Prosody and developmental dyslexia: A meta-analysis

Abstract

Purpose: A substantial body of research, including narrative and meta-analytic reviews, has established concurrent and longitudinal relationships between sensitivity to speech prosody and reading ability. This study synthesised research findings pertaining to a possible impairment of prosodic skills in dyslexia.

Method: Three-level meta-analysis was utilised to synthesise 124 effect sizes from 37 studies comprising data from 1771 participants (mean age =13.7 years, range: 6–26).

Results: A moderate-to-large impairment of prosodic competence relative to chronological-age-controls was identified. The effect emerged consistently across languages and different types of prosodic competence task, individual studies were adequately powered to detect effects of this magnitude, and there was no evidence of publication bias. There was no significant impairment relative to reading-age-controls. There was also no impairment for tasks tapping the implicit processing and mental representation of prosody. However, studies in the literature are not sufficiently powered to detect smaller differences between reading groups.

Conclusion: The evidence is consistent with a developmental delay in prosodic competence in dyslexia. One interpretation of this is that prosodic competence is in some way dependent on reading experience. Findings also support theoretical models of dyslexia which emphasise explicit processing of phonology, rather than impaired phonological representations or underlying sensory difficulties. However, fully addressing questions of causality will require additional studies that incorporate reading-age-controls and possess sufficient statistical power to detect small effects.

Prosody and developmental dyslexia: A meta-analysis

Prosody refers to the rhythmic patterning of spoken language. Prosodic patterns arise in speech due to differences in the amplitude, pitch, and duration of stressed and unstressed syllables and the pragmatic use of emphasis and intonation. Over the past decade-and-a-half reading researchers have explored children's sensitivity to speech prosody, how this develops and varies between individuals, and how it may underpin the development of written language skills (Wade-Woolley & Wood, 2006). Cross-sectional and longitudinal studies have demonstrated that prosodic skills, such as the ability to disambiguate utterances using differences in stress placement, contribute to reading development indirectly through vocabulary, phoneme awareness, and morphological awareness (Critten et al., 2021; Deacon et al., 2018), and account for unique variance in literacy ability after controlling for such skills (Enderby et al., 2021; Holliman et al., 2017; 2015). Quasi-experimental studies have also demonstrated limited prosodic skills in children and adults with reading difficulties (Mundy & Carroll, 2016; Wood & Terrell, 1998). It has also been demonstrated that prosodic skills (Harrison et al., 2018) and rhythmic ability more generally (Thomson et al., 2012) show promise as targets for intervention.

A recent narrative and theoretical review by Wade-Woolley et al. (2022) introduced the term *prosodic competence* and argued that this construct should be integrated into the Reading Systems Framework (Perfetti & Stafura, 2014) in recognition of its direct and indirect contributions to typical and atypical reading development. Wade-Woolley et al. also identified key priorities for future research. They argued there is a need for studies that rigorously test causal hypotheses about the role of prosodic competence in reading development and evaluate the contribution that rhythm-based interventions make in ameliorating reading difficulties. The authors also identified a need for research exploring the variety of tasks that have been used to measure prosodic competence. The reliability of different tasks, the extent to which they tap distinct aspects of prosodic competence (e.g., word- vs. phrase-level skills), the impact of additional task demands (e.g., memory load), and the importance (or not) of such task differences for the relationships observed with reading and reading-related skills are open empirical questions.

Wade-Woolley et al. (2022) called for increased utilisation of longitudinal panel designs and randomised controlled intervention studies to address these issues. However, an additional avenue is to synthesise some of the existing research evidence on the topic via meta-analysis. The value of this approach has already been demonstrated in meta-analyses exploring the relationship between prosodic skills and reading ability (Tong et al., 2023), children's application of prosody during fluent reading and their reading comprehension (Wolters et al., 2022), and impaired processing of auditory frequency (one of the acoustic foundations of speech prosody) in developmental dyslexia (Witton et al., 2020). These studies have been able to establish the robustness and reliability of the effects of interest, assess the extent of publication bias, and reveal moderators of these effects. For example, Tong et al. observed stronger relationships between prosodic skills and reading ability in speakers of stress-timed as opposed to syllable-timed languages, and stronger effects for tasks assessing more than one prosodic feature. Wolters et al. reported subtle differences in the contributions of different aspects of reading prosody to reading comprehension, and Witton et al. identified participant characteristics and task demands explaining variation in the magnitude of frequency processing impairments observed across individual studies.

Prosody and dyslexia

There is longstanding consensus that phonological awareness difficulties are a defining characteristic of developmental dyslexia (Snowling, 1998). Given that prosodic aspects of language form part of its phonological structure, it is reasonable to anticipate that prosodic abilities would also be impaired in dyslexia. Indeed, as noted above, there is empirical evidence to support this hypothesis. Researchers have also proposed causal mechanisms through which prosodic skills may influence the development of phonological awareness more broadly and, via this link, exert an impact on reading development. Following observations of poor performance in processing amplitude rise times in dyslexic children and adults, Goswami and colleagues (2002; 2011) argued that sensitivity to such acoustic cues is critical for syllable segmentation and developing awareness of smaller units of sound, and that underlying sensory difficulties in processing rhythmic cues in non-speech stimuli may therefore be central to understanding dyslexia. Such a deficit would also account for apparent difficulties in the processing of speech prosody, seeing them as arising from a more fundamental insensitivity to rise time. The evidence in support of the Temporal Sampling Framework (Goswami, 2011) is compelling in offering a unifying sensory explanation for diverse studies, including those that draw on the perception and production of rhythm in non-speech stimuli as well as studies demonstrating impaired processing of speech prosody.

There are two ways in which a meta-analysis of the literature on speech prosody and dyslexia can contribute to the evaluation of this theory. Firstly, dyslexia is characterised by high levels of heterogeneity and researchers argue that multiple risk factors contribute to the development of impaired reading (Moll et al., 2013). In order to be confident that a particular area of difficulty is implicated in dyslexia, and is not simply a methodological artefact or a consequence of reading problems themselves, a consistent and robust pattern of impaired performance must first be demonstrated across multiple samples and in relation to both chronological-age and reading-age control groups. Meta-analysis offers an ideal method of addressing this issue. Secondly, as Wade-Woolley et al. (2022) have observed, there is wide variation in how prosodic skills have been measured in the literature, both in terms of the type of prosodic skills being targeted and in terms of extraneous variables such as short-term memory load. This in itself raises important methodological questions over the processing demands associated with different forms of prosodic task, and whether such paradigms are in fact measuring prosody or other more general cognitive abilities. Meta-analysis, and in particular moderator analysis, is ideally suited to explore this and other questions. For example, if it can be demonstrated that the degree of prosodic impairment is relatively uniform across different forms of prosodic task, this would argue against the explanation that prosodic difficulties are artefactual of the working memory difficulties that characterise dyslexia (e.g., Beneventi et al., 2010; Menghini et al., 2011). Such an analysis could also offer further insights into the locus of phonological difficulties in dyslexia. A recent debate has centred around the question of whether the phonological difficulties that characterise dyslexia are specific to phonological awareness tasks or whether they also extend to the integrity of the underlying phonological representations, with research pointing to the conclusion that dyslexic individuals struggle to access and process phonological information in the context of meta-linguistic tasks while having intact phonological representations (e.g., Boets et al., 2013; Mengisidou & Marshall, 2019; Ramus et al., 2013; Ramus & Ahissar, 2012). A meta-analysis of the literature on prosody and dyslexia would allow the investigation of the same issues at another level of phonological processing via a moderator analysis contrasting paradigms that tap phonological representations with those that require more explicit processing of prosodic information. While Wade-Woolley et al. (2022) have argued that prosodic *competence* specifically is implicated in reading difficulties, the hypothesis that individuals with dyslexia may additionally experience difficulties in prosodic *representation*, is arguably more consistent with the Temporal Sampling Framework (Goswami, 2011).

Similarly, a moderator analysis exploring prosodic difficulties across languages would also have theoretical implications. Both Wade-Wooley et al. and Goswami imply that prosodic difficulties should be universally experienced rather than being specific to particular languages. However, languages vary in the extent to which they are syllable-timed (e.g., French and Spanish), with approximately equal time intervals elapsing between each spoken syllable, or stress-timed (e.g., English and Dutch), with approximately equal intervals between stressed syllables (Nespor et al., 2011). We may therefore expect to observe language-specific effects of prosody that suggest that it is only relevant for understanding dyslexia in stress-timed languages.

Rationale

The current study adopts a meta-analytic approach to address some of the theoretical and empirical questions outlined above in relation to speech prosody and the nature of dyslexia. This is achieved via a synthesis of quasi-experimental studies comparing the prosodic skills of participants with dyslexia to those of chronological- and/or reading-age-controls. A number of such studies are characterised by high inter-study variability, both in terms of participant characteristics, and aspects of study design, such as the nature of control groups. In addition, a variety of tasks have been used to measure prosodic skills. A quantitative synthesis is therefore merited to establish the quality of the research evidence supporting the claim that prosodic skills are impaired in dyslexia, and to provide insights into potential participant- and task-related moderators of that effect. In particular, exploring the magnitude of pooled effects relative to chronological- vs. reading-age-controls allows the question of causality to be tested with greater statistical power than has hitherto been possible within individual studies (Parrila et al., 2020). Moreover, by exploring differences in effect sizes obtained using different prosodic tasks, we can understand which tasks may be more or less challenging for those with reading difficulties and draw inferences regarding the impact of specific task demands and the extent to which tasks may tap distinct processes or skills within the prosodic domain. Such an analysis may identify more precise targets for rhythm-based interventions and improve future research by informing more efficient and higher quality task selection. Finally, there is an opportunity to explore other questions of interest, e.g., whether impairments are larger for those who are younger, those with more severe reading difficulties, or speakers of certain languages.

Scoping searches of the *Google Scholar* and *PROSPERO* databases indicated no pre-existing meta-analyses on this topic. Recent meta-analyses exploring the association between prosody and typical reading development (Tong et al., 2023; Wolters et al. 2022), theoretical review papers (Wade-Woolley et al., 2022), and longitudinal path analyses of the prosody-literacy relationship (Critten et al., 2021) suggest that a meta-analytic study is timely as researchers seek to synthesise research in this area.

Objectives

The aim of this study was to establish whether there was robust evidence of impaired prosodic skills in participants with dyslexia relative to controls. Findings of individual studies were synthesised and the resulting pooled effect adjusted to account for the undue influence of any individual studies with extreme effect sizes. The extent of publication bias and levels of statistical power in the literature were also assessed. We hypothesised that the prosodic skills of dyslexic participants would be found to be significantly below those of matched controls.

The study also explored several potential moderators of the effect. Moderators were selected a-priori based on Wade-Woolley et al.'s (2022) review, the first author's own knowledge of the research area, and common meta-analytic practice. Firstly, moderator analyses were conducted to observe the impact of different control group matches (chronological- vs. reading-age-controls) on the pooled effect to draw inferences about the potential causal role of prosodic skills in dyslexia. In addition, further moderator analyses were conducted to explore the extent to which sample characteristics (e.g., participant age and native language), aspects of the research design (e.g., the task used to measure prosodic sensitivity), and indicators of study quality (e.g., internal reliability of the prosody task) accounted for significant variance in effect size estimates across studies.

Methods

Study registration and protocol

Records of the search and screening process, data extracted from included studies, and R code used to perform the analyses are publicly available via the Open Science Framework. An anonymised link to the project page has been created for peer review purposes. A study protocol and PRISMA checklist were completed and uploaded in August 2021 prior to commencing database searches. Where deviations from the protocol have occurred, these have been noted and explained below.

Eligibility criteria

Studies were included in the meta-analysis if they reported a comparison between a group of children or adults with dyslexia and a control group, i.e., chronological- age- or reading-age-matched controls, on at least one prosodic task. Studies that took a broad definition of dyslexia and described participants as having reading difficulties, or similar, were also included provided the other eligibility criteria were met. The authors felt this approach was more in-line with current theoretical conceptualisations acknowledging the heterogeneity that characterises dyslexia, it's multifaceted causes, and the fact that reading difficulties can occur across the ability range (Moll et al., 2013).

'Prosodic tasks' were those for which successful completion relied upon the perception, production, matching, discrimination, identification, interpretation, or manipulation of one or more prosodic features. Prosodic features were defined as syllabic stress and intonation, amplitude, pitch, and duration. Studies which measured prosody as a component of reading fluency (i.e., prosodic reading) were excluded. These paradigms involved participants reading connected text aloud with subsequent analyses of pitch contours and pauses and investigations were in the context of reading comprehension processes rather than the sorts of reading skills known to be impaired in dyslexia (e.g., Wolters et al., 2022).

In a departure from the study protocol, some additional criteria were applied while screening search results. This was considered necessary to ensure that the paradigms and outcome measures included in the meta-analysis were comparable, addressing a coherent construct and to avoid overlap with other recently published meta-analyses (e.g., Witton et al., 2020). Specifically, studies exclusively utilising non-linguistic stimuli (e.g., tones) were excluded to restrict the meta-analysis to paradigms involving the processing of speech prosody. Studies utilising passive listening procedures with non-behavioural outcome measures, such as the oddball and mismatch negativity paradigms, were also excluded. Finally, studies of children with other types of reading difficulty (e.g., Poor Comprehenders) and children at-risk of developing dyslexia but without manifest reading difficulties were also excluded.

Search strategy

The following databases were searched: *ERIC, PsycArticles, PsycInfo, ProQuest Dissertations and Theses, PubMed, and Web of Science*. Additional search strategies deployed were backward and forward citation tracking of all studies ultimately selected for inclusion in the meta-analysis, using *Web of Science*, and hand searches of *Annals of Dyslexia, Dyslexia, Journal of Research in Reading, Reading Research Quarterly, Reading and Writing, and Scientific Studies of Reading*. Databases were searched using the following keywords and operators:

(prosody OR "speech rhythm" OR "prosodic processing" OR "prosodic awareness" OR "prosodic sensitivity" OR "stress awareness" OR "stress sensitivity" OR "lexical stress" OR "metrical stress" OR "linguistic stress" OR "syllabic stress" OR "syllable stress" OR "word stress") AND (dyslexi OR "reading impair*" OR "reading disab*" OR "reading diffic*" OR "reading deficit*" OR "reading disorder*")*

No search limits were imposed on publication dates. Search results were limited to peer-reviewed publications and doctoral theses written in English. For hand searches, back issues of journals from the year 2000 onwards were included.

All searches were conducted between August 2021 and April 2022. Prior to conducting the analyses, the searches were updated to check for any new articles.

Figure 1 shows the PRISMA flow diagram recording the screening and ultimate selection of 37 studies for inclusion in the meta-analysis. Screening and selection were completed by the first author. To ensure transparency, the .csv files containing all of the search records are available on the OSF project page.

Figure 1 here.

Data items

Outcomes extracted from studies were measures of accuracy and/or response time for one or more prosody tasks. Studies were also coded for relevant methodological characteristics and indicators of study quality. Methodological variables extracted included the mean age of participants in the dyslexic group (recorded in years as well as coding child vs. adult samples), severity of participants' reading difficulties (i.e., the number of standard deviation units below the population mean on a standardised reading test), and the participants' native language. When coding reading severity, a standardised test of single word reading was used. This information was available for 26 of the 37 studies. A z-score was extracted directly or calculated based on standard scores or percentiles reported by study authors. As noted previously, an inclusive definition and broad range of search terms were adopted in relation to dyslexia to avoid overlooking potentially relevant studies. This resulted in a data set where some samples of participants clearly had very severe reading impairments while others were adults with dyslexia whose reading difficulties were well compensated (Dickie et al., 2013; Leong et al., 2011; Leong & Goswami, 2014).

The name of the prosody task was also extracted, along with the response format (verbal vs. non-verbal response) and the linguistic level (word-level vs. phrase-level) and prosodic feature (syllabic stress vs. intonation) addressed by the task. Approximately two-thirds of study effect sizes originated from word-level tasks (65%). A large majority (88%) addressed syllabic stress as the prosodic feature of interest. Furthermore, effect sizes pertaining to intonation all utilised tone-based tasks conducted with Chinese speakers, so linguistic level was confounded with participant language in the data set. Finally, a large majority of effect sizes used accuracy-based outcomes (80%) and authors often did not state explicitly whether responses were verbal or non-verbal, resulting in a substantial amount of missing data for this item (17%). Therefore, in a second departure from the study protocol and to help discriminate between measures of prosodic ability more effectively, tasks were also coded as either a prosodic competence task (adopting the terminology suggested by Wade-Woolley et al., 2022) or a prosodic representations task (echoing a distinction drawn previously by Ramus et al., 2013 in their study of phonological difficulties in children with dyslexia and/or SLI).

'Prosodic competence' is defined by Wade-Woolley et al. (2022, p. 166) as "the ability to reflect on and process speech prosody", and includes several well-known measures of prosodic ability, e.g., the DEEdee task, requiring participants to attend explicitly to prosodic features in the stimuli. These tasks were further coded according to the level of analysis required:

Disambiguation: tasks in which participants needed to use prosody of a word or phrase to correctly interpret the meaning of an utterance or identify the intended focus of an utterance, such as compound-noun vs. noun phrase discrimination or contrastive stress perception (e.g., green *house* vs. *green*house; rec*ord* vs. *rec*ord; the tall *boy* vs. the *tall* boy).

Discrimination: tasks where participants heard pairs of spoken stimuli and had to make same/different judgments regarding their stress patterns. The stimuli could be words or nonwords. Variants of this task were where participants discriminated between syllables on the basis of lexical tone, and oddity tasks where participants indicated which one of a set of stimuli sounded different to the others.

Identification: participants needed to locate the syllable carrying primary stress within a target word. In the majority of cases the stimuli were printed words or nonwords and participants had to underline the stressed syllable. A small number of studies presented participants with reiterative speech stimuli (e.g., dee*DEE* vs. *DEE*dee) and participants had to indicate which spoken syllable carried primary stress.

Manipulation: this task category included suprasegmental equivalents of classic phoneme awareness paradigms, analysing and manipulating the stress pattern of a target item. The mispronunciations task where children heard a target word pronounced with incorrect stress and were then asked to produce the correct pronunciation was also included in this category.

Production: these tasks measured children's ability to produce prosodic features accurately when writing (accurate placement of stress diacritics) or reading (accurate assignment of lexical stress in words and nonwords). Speech tasks asking children to use contrastive or emphatic stress to indicate focus or disambiguate an utterance are also included in this category.

Recognition: tasks where participants were presented with reiterative speech or low-pass filtered speech stimuli and had to recognise them or match them to one of several response options based on prosodic structure (e.g., deedee*DEE*dee > cinder*ELL*a).

Repetition: tasks which required participants to listen and repeat words, nonwords, short phrases, or reiterative speech stimuli with correct stress placement.

In contrast to these prosodic competence tasks, representations tasks are implicit measures of prosodic processing which tap the encoding and mental representation of prosody directly without participants' awareness being drawn directly to prosodic constituents of the stimuli. This includes paradigms that utilised experimental manipulations of prosody (e.g., presenting items with different stress patterns) to observe their effects on reading or spelling, and tasks which manipulated other variables or demands (e.g., memory load or word frequency) to observe their effects on stress production in reading or spelling. Representations tasks included priming paradigms and other psycholinguistic effects (e.g., spelling-sound regularity, neighbourhood density, and phonological similarity) applied to the prosodic domain.

Indicators of study quality were: overall sample size, sample reliability of the prosody task (Cronbach's alpha), criteria used to confirm dyslexia status (diagnosis from a clinician, researcher defined cut-off on a standardised reading measure, or self-report), the nature of the control group (chronological-age match, reading-age match, or both), and whether or not the dyslexic and control groups were matched for non-verbal ability.

Effect measures

The effect sizes calculated were standardised mean differences. Hedges' *g* was chosen as the measure of effect size to correct for bias arising from the small sample sizes of at least some of the studies (Hedges & Olkin, 1985). For prosodic competence, effect sizes were calculated based on the descriptive statistics reported by study authors for each group of participants. To aid comparison across studies, effect sizes for response time and error rates were sign reversed ensuring that negative effect sizes consistently indicated a deficit for the participants with dyslexia relative to matched controls. When study authors had plotted the statistical values in figures and exact values were not provided in the text or tables, means and standard deviations were extracted using WebPlotDigitizer (Rohatgi, 2022). A total of 18 effects were extracted in this manner.

A two-step process was used for calculating effect sizes for representations tasks. Firstly, mean differences between two experimental conditions of interest (e.g., a stress-congruent and a stress-incongruent prime) were calculated separately within groups and standard deviations of the differences imputed based on the standard deviations reported for the individual conditions (Abrams et al., 2005; Higgins et al., 2022). The mean differences and standard deviations of the differences were then used to calculate a Hedges' *g* effect size between groups. As such, effect sizes for representations tasks indicate the difference in the magnitude of an experimental effect in participants with dyslexia versus matched controls, with negative values signalling a weaker effect in the group with dyslexia.

Synthesis methods

A fundamental assumption of traditional meta-analytic techniques is that effect sizes are independent (Bornstein et al., 2009). When multiple effect sizes are extracted from individual studies one common approach to the problem of the resulting non-independence of the effect sizes is to average within studies and ensure that each study contributes only one effect size to the estimate of the pooled effect. A disadvantage of this approach is that information is lost when averaging within studies. More recently, novel approaches to meta-analysis now allow researchers to account for non-independence of effect sizes and permit individual studies to contribute multiple effect sizes to the synthesis (see Scammacca et al., 2014, for review). We utilized one such approach, three-level meta-analysis, which adopts a hierarchical approach and accounts for non-independence by allowing the nesting of individual effect sizes within studies (Cheung, 2014; 2019).

All analyses were conducted in R version 4.1 (R Core Team, 2021). Random effects meta-analytic models were calculated using the Metafor package (Viechtbauer, 2010) and τ^2 heterogeneity estimates were produced using restricted-maximum-likelihood estimation (REML: Viechtbauer, 2005). Heterogeneity was tested with Cochran's Q and quantified using the I^2 metric, which indicates the proportion of the variance in effect sizes that is systematic and not solely due to sampling error. A significant result in the test of heterogeneity combined with a larger value of I^2 indicates that variations in effect sizes across studies are likely to be attributable to methodological differences between those studies rather than to chance. The impact of such methodological factors can then be explored in a moderator analysis. The I^2 metric has been adapted for three-level meta-analysis and is able to provide estimates of heterogeneity between studies as well as within studies (Cheung, 2014).

The fit of the three-level model was compared to a restricted two-level model with between-study heterogeneity constrained to zero (equivalent to a standard meta-analytic approach which assumes independence of effect sizes) to test whether the added complexity of the three-level model was merited and had resulted in an improved estimate of the pooled effect (Harrer et al., 2021). The *t* estimator was used for calculating model coefficients. This test is analogous to the Knapp-Hartung method used within traditional two-level meta-analyses and has been shown to reduce the likelihood of type-1 errors (Langan et al., 2019).

The robustness of the pooled effect size estimate was assessed with a sensitivity analysis. Effects were identified as potential outliers if the confidence interval for the effect did not overlap the confidence interval for the pooled effect (Viechtbauer & Cheung, 2010). Influence diagnostics were subsequently calculated. The Metafor package allows the calculation of three influence indicators for three-level models; HAT values which indicate how heavily an effect is weighted in the analysis and DFBETA and Cook's *D* values which indicate the extent to which the pooled effect size changes when a given effect is omitted from the model. Diagnostics were interpreted using guidelines suggested by Harrer et al. (2021) and the meta-analysis repeated with influential cases excluded.

The impact of different study characteristics and indicators of study quality were explored by adding them to the three-level meta-analytic model as moderator variables. Potential moderators were tested individually.

Reporting bias assessment

A moderator analysis contrasting effect sizes arising from published studies with those originating from grey literature (i.e., unpublished doctoral theses) was conducted as a coarse test of publication bias. However, as is the case with traditional synthesis methods, Egger's test and other commonly used techniques for assessing publication bias in meta-analyses assume independence of effect sizes and using them in the context of a multi-level meta-analysis is therefore advised against (Rodgers & Pustejovsky, 2021). Therefore, an approach analogous to the PET-PEESE method (Stanley & Doucouliagos, 2014) was used to assess publication bias while accounting for the non-independence of effect sizes in the data set. This approach offers a rigorous test of publication bias while continuing to account for dependency among effect sizes (e.g., Soveri et al., 2017). Firstly, a three-level meta-analytic model including any significant moderator variables was fit to the data and the residuals plotted against the standard errors and variances of the individual effect sizes. The symmetry of the resulting funnel plots allows for an assessment of publication bias after accounting for the impact of the significant moderators. Asymmetric clusters of points focused within certain regions of the funnel plot, and in particular concentrations of studies with larger effect sizes and larger standard errors, indicate that publication bias has likely inflated the estimate of the pooled effect (Harrer et al., 2021). To test for publication bias, the standard errors of the effect sizes were then added to the model as an additional moderator. In the resulting model a significant beta coefficient for the standard error variable indicates publication bias. Furthermore, the intercept of this model provides an estimate of the pooled effect when the standard error is zero. This is known as the limit effect and can be interpreted as an estimate of the pooled effect corrected for the impact of publication bias. Where the pooled effect is statistically significant, study variances provide a more accurate estimate of the limit effect (Stanley & Doucouliagos, 2014). Therefore, in the event that the intercept of the PET model is statistically significant a final model is calculated with the study variances replacing the standard errors.

Study quality and power assessment

Study quality was assessed with regards to overall sample size, internal reliability of the prosody measure, the method of identifying participants with dyslexia, and whether or not control participants were matched for non-verbal ability. Due to the quasi-experimental nature of the research designs, rubric scoring of individual studies was not undertaken. Variables were coded as described under Data Items, above, and then tested individually as moderators.

A power analysis was conducted using the R packages Metaviz (Kossmeier et al., 2020) and Metameta (Quintana, 2023). Assuming a true population effect size equivalent to the limit effect identified by the PET-PEESE analysis, the median power of individual studies was calculated.

Results

Study selection and characteristics

In total 124 effect sizes were extracted from 37 studies. Characteristics of the individual studies are summarised in supplementary materials (Online Resource 1). The studies report data from 1771 participants. The mean age of participants with dyslexia was 13.7 years (range: 6–26 years).

Power et al. (2015) and Goswami et al. (2013) report overlapping data from the same longitudinal study. Therefore, Experiment 2 from Goswami et al. was excluded to avoid double counting. Where studies reported response time and accuracy for the same task, effect sizes were extracted for both of these performance indicators. Where multiple measures of accuracy (e.g., d’ vs. percentage correct) or response time (e.g., overall vs. per item) were reported only one of these was selected for inclusion.

Results of individual studies

The individual effect sizes are summarised in Figure 2.

Figure 2 here.

The most common study design involved comparing participants with dyslexia to a chronological-age matched control group using one or more prosodic competence tasks. The majority of all effect sizes fell below zero, confirming that participants with dyslexia consistently performed worse than matched controls on measures of prosodic ability.

Results of syntheses

A three-level meta-analysis showed a significant pooled effect, g=-.60 [-.78, -.41], p<.0001, in favour of the control participants, i.e., participants with dyslexia were significantly and moderately impaired on measures of prosodic ability. The three-level model was a superior fit when compared to a restricted two-level model with between-study heterogeneity constrained to zero, $\chi^2(1) = 9.59$, $p = .002$. Cochran's test of heterogeneity indicated significant variability in the reported effect sizes, $Q(123)=474.36$, $p < .0001$, and approximately 80% of this variability was not due to sampling error, $I^2=80.27$. There was more variability in the effect sizes between studies, $I^2=43.90$, $\tau^2=.21$, than within studies, $I^2=36.37$, $\tau^2=.18$.

Twenty-two effect sizes met the coarse criterion for outliers. When inspecting influence diagnostics for all 124 effect sizes, all of the HAT values were in the acceptable range indicating that none of the individual effects were excessively weighted when estimating the pooled effect. However, one effect (Mehta et al., 2018: g=-4.76, SE=.50) had a DFBETA value outside of the acceptable range (DFBETA=-.65). The same effect had a

Cook's *D* value of .25 which, while within the acceptable range, was approximately 50 times greater than the mean Cook's *D* value for all effects. Together these values indicate that this effect, which was among the 22 potential outliers, had exerted a large impact on the pooled effect size estimate. The DFBETA value indicates that excluding this effect adjusted the estimate of the pooled effect size by more than half a standard deviation. Consequently, it was decided to exclude this one influential case and retain the other 21 effects with non-overlapping confidence intervals.

A further three-level meta-analysis conducted on the remaining 123 effect sizes still showed significant pooled effect, g=-.55 [-.69, -.41], p <.0001, in favour of the control participants. The three-level model remained a superior fit when compared to a restricted two-level model with between-study heterogeneity constrained to zero, $\chi^2(1)=7.67$, p=.006. Cochran's test of heterogeneity continued to indicate significant variability in the reported effect sizes, $Q(122)=405.96$, p<.0001, and approximately 72% of this variability was not due to sampling error, $I^2=71.95$. Following removal of the influential case, variability within studies, $I^2=46.12$, $\tau^2=.16$, now accounted for more of this systematic heterogeneity than variability between studies, $I^2=25.82$, $\tau^2=.09$.

Moderator analyses

Moderator analyses were conducted with the remaining 123 effect sizes. Two significant moderators were identified (Figure 3). Firstly, there was a significant effect of the type of control to which participants with dyslexia were compared, $F(1, 121)=12.96$, p=.0005, with effect sizes being significantly larger when the control group was matched for chronological age, g=-.61, p<.0001, as opposed to reading age, g=-.16, p=.22. In addition, there was a significant effect of task type, $F(1, 121) = 42.78$, p<.0001, with effect sizes being significantly larger for prosodic competence tasks, g=-.70, p<.0001, than for representations tasks. In fact, for the representations tasks, there was a small effect *in favour* of the participants with dyslexia, *g*=.28, *p*=.05.

Figure 3 here.

A fine-grained analysis of the different task types revealed similar pooled effect sizes for all seven types of prosodic competence task (Table 1). These all differed significantly from the pooled effect for the representations tasks.

Table 1 here.

There was also a significant moderator effect obtained for language, $F(8, 114)=3.52$, $p=.001$, but further exploration suggested that this was likely confounded with task type and an artefact of the small numbers of effects recorded for languages other than English. As most of the effects were from studies of English speakers, English was chosen as the referent category for this analysis. The pooled effect for studies of English speakers differed significantly from zero, $g=-.45$, $p<.0001$. As shown in Table 2, four languages had a significantly larger pooled effect size than English (i.e., there were significantly larger deficits for participants with dyslexia in these languages); Chinese, Greek, Spanish, and Japanese. However, the pooled effects for these four languages were based exclusively on prosodic competence tasks while approximately 14% of the effect sizes obtained from English speakers were for representations tasks on which participants with dyslexia were not impaired. In addition, there was only one effect size relating to Japanese speakers and this was larger than the average effect size.

Table 2 here.

Similarly, there was a smaller deficit evident for participants with dyslexia in studies of Italian speakers and, to a slightly lesser extent, French speakers. This can again be explained by the relatively small number of effects recorded for speakers of these languages and the fact that approximately 43% and 50% of those effects pertained to representations tasks for Italian and French speakers respectively. Perhaps the main insight provided by this analysis, therefore, is that the deficit in prosodic *competence* is actually observed very reliably across all of the languages studied to date (Figure 4). Indeed, when the moderation analysis was repeated with effect sizes for representations tasks excluded, there were no significant differences in the magnitude of the pooled effect across languages, $F(8, 99)=1.46$, $p=.18$.

Figure 4 here.

None of the other theoretically or methodologically motivated moderators tested were significant, with *p* values ranging from .32 -.70.

Reporting biases

A three-level meta-analytic model with task type (competence vs. representations) and control group (chronological– vs. reading-age) both included as moderators was fit to the data and the residuals plotted against the standard errors and variances of the individual effect sizes (Figure 5). The resulting funnel plots revealed a high degree of symmetry and thus no evidence of publication bias.

Figure 5 here.

The standard errors of the effect sizes were then added to the model as a third moderator. The standard errors were not a significant predictor of effect size magnitude and therefore this model provided no evidence of publication bias, beta = -.38, p = .72. The intercept was statistically significant and therefore a final model was calculated with the study variances replacing the standard errors. The study variances were not significantly related to the magnitude of the effect sizes, beta = -.10, p = .95, and thus there continued to be no evidence of publication bias.

Prosodic competence and the chronological-age control were used as referent categories in these models and as such the intercept provides an estimate of the pooled effect size for prosodic competence tasks against a chronological-age control at infinite precision. The intercept of this model was statistically significant (p<.0001) and estimated a pooled effect size of -.78 [-1.11, -.46] for prosodic competence tasks against a chronological-age control.

As an additional, coarser test of publication bias, a separate moderator analysis was conducted to contrast effect sizes from published studies (k=111) with those originating from grey literature (i.e., unpublished doctoral theses). This analysis revealed no evidence of publication bias, with published and unpublished effect sizes appearing to be of a very similar magnitude, p = .950.

Study quality and power

One point that is evident from the standard errors, variances, and precision estimates for the individual effect sizes is that most of the studies in the literature had very similar sample sizes. The clustering of the large majority of data points in a single band across the centre of each of the funnel plots indicates this most clearly. While there were relatively few studies with very small samples there was also an absence of larger studies. The mean sample size for the 123 effects included in the meta-analysis was approximately 46 with 21 being the average size of the group with dyslexia.

Quantifying the statistical power of the individual effects within the data set was complicated by the significant moderator effects. Essentially, there was more than one true population effect size depending on the nature of the control and the type of task being utilized. Therefore, the power analysis focused in particular on the effects that used prosodic competence tasks and a chronological-age control. This approach was taken because the fundamental claim made in this literature is that participants with dyslexia are impaired relative to chronological-age-controls on prosodic competence tasks, and the majority of the effects in the meta-analysis tested this hypothesis. Additionally, the limit effect provided by the PET-PEESE analysis gave a bias-corrected estimate of the true population effect size for these specific effects. Figure 6 plots the median power for the 85 effects that used prosodic competence tasks and a chronological-age control over a range of hypothetical population effect sizes. When assuming a true population effect size of -.78 (the limit effect identified by the PEESE test) the analysis indicated a respectable median power of 73%.

Figure 6 here.

The other variables considered in relation to study quality were the internal reliability (i.e., Cronbach's alpha) of the prosody tasks, the method used to identify participants with dyslexia, and the matching of dyslexic and control groups for non-verbal IQ. Approximately 77% of the effects came from studies where groups had been matched for non-verbal skills and only one of the studies (the extreme outlier excluded following the influence analysis) relied solely on self-report methods for identifying participants with dyslexia. However, only

31% of the extracted effect sizes had a Cronbach's alpha value reported. The median Cronbach's alpha value was .74 indicating good reliability for many of the prosody tasks. However, there was a wide range of alpha values (.36-.98) and some measures showed poor reliability. In particular, tasks using reiterative speech (e.g., the DEEdee task) accounted for eight of the 10 smallest alpha values (range .37-.52).

Discussion

The aim of this study was to synthesise existing research exploring the extent and nature of a hypothesised prosodic deficit in developmental dyslexia, establish the robustness of any group differences in prosodic skills, and investigate potential moderators of this effect. An initial meta-analysis of 124 effect sizes from 37 studies revealed a moderate and statistically significant impairment of prosodic skills in participants with dyslexia. Subsequent analyses found the effect to be robust, remaining after assessing for influential studies and publication bias.

Dyslexic participants performed approximately three quarters of a standard deviation below the level of chronological-age-controls on measures of prosodic competence. Individual studies were found generally to be adequately powered to detect effects of this magnitude. The magnitude of this effect is somewhat smaller than that observed in the domains of phoneme awareness, RAN, and morphological awareness (Araújo & Faísca, 2019; Georgiou et al. 2023; Melby-Lervåg et al., 2012). Nevertheless, these findings indicate a moderate-to-large deficit in an aspect of spoken language processing that remains neglected in theoretical models of reading development and dyslexia as well as in remediation programmes. Furthermore, moderator analyses suggest that the impairment in prosodic competence is obtained consistently across all nine languages studied to date, including those with alphabetic and non-alphabetic writing systems. While more studies of non-English speakers are required, those synthesised in this analysis indicate that prosodic competence is associated with reading development across languages and supports the proposal that prosodic competence should be incorporated into theoretical accounts of reading development (Wade-Woolley et al., 2022).

Moderator analyses also contrasted the pooled effects obtained relative to chronological- and reading-age-controls to assess the quality of evidence for a causal link between prosodic sensitivity and dyslexia. In contrast to the moderate-to-large effect found in relation to a chronological-age-control, performance of individuals with dyslexia did not differ significantly from that of reading-age-controls, a finding which would suggest that the prosodic abilities captured by these tasks are not causally implicated in dyslexia. However, while the pooled effect in comparison to reading-age-controls was small, it was larger than that obtained in a recent meta-analysis of morphological ability (g=-.08: Georgiou et al., 2023). Participants with dyslexia performed approximately one sixth of a standard deviation below the level of reading-age matched controls. Furthermore, the t estimator used in the calculation of model coefficients is conservative (Langan et al., 2019), our study included a relatively small number of effect sizes comparing individuals with dyslexia to reading-age-controls (k = 24), and the power analysis further suggests that sample sizes typically used in the literature may not be sufficient for individual studies to detect the smaller effects that might be expected against a reading-age-control group.

In contrast to the findings for prosodic competence, there was no impairment at all observed for the prosodic representations tasks. A small number of primary studies have contrasted performance in competence and representations type tasks (Mundy & Carroll, 2012) and the overall finding of the meta-analysis supports the argument that the implicit, online processing of prosody is unaffected in dyslexia and underlying lexical representations encode prosody accurately in these individuals. More broadly, this finding is also consistent with a theoretical conceptualisation of dyslexia that places the phonological deficit at the level of access rather than representation (e.g., Boets et al., 2013; Mengisidou & Marshall, 2019; Ramus et al., 2013; Ramus & Ahissar, 2012).

The pooled effect for representations tasks indicated a small advantage for the participants with dyslexia, i.e., experimental manipulations tapping the implicit processing and underlying representation of prosody without conscious awareness elicited larger effects (e.g., more priming) among participants with dyslexia than among controls. This is likely to be an artefact of the longer response times and higher error rates of participants with dyslexia and the greater variability commonly observed within special populations inflating mean differences between experimental conditions. An alternative approach to the representations analysis would therefore be to first calculate *g* within-groups to correct for any main effects in the study data and use these effect sizes along with estimates of their variances in the synthesis in place of the raw mean differences and their standard deviations. This alternative approach was not taken because it was preferable to adopt an approach that was more directly comparable to the process used in calculating the effect sizes for prosodic competence tasks. Individual studies that have calculated standardised effects for dyslexic and control groups have reported null effects with nearly identical magnitudes for participants with dyslexia and controls (Mundy & Carroll, 2012; 2013).

So, what does this meta-analysis tell us about the nature of developmental dyslexia? We argue that individuals with dyslexia experience difficulties in prosodic competence that are consistent with a developmental lag in the ability to reflect on and process prosody, rather than a specific prosodic deficit being causally implicated in dyslexia (although, we do so cautiously, given the limited statistical power of the studies that included reading-age-controls). This supports the argument made elsewhere that difficulties in prosodic competence are reliably associated with reading difficulties (Wade-Woolley et al., 2022).

Additionally, moderator analyses found no evidence of a deficit in forming robust prosodic representations. This pattern of results potentially challenges the claim that the prosodic difficulties experienced by individuals with dyslexia are a consequence of a more fundamental sensory difficulty with processing amplitude rise times and other rhythmic cues, as a core contention of the Temporal Sampling Framework is that impaired rise time perception disrupts the formation of appropriate phonological representations (Goswami, 2011, pg. 5). Instead, this pattern of results is consistent with the wider literature on phonological abilities in dyslexia, which has similarly demonstrated that dyslexia is characterised by difficulties not with representation but with the explicit processing of segmental phonology (Snowling, 2003). Given that we also observe this pattern with respect to suprasegmental phonology, there is scope to argue that dyslexia may be characterised by analogous difficulties at various levels of phonological processing. The fact that representations are intact implies that the encoding of phonology is adequate, but for dyslexic readers access to and retrieval of that information is compromised (e.g., Boets et al., 2013; Mengisidou & Marshall, 2019; Ramus et al., 2013; Ramus & Ahissar, 2012). Future research could seek to understand the precise nature of these processing difficulties.

Another idea tested by our analyses was that the apparent deficit in prosodic competence is simply an expression of the limits on working memory that are observed as a hallmark of dyslexia (e.g., Beneventi et al., 2010; Menghini et al., 2011). That is, dyslexic participants find the cognitive demands of prosodic competence assessments challenging and it is the limits on conscious cognitive processing that are tapped by such measures, rather than a problem with prosody per se. If this were the case, we might expect to see some substantial variation in performance across different types of prosodic competence task, in line with the differing levels of cognitive demand. However, the moderator analysis reported in Table 1 confirms the severity of the difficulties experienced by participants with dyslexia is remarkably similar across prosodic competence tasks, which would rule out this interpretation.

Arguably the most parsimonious interpretation of the current findings is that prosodic competence is somewhat dependent on reading experience. That is, individuals with dyslexia may be at a disadvantage because there is a level of reading experience, perhaps involving mastery of more complex words and texts, that facilitates conscious access to prosody, and the reading difficulties experienced by dyslexic individuals mean that they achieve this experience later than their same-age peers. This interpretation can be tested experimentally through an appropriately powered comparison of dyslexic readers vs. chronological-age-controls vs. reading-age-controls, and by intervention studies where the degree of reading experience is accelerated for one group of dyslexic readers and compared to another group engaging in normal reading practice. Such studies, ideally conducted across languages and using a variety of prosodic competence tasks, would represent a valuable addition to the literature.

Finally, as well as the theoretical implications discussed above, the current study highlights some practical and methodological issues for research in the field. Firstly, the finding that sizeable impairments were observed across languages for all prosodic competence tasks has implications for intervention. Specifically, providing support for prosodic skills in the context of text-based activities may be beneficial. This could include distinguishing between different stress patterns and intonation contours, identifying the location of syllabic stress in words, the metalinguistic manipulation of prosody, and the correct stress production. The relatively uniform impairments observed across prosodic competence task types also suggests that researchers can reasonably select a small number of tasks to serve as a proxy for prosodic competence more broadly. Finally, the analysis of

Cronbach's alpha values suggests that researchers should avoid the use of reiterative speech tasks such as the DEEdee task, which tend to have poor reliability.

Conclusions

There is robust evidence of an impairment of prosodic competence in dyslexia relative to chronological-age-controls while the implicit processing and mental representation of prosodic structures appears to be intact. These findings support theories of dyslexia that place the locus of impairment at the level of phonological access and awareness and not those that argue for impaired phonological representations arising from underlying sensory processing difficulties. In relation to reading-age-controls, there is a need for additional studies with the statistical power to reliably detect small effects and help resolve questions of causality. Finally, the literature remains Anglocentric and more studies are required in languages other than English. We also need to address measurement issues and move away from tasks which have poor reliability. Improving the reporting of reliability and other psychometric properties of prosodic tasks and pursuing larger multi-language studies with reading-age and chronological-age-controls are key priorities for future research.

References

Abrams, K.R., Gillies, C.L., & Lambert, P.C. (2005). Meta-analysis of heterogeneously reported trials assessing change from baseline. *Statistics in Medicine, 24(24),* 3823-3844. DOI: 10.1002/sim.2423.

*Anastasiou, D., & Protopapas, A. (2015). Difficulties in lexical stress versus difficulties in segmental phonology among adolescents with dyslexia. *Scientific Studies of Reading, 19(1),* 31-50. DOI: 10.1080/10888438.2014.934452.

*Anderson, A., Lin, C.Y., & Wang, M. (2013). Native and novel language prosodic sensitivity in English-speaking children with and without dyslexia. *Dyslexia, 19(2),* 92-112. DOI: 10.1002/dys.1451.

Araújo, S., & Faísca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading, 23(5),* 1–20. https://doi.org/10.1080/10888438.2019.1572758.

*Barry, J.G., Harbodt, S., Cantiani, C., Sabisch, B., & Zobay, O. (2012). Sensitivity to lexical stress in dyslexia: A case of cognitive not perceptual stress. *Dyslexia, 18(3),* 139-165. DOI: 10.1002/dys.1440.

Beneventi, H., Tønnessen, F.E., Ersland, L. & Hugdahl, K. (2010) Working Memory Deficit in Dyslexia: Behavioral and fMRI Evidence, *International Journal of Neuroscience, 120 (1),* 51-59. DOI: 10.3109/00207450903275129

Boets, B., Op de Beeck, H.P., Vandermosten, M., Scott, S.K., Gillebert, C. R., Mantini, D., Bulthe, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science, 342(6163),* 1251-1254. DOI: 10.1126/science.1244333.

*Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2020). Processing of rhythm in speech and music in adult dyslexia. *Brain Sciences, 10(5),* 261. DOI: 10.3390/brainsci10050261.

Bornstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-Analysis.* Chichester: Wiley.

*Caccia, M., & Lorusso, M.L. (2019). When prosody meets syntax: the processing of the syntax-prosody interface in children with developmental dyslexia and developmental language disorder. *Lingua*, *224*, 16-33. DOI: 10.1016/j.lingua.2019.03.008.

*Caccia, M., & Lorusso, M.L. (2021). The processing of rhythmic structures in music and prosody by children with developmental dyslexia and developmental language disorder. *Developmental Science*, *24(1)*, e12981. DOI: 10.1111/desc.12981.

*Caccia, M., Presti, G., Toraldo, A., Radaelli, A., Ludovico, L. A., Ogliari, A., & Lorusso, M. L. 2019). Pitch as the main determiner of Italian lexical stress perception across the lifespan: Evidence from typical development and dyslexia. *Frontiers in Psychology*, *10*, 1458. DOI: 10.3389/fpsyg.2019.01458.

*Calet, N., Gutiérrez-Palma, N., Defior, S., & Jiménez-Fernández, G. (2019). Linguistic and non-linguistic prosodic skills in Spanish children with developmental dyslexia. *Research in Developmental Disabilities*, *90*, 92-100. DOI: 10.1016/j.ridd.2019.04.13.

Cheung, M.W.L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modelling approach. *Psychological Methods*, *19(2)*, 211-229. DOI: 10.1037/a00329968.

Cheung, M.W.L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, *29*, 387-396. DOI: 10.1007/s11065-019-09415-6.

Critten, S., Holliman, A.J., Hughes, D. J., Wood, C., Cunnane, H., Pillinger, C., & Deacon, H. (2021). A longitudinal investigation of prosodic sensitivity and emergent literacy. *Reading and Writing*, *34*, 371-389. DOI: 10.1007/s11145-020-10077-7.

*Cuetos, F., Martínez-García, C., & Suárez-Coalla, P. (2018). Prosodic perception problems in Spanish dyslexia. *Scientific Studies of Reading, 22(1),* 41-54. DOI: 10.1080/10888438.2017.1359273.

*Cunningham, T.R. (2011). *Understanding synthetic speech and language processing of students with and without a reading disability* (Doctoral Dissertation). University of Toronto.

Deacon, S.H., Holliman, A.J., Dobson, G.J., & Harrison, E.C.J. (2018). Assessing direct contributions of morphological awareness and prosodic sensitivity to children's word reading and reading comprehension. *Scientific Studies of Reading, 22(6),* 527-534. DOI: 10.1080/10888438.2018.1483376.

*Dickie, C., Ota, M., & Clark, A. (2013). Revisiting the phonological deficit in dyslexia: Are implicit nonorthographic representations impaired? *Applied Psycholinguistics, 34(4),* 649-672. DOI: 10.1017/S0142716411000907.

Enderby, J.L., Carroll, J.M., Tarczynski-Bowles, M.L., & Breadmore, H.L. (2021). The roles of morphology, phonology, and prosody in reading and spelling multisyllabic words. *Applied Psycholinguistics, 42(4),* 865-885. DOI: 10.1017/S0142716421000096.

*Fratantonio, A., Rappo, G., & Pepi, A. (2013). Syllabic composition and use frequency: how do they affect stress assignment? A comparison between slow readers and fluent readers. *Life Span and Disability, 16(2),* 169-186.

*Geiser, E., Kjelgaard, M., Christodoulou, J.A., Cyr, A., & Gabrieli, J.D. (2014). Auditory temporal structure processing in dyslexia: processing of prosodic phrase boundaries is not impaired in children with dyslexia. *Annals of Dyslexia, 64,* 77-90. DOI: 10.1007/s11881-013-0087-7.

Georgiou, G.K., Vieira, A.P.A., Rothou, K.M., Kirby, J.R., Antoniuk, A., Martinez, D., &

Guo, K. (2023). A Meta-analysis of Morphological Awareness Deficits in Developmental Dyslexia. *Scientific Studies of Reading*, *27(3),* 253-271. DOI: 10.1080/10888438.2022.2155524.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15(1)*, 3-10. DOI: 69527724/j.tics.2010.10.0012021

*Goswami, U., Barnes, L., Mead, N., Power, A.J., & Leong, V. (2016). Prosodic similarity effects in short-term memory in developmental dyslexia. *Dyslexia*, *22(4),* 287-304. DOI: 10.1002/dys.1535.

*Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing, 23*, 995-1019. DOI: 10.1007/s11145-009-9186-6.

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szűcs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: the Ba–Wa distinction in developmental dyslexia. *Developmental Science, 14(1),* 34-43. DOI: 10.1111/j.1467-7687.2010.00955.x

*Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, *69(1),* 1-17. DOI: 10.1016/j.jml.2013.03.001.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, *99*(16), 10911-10916. DOI: 10.1073/pnas122368599.

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall / CRC Press.

Harrison, E., Wood, C., Holliman, A.J., & Vousden, J.I. (2018). The immediate and longer-term effectiveness of a speech-rhythm-based reading intervention for beginning readers. *Journal of Research in Reading, 41(1),* 220-241. DOI: 10.1111/1467-9817.12126.

Hedges, L.V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* Cambridge, Massachusetts: Academic Press.

Higgins. J.P.T., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., Welch V.A. (Eds.). (2022). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.3). Retrieved from www.training.cochrane.org/handbook.

Holliman, A.J., Mundy, I.R., Wade-Woolley, L., Wood, C., & Bird, C. (2017). Prosodic awareness and children's multisyllabic word reading. *Educational Psychology, 37(10),* 1222-1241. DOI: 10.1080/01443410.2017.1330948.

*Holliman, A.J., Wood, C., & Sheehy, K. (2012). A cross-sectional study of prosodic sensitivity and reading difficulties. *Journal of Research in Reading*, *35(1),* 32-48. DOI: 10.1111/j.1467-9817.2010.01459.x.

Holliman, A.J., Wood, C., & Sheehy, K. (2015). Does speech rhythm sensitivity predict children's reading ability one year later? *Journal of Educational Psychology*, *102(2),* 356-366. DOI: 10.1037/a0018049.

*Jiménez-Fernández, G., Gutiérrez-Palma, N., & Defior, S. (2015). Impaired stress awareness in Spanish children with developmental dyslexia. *Research in Developmental Disabilities*, *37*, 152-161. DOI: 10.1016/j.ridd.2014.11.002.

*Kelić, M. (2017). *Phonological representations and phonological skills in children with dyslexia in Croatian* (Doctoral Dissertation). University of Zagreb.

*Kitzen, K.R. (2001). *Prosodic sensitivity, morphological ability, and reading ability in young adults with and without childhood histories of reading difficulty* (Doctoral Dissertation). Columbia University.

Kossmeier, M., Tran, U.S., & Voracek, M. (2020). *Visualising meta-analytic data with R package metaviz* (Version 3.1) [Computer Software]. Retrieved from https://cran.r-project.org/web/packages/metaviz.

Langan, D., Higgins, J.P., Jackson, D., Bowden, J., Veroniki, A.A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods, 10(1),* 83-98. DOI: 10.1002/jrsm.1316.

*Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience, 8,* 96. DOI: 103389/fnhum.2014.00096.

*Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language, 64(1),* 59-73. DOI: 10.1016/j.jml.2010.09.003.

*Li, W.S., & Ho, C.S.H. (2011). Lexical tone awareness among Chinese children with developmental dyslexia. *Journal of Child Language*, *38(4),* 793-808. DOI: 10.1017/S0305000910000346.

*Majoos, K.J. (2012). *Factors affecting how children hear words and their relation to reading ability* (Doctoral Dissertation). University of New England.

*Marshall, C.R., Harcourt-Brown, S., Ramus, F., & Van Der Lely, H.K.J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language and Communication Disorders, 44(4),* 466-488. DOI: 10.1080/13682820802591643.

Menghini, D., Finzi, A. , Carlesimo, G.A. & Vicari, S. (2011) Working Memory Impairment in Children with Developmental Dyslexia: Is it Just a Phonological Deficit?, *Developmental Neuropsychology, 36 (2),* 199-213. DOI: 10.1080/87565641.2010.549868.

*Mehta, S., Ding, Y., Ness, M., & Chen, E.C. (2018). Invented spelling, word stress, and syllable awareness in relation to reading difficulties in children. *Journal of Psycholinguistic Research*, *47,* 585-606. DOI: 10.1007/s10936-017-9547-2.

Melby-Lervåg, M., Lyster, S.-A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138(2),* 322–352. https://doi.org/10.1037/a0026744.

Mengisidou, M., & Marshall, C.R. (2019). Deficient explicit access to phonological representations explains phonological fluency difficulties in Greek children with dyslexia and/or developmental language disorder. *Frontiers in Psychology*, *10,* 638. DOI: 10.3389/fpsyg.2019.00638.

Moll, L., Loff, A., & Snowling, M. J. (2013). Cognitive endophenotypes of dyslexia. *Scientific Studies of Reading, 17(6),* 385-397. DOI: 10.1080/10888438.

*Mundy, I.R., & Carroll, J.M. (2012). Speech prosody and developmental dyslexia: Reduced phonological awareness in the context of intact phonological representations. *Journal of Cognitive Psychology, 24(5),* 560-581, DOI: 10.1080/20445911.2012.662341.

*Mundy, I.R., & Carroll, J.M. (2013). Spelling–stress regularity effects are intact in developmental dyslexia. *Quarterly Journal of Experimental Psychology, 66(4),* 816-828. DOI: 10.1080/17470218.2012.719530.

Mundy, I.R., & Carroll, J.M. (2016). Which prosodic skills are related to reading ability in adulthood? In J. M. Thomson & L. Jarmulowicz (Eds.), *Trends in Language Acquisition Research: Linguistic Rhythm and Literacy* (Chapter 3, pp. 51-76). Philadelphia: John Benjamins Publishing Company.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hulme, and K. Rice (Eds.), *The Blackwell Companion to Phonology* (Chapter 48, pp. 1147-1159). Malden: Wiley-Blackwell.

Parrila, R., Dudley, D., Song, S., & Georgiou, G.K. (2020). A meta-analysis of reading-level match dyslexia studies in consistent alphabetic orthographies. *Annals of Dyslexia, 70(1),* 1-26. DOI: 10.1007/s11881-019-00187-5.

*Paizi, D., Zoccolotti, P., & Burani, C. (2011). Lexical stress assignment in Italian developmental dyslexia. *Reading and Writing*, *24*, 443-461. DOI: 10.1007/s11145-010-9236-0.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18(1),* 22–37. DOI:10.1080/10888438.2013.827687.

*Power, A.J., Colling, L.J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1-10. DOI: 10.1016/j.bandl.2016.06.006.

Quintana, D.S. (2023). A guide for calculating study-level statistical power for meta-analyses. *Advances in Methods and Practices in Psychological Science, 6(1).* DOI:10.1177/25152459221147260

R Core Team (2021). *R: A language and environment for statistical computing* (Version 4.1) [Computer Software]. Retrieved from https://cran.r-project.org.

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, *29(1-2),* 104-122. DOI: 10.1080/02643294.2012.677420.

*Ramus, F., Marshall, C.R., Rosen, S., & van der Lely, H.K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain, 136(2),* 630-645. DOI: 10.1093/brain/aws356.

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods, 26,* 141-160. DOI: 10.1037/met0000300.

Rohatgi, A. (2022). *WebPlotDigitizer* (Version 4.6) [Computer Software]. Retrieved from https://automeris.io/WebPlotDigitizer.

*Sakono, S., Ito, T., Fukuda, S.E., & Fukuda, S. (2011). The effect of word accent production on reading performance in Japanese young children. *Asia Pacific Journal of Speech, Language and Hearing, 14(1),* 51-59. DOI: 10.1179/136132811805334948.

Scammacca, N., Roberts, G., & Stuebing, K.K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84(3),* 328-364. DOI: 10.3102/0034654313500826.

*Soroli, E., Szenkovits, G., & Ramus, F. (2010). Exploring dyslexics' phonological deficit III: foreign speech perception and production. *Dyslexia, 16(4),* 318-340. DOI: 10.1002/dys.415.

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review, 24,* 1077–1096. DOI: 10.3758/s13423-016-1217-0.

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

Stanley, T.D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5(1),* 60-78. DOI: doi.org/10.1002/jrsm.1095.

Thomson, J.M., Leong, V., & Goswami, U. (2012). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing, 26,* 139-161. DOI: 10.1007/s111145-012-9359-6.

Tong, S.X., Lentejas, K., Deng, Q., An, N., & Cui, Y. (2023). How prosodic sensitivity contributes to reading comprehension: A meta-analysis. *Educational Psychology Review*, *35*(3), 1-28. DOI: 10.1007/s10648-023-09792-8.

*Tong, X., Tong, X., & King Yiu, F. (2018). Beyond auditory sensory processing deficits: lexical tone perception deficits in Chinese children with developmental dyslexia. *Journal of Learning Disabilities, 51(3),* 293-301. DOI: 10.1177/0022219417712018.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30(3),* 261-293. DOI: 10.3102/10769986030003261.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36(3),* 1–48. DOI: 10.18637/jss.v036.i03.

Viechtbauer, W., & Cheung, M.W-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1(2),* 112-125. DOI: 10.1002/jrsm.11.

Wade-Woolley, L., & Wood, C. (2006). Editorial: Prosodic sensitivity and reading development. *Journal of Research in Reading, 29(3),* 253-257. DOI: 10.1111/j.14679817.2006.00306.x.

Wade-Woolley, L., Wood, C., Chan, J., & Weidman, S. (2022). Prosodic competence as the missing component of reading processes across languages: Theory, evidence and future research. *Scientific Studies of Reading, 26(2),* 165-181. DOI:

10.1080/10888438.2021.1995390.

*Wang, L.C., Liu, D., Chung, K.K.H., & Yang, H.M. (2017). Development of lexical tone awareness in Chinese children with and without dyslexia. *Contemporary Educational Psychology, 49*, 203-214. DOI: /10.1016/j.cedpsych.2017.02.002.

*Wang, H.L.S., Wang, N.Y.H., Chen, I.C., & Tsao, Y. (2019). Auditory identification of frequency-modulated sweeps and reading difficulties in Chinese. *Research in Developmental Disabilities, 86*, 53-61. DOI: 10.1016/j.ridd.2019.01.006.

Witton, C., Swoboda, K., Shapiro, L.R., & Talcott, J.B. (2020). Auditory frequency discrimination in developmental dyslexia: A meta-analysis. *Dyslexia, 26(1),* 36-51. DOI: 10.1002/dys.1645.

Wolters, A.P., Kim, Y.S.G., & Szura, J.W. (2022). Is reading prosody related to reading comprehension? A meta-analysis. *Scientific Studies of Reading 26(1)*, 1-20. DOI: 10.1080/10888438.2020.1850733.

*Wood, C., & Terrell, C. (1998). Poor readers' ability to detect speech rhythm and perceive rapid speech. *British Journal of Developmental Psychology, 16(3),* 397-413. DOI: 10.1111/j.2044-835X.1998.tb00760.x.

Table 1.

Moderator analysis demonstrating the effect of task type (referent = Representations)

Task	***k***	***g***	***beta***	***SE***	***p***
Representations	15	0.28	--	--	--
Disambiguation	18	-0.83	-1.11	0.19	< .0001
Discrimination	23	-0.65	-0.93	0.19	< .0001
Identification	12	-0.63	-0.91	0.22	< .0001
Manipulation	6	-0.82	-1.10	0.27	< .0001
Production	17	-0.70	-0.98	0.20	< .0001
Recognition	23	-0.63	-0.91	0.18	< .0001
Repetition	9	-0.95	-1.23	0.22	< .0001

Table 2.

Moderator analysis demonstrating the effect of language (referent = English)

Task	***k***	***g***	***beta***	***SE***	***p***
English	71	-0.45	--	--	--
Chinese	12	-0.83	-0.38	0.19	0.05
Croatian	2	-1.02	-0.57	0.39	0.14
French	4	-0.16	0.29	0.33	0.38
German	5	-0.76	-0.31	0.28	0.27
Greek	8	-1.09	-0.64	0.25	0.01
Italian	7	0.07	0.52	0.23	0.02
Japanese	1	-1.61	-1.16	0.57	0.04
Spanish	13	-0.83	-0.38	0.18	0.04

Figure 1.

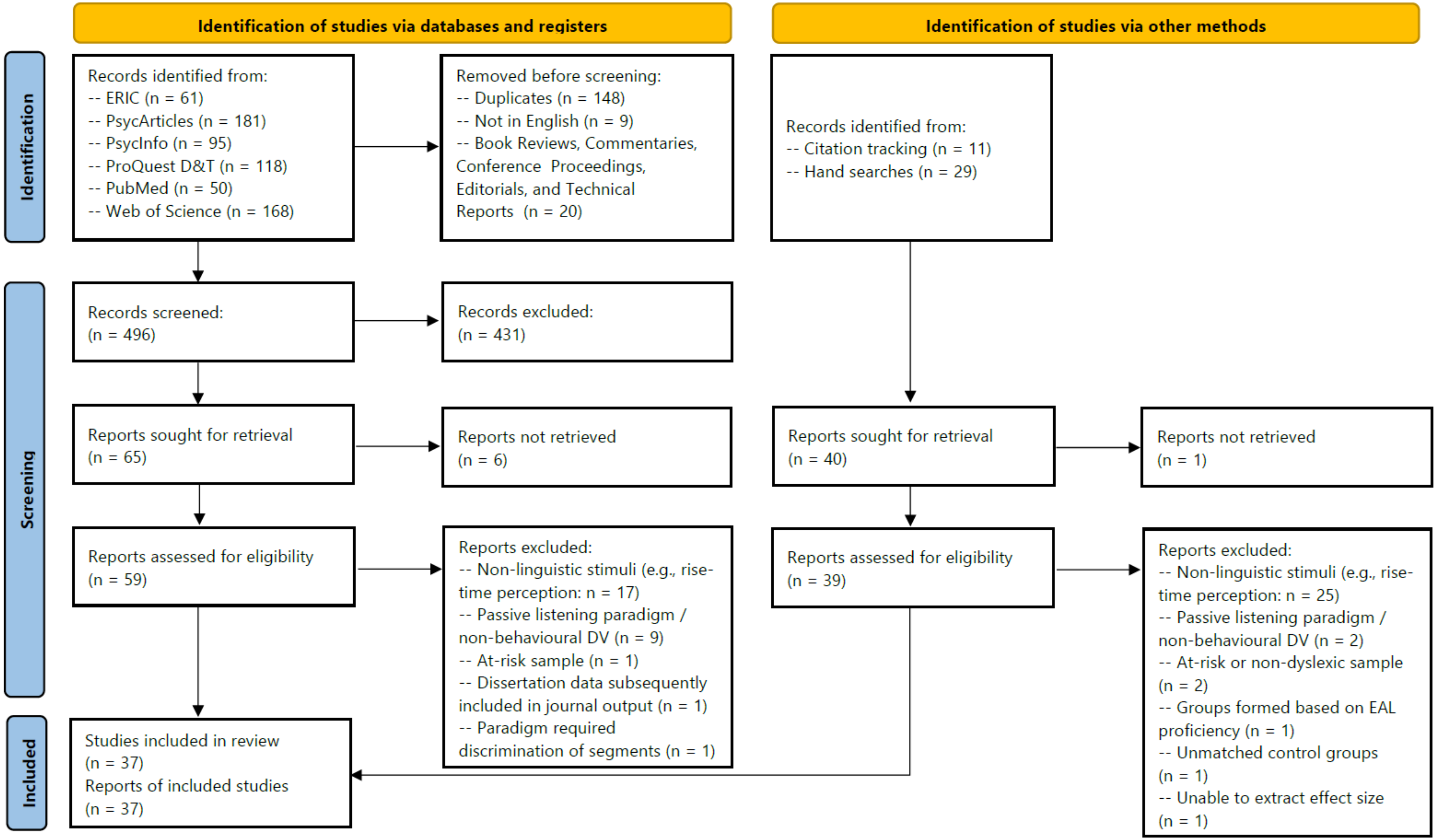
Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
-- ERIC (n = 61)
-- PsycArticles (n = 181)
-- PsycInfo (n = 95)
-- ProQuest D&T (n = 118)
-- PubMed (n = 50)
-- Web of Science (n = 168)
Removed before screening:
-- Duplicates (n = 148)
-- Not in English (n = 9)
-- Book Reviews, Commentaries, Conference Proceedings, Editorials, and Technical Reports (n = 20)
Records identified from:
-- Citation tracking (n = 11)
-- Hand searches (n = 29)
Records screened:
(n = 496)
Records excluded:
(n = 431)
Reports sought for retrieval
(n = 65)
Reports not retrieved
(n = 6)
Reports sought for retrieval
(n = 40)
Reports not retrieved
(n = 1)
Reports assessed for eligibility
(n = 59)
Reports excluded:
-- Non-linguistic stimuli (e.g., rise-time perception: n = 17)
-- Passive listening paradigm / non-behavioural DV (n = 9)
-- At-risk sample (n = 1)
-- Dissertation data subsequently included in journal output (n = 1)
-- Paradigm required discrimination of segments (n = 1)
Reports assessed for eligibility
(n = 39)
Reports excluded:
-- Non-linguistic stimuli (e.g., rise-time perception: n = 25)
-- Passive listening paradigm / non-behavioural DV (n = 2)
-- At-risk or non-dyslexic sample (n = 2)
-- Groups formed based on EAL proficiency (n = 1)
-- Unmatched control groups (n = 1)
-- Unable to extract effect size (n = 1)
Studies included in review
(n = 37)
Reports of included studies
(n = 37)

Figure 2.

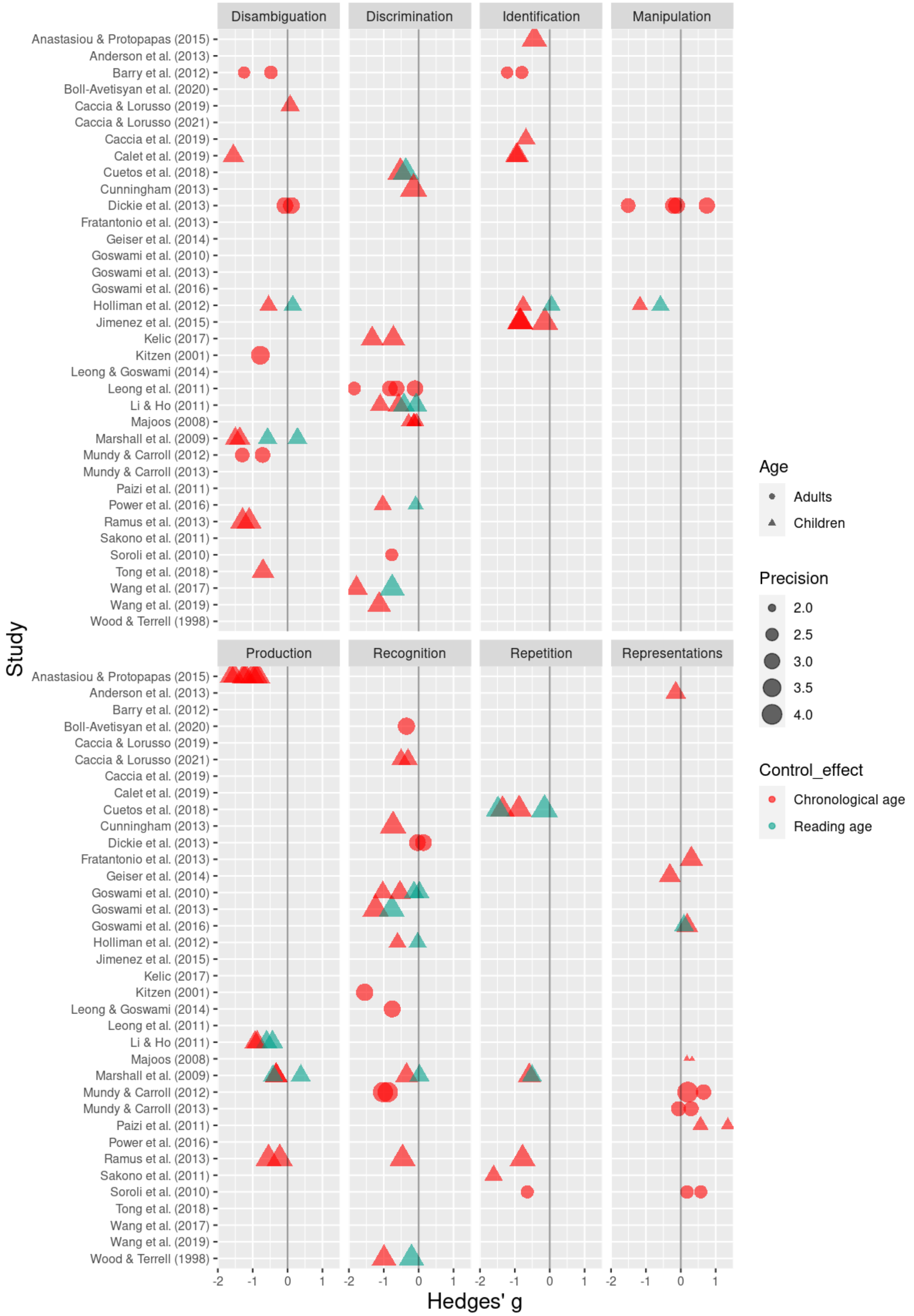
Disambiguation
Discrimination
Identification
Manipulation
Production
Recognition
Repetition
Representations
Study
Anastasiou & Protopapas (2015)
Anderson et al. (2013)
Barry et al. (2012)
Boll-Avetisyan et al. (2020)
Caccia & Lorusso (2019)
Caccia & Lorusso (2021)
Caccia et al. (2019)
Calet et al. (2019)
Cuetos et al. (2018)
Cunningham (2013)
Dickie et al. (2013)
Fratantonio et al. (2013)
Geiser et al. (2014)
Goswami et al. (2010)
Goswami et al. (2013)
Goswami et al. (2016)
Holliman et al. (2012)
Jimenez et al. (2015)
Kelic (2017)
Kitzen (2001)
Leong & Goswami (2014)
Leong et al. (2011)
Li & Ho (2011)
Majoos (2008)
Marshall et al. (2009)
Mundy & Carroll (2012)
Mundy & Carroll (2013)
Paizi et al. (2011)
Power et al. (2016)
Ramus et al. (2013)
Sakono et al. (2011)
Soroli et al. (2010)
Tong et al. (2018)
Wang et al. (2017)
Wang et al. (2019)
Wood & Terrell (1998)
Hedges' g
-2
-1
0
1
Age
Adults
Children
Precision
2.0
2.5
3.0
3.5
4.0
Control_effect
Chronological age
Reading age

Figure 3.

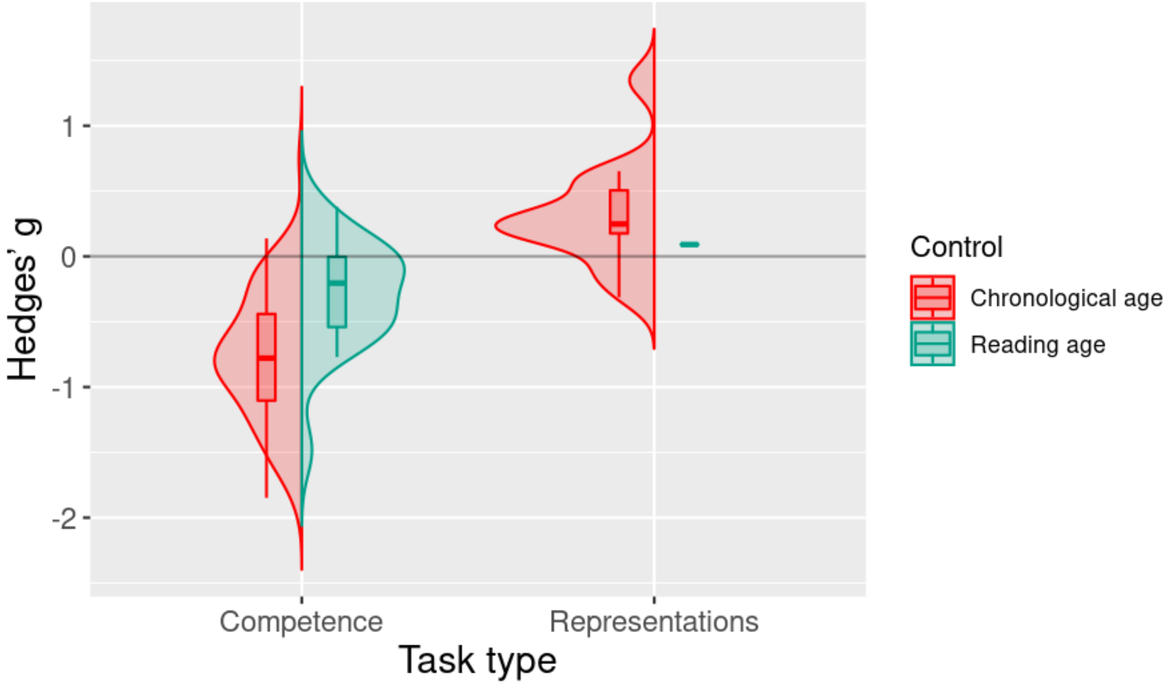
Hedges' g
1
0
-1
-2
Competence
Representations
Task type
Control
Chronological age
Reading age

Figure 4.

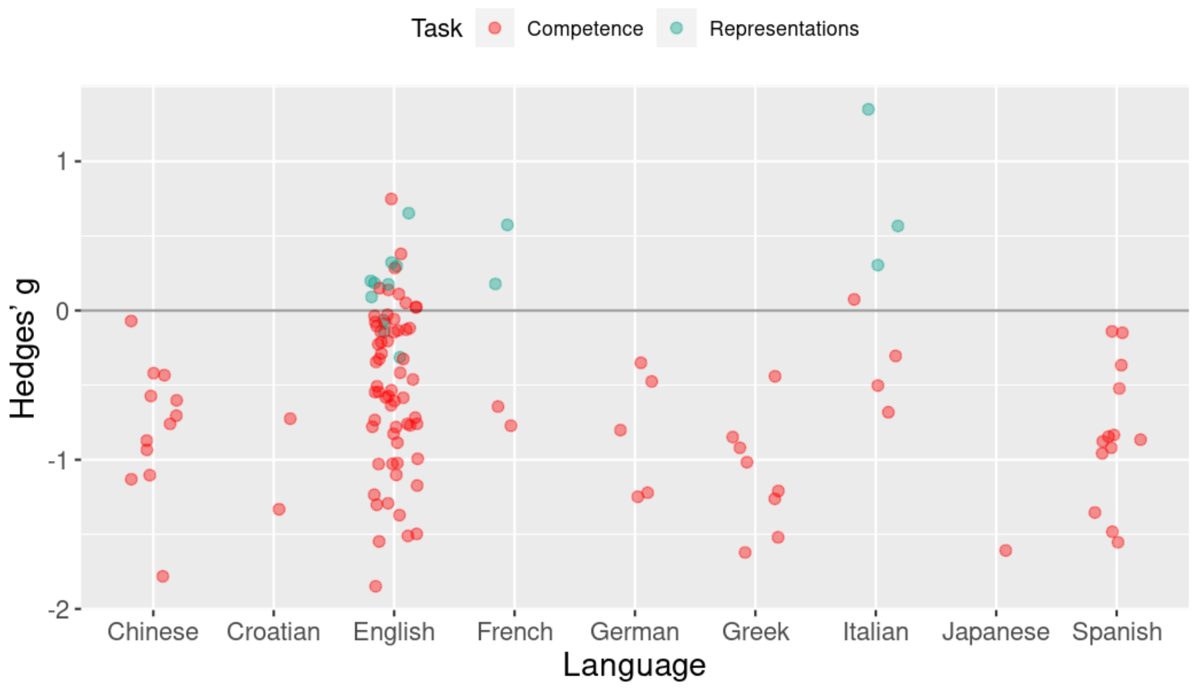
Task
Competence
Representations
Hedges' g
1
0
-1
-2
Chinese
Croatian
English
French
German
Greek
Italian
Japanese
Spanish
Language

Figure 5.

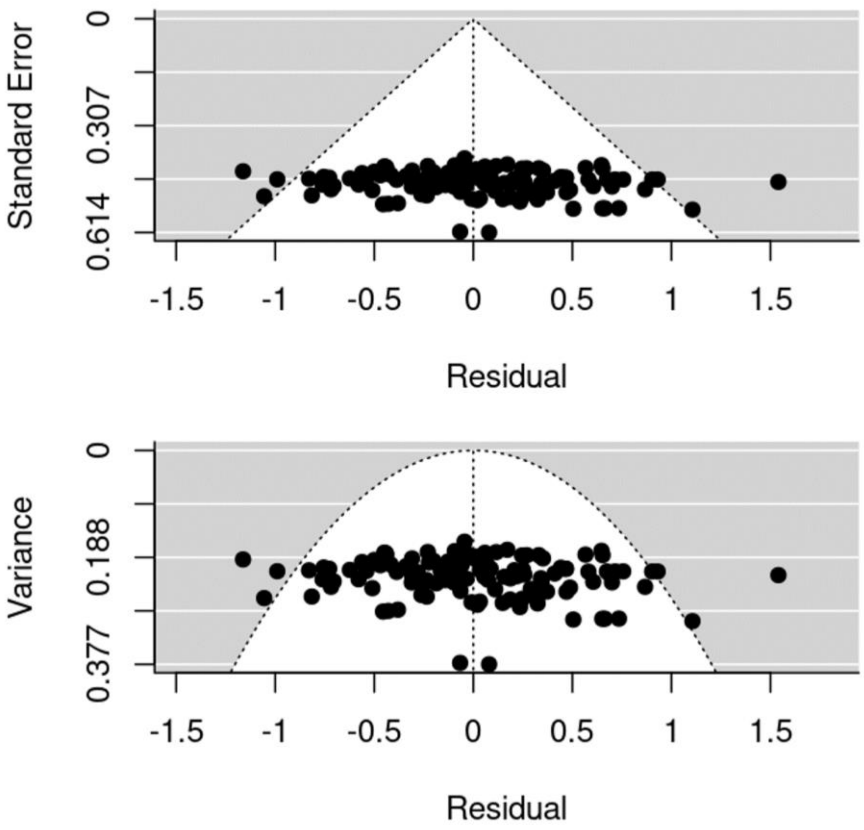
Standard Error
0
0.307
0.614
-1.5
-1
-0.5
0
0.5
1
1.5
Residual
Variance
0
0.188
0.377
-1.5
-1
-0.5
0
0.5
1
1.5
Residual

Figure 6.

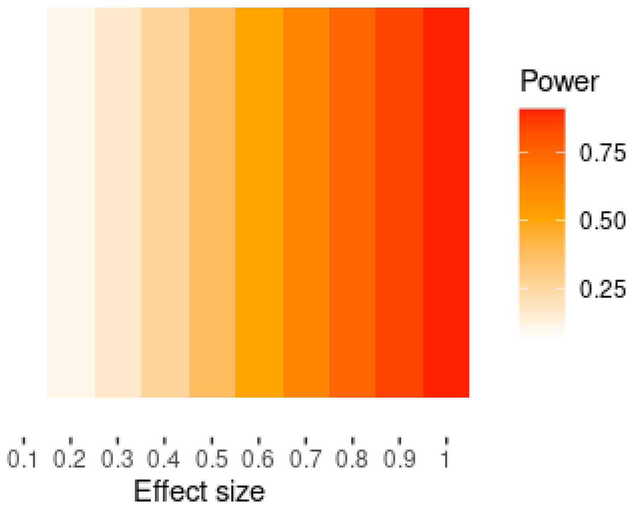

Figure Captions

Figure 1.

PRISMA flow diagram

Figure 2.

Hedges' g and precision estimates for individual study effect sizes (k = 123)

Note: One outlier not plotted (Mehta et al., 2018: identification task, g = -4.76, Precision = 1.99)

Figure 3.

Effect size distributions by task type and control group (k = 123)

Figure 4.

Effect sizes for representations and competence tasks across languages (k = 123)

Figure 5.

Funnel plot of residuals against effect size standard errors (top) and variances (k = 123)

Figure 6.

Firepower plot showing median power of studies to detect prosodic competence impairments against a chronological-age control for a range of population effect sizes (k = 85)