purpose of pedigree analysis. *Journal of Speech and Hearing Research, 35*(4), 832–843. https://doi.org/10.1044/jshr.3504.832

***Torkildsen, J. v. K.** (2010, November). *Event-related potential correlates of artificial grammar learning in preschool children with specific language impairment and controls*. Poster session presented at the Second Annual Neurobiology of Language Conference, San Diego, CA.

Torkildsen, J. v. K., Dailey, N. S., Aguilar, J. M., Gómez, R. L., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research, 56,* 618–629. https://doi.org/10.1044/1092-4388(2012/11-0125)

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analysis: Toward cumulative data assessment. *Perspectives on Psychological Science, 9*(6), 661–665. https://doi.org/10.1177/1745691614552498

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41,* 399–433.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*(3), 1048. Retrieved from http://www.jstatsoft.org/v36/i03/

Wallace, G., & Hammill, D. D. (1997). *The Comprehensive Receptive and Expressive Vocabulary Test: Adult*. Austin, TX: Pro-Ed.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised*. San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (2003). *The Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: The Psychological Corporation.