NORTHWESTERN UNIVERSITY

Longitudinal and Concurrent Correlates of Reading Ability:
Two Meta-Analytic Approaches

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders

By
Sean McWeeny

EVANSTON, ILLINOIS

June 2022

Abstract

Learning to read is a complex process that requires integration across sensory, cognitive, and linguistic processes. Accordingly, there are numerous process that may lead to early reading difficulties. The earlier these difficulties are found, the more effective interventions can be, and the deleterious effects of falling behind in reading cannot be understated. This dissertation tested correlates of early reading ability with two meta-analytic approaches to address the considerable heterogeneity present in reading research.

Study 1 investigated rapid automatized naming (RAN), a predictor of future reading across different ages, ability levels, and languages, which may be useful in literacy screening for reading disability (RD). To investigate the longitudinal relationship between preschool/kindergarten RAN and future reading, we conducted a systematic review and meta-analysis (N = 60 samples; k = 373 effect sizes; n = 10,513 participants), in which we also tested whether characteristics of the RAN tasks, reading measures, or sample demographics moderate this relationship. Our results show that kindergarten/preschool RAN is correlated with grade-school reading at $r = -.38$, similar in magnitude to previous concurrent meta-analyses. We also found that RAN has independent predictive ability above and beyond phonological awareness (PA), which has clear theoretical and practical impacts. This meta-analysis was the first to measure RAN’s unique effect on reading, as well as the first to test practical and theoretical moderators longitudinally.

Study 2 tested another early correlate of reading ability, auditory processing. Several hypotheses exist regarding the link between RD and auditory processing impairments, but none fully account for the range of impairments reported. These impairments have been primarily summarized by qualitative reviews, but these reviews fall short in numerous key domains. To understand the full range and size of deficits in individuals with RD, we conducted a systematic review and meta-analysis (N = 63, k = 135; n = 3,545) on four auditory task domains: frequency, duration, and intensity discrimination, as well as gap detection. Our results show large impairments (*g* = .6 to *g* = .8) in each domain for individuals with RD, undermining causal hypotheses of RD from highly specific deficits. These results motivate future testing of auditory processing abilities as a correlate of reading ability, as our meta-analysis was the first to quantify deficits in duration and intensity discrimination, as well as gap detection. These studies have clear implications relating to universal screening in reading research and meta-science more broadly.

Acknowledgements

I want to begin by thanking all the faculty who helped make this dissertation possible. First, I would like to thank my committee chair and advisor, Dr. Elizabeth Norton, for accepting me into her lab on little notice when she had already welcomed 2 PhD students into her lab that year. Her trust allowed me to pursue anything and everything that would help advance my science. The lab culture was open and inviting and allowed each of us to grow in the ways we needed, and I couldn’t ask for much more than that.

My committee, Dr. Megan Roberts and Dr. Sumit Dhar, provided invaluable feedback in completing this dissertation. Their guidance, both technical and theoretical, shaped this dissertation immensely from design to completion. Their support allowed me to approach unique projects that typically fall outside the purview of a dissertation but nonetheless help advance the field tremendously. Dr. Beth Tipton provided unmatched statistical guidance with the exact answers I needed with unparalleled speed. Dr. Bev Wright and Dr. Greico-Calub challenged my ideas and made me think deeply about the subject material in ways that were essential for my growth as a scientist. Christina Maimone provided supreme technical support at numerous junctures in completing this dissertation.

I would also like to thank my meta-analysis coding teams in the LEARN Lab: Jinnie Choi, June Choe, Sandy LaTourette, Alice Wang, Anu Raife, Meakaylin Phillips, Cadence Reed-Bippen, Rick Qian, Gabriella Leibowitz, Tessneem Shahbandar, and Smitha Ramesh. Without them, this science would not have been possible. Jinnie Choi, Julia Nikolaeva, Silvia Lam, Brittany Manning, and Alex Clain have always been there when I needed to talk science, life, and everything in between. Their support was the foundation of my health and wellbeing the last 6 years.

I would like to thank my family for their support throughout the years. I often took for granted the supportive environment they created by allowing me to have the space to explore my ideas and interests. They undoubtedly instilled a belief that I was capable of achieving whatever I threw myself into. Last and surely not least, I would like to thank my life partner, Lauren Attewell, for being who she is, which is the most considerate, caring, loving person I have ever known. We have shaped each other into the people we want to be, and I could ask for nothing more than that.

Table of Contents

List of Tables and Figures

1. Introduction

1.1 The Central Problem in Reading Research

Reading is a complex process that requires the automatic integration of sensory, cognitive, and linguistic abilities. Developing the skills required to read accurately and efficiently takes years and reading development can be disrupted by deficits in the fundamental skills that underlie reading, many of which develop prior to reading age. These pre-reading skills, such as rapid automatized naming (RAN), phonological awareness (PA), and letter knowledge, explain a significant amount of variance of concurrent and longitudinal reading abilities (Clayton et al., 2020; Schatschneider et al., 2004). A child's profile of strengths and weaknesses in these pre-reading skills, and potentially many others, can roughly predict whether they may be at-risk for developing a reading disability (RD).

When a child has an unexpected difficulty in learning to read, they may be diagnosed with RD. RD is the most common type of developmental disorder, affecting 4-9% of the population (American Psychiatric Association, 2013), and is a key public health issue with a potential for early intervention. Extensive evidence exists that RD exists before it can presently be diagnosed (Beelen et al., 2019; Raschle et al., 2012; Saygin et al., 2013), suggesting that RD is a prime target for earlier diagnosis and treatment. However, despite tremendous progress in understanding features of RD, our measures are insufficiently refined to be able to diagnose RD before a child has already fallen behind (Norton & Wolf, 2012). This central problem is the primary motivation for the present dissertation, which will approach several behavioral correlates of reading and reading disability, namely RAN, PA, and auditory processing, in an applied meta-analytic framework. Understanding how these behavioral correlates concurrently and longitudinally relate to reading is key not only to addressing the central problem with RD diagnosis, but also to shaping unifying neurobiological theories of RD, which currently do not address the full range of impairments experienced by individuals with RD.

A number of attempts have been made to summarize the relationship between reading and its behavioral correlates, with some correlates being better documented and understood than others. For example, a number of meta-analyses have summarized RAN's relationship with reading (Araújo et al., 2015; Chen et al., 2021; Hjetland et al., 2017) and RD (Araújo & Faísca, 2019; Reis et al., 2020), whereas only one meta-analysis has been published for one specific auditory processing impairment, frequency discrimination, despite the large range of impairments indicated by a substantial literature (Witton et al., 2020). Therefore, I approach the respective literatures with the appropriate level of specificity. Accordingly, the present meta-analysis of RAN and reading addresses the relationship between RAN measured before reading age (i.e., in kindergarten and preschool) and longitudinal reading outcomes. This meta-analysis has policy-level impact on universal screening due to the specificity of the questions asked. On the other hand, the meta-analysis of auditory processing deficits includes a broad range of ages and uses concurrent data, as there are comparatively few longitudinal studies that measure auditory processing before reading age and follow participants for several years. The impact of this study will not immediately affect policy but has strong implications for researchers and clinicians alike who are trying to understand the relationship between RD and auditory processing.

1.2 Why Meta-Analysis?

The modern scientific method is founded on the principle that scientific knowledge is acquired through the cyclical testing and refining of hypotheses. Key to this cycle is the idea of reproducibility, or the assumption that one will obtain the same results from a study repeated in the same way. As science is a global enterprise, thousands of attempted replications are created each year; whether their authors describe their studies as a replication is not pertinent. The body of literature that has been created from this global enterprise contains a large number of studies whose effects do not replicate, up to more than half of all published research (Ioannidis, 2005). The many reasons this situation, referred to as the Replication Crisis, has arisen include low study power, reliance on *p*-values for interpretation, a lack of correction for multiple comparisons, and within- and between-study bias (Benjamin et al., 2018; Loken & Gelman, 2017). Though its causes can be addressed directly in new studies by pre-registering study designs and analytical plans that maximize the probability that the individual study will make an accurate and replicable conclusion (Ansari & Gervain, 2018), it is also necessary and economical to summarize the extant literature in a meaningful way. This systematic and quantitative summary is meta-analysis, which is a methodology designed to make conclusions about a body of research.

A meta-analysis is comprised of individual studies that measure the same effect, such as the relative success of one treatment over another, the strength of relationships among multiple variables, or summarize mean impairments in one population versus another. In meta-analysis individual studies' effects are not compared against each other, but rather analyzed together under the assumption that each individual study's effects come from a larger underlying distribution that the included studies share. In meta-analysis, replication is not a dichotomous outcome (i.e., one study does or does not replicate another), but rather a concept that is quantified in terms of effect sizes and their respective variances.

Though the usual reason meta-analyses are run is to calculate an aggregate effect size, their value extends to analyses of publication bias and to the examination of within-study factors may lead to larger or smaller effect size. Publication bias analyses can describe whether the aggregate effect size is systematically publication biased, and whether the aggregate effect size should be appropriately corrected. Meta-regression can test whether design-related factors, such as experimental design and stimuli characteristics, or participant-related factors, such as the participants' age and native language, systematically change effect sizes. Both between- and within- study analyses are key in understanding why two or more studies may come to different conclusions about the same effect.

1.3 The Present Dissertation

The present meta-analyses explore many of the aforementioned purposes of meta-analyses in an attempt to answer both theoretical and practical questions relevant to the field of reading research. These goals are listed extensively in their respective chapters, but a summary view of how each chapter approaches the goals of meta-analyses is presented here.

Chapter 2 is a systematic review and meta-analysis of the relationship between kindergarten and preschool RAN and future reading outcomes. In this meta-analysis, I quantify several summary effect sizes that describe the longitudinal relationship between RAN measured in kindergarten or preschool and reading measured in grade school. The primary analysis of our study quantifies summary effect sizes for different measures of RAN (e.g., colors, letters, etc.) and reading (e.g., comprehension, fluency, etc.). A secondary set of analyses describes the unique effect of early RAN on future reading, controlling for PA, creating a summary effect size that reveals RAN's independence from PA. Meta-regression on both sets of summary effect sizes tests specific hypotheses about RAN's role in reading development, and their implications for practice and theory. Finally, within- and between-study bias analyses describe the RAN-reading literature and indicators of its overall health.

Chapter 3 is a systematic review and meta-analysis of auditory processing impairments in children with RD. Its protocol was approved as a Registered Report, which is a recently developed format that requires registration of *a priori* methodological and analytic plans in an effort to combat the Replication Crisis. In this meta-analysis, I quantify four summary effect sizes relevant to auditory processing and RD. These summary effect sizes describe the mean impairment for individuals with RD in four categories: frequency discrimination, duration discrimination, intensity discrimination, and gap detection. These summary effect sizes are particularly important to this literature, as even a well-cited recent review (Hämäläinen et al., 2013) resorts to "vote counting" in which *p*-values are compared directly in order to determine whether the aforementioned categories are impaired in RD. Meta-regression analyses test whether psychophysical task design systematically affect a given study's effect size. Exploratory meta-regression analyses also test whether the auditory processing-RD relationship changes over the course of development. Finally, as in the RAN-reading meta-analysis, within- and between-study bias analyses describe the auditory processing and RD literature and indicators of its overall health.

The final chapter of this dissertation, Chapter 4, compares and contrasts the respective meta-analyses and the literatures they summarize. This qualitative discussion provides insights pertaining to outstanding questions, refining hypotheses, and future directions. Most importantly, this chapter explores the need for expanding the framework that describes reading disabilities so that they can be understood, diagnosed, and treated as early as possible.

1.4 References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ansari, D., & Gervain, J. (2018). Registered Reports: Introducing a new article format in Developmental Science. *Developmental Science*, *21*(1), e12650. https://doi.org/10.1111/desc.12650

Araújo, S., & Faísca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading*, *23*(5), 349–368. https://doi.org/10.1080/10888438.2019.1572758

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Beelen, C., Vanderauwera, J., Wouters, J., Vandermosten, M., & Ghesquière, P. (2019). Atypical gray matter in children with dyslexia before the onset of reading instruction. *Cortex*, *121*, 399–413. https://doi.org/10.1016/j.cortex.2019.09.010

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., Boeck, P. D., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., … Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Chen, Y.-J. I., Thompson, C. G., Xu, Z., Irey, R. C., & Georgiou, G. K. (2021). Rapid automatized naming and spelling performance in alphabetic languages: A meta-analysis. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10160-7

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading*, *24*(2), 91–107. psyh. https://doi.org/10.1080/10888438.2019.1622546

Hämäläinen, J. A., Guttorm, T. K., Richardson, U., Alku, P., Lyytinen, H., & Leppänen, P. H. T. (2013). Auditory event-related potentials measured in kindergarten predict later reading problems at school age. *Developmental Neuropsychology*, *38*(8), 550–566. https://doi.org/10.1080/87565641.2012.718817

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–155. https://doi.org/10.4073/csr.2017.14

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584–585. https://doi.org/10.1126/science.aal3618

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Raschle, N. M., Zuk, J., & Gaab, N. (2012). Functional characteristics of developmental dyslexia in left-hemispheric posterior brain regions predate reading onset. *Proceedings of the National Academy of Sciences*, *109*(6), 2156–2161. https://doi.org/10.1073/pnas.1107721109

Reis, A., Araújo, S., Morais, I. S., & Faísca, L. (2020). Reading and reading-related skills in adults with dyslexia from different orthographic systems: A review and meta-analysis. *Annals of Dyslexia*, *70*(3), 339–368. https://doi.org/10.1007/s11881-020-00205-x

Saygin, Z. M., Norton, E. S., Osher, D. E., Beach, S. D., Cyr, A. B., Ozernov-Palchik, O., Yendiki, A., Fischl, B., Gaab, N., & Gabrieli, J. D. E. (2013). Tracking the roots of reading ability: White matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *Journal of Neuroscience*, *33*(33), 13251–13258. https://doi.org/10.1523/JNEUROSCI.4383-12.2013

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology*, *96*(2), 265–282. psyh. https://doi.org/10.1037/0022-0663.96.2.265

Witton, C., Swoboda, K., Shapiro, L. R., & Talcott, J. B. (2020). Auditory frequency discrimination in developmental dyslexia: A meta-analysis. *Dyslexia: An International Journal of Research and Practice*, *26*(1), 36–51. https://doi.org/10.1002/dys.1645

2. Rapid automatized naming (RAN) as a kindergarten predictor of future reading in English: A systematic review and meta-analysis

2.1 Introduction

Reading is a complex process that requires the automatic integration of multiple cognitive and linguistic abilities. Reading-related skills such as rapid automatized naming (RAN), phonological awareness, and letter knowledge can all be measured at the pre-reading stage and predict later reading ability (Byrne et al., 1997; Pennington & Lefly, 2001; Scarborough, 1998; Schatschneider et al., 2004). However, it is currently a major challenge to accurately identify reading difficulties early in reading development, when intervention is likely more effective (Al Otaiba et al., 2014; Blachman et al., 2014; Cavanaugh et al., 2004; Lovett et al., 2017; Torgesen, 2004; Vellutino et al., 1998). Optimizing screening batteries that allow early identification of reading problems at the outset of schooling, and therefore earlier intervention, is critical to optimizing long-term outcomes for children with reading difficulties (Connor et al., 2014).

Numerous studies have examined pre-school and kindergarten-age predictors of later reading ability and how various factors can modify the relationship between predictors and reading outcomes (e.g., Hjetland et al., 2017). Across studies, the measures that are most commonly identified as strong predictors of later reading in English include phonological awareness (PA), RAN, letter name and sound knowledge, and language ability (for reviews, see National Early Literacy Panel, 2008; Ozernov-Palchik & Gaab, 2016). Though RAN shares some processes with these other predictors, it has consistently been shown to uniquely relate to reading, beyond the contribution of phonological awareness (Kirby et al., 2003; Manis et al., 2000; Wolf & Bowers, 1999), and beyond similar measures of general processing speed and single (discrete) item naming (Altani et al., 2020; Logan et al., 2011). Whereas some measures such as letter knowledge are only predictive of reading for a short interval until they are mastered

(Paris, 2005), RAN retains its concurrent and predictive relation with reading over time (Wagner et al., 1997). Further, early RAN predicts reading over long time intervals, at least a decade into the future (Adlof et al., 2010; Mazzocco & Grimm, 2013). Importantly, the RAN-reading relationship persists across varying ages, reading abilities and alphabetic and non-alphabetic languages and orthographies of varying depth (Araújo et al., 2015; Araújo & Faísca, 2019; Caravolas et al., 2019; Furnes & Samuelsson, 2011).

Gaining a nuanced understanding of the relation between RAN and reading ability is important for two major reasons: informing educational/clinical practice and informing theory. In terms of informing practice, understanding the circumstances under which RAN best predicts later reading is crucial for screening and early identification of reading difficulties. For example, little is known about when the optimal time is to screen and whether the exact type of RAN test matters (in terms of number of items, type of items, use of raw or standardized score, and more). Identifying children with reading difficulties as early as possible, when intervention is more effective, would mitigate the compounding negative consequences that poor readers face under the predominant "wait to fail" model, such as reduced educational attainment, poorer socio-emotional well-being, and higher rates of entry into the juvenile justice system (Humphrey & Mullins, 2002; Richardson & Wydell, 2003; Svensson et al., 2001; Torgesen, 2004).

Understanding the nature of the RAN-reading relationship also informs understanding of the nature of reading ability and development as well as theory related to reading. Multiple-deficit models, pioneered by Wolf and Bowers' (1999) Double Deficit Hypothesis, consider naming speed to be one causal factor in reading ability (Menghini et al., 2010; Pennington, 2006; Pennington et al., 2012). However, in other prominent accounts such as the Simple View of Reading (Gough & Tunmer, 1986), the constructs of speed and automaticity as measured by

RAN are considered to play a minor role at best (as part of the decoding component, Johnston & Kirby, 2006). Another longstanding question in the field is how unique RAN is as a predictor, and its relationship to phonological processing (a construct that includes PA; e.g., Wagner et al., 1994; 1997). Many individual studies find that RAN is a unique predictor of reading, distinct from or beyond the contributions of phonological and letter knowledge or orthographic measures (Landerl et al., 2019; Norton & Wolf, 2012), and that they have distinct neural correlates (Norton et al., 2014, 2021). However, no meta-analysis to date has directly tested RAN's unique contribution above and beyond other pre-reading measures. Understanding the relationship between RAN, reading, and other pre-reading variables is thus key to clarifying RAN's role in reading development.

2.1.1 Defining RAN Tasks

RAN is measured as the time it takes a child to name an array of familiar items, such as objects, colors, numbers, or letters (Denckla & Rudel, 1976; Norton & Wolf, 2012), reflecting the automaticity of the multiple processes that are involved in this process (Wolf et al., 2000). There are several important parameters that define a true RAN task. First, the items to be named must be highly familiar or *automatized*. For example, when children are typically still learning their letters in kindergarten, the RAN letters task may not relate closely to reading because the naming is not automatized. However, once children have learned the names of letters and numbers with automaticity, these alphanumeric RAN tasks are completed faster than non-alphanumeric tasks (such as objects or colors) and are more strongly related to reading (Cardoso-Martins & Pennington, 2004; Schatschneider et al., 2004; Torgesen et al., 1997). Second, the items must be arranged in an array or grid and named in the left-to-right, row-by-row fashion that is analogous to reading in English. In rare cases, the items can be named top-to-bottom in columns, (e.g., van den Bos et al., 2002). Naming items that are presented one at a time in a speeded manner (discrete naming) is not the same as the serial process of a true RAN task (Altani et al., 2020; de Jong, 2011; Logan et al., 2011; Protopapas et al., 2013), even though some studies call this “discrete RAN.” Third, the RAN measure is usually based on time to complete the task. Some studies use the number of items/second or seconds/item (e.g., Schatschneider et al., 2004). Errors and self-corrections are not typically used in calculating a RAN score, but they may increase the time to name the array and thus be reflected in the naming time. Other factors can be calculated from a RAN task, such as pause time or change row-by-row (Amtmann et al., 2007; Georgiou et al., 2006; 2008), but these are less widely used in practice.

2.1.2 Theories of Mechanisms Underlying the RAN-Reading Relationship

Many potential explanations for why RAN relates so strongly to reading have been posited, including their shared processes of global processing speed (e.g., Kail & Hall, 1994), phonological processing (e.g., Wagner et al., 1997), serial visual processing and orthographic access (Sunseth & Bowers, 2002), and articulation (Papadopoulos et al., 2016; Wolf et al., 2000). These variables, along with many other shared cognitive processes, change over the course of development, and therefore the model explaining the relationship between RAN and reading must account for this. For example, as children gain accuracy and automaticity in reading, RAN speed becomes more strongly correlated with reading speed (Juul et al., 2014). This relationship varies depending on orthographic transparency, with accuracy measures plateauing much earlier in transparent than opaque orthographies (Seymour et al., 2003).

No matter how dynamic and multi-faceted the model between RAN and reading can be, there are specifications of how variables such as processing speed, serial processing, and articulation may relate to RAN and reading. Path models have been extensively tested, with each study finding slightly different model specifications (Cutting & Denckla, 2001; Georgiou et al., 2016; Papadopoulos et al., 2016). For example, the relationships among general processing speed, RAN, phonological processing, and orthographic processing change based on whether the orthographic processing measures are speeded or not (Georgiou et al., 2016). Another key specification is that the RAN-reading relationship is driven by not only serial processing or left-to-right eye movements (Protopapas et al., 2013), but cascading processing (i.e., processing multiple items simultaneously in overlapping fashion and effectively looking ahead at items to be named next; Gordon & Hoedemaker, 2016; Nayar et al., 2018). RAN may also have a unique relationship with oral reading fluency as opposed to silent word reading fluency (i.e., word-chains), suggesting that articulation plays an important role in the relationship between RAN and oral reading fluency (Georgiou et al., 2013; Papadopoulos et al., 2016). Though these studies were in Greek, it may hold that these models would replicate in English, as RAN shows similar patterns of relation with reading across languages (Araújo et al., 2015) and is considered more general to cognition than specific to a given language (Papadopoulos et al., 2016).

Ultimately, most current models suggest that RAN and reading are related because they share multiple underlying linguistic and non-linguistic cognitive processes (Georgiou & Parrila, 2020; Norton & Wolf, 2012; Wolf et al., 2000). The paths of these models may be “common cause” with RAN and reading both directly affected by processes like working memory, or through mediation, in which RAN ability may affect reading indirectly through improved orthographic processing or phonological awareness (Papadopoulos et al., 2016). Thus, within an individual, a profile of strengths and weaknesses of underlying cognitive processes will affect both RAN, reading, and other mediating variables to account for their relationship. Although the exact role of some processes such as articulation is debated (Cutting & Denckla, 2001; Georgiou

& Parrila, 2020; Lervåg & Hulme, 2009), it is agreed that multiple shared neural and cognitive processes underlie both RAN and reading (as demonstrated with fMRI; Cummine et al., 2015).

2.1.3 Insights on how RAN Relates to Reading from Meta-Analyses

Previous meta-analyses have documented the significant correlation between RAN and reading across various reading constructs and languages. In the first published meta-analysis of RAN and reading, Swanson et al. (2003) found a strong concurrent relationship between RAN and single word reading (*r* = -.41), when looking across a range of ages, reading abilities, and languages[1]. Two subsequent meta-analyses have found a similar magnitude of relationship between RAN and reading, while providing new contributions. Araújo et al. (2015) found the overall concurrent RAN-reading relationship across languages to be *r* = -.43, with slightly higher correlations in opaque orthographies like English. Their analyses included substantially more studies, and thus provided greater statistical power than earlier work by Swanson and colleagues. In turn, Hjetland et al. (2017) found the longitudinal correlation from early RAN to later reading to range from *r* = -.34 to -.37, depending on the reading measures used. Thus, they demonstrated that longitudinal correlations with RAN have similar effect sizes to concurrent correlations.

Differences in RAN ability have also been identified in two meta-analyses of children with reading difficulties. In a meta-analysis of various cognitive and reading-related skills, Kudo et al. (2015) found that the effect size difference for RAN in children without versus with reading difficulties was *d* = 0.89 (equivalent to *r* = .41), however only 10 samples were included in that analysis. In a much larger meta-analysis with 216 effect sizes analyzed, Araújo et al. (2019) documented an even larger RAN deficit in individuals with dyslexia (*d* = 1.19, equivalent

---

[1] Note that here, we present all correlations as negative, despite factors like raw versus standard scores, indicating that faster RAN is associated with better reading, as this is usually the observed direction of the relation.

to $r = .51$). These documented RAN deficits in children with reading difficulties/dyslexia support its use as an early screener.

In addition to demonstrating consistent correlations between RAN and reading, these meta-analyses also demonstrated that various factors (i.e., moderators), such as the type of stimuli used, the orthographic depth of the language studied, and the type of reading measure, affect the strength of the RAN-reading correlation. Swanson et al. (2003) found that of 11 possible moderators, children's grade when RAN and reading were assessed was the only significant moderator, with older children showing a stronger relationship between RAN and reading. However, these analyses were likely underpowered due to the limited published literature available in 2003. With more available literature, Araújo et al. (2015) found another moderator: the RAN-reading relationship is stronger in opaque vs. transparent alphabetic orthographies. They also found that the concurrent RAN-reading correlation was moderated by the type of RAN stimuli (alphanumeric stimuli had a stronger relationship with reading than non-alphanumeric), and by the type of reading measure (e.g., RAN had a stronger relationship with real word reading versus nonword reading). RAN's relationship with real word versus nonword reading was also extended to nonword versus real word spelling (Chen et al., 2021).

As noted above, only one meta-analysis has examined the longitudinal RAN-reading relationship; the broader focus of Hjetland et al. (2017) was to assess a variety of longitudinal predictors of reading comprehension, such as vocabulary and grammar, as well as RAN, across languages. As a result, they did not assess many potential moderators of the RAN-reading relationship. They found mean effect sizes for RAN predicting later single word reading of $r = -.37$ and predicting reading comprehension of $r = -.34$. These correlations are slightly lower than those found by Araújo et al. (2015), perhaps due to Hjetland et al.'s inclusion of only studies with reading comprehension measures and much smaller sample size overall, or the fact that this analysis included only longitudinal studies. Furthermore, in Hjetland et al.'s analyses, one study was an extreme outlier and was included with a positive rather than negative correlation with RAN[2]; thus, the effect sizes from this study may even be under-estimated.

2.1.4 Motivations and Goals for the Current Study

The purpose of this meta-analysis is to assess the longitudinal relationship from RAN measured in kindergarten or preschool to later reading abilities in English. Measuring the longitudinal relationship, as opposed to the concurrent relationship, is essential not only for investigating RAN's utility as an early screener for reading difficulties, but also essential for understanding the changing relationship between RAN and reading as reading transitions from a focus on accuracy to efficiency (Seymour et al., 2003). We consider a variety of reading constructs, including measures of nonword decoding (i.e., reading nonsense words like "sorp"), sight word reading (i.e., reading single words that can be recognized without decoding), reading comprehension (i.e., reading paragraphs or sentences and being able to answer questions about the writing's content) and reading fluency (i.e., reading sentences or paragraphs aloud as accurately and quickly as possible). This work thus extends a previous meta-analysis (Hjetland et al., 2017) to include articles that use all reading constructs rather than only reading comprehension as an outcome. We also directly test early RAN's unique contribution to later reading, above and beyond the contribution of PA. PA and RAN share considerable variance and

---

[2] Bishop & League (2006) reported a positive correlation between RAN time and reading ability (it appears the authors used raw time measures of RAN). However, in an earlier report from the same sample, Bishop (2003) reported positive correlations using standard scores (in the expected direction of this relationship). The RAN-reading correlations from this paper should likely have been treated as negative in this case for Hjetland's analyses, as all other measures in this and other meta-analyses were negative.

interest in parsing their respective effects has only grown since the formulation of the double deficit hypothesis (Norton & Wolf, 2012). This question serves practical and theoretical purposes in understanding how much RAN contributes to our understanding of early reading development. Finally, we perform extensive forward and backward snowball searching, as more papers were available to include beyond those identified in the Hjetland et al. (2017) dataset.

2.1.4.1 Practical Motivations

The key considerations for this design, including its focus on work in English-speaking children, early measures of RAN, and longitudinal relationships, are driven by a goal for this meta-analysis to inform specific policy recommendations for educators and administrators. It is clear that state- and local-level policymakers are looking for ways to best implement RAN in screening, as evidenced by the creation of measures such as the Arkansas Rapid Naming Screener and its use by other states (Arkansas Department of Education, 2017). As in previous meta-analyses examining the concurrent RAN-reading relationship, we also test several potential moderators, which address key practical questions. Practical questions, such as “how many items should a RAN task include?” and “at what age should I evaluate RAN?”, may help educators and clinicians choose effective screening measures. Policymakers are also interested in RAN’s unique contribution to predicting reading outcomes, which is why we have considered it alongside PA measures.

2.1.4.2 Theoretical Motivations

Most meta-analyses of RAN focus on documenting the relationship between RAN and reading while generally not trying to explain why RAN and reading are related. Here, we will test several questions related to why RAN and reading are correlated. Theoretical questions, such as “do timed reading measures more strongly relate to RAN than untimed reading measures?”

and “do nonword decoding tasks relate less strongly to RAN than sight word tasks?” may help researchers further converge on theory for why RAN and reading relate.

2.1.5 Summary

Because RAN ability develops considerably during the school-age years (Denckla & Rudel, 1974; Georgiou et al., 2006), its relationship to later reading ability may be different than the concurrent relations between RAN and reading at older ages. However, if early RAN reliably predicts later reading, it further increases the motivation to include RAN in kindergarten or preschool literacy screening. However, there is a lack of understanding of the theoretical and practical questions about how early RAN task performance relates to later reading abilities. As such, quantifying the average relationship between early RAN and later reading is the primary research question in this meta-analysis. Secondary questions are whether factors related to the RAN task, reading measure, or child participant sample, moderate the RAN-reading relationship. These specific questions and their rationale are explained in depth, and specific analyses are proposed in the Method section.

2.2 Method

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). Data collection and extraction processes are described in text and in **Figure 1**. The PRISMA checklist is provided as Supplemental Material. Our data, protocols, processing and analyses scripts, and other related documents are available via Open Science Framework: https://osf.io/awpqk/. This meta-analysis was considered exempt by the Institutional Review Board at Northwestern University.

2.2.1 Study Inclusion Criteria

For the present study, we focused on articles in which English was the primary language of the participants, as consistency of orthography can moderate the RAN-reading relationship

(Araújo et al., 2015) and the largest number of published studies are in English. We acknowledge that English is not a representative orthography (Share, 2008; 2021), but that this analysis serves as a starting point and allows specific conclusions to be drawn in at least this one language. As we were interested in early predictors of reading ability, we only included articles in which the initial timepoint with RAN assessment was in (the US equivalent of) kindergarten or preschool (the earliest stage at which RAN can be measured reliably) and reading was subsequently measured at some point in Grades 1-5. Thus, we only included studies that spanned at least one school year. For studies that only reported the sample's age rather than grade, we included the study if the mean age was ≤78 months (age 6.5 years, or the middle of first grade in the US). Studies with children who spoke other languages were excluded; however, studies with bilingual children were included if a) the language of instruction was English and b) the children were described as fluent in English. All eligibility criteria can be found in **Table 1**. Examples of specific decisions regarding inclusion/exclusion can be found in the Supplemental Materials.

2.2.2 Data Collection

On September 26, 2019, we identified possible sources through full-text database searches of EBSCO (PsychINFO, PsycARTICLES, and ERIC) and PubMed. We used the search terms: (reading OR dyslexia) AND ("rapid naming" OR "naming speed" OR "rapid automat* naming" OR "RAN" OR "rapid serial naming") AND ("preschool*" OR "kindergart*" OR "pre-school*" OR "pre k*" OR "pre-k*" OR "prek*" OR "child*"), see **Table 1**. This search returned 4497 titles, 4088 of which were unique. We re-ran this search on November 8, 2021, to include articles published since September 2019. **Figure 1** shows the number of articles at each stage.

2.2.2.1 Abstract and Title Screening

As a first step, one of two authors reviewed the title of each article from the database search; titles that were deemed to be clearly irrelevant were screened out. This title screening step resulted in 2098 potentially relevant articles with abstracts to be screened. These abstracts were then each reviewed by two different screeners. Three individuals contributed to abstract screening and consensus was reached in all cases of conflict[3]. Abstract screening for full-text inclusion agreement was 85% and all disagreements were resolved with consensus of three coders. 437 of these articles were deemed relevant and were then full-text screened. Seven trained coders screened full texts for inclusion, with 89% agreement and resolution of all disagreements. From these, 94 articles met the eligibility criteria. After contacting authors to obtain some that were not included in articles, 52 had relevant effect sizes. These articles were then each coded for various measures of interest twice, by two of five trained coders. There was 94% agreement across all variables and any disagreements were reviewed by the first and second author and resolved through consensus.

2.2.2.2 Snowball Search

After the database search and screening, a snowball search was conducted using references and citations of the 52 included studies with relevant effect sizes. For this snowball search, we used Microsoft Academic Graph (Wang et al., 2019), which is a database that tracks connections between published papers, such that every backward reference is also a forward citation, similar to Web of Science. All articles that were identified by the snowball search were title and abstract screened using the same processes as those described above. Snowball searching returned 43 articles that met the eligibility criteria. 15 of these studies had relevant effect sizes (after contacting authors) and added 10 unique samples. The search also returned 28

---

[3] (Some articles were triple-screened during training, but all other articles were double-screened.)

studies without relevant effect sizes, 14 of which were related to samples already contained in the corpus.

2.2.2.3 Contacting Authors for Additional Information/Data

Authors from either the database search or snowball search whose paper had no relevant effect sizes (e.g., because of reporting regressions or grouped analyses rather than correlations) were contacted via email to request raw data or correlation matrices so that the information could be included in the current analysis. For the papers where this was the case, 9 authors responded to our request, providing data on 10 unique samples.

2.2.3 Data Extraction

Data for this study were collected and managed using REDCap (Research Electronic Data Capture) tools hosted at Northwestern University (Harris et al., 2009, 2019). REDCap is a secure, web-based software platform designed to support data capture for research studies.

Data (including relevant information on the sample, tasks, and Pearson correlations) from each paper/sample were entered in REDCap by two independent coders, and consensus was reached in case of any discrepancy. For longitudinal studies that measured RAN and/or reading at multiple timepoints, we extracted only one kindergarten/preschool time point and only one grade school timepoint. This design consideration intentionally minimizes variance, as our primary question is focused on the utility of RAN as an early screener. However, a side effect of this approach is that it limits the variability that can be explained by age of testing. Timing of initial and follow-up assessments were coded in terms of the sample’s grade, as papers predominantly reported grade rather than age. Exceptions and further details are listed in Supplemental Materials.

2.2.3.1 Effect Size Extraction

The scoring of the RAN task affected whether the Pearson correlation with reading would be positive or negative. If a raw score (i.e., time) or rate (time/item) was used, the correlation was entered as negative. If a standard score or rate (item/time) was used, this value was multiplied by -1. There were a few exceptions to this rule, in which a reading measure was either based on time or rate (e.g., Wolf et al., 1986) or expressed as a chronological age lag (Heath & Hogben, 2004). In addition, there were several ambiguous cases that were carefully considered, see details in Supplemental Materials.

Many studies assess RAN as part of a large battery of reading-related measures that potentially predict later reading. Due to the many constructs measured in these large and longitudinal studies, many researchers created latent RAN or reading measures through factor or principal components analysis (Dally, 2006; Macdonald et al., 2013). We decided to extract these correlations between one or two latent variables as they qualify as Pearson correlations, and later test whether including them would change our results.

2.2.3.2 RAN Measure Categories

The stimuli used in a RAN task are typically restricted to one of five types: colors, objects, letters, digits, or occasionally animals. Even more rarely, studies have used colored animals (e.g., Catts et al. 1999). The ‘colored animals’ task (e.g., naming “blue cow,” “red dog,” etc.) is included here as a RAN task, but not compared with other stimulus types in moderator analyses due to the very few studies that employed it. We also excluded tasks with multiple stimulus types in the array, such as letters and numbers, in order to focus on the classic RAN task. Previous meta-analyses have found that the relationship with reading is stronger between alphanumeric (i.e., letters or numbers) than non-alphanumeric stimuli (such as colors or objects; Araújo et al., 2015). However, this was assessed concurrently, whereas different results may be seen with early RAN predicting later reading. Further, many children do not know their letters accurately or automatically in kindergarten or preschool, making a RAN letters task inappropriate for these younger children. Thus, in the current study **we quantified each RAN task's relationship with later reading** and **whether alphanumeric RAN tasks are a stronger predictor of later reading than non-alphanumeric RAN**.

2.2.3.3 Reading Measure Categories

Here, we operationalized three primary types of reading measures: reading fluency, reading comprehension, and single word reading measures. Fluency measures had to measure either a rate or total number of words read correctly in a pre-determined time limit in connected text (sentences or passages). This definition differs from fluency measures in Araújo et al. (2015), who used "items per second" as a measure of fluency. Single word reading included real and nonword reading tasks and was further broken down into single word efficiency (i.e., timed single word and nonword reading) and single word accuracy (i.e., untimed single word and nonword reading) measures. The full categorization of each reading measure is located in files available on the Open Science Framework site for this project.

Previous meta-analysis of children of all ages indicates that RAN is associated with single word reading accuracy (i.e., word ID) at $r = -.41$ and reading comprehension at $r = -.45$ (Swanson et al., 2003). Hjetland and colleagues (2017) found mean effect sizes of $r = -.37$ for word reading and $r = -.34$ for reading comprehension with earlier RAN measures. However, the specific correlations between RAN and reading vary considerably between and within studies. For example, in one study (Cronin & Carver, 1998), kindergarten RAN scores related to Grade 1 Word ID scores at $r = -.37$ to $-.60$, depending on the RAN task, and to passage comprehension at $r = -.31$ to $-.57$. Thus, we quantified RAN's relationship with 3 primary types of reading:

fluency, comprehension, and single word reading. Single word reading was further analyzed as accuracy versus efficiency measures.

***2.2.3.3.1* Timed Measures.** Because RAN is a speeded task, it is typically more closely related to timed or speeded reading measures (Savage & Frederickson, 2005; Schatschneider et al., 2004). This is evident in studies of older students; for example, RAN speed in grade 3 significantly predicted performance on a timed single word reading task in grades 3, 4, and 5, but did not reliably predict untimed single word reading (Georgiou et al., 2009). Further, one theoretical account posits that processes underlying RAN constrain the development of reading fluency (Lervåg & Hulme, 2009). Thus, we quantified RAN's relationship with timed and untimed reading measures.

***2.2.3.3.2* Nonword Reading.** Nonword reading task have extra phonological demands that sight words do not. Previous meta-analyses (Araújo et al., 2015) found a weaker correlation between nonword reading and RAN than real word reading and RAN. This difference may exist because nonword reading is much less automatic than real word reading, even early in reading development. Therefore, we quantified RAN's relationship with real word reading and nonword reading, with the prediction that the relationship between RAN and nonword measures would be weaker than RAN and real word reading.

2.2.3.4 Participant Characteristics

2.2.3.4.1 Reading Ability. Among older students, there is mixed evidence regarding whether RAN is a stronger correlate or predictor of reading ability among children who are poor readers than typical or skilled readers. Some studies find a stronger concurrent RAN-reading relation in poor readers (Araújo et al., 2011; Bowers et al., 1988; Felton & Brown, 1990; McBride-Chang & Manis, 1996). One study found that RAN in 3rd grade significantly predicted later single word reading in 8th grade among poor readers, but that there was no such significant relation in good readers (Meyer et al., 1998). On the other hand, meta-analyses of concurrent RAN-reading relations in older children reveal that the correlation between RAN and reading is similar in samples of typical readers and poor readers; Swanson et al. (2003) found correlations of $r = -.41$ for typical readers and -.43 for poor readers, and Araújo et al. (2015) found no significant differences in the magnitude of the concurrent relations between RAN and reading whether the sample of readers was poor/impaired ($r = -.49$), typical/average ($r = -.45$), or unselected ($r = -.43$). It is not known whether these differences across studies are due to a restricted range or "ceiling" effect in RAN among good readers with greater variability among poor readers (McBride-Chang & Manis, 1996) or whether differential relations truly exist in good versus poor readers.

Due to the focus here on young children, we are not able to examine the full range of reading ability and how it may correlate with RAN. We can probe whether children at risk for dyslexia may have a different RAN-reading relationship than peers without risk for dyslexia. Children with familial risk for dyslexia tend to have poorer RAN skills than their peers (Pennington & Lefly, 2001; van Bergen et al., 2012), yet not all children with familial risk or poor RAN scores go on to be poor readers. Some studies find a weaker RAN-reading relationship in those at risk for dyslexia; for example, Heath and Hogben (2004) found that pre-kindergarten RAN correlated with Grade 2 Word ID at $r = -.03$ for children with poor PA skills, compared with $r = -.38$ for children with good PA skills. Other studies find quite similar effect sizes across risk status; for example, Hulme et al. (2015) found children with vs. without risk for dyslexia had correlations between kindergarten RAN Objects and Grade 3 reading of $r = -.21$ and $r = -.22$, respectively. Here, we used a three-tier classification system of risk: low, medium, and high-risk. Any sample from the general population or an explicitly low-risk group was considered low-risk (e.g., Cardoso-Martins & Pennington, 2004). A medium risk sample was one where the study oversampled for dyslexia risk using family history and/or poor performance on pre-reading measures, but still included many low-risk participants (e.g., Ozernov-Palchik et al., 2017). High-risk samples were explicitly stated as such, categorized using family history and pre-reading measure performance, and were often analyzed as sub-groups in studies (e.g., Cardoso-Martins & Pennington, 2004). Thus, we tested whether early RAN is a better predictor in samples of primarily typically developing children as opposed to samples with larger proportions of children identified as at-risk for reading difficulties.

2.2.3.5 Practical Considerations

***2.2.3.5.1* RAN Task Publication, Standardization and Test Length.** There are a number of published, standardized and normed RAN measures that are used widely, including the Comprehensive Test of Phonological Processing (CTOPP and CTOPP-2; Wagner et al., 2013) and the RAN/RAS Tests (Wolf & Denckla, 2005), among others. However, many studies use researcher-created RAN tasks that have not necessarily been standardized or normed. Among these tests, the format of the RAN task, including how many different unique items (types) and total number of items included (tokens), also varies. A previous meta-analysis found no moderating effect for the total number of items in a RAN task on concurrent relations with reading (Araújo et al., 2015). Thus, we tested whether using a published, standardized measure influenced the RAN-reading relationship, as well as whether RAN measures with different numbers of items per set or total items, were more strongly related to reading.

***2.2.3.5.2* Timing of Initial RAN Assessment and Later Reading Assessment.** Dyslexia is typically not diagnosed before the end of grade 2 because the heterogeneity of reading development profiles makes it difficult to reliably identify children who will have ongoing reading difficulty. Thus, it would be helpful to know when RAN assessment is effective for predicting later reading. In the US, kindergarten screening often includes literacy; thus, many studies that investigate longitudinal relations with RAN measure it at the start of kindergarten. However, some studies have assessed RAN in children as young as age 3.5 (McBride-Chang & Kail, 2002; Su et al., 2017). Widely used normed measures of RAN are available for children age 4 and up (e.g., CTOPP-2). Thus, we tested how the timing of RAN assessment (i.e., preschool versus kindergarten) differentially impacts the RAN-reading correlation.

Another important consideration is the timing of the later or "outcome" reading measure, as the nature of the relations between early RAN and subsequent reading may change over the course of reading development. For example, early in reading development, children are developing accuracy in reading, and over time, they become accurate and build automaticity; thus, RAN may relate to fluency-based reading more strongly when reading is more automatized. In a practical sense, for early identification of reading problems, it may be important to know when this relation becomes stable. Wolf and colleagues (2000) suggested that RAN may play an attenuated role in predicting reading for typical readers after grade 2, because so many children achieve automaticity in naming and reading. Thus, we tested the extent to which the timing of reading assessment moderated the RAN-reading relationship.

2.2.3.6 Distinct Associations with Reading from Phonological Awareness

There is substantial shared variance between RAN and PA; thus, understanding each one's unique longitudinal relationship with reading is essential to understand the broader picture of how pre-reading skills relate to reading ability (Schatschneider et al., 2004; Vander Stappen & Reybroeck, 2018). The double deficit hypothesis (Wolf & Bowers, 1999) generated considerable interest in this topic. Sufficient studies exist to extract and meta-analyze their intercorrelations, yet no meta-analysis has done so. We operationalized PA measures as any task that required a participant to manipulate or isolate phonemes in words or nonwords (phonological memory tasks such as nonword repetition were excluded). Our categories of PA measures were thus elision/deletion, isolation, blending, and matching/rhyming, as well as composite PA measures testing these subcategories. Thus, we tested the unique relationship between RAN and reading controlling for PA, using semipartial correlations.

2.2.4 Outlier Handling

Due to the nature of nested effect sizes, we examined outliers at the study level. We did this by taking the mean of each effect size and moderator variable at the study level and then testing whether any observations fell above the 97.5%ile or below the 2.5%ile. If a study fell outside of these values, it was further investigated and considered for inclusion on a case-by-case basis; importantly, this was done before analysis so as not to bias results. All studies/samples were retained for intercept-only models. For moderator analyses, several studies were excluded as they were outliers for the variable of interest. These cases are described in Supplemental Materials.

2.2.5 Study Quality and Risk of Bias

Study quality measures can be helpful in identifying whether certain designs, such as double-blind randomized control trials, yield less-biased estimates of effect sizes. Features that reflect study quality are less clear for correlational, longitudinal research designs. Here, we use three measures of study quality and risk of bias: use of a published standardized RAN test, use of latent variables, and the study’s sample size. These were all separately analyzed as moderators of the RAN-reading relationship, as there is no gold-standard or guidance for doing so, we felt it was not appropriate to create a composite study quality and risk of bias measure.

2.2.6 Statistical Power

Power was calculated for each moderator analysis and is reported alongside each moderator analysis. As in Araújo et al. (2015), we used the value of 0.1 difference between Fisher's z values as the smallest difference that would be meaningful. For the sample risk proportion analysis (e.g., low, medium, and high risk proportion), we used .1 Fisher's z difference on either side of $z = .4$, as this is a typical RAN-reading correlation reported in other meta-analyses. As there is no widely accepted methodology for calculating moderator analyses' power in robust variance estimation (RVE) models, we used the degrees of freedom from each moderator analysis (rounded to the nearest integer, which is effectively a sample size). We used the *metapower* package (Griffin, 2020, 2021) to calculate power for each moderator tested, using the mean sample size of $n = 176$ and an I^2 value of 75%. Because this uses an *a priori* effect size estimate, this is not a *post hoc* power calculation. Power values for each analysis are presented alongside each model in **Table 4.** To calculate power for moderator analyses of semipartial correlations, we used a nearly identical procedure to the Pearson correlation power calculation. The only difference was that instead of using an I^2 value of 75%, we used an I^2 value of 50%, as this was much closer to the I^2 of the intercept-only model of the semipartial correlations.

2.2.7 Analysis Process and Plan

2.2.7.1 Meta-Analysis of RAN-Reading Correlations

Reported effects in the literature were transformed from Pearson correlations to Fisher's z-scores, which normalizes their distribution for analysis. They were then transformed back to Pearson correlations in results here, for ease of interpretation and comparison with other meta-analyses. To accommodate multiple effect sizes per study, we used correlated effects models using robust variance estimation (RVE) with the R (R Core Team, 2013) package *robumeta* (Fisher et al., 2017; Hedges et al., 2010). These models allow for correlated effects within a study, maximizing data retention. Furthermore, these models allow the grouping of multiple studies that share a sample (e.g., the International Longitudinal Twin Study; Furnes & Samuelsson, 2009, 2011). Intercept-only and moderator analyses were performed using the *robu* function. Moderators were tested in separate meta-regression models (e.g., separate models testing alphanumeric stimuli as a moderator and testing dyslexia risk as a moderator), except for time of assessment, in which the initial and outcome timepoints were considered together.

2.2.7.2 Meta-analysis of RAN-PA-Reading Semipartial Correlations

To address the practical question of RAN's unique contribution to reading, we coded the associations among PA, RAN, and reading. Correlation matrices from included studies were examined and the correlations between RAN-PA, PA-reading, and RAN-reading were extracted. For the semipartial analyses, correlations were not z-transformed, as semipartial correlations cannot be z-transformed (Aloe & Thompson, 2013). Pearson correlations (RAN-PA, PA-reading, RAN-reading) were used to calculate the semipartial correlations between RAN and reading, with the variance of PA partialled out. In order to pool these semipartial correlations, there needed to be equal numbers of RAN, PA, and reading measures per matrix. Because each study varied greatly in the number of measures for each construct, the simplest case of one measure for each construct (e.g., RAN, PA, or reading) was used to calculate each semipartial correlation. If multiple RAN, PA, or reading measures were used, the number of semipartial correlations calculated for each study could be represented by the formula $n_{sp} = n_{ran} * n_{pa} * n_{reading}$. These semipartial correlations were then pooled using the methods outlined by Aloe and Becker (2012). The variance component for each semipartial correlation was calculated using Equation 5 from

Aloe and Becker (2012).

2.2.7.3 Risk of Bias

To test for funnel plot asymmetry, which is indicative of publication or reporting bias, we used a technique that allows for multiple effect sizes per study. Traditional methods for examining funnel plot asymmetry, such as Egger's Regression or trim-and-fill analyses, only accommodate one effect size per study. "Sandwich" estimators (Rodgers & Pustejovsky, 2020) expand these methods to correlated effects models. We therefore used an "Egger's Sandwich Regression" to test for funnel plot asymmetry. As our data came from a variety of sources, we also ran a moderator analysis to test whether published effect sizes were larger than unpublished effect sizes (e.g., an unpublished dissertation, data emailed from authors).

2.3 Results

2.3.1 Sample Description

The final analytic sample (*n* = 10,513) was drawn from 60 independent samples across 67 papers. Whereas the largest sample size in the Hjetland et al. (2017) longitudinal RAN analyses was 3,746, the current sample is thus nearly three times greater, even though we restricted the language of the participants to English and the initial timepoint to before grade 1. For studies that reported age of participants at the initial timepoint, the mean age was 67.51 months (SD of 4.02) and a range of mean ages from 54-75 months across studies. The mean interval between initial and final timepoint was 27.41 months, which is consistent with our prioritization of the Grade 2 timepoint. Other descriptive statistics for the samples included are presented in **Table 2**.

2.3.2 Intercept-only Models

We calculated an intercept-only model to assess our main research question, the overall correlation between preschool/kindergarten RAN scores and later reading scores. The intercept-only model yielded an mean effect size of $z = -.40$ (95% CI: -.37 to -.44, $p < .001$), equivalent to a Pearson correlation of $r = -.38$. This indicates that on average, children with faster RAN time before grade school have stronger grade school reading performance. The forest plot for the overall intercept-only model is presented in Supplemental Material. Excluding studies that reported latent variables for RAN or reading resulted in nearly identical model results ($r = -.38$). There was substantial variability in studies' effect sizes ($I^2 = 74.09$; $\tau^2 = .018$), indicating that analysis of moderators may further clarify the RAN-reading relationship. We also tested intercept-only models including only a subset of studies based on what types of RAN tasks and reading measures the study used. These results are presented in **Table 3.** All models were significant at $p < .001$, indicating that the relationship between various RAN and reading measures is quite robust.

Many papers that report a RAN-reading correlation also measured PA and reported its correlations with RAN and reading. The meta-analysis of the semipartial correlations (r_{sp}) calculated from these matrices had large samples (N = 32; k = 353; n = 5,452). The intercept-only model of the semipartial correlations yielded an effect of $r_{sp} = -.25$; 95% CI -.28 to -.22.

2.3.3 Moderators and Meta-Regression

Primary practically and theoretically motivated moderators were analyzed and are presented in **Table 4**. We also tested whether partialling PA out of the RAN-reading relationship changed the theoretically motivated moderator effects; these analyses will be referred to as semipartial moderator analyses, as opposed to the primary moderator analyses, and are presented in **Table 5**. Several moderators changed considerably when PA was partialled out. To ensure that these changes were not due to the specific subset of studies included in semipartial analysis, the primary meta-analysis models were re-run with the same subset of studies as the semipartial correlation analyses. This subset of studies will be referred to as the subset of semipartial studies, for which the sample size is n = 5,452 compared to n = 10,513 for the full sample.

2.3.3.1 Practical Moderators

2.3.3.1.1 Unique RAN Items and Total RAN Items. We tested whether specific features of the RAN task administered in each study, such as the number of total items or the number of unique items, were differentially predictive of later reading. We found that neither the number of total items, nor the number of unique items moderated the RAN-reading relationship (all *p*s > .26). This indicates that RAN test length and item composition, within the limits of what has been studied, does not meaningfully modify the RAN-reading relationship.

2.3.3.1.2 Standardized RAN Measure. Next, we tested whether using published assessments that are standardized and normed, such as the RAN/RAS Tests or the RAN subtests from the CTOPP, affected the RAN-reading relationship. We found that using a published assessment had no effect ($\Delta r = .06$; $p = .18$) on the strength of the RAN-reading relationship. This also was an indicator of risk of study bias, indicating that study quality may be less likely to bias these results.

2.3.3.1.3 Age at Assessments. We tested whether the timing of the RAN or reading assessments (e.g., earlier or later than initial assessment at early kindergarten for RAN assessment or than Grade 2 for reading assessment) moderated the RAN-reading relationship. We found that age at reading assessment had no moderating effect ($\Delta r = 0.00$; $p = .97$), but that age at RAN assessment did have a marginally significant effect ($\Delta r = -.01$; $p = .07$), in the direction of later assessment having a stronger RAN-reading relationship. We considered that this result may be conflated with whether alphanumeric RAN was assessed or not, as younger children are less likely to be able to complete alphanumeric RAN, and alphanumeric RAN has been a stronger predictor than non-alphanumeric RAN in previous meta-analyses (Araújo et al., 2015). After controlling for whether the RAN task was alphanumeric or not, there was no effect of age at initial assessment ($\Delta r = 0.00$; $p = .15$). This result indicates that the exact timing of early RAN measurement does not differentially affect the RAN-reading relationship.

2.3.3.2 Theoretical Moderators

2.3.3.2.1 Alphanumeric versus Non-alphanumeric RAN. The correlations for RAN letters and RAN digits with reading were nearly identical ($r = -.46$ and $r = -.45$, respectively), as were correlations for RAN colors and RAN objects with reading ($r = -.32$ and $r = -.34$, respectively). Based on these values, the fact that studies find RAN digits to be automatized even earlier than letters (Åvall et al., 2019) and to be consistent with previous meta-analyses that combined these categories (e.g., Araújo et al., 2015), we collapsed the RAN types into alphanumeric and non-alphanumeric RAN. We then directly tested whether alphanumeric RAN was a better predictor of reading than non-alphanumeric RAN. We found that alphanumeric RAN is a significantly stronger predictor of reading ($\Delta r = .13$; $p = .01$), meaning that RAN tasks with letters or numbers had a stronger correlation with reading than did tasks with colors or objects. To consider the possibility that this relationship was conflated with initial age (because younger children may be less likely to have completed an alphanumeric task successfully), we ran the same analysis controlling for initial age, and the effect was unchanged ($\Delta r = .13$; $p = .01$). In sum, for our samples' ages, alphanumeric RAN was a stronger predictor of future reading regardless of age. However, it may be the case that studies considered age when selecting their RAN measures and tended to administer alphanumeric measures for children who were already automatic with those stimuli, as is intended. To test whether partialling out PA affected this relationship, we tested the moderator effect for Pearson correlations in the subset of semipartial studies ($\Delta r = .09$; $p = .02$), which was again significant. With PA partialled out, whether the RAN task was alphanumeric or not had a marginal effect on reading ability ($\Delta r_{sp} = .07$; $p = .07$).

2.3.3.2.2 Real versus Nonword Reading. Next, we directly tested whether measures of nonword reading had a weaker relationship with RAN than measures of single, real word reading. We found a significant effect ($\Delta r = -.09$; $p < .001$), with measures of nonword reading having a weaker relationship with RAN than measures of single, real word reading. This effect was unchanged in the subset of semipartial studies ($\Delta r = -.10$; $p = .01$). However, with PA partialled out, real word and nonword reading did not have a differential relationship with RAN ($\Delta r_{sp} = .04$; $p = .19$).

2.3.3.2.3 Timed versus Untimed Reading. We then tested whether timed reading measures were more related to RAN than untimed measures. We found no difference ($\Delta r = .00$; $p = .90$) between timed and untimed reading measures as they relate to RAN. In the subset of semipartial studies ($\Delta \underline{r} = .01$; $p = .88$), as well as with PA partialled out, timed and untimed reading tasks had no significant moderating effect ($\Delta r_{sp} = .05$; $p = .11$).

2.3.3.2.4 Reading Efficiency versus Reading Accuracy. As there were no differences in timed versus untimed reading measures, we also tested whether measures of reading efficiency were more related to RAN than measures of reading accuracy only. We found no difference ($\Delta r = -.01$; $p = .73$) between how measures of reading efficiency and reading accuracy relate to RAN. This effect was unchanged in the subset of semipartial studies ($\Delta r = .03$; $p = .57$). However, with PA partialled out, reading efficiency measures had a significantly stronger relationship with RAN than reading accuracy measures ($\Delta r_{sp} = .08$; $p = .03$).

2.3.3.2.5 Dyslexia Risk Proportion in the Sample. Using the three-level classification of dyslexia risk of the sample (low, medium, or high proportion of children at risk) in a single model, we tested whether the RAN-reading relationship was affected by dyslexia risk. There was no significant moderating effect of level of dyslexia risk (all $\Delta r \leq .05$; all *p*s > .46). In order to ensure that this was not specific to this grouping categorization, we also ran a model using a dichotomous categorization of risk (i.e., general population versus any type of high-risk sample) and found highly similar results ($\Delta r = .05$; $p = .31$). There was also no effect of dichotomized risk in the subset of semipartial studies ($\Delta r = .07$; $p = .11$), and there was no significant moderating effect of dyslexia risk with PA partialled out ($\Delta r_{sp} = .04$; $p = .21$). These results indicate that the RAN is a similar predictor of reading across samples of children that vary in risk for dyslexia.

2.3.4 Risk of Bias Analysis

To assess risk of bias, we ran an Egger's Sandwich Regression, in which the standard deviation estimates from each study were used as the moderator. We found no risk of bias in our effect size estimates ($p = .32$). Sample size is often used as a study quality measure as well; this result indicates that sample size has no significant effect on effect size estimates. However, because our data were composed of peer-reviewed studies, unpublished theses, and emailed data from published studies, we also ran moderator analyses with whether data were from a published paper or not (i.e., an unpublished dissertation or emailed data). These analyses revealed strong evidence of reporting bias, with published effect sizes being stronger than unpublished effect sizes ($\Delta r = .09$; $p = .02$). This effect was not driven by the inclusion of dissertation manuscripts ($\Delta r = .003$; $p = .97$), but rather by other types of unpublished data (e.g., emailed data). Due to the highly nested nature of these data, a funnel plot visualization is not provided, given that plotting up to 27 effect sizes with the same standard error would result in essentially a horizontal line on the funnel plot and be difficult to interpret.

2.4 Discussion

This meta-analysis expands on previous findings by documenting the longitudinal relationship between early RAN and various measures of later reading abilities in English-speaking children. Consistent with previous research and meta-analyses, RAN tasks were found to be a strong predictor of all types of reading. The mean effect size found here for RAN predicting reading overall ($r = -.38$) is similar to meta-analyses of concurrent RAN-reading correlations, with r ranging from -.38 to -.45 depending on reading measure in Swanson et al. (2003), $r = -.43$ Araújo et al. (2015), $r = -.34$ for reading comprehension, and $r = -.37$ for Word ID in Hjetland et al. (2017). We also estimated the semipartial correlation of early RAN on future reading controlling for PA ($r_{\text{sp}} = -.25$), distilling decades of research that has studied RAN unique effect on reading beyond the contribution of PA.

Our meta-analysis adds uniquely to the literature assessing the links between RAN and reading by highlighting the relevance of assessing RAN in kindergarten or preschool, and the robustness of this relationship over time and across various RAN and reading measures. The only existing longitudinal meta-analysis between RAN and reading was limited in its coverage of the literature and theoretical scope, with no moderators assessed (Hjetland et al., 2017). Our database searching, in conjunction with a snowball search strategy, yielded many more included articles, resulting in a sample size nearly three times larger. This much larger sample was ascertained despite restricting our age range to kindergarten and preschool and restricting our language to English.

Though RAN has long been considered independent of PA (Bowers & Wolf, 1993; Wolf & Bowers, 1999), the shared variance between the two is considerable, and parsing their independent effects is essential to understand their respective contributions to reading outcomes (Norton & Wolf, 2012; Vander Stappen & Reybroeck, 2018). We have therefore meta-analytically demonstrated for the first time the unique contribution of early RAN to later reading above early PA. This was ascertained by meta-analyzing semipartial correlations that were derived from correlation matrices. This analysis is the first step toward creating longitudinal meta-analytic path models of cognitive, pre-reading, and reading variables. We thus strongly advocate for researchers to share correlation matrices (and/or raw data), such as through supplementary materials and platforms such as Open Science Framework.

Another major contribution of the present study is the analysis of a variety of potential practical and theoretical moderators of the relationship between early RAN and later reading. For practical moderators, our analyses show that number of total items, and how many unique items were included in each set did not moderate the RAN-reading relationship align with and extend previous concurrent findings from Araújo and colleagues (2015). Our study is the first, to our knowledge, to examine RAN tasks that were published and standardized versus researcher-created; these variations also did not significantly alter the predictive relation of RAN with reading. In sum, these results show that RAN's relationship to reading is robust, regardless of how the measure is constructed. Whereas educators may not always have access to published, standardized measures, these data suggest that some RAN information is better than nothing.

For theoretical moderators, we found that RAN has a significantly stronger relation with reading when alphanumeric stimuli are used. This replicates and extending a previous concurrent meta-analysis across ages (Araújo et al., 2015), even despite the young age of the RAN assessments analyzed here. Partialling PA out slightly changed the moderating effect of alphanumeric stimuli from significant ($\Delta r = .09$; $p = .02$) to marginally significant ($\Delta r_{\text{sp}} = .07$; $p = .07$), but these small changes do not meaningfully change our interpretation. In considering different reading measures as outcomes, we found only a significant difference for RAN better predicting real word than nonword reading from the primary moderator analyses. However, with early PA partialled out, RAN correlated similarly with nonword and real word reading (Δr_{sp} = .04; p = .19). We also found differences in reading efficiency measures versus reading accuracy measures, only with PA partialled out. In contrast, Araújo and colleagues found differences between timed and untimed measures across orthographies and ages, without partialling out PA. We discuss the implications of these findings for theory and for practice, below.

2.4.1 Insights to the Nature of the RAN-Reading Relationship

Our primary moderator analyses show that alphanumeric RAN has a significantly stronger relationship with later reading than does non-alphanumeric RAN, as well as that nonword reading is significantly less related to RAN than real word reading. These results, taken together, support shared cognitive processes models, such that the more similar the processes that RAN and a given reading task tap, the more strongly that they will be correlated (Georgiou & Parrila, 2020). In the case of nonword reading, there is a heavy phonological decoding (letter-to-sound correspondence) component that RAN does not share, which is why partialling out PA reduces this effect. In other words, when PA was controlled for, RAN had no differentiable relationship to real word versus nonword reading. In the case of alphanumeric RAN, symbolic representation is required for both alphanumeric RAN and reading. Individual studies have found that alphanumeric RAN and non-alphanumeric RAN correlate equally well with later reading (e.g., van den Bos et al., 2002) or that both load on the same latent factor (Papadopoulos et al., 2016). However, our meta-analysis in young children shows that alphanumeric RAN is stronger than non-alphanumeric RAN regardless of whether RAN was measured in preschool or kindergarten, and that age on its own had no effect on the RAN-reading relationship once the alphanumeric stimulus type effect was accounted for. This is strongly consistent with meta-analytic findings from Araújo et al. (2015). The effect is large in both the current and Araújo et al.’s meta-analyses, but not so large that it would be unexpected for an individual study to find similar correlation sizes between reading and alphanumeric and non-alphanumeric RAN.

Our results also show an interesting pattern for timed measures versus untimed measures, as well as reading efficiency versus reading accuracy. In the primary moderator analyses, neither timed versus untimed nor efficiency versus accuracy showed significant results. However, by partialling out the effect of PA, it is clear that RAN alone has a stronger relationship with reading efficiency than reading accuracy measures. Though the semipartial moderator analysis for timed versus untimed measures did not reach significance ($\Delta r_{\text{sp}} = .05$; $p = .11$), there was a moderate change from the primary moderator analyses which show the same RAN-reading correlations for timed and untimed measures ($\Delta r = .00$; $p = .90$).

One potential reason the semipartial moderator analyses did not reach significance for timed versus untimed measures is that in the early years of reading development, accuracy-based and time-based measures are strongly correlated (e.g., Schatschneider et al., 2004). A difference emerges in intermediate and advanced readers once children build reading automaticity, but it is not present in beginning readers in either our sample or in the beginning and pre-readers included in the meta-analysis from Araújo et al. (2015). This may be particularly true for the English-speaking samples used here, as reading accuracy takes longer to transition to reading efficiency in opaque orthographies (Seymour et al., 2003). These findings are consistent with the idea that reading accuracy is not yet automatic in early grades in English (Chall, 1983; Samuels & Flor, 1997), and as a result, various reading measures may be more highly correlated early in schooling (i.e., less differentiable) than they are at later stages when most children have developed automaticity. More highly correlated reading measures in our earlier outcome timepoint (centered around 2^{nd} grade) would likely result in weaker moderating effects when comparing different types of reading measures.

Consistent with other meta-analyses' findings of no differences in relations with RAN between good versus poor readers, we found no difference between samples with a large proportion of children at-risk for dyslexia and those with very few at risk. This may indicate that children at-risk and children not at risk are using similar cognitive processes, even if these processes are impaired in children at risk. Although we are not fully able to explore the lower tail of RAN and reading performers, these results further support the idea that RAN is a continuous ability and dimensionally predicts of reading, rather than a dichotomous "present or absent" skill.

2.4.2 Practical Insights for Using RAN as a Screener

These results provide practical insights into using RAN for effective screening for later reading difficulties. Importantly, RAN should always be assessed as part of a battery of screening measures, as RAN alone only predicts 14% of variance in future reading scores. No screening battery is perfectly accurate (with no false positives or negatives), but a nuanced understanding of a child's profile will provide educators with the clearest path forward. Nonetheless, our results indicate that the relation between early RAN and later reading is remarkably consistent. The particular characteristics of the RAN measure, such as number of items and whether the task was from a published test, did not significantly alter the strength of the RAN-reading relationship. These facets of RAN as a predictor had not been assessed in previous meta-analyses, yet they provide concrete guidance for researchers and educators in planning RAN measures for screening. There was not a significant difference between RAN measures conducted in preschool versus kindergarten in terms of their relationship with later reading; there was a trend toward stronger predictive power, but the trend was reduced when controlling for alphanumeric RAN, which is often administered in later years. The advantage of earlier identification of potential reading difficulties, so that earlier intervention can be provided, suggests that it would be optimal to employ RAN tasks in screening in pre-school or pre-kindergarten, as soon as RAN can be assessed validly.

The stimulus type used in early RAN assessment is a relevant consideration, as alphanumeric RAN measures were more strongly related to later reading than were non-alphanumeric measures. An important caveat is that RAN tasks, by definition, depend on the child being able to name items with automaticity, and many articles noted that many children could not perform a RAN Letters task in kindergarten, as their letter name knowledge was not yet accurate and automatic (e.g., Catts et al., 1999). Thus, for children in kindergarten or preschool who do not yet know the names of letters or digits automatically, a RAN task using colors or objects would be a better choice; once letters or digits are known with automaticity, those are a better choice for later reading prediction. To what degree a speeded naming task is automatized in young children has long been debated (e.g., Åvall et al., 2019; Wolf et al., 1986) and is not particularly testable in a meta-analysis. Nonetheless, our results clearly demonstrate that RAN, when measured at a young age, maintains its robust relationship with reading.

The question that frequently follows after RAN screening is “what RAN time or score is worrisome?” Unfortunately, research has not yet determined a single cutoff score for “dyslexia risk” or what is “good” versus “poor” RAN; in fact, this may not be possible given that RAN is both a continuous measure and one aspect of the constellation of reading-related abilities. At this point, using a published, standardized RAN measure that provides standard scores or percentiles provides the advantage that it may help educators and clinicians understand where a child’s RAN ability falls relative to their peers as an indicator of risk for dyslexia, even though our data showed that researcher-created measures equally predicted later RAN. It is important to note that administering a RAN task according to any standardized instructions and minimizing distractions so as to obtain the child's best performance is crucial to obtaining a valid score.

Educators and clinicians should also recognize that an effective screening battery for dyslexia and reading difficulties must include RAN alongside other indicators such as phonological awareness (see Petscher et al., 2019, for recommendations). Even using the most evidence-based screening tools in combination with assessment of the child's family or neuroimaging measures, there is still uncertainty about which children will develop reading difficulty (Norton et al., 2019; Zuk et al., 2020). As the field moves forward in understanding early indicators of reading difficulties, RAN will undoubtedly play a role, given its universal and robust relation with reading.

2.4.3 Limitations

There are several limitations of this study to consider. The primary limitation was that we restricted our sample to only English-speaking students. As English is an outlier orthography, many of our findings about the transition from accurate to efficient reading are not generalizable to more transparent orthographies. Specifically, the children in our study likely acquire reading efficiency later than those learning transparent orthographies, which would affect many of our analyses, such as RAN's relationship to timed measures. We plan to address this shortcoming in future studies that include cross-language comparisons.

Another potential limitation of our study was our decision to not create composite measures of study quality. Instead, we chose to analyze study quality in terms of moderators, based on the concern over validity of using simple sums to describe study quality (Shamliyan et al., 2010; Whiting et al., 2005). Similarly, Hjetland et al. (2017) found no effect of study quality in an overlapping sample of papers, which aligns with our results that sample size, latent variables, and use of published/standardized tests do not predict variation in effect sizes. These variables functionally comprise study quality in longitudinal designs capturing the relationship between RAN and reading.

Another limitation is the limited statistical power for moderator analyses. Although we found no differences for unique RAN items or total RAN items, we had limited power to detect possible effects for a multitude of reasons. Araújo et al. (2015) noted similar difficulties, even with a larger corpus of sources and subjects. We offer the same caution in interpreting our moderator analysis results with low power.

Other limitations relate to the RAN tasks themselves. One limitation was the fact that there were incomplete descriptions of the measures in many studies, which was particularly common for researcher-created RAN tasks. Despite our effort to carefully review all available information in the published papers (and in many cases, request additional details from authors via email), many papers had incomplete descriptions of their RAN tasks, particularly relating to how many unique items and how many total items the task had. Furthermore, there was not much variability in the number of unique items, as many articles used Denckla and Rudel’s (1976) version or the updated RAN-RAS tests (Wolf & Denckla, 2005) each with 5 unique items per task, or the CTOPP that has 6 unique items. Despite the incomplete information from a number of studies, we believe we had sufficient power to detect these effects if they truly existed, as 288 (of 373) effect sizes were analyzed for the model that tested unique and total items as moderators.

The definition of at-risk in samples also varied greatly across studies and could limit interpretation of our results. For example, Cardoso-Martins and Pennington (2004) recruited a high-risk group from the children whose one of the parents has reading problems and a low-risk group from the children with no family history of reading problems. Hulme et al. (2015) also divided groups based on family history; however, they included another criterion of whether children have language impairment or not. In contrast, Heath and Hogben (2004) divided groups only based on poor and good phonological awareness abilities. Felton (1992) used teacher ratings of children's expected reading ability. There is strong evidence for different subtypes or component skills in dyslexia even beyond the double deficit (e.g., O'Brien et al., 2012), and pooling these samples could miss whether the RAN-reading relationship changes with the etiology for a given subgroup. Furthermore, examining the lower end of the RAN distribution through the lens of dyslexia risk does not directly test nonlinearities in the relationship between RAN and reading. Nonetheless, the heterogeneity present in our coding reflects the real-world heterogeneity of risk definitions, and our categories were designed to reflect that.

A final limitation is that the studies selected for the semipartial analyses may have some bias. Specifically, the reporting of correlation matrices in supplementary or primary data has become somewhat standard practice for large studies. The results from primary moderator and semipartial moderator analyses appeared highly similar, but we cannot rule out that some bias may be present in selecting these studies for a semipartial correlational meta-analysis.

2.4.4 Future Directions

We chose to focus on only traditional RAN tasks at certain timepoints in the English language in order to maximize practical and policy impact. As a result, there are several clear directions for future research to expand upon our study by broadening the scope. Future studies may consider different designs, such as meta-analytic path modeling of the relationships among cognitive, pre-reading, and reading variables. Though the majority of studies and all published tests focus on RAN total time, aspects of RAN such as analyses of inter-item pause times as a predictor would be promising to investigate, as pause times have been shown to relate highly with reading fluency (Lervåg & Hulme, 2009).

Given that we focused on a single outcome timepoint in each study that was close to the end of Grade 2, another potential future direction would be to test how longitudinal RAN-reading relationships change within studies and more broadly over time. As we prioritized collecting only one time point per study, we were not able to analyze whether correlations from early RAN to later reading changed over time within a study, as is suggested by a number of authors (de Jong & van der Leij, 2002; Wagner et al., 1997). To our knowledge, correlated effects RVE models have not been used to analyze longitudinal, within-study data. Many of the papers collected for the present analysis would be ideal to use in testing whether RVE is suitable for longitudinally dependent effect sizes and provide further insight into how RAN relates to reading over time.

Another clear direction for future research is to include multiple languages, as well as individuals who speak multiple languages, to assess similarities and differences of RAN as a predictor reading ability (Gottardo et al., 2021). In the past, other authors had suggested that RAN is a better predictor in more transparent languages (see Georgiou et al., 2008). In their meta-analysis, Araújo et al. (2015) reported that orthographically opaque orthographies such as English have a stronger concurrent correlation between RAN and reading than do transparent orthographies, but we do not have meta-analytic evidence of this effect longitudinally. Cross-linguistic studies have provided evidence that kindergarten RAN may be a stronger longitudinal predictor in opaque orthographies than more transparent orthographies, but there are no significant differences across languages for RAN measured in grade 1 (Furnes & Samuelsson, 2011; Landerl et al., 2021). Other studies have found equally strong correlations in transparent orthographies such as Czech (Caravolas et al., 2013), and qualitative reviews have noted that the longitudinal, cross-linguistic effect is likely small (Landerl et al., 2021). Taken together, this further highlights the need for a larger systematic approach that is sensitive to the many between-study differences in cross-linguistic research, such as the selection of developmentally appropriate reading measures across languages (see Papadopoulos et al., 2021 for a review).

Finally, given that a major focus was the utility of using RAN as a screener, future research should endeavor to provide concrete recommendations of what RAN performance indicates meaningful risk for reading difficulties and dyslexia. Few studies have provided clear formulas or cutoffs about which children are at greatest risk (Catts et al., 2001 is a notable exception). Even fewer studies have examined how best to provide intervention specific to children who have RAN difficulties that impact their reading, as it seems that training RAN itself is not effective in improving reading (de Jong & Vrielink, 2004; Kirby et al., 2010). Indeed, early measures of RAN may be an important, easy-to-collect early indicator of reading problems, akin to a “check engine light” that signals the need for further assessment and monitoring (Norton, 2020).

2.5 References

*Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. https://doi.org/10.1177/0022219410369067

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children*, *81*(1), 11–27. https://doi.org/10.1177/0014402914532234

Aloe, A. M., & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*, *37*(2), 278–297. https://doi.org/10.3102/1076998610396901

Aloe, A. M., & Thompson, C. G. (2013). The synthesis of partial effect sizes. *Journal of the Society for Social Work and Research*, *4*(4), 390–405. https://doi.org/10.5243/jsswr.2013.24

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). Tracking the serial advantage in the naming rate of multiple over isolated stimulus displays. *Reading and Writing*, *33*(2), 349–375. https://doi.org/10.1007/s11145-019-09962-7

Amtmann, D., Abbott, R. D., & Berninger, V. (2007). Mixture growth models of RAN and RAS row by row: Insight into the reading system at work over time. *Reading and Writing*, *20*(8), 785–813. https://doi.org/10.1007/s11145-006-9041-y

Araújo, S., & Faísca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading*, *23*(5), 349–368. https://doi.org/10.1080/10888438.2019.1572758

Araújo, S., Inacio, F., Francisco, A., Faísca, L., Petersson, K. M., & Reis, A. (2011). Component processes subserving rapid automatized naming in dyslexic and non-dyslexic readers. *Dyslexia*, *17*(3), 242–255. https://doi.org/https:doi.org/10.1002/dys.433

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Arkansas Department of Education. (2017). *Dyslexia Resource Guide*. Arkansas Department of Education.http://dese.ade.arkansas.gov/public/userfiles/Learning_Services/Dyslexia/DRG-Final-12-13-17-JS1.pdf

Åvall, M., Wolff, U., & Gustafsson, J. E. (2019). Rapid automatized naming in a developmental perspective between ages 4 and 10. Dyslexia, 25(4), 360–373. https://doi.org/10.1002/dys.1631

*Badian, N. A. (1998). A validation of the role of preschool phonological and orthographic skills in the prediction of reading. *Journal of Learning Disabilities*, *31*(5), 472–481. https://doi.org/10.1177/002221949803100505

*Badian, N. A., Duffy, F. H., Als, H., & McAnulty, G. B. (1991). Linguistic profiles of dyslexic and good readers. *Annals of Dyslexia*, *41*, 221–245. https://doi.org/10.1007/BF02648088

*Badian, N. A., McAnulty, G. B., Duffy, F. H., & Als, H. (1990). Prediction of dyslexia in kindergarten boys. *Annals of Dyslexia*, *40*, 152–169. https://doi.org/10.1007/BF02648146

*Biddle, K. R. (1996). *Timing deficits in impaired readers: An investigation of visual naming speed and verbal fluency*. Thesis, Tufts University.

*Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly*, 26(3), 189-200.

*Bishop, A. G., & League, M. B. (2006). Identifying a multivariate screening model to predict reading difficulties at the onset of kindergarten: A longitudinal analysis. *Learning Disability Quarterly*, 29(4), 235-252.

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, *106*(1), 46–57. https://doi.org/10.1037/a0033663

Bowers, P. G., Steffy, R., & Tate, E. (1988). Comparison of the effects of IQ control methods on memory and naming speed predictors of reading disability. *Reading Research Quarterly*, *23*(3), 304–319. https://doi.org/10.2307/748044

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing*, *5*(1), 69–85. https://doi.org/10.1007/BF01026919

*Burgoyne, K., Malone, S., Lervag, A., & Hulme, C. (2019). Pattern understanding is a predictor of early reading and arithmetic skills. *Early Childhood Research Quarterly*, *49*, 69–80. https://doi.org/10.1016/j.ecresq.2019.06.006

Byrne, B., Fielding-Barnsley, R., Ashley, L., & Larsen, K. (1997). Assessing the child’s and the environment’s contribution to reading acquisition: What we know and what we don’t know. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 265–285). Lawrence Erlbaum.

*Caravolas, M., Lervag, A., Defior, S., Seidlova Malkova, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*(8), 1398–1407. https://doi.org/10.1177/0956797612473122

*Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, *23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

*Cardoso-Martins, C., & Pennington, B. F. (2004). The relationship between phoneme awareness and rapid serial naming skills and literacy acquisition: The role of developmental period and reading ability. *Scientific Studies of Reading*, *8*(1), 27–52. https://doi.org/10.1207/s1532799xssr0801_3

*Carroll, J. M., Solity, J., & Shapiro, L. R. (2016). Predicting dyslexia using prereading skills: the role of sensorimotor and cognitive abilities. *Journal of Child Psychology and Psychiatry*, 57(6), 750-758.

*Catts, H. W. (1991). Early identification of dyslexia: Evidence from a follow-up study of speech-language impaired children. *Annals of Dyslexia*, *41*(1), 163–177. https://doi.org/10.1007/BF02648084

*Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, *36*(5), 948–958. https://doi.org/10.1044/jshr.3605.948

*Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*(4), 331–361. https://doi.org/10.1207/s1532799xssr0304_2

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children. *Language, Speech, and Hearing Services in Schools, 32*(1), 38–50. https:// doi.org/10.1044/0161-1461(2001/004)

*Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407–1425. psyh. https://doi.org/10.1007/s11145-015-9576-x

*Catts, H. W., McIlraith, A., Bridges, M. S., & Nielsen, D. C. (2017). Viewing a phonological deficit within a multifactorial model of dyslexia. *Reading and Writing*, *30*(3), 613–629. psyh. https://doi.org/10.1007/s11145-016-9692-2

*Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, *48*(3), 281–297. https://doi.org/10.1177/0022219413498115

Cavanaugh, C. L., Kim, A.-H., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal*, *2*(1), 13.

Chall, J. (1983). *Stages of reading development*. McGraw-Hill.

Chen, Y.-J. I., Thompson, C. G., Xu, Z., Irey, R. C., & Georgiou, G. K. (2021). Rapid automatized naming and spelling performance in alphabetic languages: A meta-analysis. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10160-7

*Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education*, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

*Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B. F., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2015). Genetic and environmental etiologies of the longitudinal relations between prereading skills and reading. *Child Development*, *86*(2), 342–361. https://doi.org/10.1111/cdev.12295

*Cirino, P. T., Child, A. E., & Macdonald, K. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002

Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute of Education Sciences Research Centers. NCSER 2014-3000*. National Center for Special Education Research.

*Cronin, V., & Carver, P. (1998). Phonological sensitivity, rapid naming, and beginning reading. *Applied Psycholinguistics*, *19*(3), 447–461. https://doi.org/10.1017/S0142716400010262

*Cronin, V. S. (2013). RAN and double-deficit theory. *Journal of Learning Disabilities, 46*(2), 182–190. https://doi.org/10.1177/0022219411413544

Cummine, J., Chouinard, B., Szepesvari, E., & Georgiou, G. K. (2015). An examination of the rapid automatized naming–reading relationship using functional magnetic resonance imaging. *Neuroscience*, *305*, 49–66. https://doi.org/10.1016/j.neuroscience.2015.07.071

Cutting, L. E., & Denckla, M. B. (2001). The relationship of rapid serial naming and word reading in normally developing readers: An exploratory model. *Reading and Writing*, *14*(7), 673–705.

*Dally, K. (2006). The influence of phonological processing and inattentive behavior on reading acquisition. *Journal of Educational Psychology*, *98*(2), 420–437. https://doi.org/10.1037/0022-0663.98.2.420

de Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading*, *15*(4), 314–337. https://doi.org/10.1080/10888438.2010.485624

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, *6*(1), 51–77. https://doi.org/10.1207/S1532799XSSR0601_03

de Jong, P. F., & Vrielink, L. O. (2004). Rapid automatic naming: Easy to measure, hard to improve (quickly). *Annals of Dyslexia*, *54*(1), 65–88. https://doi.org/10.1007/s11881-004-0004-1

Denckla, M. B., & Rudel, R. G. (1976). Rapid automatized naming (R.A.N): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479.

*Dittman, C. K. (2016). The impact of early classroom inattention on phonological processing and word-reading development. *Journal of Attention Disorders*, *20*(8), 653–664. https://doi.org/10.1177/1087054713478979

*Ellis, N., & Large, B. (1988). The early stages of reading: A longitudinal study. *Applied Cognitive Psychology*, *2*(1), 47–76. https://doi.org/10.1002/acp.2350020106

*Evans, M. A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds and phonological awareness: An examination of kindergarten children across letters and of letters across children. *Reading and Writing: An Interdisciplinary Journal*, *19*(9), 959–989.

*Evans, M. A., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *54*(2), 65–75.

Felton, R. H. (1992). Early identification of children at risk for reading disabilities. *Topics in Early Childhood Special Education, 12*(2), 212– 229. https://doi.org/10.1177/027112149201200206

*Felton, R. H., & Brown, I. S. (1990). Phonological processes as predictors of specific reading skills in children at risk for reading failure. *Reading and Writing*, *2*(1), 39–59. https://doi.org/10.1007/BF00383373

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: An R-package for robust variance estimation in meta-analysis*.

*Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading*, *32*(3), 275–292. https://doi.org/10.1111/j.1467-9817.2009.01393.x

*Furnes, B., & Samuelsson, S. (2011). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, *21*(1), 85–95. https://doi.org/10.1016/j.lindif.2010.10.005

Georgiou, G. K., & Parrila, R. (2020). What mechanism underlies the rapid automatized naming–reading relation? *Journal of Experimental Child Psychology*, *194*, 104840. https://doi.org/10.1016/j.jecp.2020.104840

*Georgiou, G. K., Parrila, R., & Kirby, J. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading*, *10*(2), 199–220. https://doi.org/10.1207/s1532799xssr1002_4

Georgiou, G. K., Parrila, R., Kirby, J. R., & Stephenson, K. (2008). Rapid naming components and their relationship with phonological awareness, orthographic knowledge, speed of processing, and different reading outcomes. *Scientific Studies of Reading*, *12*(4), 325–350. https://doi.org/10.1080/10888430802378518

Georgiou, G. K., Parrila, R., & Liao, C. H. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing*, *21*(9), 885–903. https://doi.org/10.1007/s11145-007-9096-4

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2016). The anatomy of the RAN-reading relationship. *Reading and Writing*, *29*(9), 1793–1815. https://doi.org/10.1007/s11145-016-9653-9

*Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology, 115*(1), 218–225. https://doi.org/10.1016/j. jecp.2012.10.015

Gottardo, A., Chen, X., & Huo, M. R. Y. (2021). Understanding within- and cross-language relations among language, preliteracy skills, and word reading in bilingual learners: Evidence from the science of reading. *Reading Research Quarterly*, *56*(S1), S371–S390. https://doi.org/10.1002/rrq.410

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Griffin, J. W. (2020). *metapower: Power Analysis for Meta-Analysis* (0.2.1) [Computer software]. https://CRAN.R-project.org/package=metapower

Griffin, J. W. (2021). Calculating statistical power for meta-analysis using metapower. *The Quantitative Methods for Psychology*, *17*(1), 24–39. https://doi.org/10.20982/tqmp.17.1.p024

*Gunn, B., Smolkowski, K., & Vadasy, P. (2010). Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, *4*(1), 53–86. https://doi.org/10.1080/19345747.2010.488716

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

*Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology*, *101*(2), 388–402. https://doi.org/10.1037/a0015115

*Heath, S. M., & Hogben, J. H. (2004). Cost-effective prediction of reading difficulties. *Journal of Speech, Language, and Hearing Research*, *47*(4), 751–765. https://doi.org/10.1044/1092-4388(2004/057)

*Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing*, *12*(1), 99–128. https://doi.org/10.1023/A:1008033824385

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. https://doi.org/10.1002/jrsm.5

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–155. https://doi.org/10.4073/csr.2017.14

Gordon, P. C., & Hoedemaker, R. S. (2016). Effective scheduling of looking and talking during rapid automatized naming. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 742. https://doi.org/10.1037/xhp0000171

*Hood, M. H. (2005). *The Role of Temporal and Phonological Processing In Early Reading Development: A Longitudinal Study.* Thesis, Griffith University. https://doi.org/10.25904/1912/2673

*Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Humphrey, N., & Mullins, P. M. (2002). Self-concept and self-esteem in developmental dyslexia. *Journal of Research in Special Educational Needs*, *2*(2), 1–13. https://doi.org/10.1111/j.1471-3802.2002.00163.x

*Inoue, T., Georgiou, G. K., Parrila, R., & Kirby, J. R. (2018). Examining an extended home literacy model: The mediating roles of emergent literacy skills and reading fluency. *Scientific Studies of Reading, 22*(4), 273-288. doi.org/10.1080/10888438.2018.1435663

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19,* 339–361. https://doi.org/10.1007/s11145-005-4644-2

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology, 106*(4), 1096–1106. https://doi.org/10.1037/a0037100

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30*(6), 949–954. https://doi.org/10.1037/0012-1649.30.6.949

*Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P. G., & Landerl, K. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*(3), 341–362. https://doi.org/10.1598/RRQ.45.3.4

*Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*(3), 453–464. https://doi.org/10.1037/0022-0663.95.3.453

Kudo, M. F., Lussier, C. M., & Swanson, H. L. (2015). Reading disabilities in children: A selective meta-analysis of the cognitive literature. *Research in Developmental Disabilities, 40*, 51–62. https://doi. org/10.1016/j.ridd.2015.01.002

*Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*(3), 195–216. psyh. https://doi.org/10.1016/j.lindif.2005.12.001

Landerl, K., Castles, A., & Parrila, R. (2021, epub ahead of print). Cognitive precursors of reading: A cross-linguistic perspective. *Scientific Studies of Reading*. https://doi.org/10.1080/10888438.2021.1983820

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., Parrila, R., & Georgiou, G. K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading*, *23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

*Ledesma, H. M. L. (2002). *Language factors influencing early reading development in bilingual (Filipino-English) boys*. Thesis, Georgia State University.

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*(8), 1040–1048. https://doi.org/10.1111/j.1467-9280.2009.02405.x

*Lewis, B. A., Avrich, A. A., Freebairn, L. A., Hansen, A. J., Sucheston, L. E., Kuo, I., Taylor, H., Iyengar, S., & Steina, C. (2011). Literacy outcomes of children with early childhood speech sound disorders: Impact of endophenotypes. *Journal of Speech, Language, and Hearing Research*, *54*(6), 1628–1643. https://doi.org/10.1044/1092-4388(2011/10-0124)

Logan, J. A. R., Schatschneider, C., & Wagner, R. K. (2011). Rapid serial naming and reading ability: The role of lexical access. *Reading and Writing*, *24*(1), 1–25. https://doi.org/10.1007/s11145-009-9199-1

Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology*, *109*(7), 889–914. https://doi.org/10.1037/edu0000181

*Macdonald, H. H., Sullivan, A. L., & Watkins, M. W. (2013). Multivariate screening model for later word reading achievement: Predictive utility of prereading skills and cognitive ability. *Journal of Applied School Psychology*, *29*(1), 52–71. https://doi.org/10.1080/15377903.2013.751476

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33*(4), 325–333. https://doi.org/10.1177/002221940003300405

*Mazzocco, M. M. M., & Grimm, K. J. (2013). Growth in rapid automatized naming from grades K to 8 in children with math or reading disabilities. *Journal of Learning Disabilities*, *46*(6), 517–533. https://doi.org/10.1177/0022219413477475

McBride–Chang, C., & Kail, R. V. (2002). Cross–cultural similarities in the predictors of reading acquisition. *Child Development*, *73*(5), 1392–1407. https://doi.org/10.1111/1467-8624.00479

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal*, *8*(4), 323–339. https://doi.org/10.1007/BF00395112

*McIlraith, A. L. (2018). Predicting word reading ability: A quantile regression study. *Journal of Research in Reading*, *41*(1), 79–96. https://doi.org/10.1111/1467-9817.12089

*McMillen, S. M. (2018). Pescado or fish? Rapid automatic naming performance for young Spanish-speaking English language learners. Thesis, University of Memphis.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffino, M., & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: A comparative study. *Neuropsychologia*, *48*(4), 863–872. https://doi.org/10.1016/j.neuropsychologia.2009.11.003

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Selective predictive value of rapid automatized naming in poor readers. *Journal of Learning Disabilities*, *31*, 106–117. https://doi.org/10.1177/002221949803100201

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting Items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

National Early Literacy Panel. (2008). *Developing Early Literacy: Report of the National Early Literacy Panel: A Scientific Synthesis of Early Literacy Development and Implications of Intervention*. National Institute for Literacy.

Nayar, K., Gordon, P. C., Martin, G. E., Hogan, A. L., La Valle, C., McKinney, W., Lee, M., Norton, E. S., & Losh, M. (2018). Links between looking and speaking in autism and first-degree relatives: Insights into the expression of genetic liability to autism. *Molecular Autism*, *9*(1), 51. https://doi.org/10.1186/s13229-018-0233-5

Norton, E. S. (2020). What educators need to know about rapid automatized naming (RAN). *LDA Bulletin*, *52*(1), 25–28.

Norton, E. S., Beach, S. D., Eddy, M. D., McWeeny, S., Ozernov-Palchik, O., Gaab, N., & Gabrieli, J. D. E. (2021). ERP mismatch negativity amplitude and asymmetry reflect phonological and rapid automatized naming skills in English-speaking kindergartners.

Frontiers in Human Neuroscience, *15*, 624617. https://doi.org/10.3389/fnhum.2021.624617

Norton, E. S., Black, J. M., Stanley, L. M., Tanaka, H., Gabrieli, J. D. E., Sawyer, C., & Hoeft, F. (2014). Functional neuroanatomical evidence for the double-deficit hypothesis of developmental dyslexia. *Neuropsychologia*, *61*, 235–246. https://doi.org/10.1016/j.neuropsychologia.2014.06.015

Norton, E. S., Gabrieli, J. D. E., & Gaab, N. (2019). Neural predictors of dyslexia. In L. Verhoeven, C. Perfetti, & K. Pugh (Eds.), *Developmental Dyslexia across Languages and Writing Systems: A Handbook* (pp. 253–276). Cambridge University Press.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

O’Brien, B. A., Wolf, M., & Lovett, M. W. (2012). A taxometric investigation of developmental dyslexia subtypes. *Dyslexia: An International Journal of Research and Practice*, *18*(1), 16–39. https://doi.org/10.1002/dys.1431

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the ‘dyslexia paradox’: Reading brain and behavior for early markers of developmental dyslexia. *WIREs Cognitive Science*, *7*(2), 156–176. https://doi.org/10.1002/wcs.1383

*Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental Science*, *20*(5). https://doi.org/10.1111/desc.12471

Papadopoulos, T. C., Csépe, V., Aro, M., Caravolas, M., Diakidoy, I.-A., & Olive, T. (2021). Methodological issues in literacy research across languages: Evidence from alphabetic orthographies. *Reading Research Quarterly, 56*(S1), S351–S370. https://doi.org/10.1002/rrq.407

Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in Psychology, 7*, 1217. https://doi.org/10.3389/fpsyg.2016.01217

*Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*(1), 3–26. https://doi.org/10.1207/s1532799xssr0801_2

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202. https://doi.org/10.1598/RRQ.40.2.3

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

Pennington, B., & Lefly, D. (2001). Early reading development in children at family risk for dyslexia. *Child Development*, *72*(3), 816–833. https://doi.org/10.1111/1467-8624.00317

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of dyslexia by single vs. Multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212–224. https://doi.org/10.1037/a0025823

*Petersen, D. B., & Gillam, R. B. (2013). Accurately predicting future reading difficulty for bilingual Latino children at risk for language impairment. *Learning Disabilities Research & Practice*, 28(3), 113–128. https://doi.org/10.1111/ldrp.12014

*Peterson, R. L., Arnett, A. B., Pennington, B. F., Byrne, B., Samuelsson, S., & Olson, R. K. (2018). Literacy acquisition influences children’s rapid automatized naming. *Developmental Science*, *21*(3), 1–9. https://doi.org/10.1111/desc.12589

Petscher, Y., Fien, H., Stanley, C., Gearin, B., Gaab, N., Fletcher, J. M., & Johnson, E. (2019). *Screening for Dyslexia*. National Center on Improving Literacy. Retrieved from https://improvingliteracy.org/sites/improvingliteracy2.uoregon.edu/files/whitepaper/screening-fordyslexia.pdf

Protopapas, A., Angeli, A., & Georgiou, G. (2013). RAN backward: A test of the visual scanning hypothesis. *Scientific Studies of Reading*, *17*(6), 453–461. https://doi.org/10.1080/10888438.2013.769556

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Richardson, J. T. E., & Wydell, T. N. (2003). The representation and attainment of students with dyslexia in UK higher education. *Reading and Writing*, *16*(5), 475–503. https://doi.org/10.1023/A:1024261927214

Rodgers, M., & Pustejovsky, J. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *MetaArXiv*, 1–25. https://doi.org/10.31222/osf.io/vqp8u

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly*, *13*(2), 107–121. https://doi.org/10.1080/1057356970130202

Savage, R., & Frederickson, N. (2005). Evidence of a highly specific relationship between rapid automatic naming of digits and text-reading speed. *Brain and Language*, *93*(2), 152–159. https://doi.org/10.1016/j.bandl.2004.09.005

*Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities Research Reviews*, *2*(1), 54–63. https://doi.org/10.1002/(SICI)1098-2779(1996)2:1<54::AID-MRDD9>3.0.CO;2-X

Scarborough, H. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, *48*(1), 115–136. https://doi.org/10.1007/s11881-998-0006-5

*Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Shamliyan, T., Kane, R. L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, *63*(10), 1061–1070. https://doi.org/10.1016/j.jclinepi.2010.04.014

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography *Psychological Bulletin, 134*(4), 584–615. https://doi.org/10.1037/ 0033-2909.134.4.584

Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly*, *56*(S1), S391–S402. https://doi.org/10.1002/rrq.401

*Smith, S. L., Scott, K. A., Roberts, J., & Locke, J. L. (2008). Disabled readers’ performance on tasks of phonological processing, rapid naming, and letter knowledge before and after kindergarten. *Learning Disabilities Research & Practice*, *23*(3), 113–124. https://doi.org/10.1111/j.1540-5826.2008.00269.x

*Snowling, M. J., Lervåg, A., Nash, H. M., & Hulme, C. (2019). Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Developmental Science, 22*(1).

Su, M., Peyre, H., Song, S., McBride, C., Tardif, T., Li, H., Zhang, Y., Liang, W., Zhang, Z., Ramus, F., & Shu, H. (2017). The influence of early linguistic skills and family factors on literacy acquisition in Chinese children: Follow-up from age 3 to age 11. *Learning and Instruction*, *49*, 54–63. https://doi.org/10.1016/j.learninstruc.2016.12.003

Sunseth, K., & Bowers, P. G. (2002). Rapid naming and phonemic awareness: Contributions to reading, spelling, and orthographic knowledge. *Scientific Studies of Reading*, *6*, 401–429.

Svensson, I., Lundberg, I., & Jacobson, C. (2001). The prevalence of reading and spelling difficulties among inmates of institutions for compulsory care of juvenile delinquents. *Dyslexia*, *7*(2), 62–76. https://doi.org/10.1002/dys.178

Swanson, H. L., Trainin, G., Necoechea, D., & Hammill, D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research*, *73*(4), 407–440. https://doi.org/10.3102/00346543073004407

Torgesen, J. K. (2004). Preventing early reading failure. *American Educator*, *28*(3), 6–9.

*Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5), 276–286. https://doi.org/10.1177/002221949402700503

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading, 1*(2), 161–185. https://doi.org/10.1207/s1532799xssr0102_4

*Uhry, J. K. (2002). Kindergarten phonological awareness and rapid serial naming as predictors of Grade 2 reading and spelling. In E. Witruk, A. D. Friederici, & T. Lachmann (Eds.), *Basic functions of language, reading and reading disability*. (2002-01144-017; pp. 299–313). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-1011-6_18

*Vadasy, P. F., & Sanders, E. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading and Writing*, *21*(9), 929–963. https://doi.org/10.1007/s11145-008-9119-9

van Bergen, E., de Jong, P. F., Plakas, A., Maassen, B., & van der Leij, A. (2012). Child and parental literacy levels within families with a history of dyslexia. *Journal of Child Psychology and Psychiatry*, *53*(1), 28–36. https://doi.org/10.1111/j.1469-7610.2011.02418.x

Van den Bos, K. P., Zijlstra, B. J., & Lutje Spelberg, H. C. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading, 6*(1), 25–49. https://doi.org/10.1207/S1532799XSSR0601_02

Vander Stappen, C., & Reybroeck, M. V. (2018). Phonological awareness and rapid automatized naming are independent phonological competencies with specific impacts on word reading and spelling: An Intervention Study. *Frontiers in Psychology, 9,* 320. https://doi.org/10.3389/fpsyg.2018.00320

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology*, *36*(4), 367–397.

*Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*(1), 73–87. https://doi.org/10.1037/0012-1649.30.1.73

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479. https://doi.org/10.1037//0012-1649.33.3.468

Wagner, R. K., Torgsen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing, 2nd edition (CTOPP-2)*. Pro-Ed.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., & Rogahn, R. (2019). A review of Microsoft Academic Services for science of science studies. *Frontiers in Big Data*, *2*. https://doi.org/10.3389/fdata.2019.00045

Whiting, P., Harbord, R., & Kleijnen, J. (2005). No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodology*, *5*(1), 19. https://doi.org/10.1186/1471-2288-5-19

*Wolf, M. (1984). Naming, reading, and the dyslexias: A longitudinal overview. *Annals of Dyslexia*, *34*(1), 87–115. https://doi.org/10.1007/BF02663615

*Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. *Child Development, 57*(4), 988–1000. https://doi.org/10.1111/j.1467-8624.1986.tb00260.x

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415. https://doi.org/10.1037/0022-0663.91.3.415

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities*, *33*(4), 387–407. https://doi.org/10.1177/002221940003300409

Wolf, M., & Denckla, M. B. (2005). *RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests*. Pro-Ed.

*Wolter, J. A, Self, T., & Apel, K. (2011). Initial mental graphemic representation acquisition and later literacy achievement in children with language impairment: A longitudinal study. *Journal of Learning Disabilities*, *44*(6), 543–555. https://doi.org/10.1177/0022219410392042

Zuk, J., Dunstan, J., Norton, E. S., Yu, X., Ozernov-Palchik, O., Wang, Y., Hogan, T. P., Gabrieli, J. D. E., & Gaab, N. (2020). Multifactorial pathways facilitate resilience among kindergarteners at risk for dyslexia: A longitudinal behavioral and neuroimaging study. *Developmental Science*, *24*(1), e12983. https://doi.org/10.1111/desc.12983

2.6 Tables and Figures

Table 1. Search Terms and Study Eligibility Criteria for Study 1

Variable	Search Terms	Inclusion Criteria	Exclusion Criteria
Initial Sample Point	"preschool*" OR "kindergart*" OR "pre-school*" OR "pre k*" OR "pre-k*" OR "prek*" OR "child*"	If no grade listed, ≤78 months mean age	If US/CAN sample: Called Grade 1 (or later) If UK/AUS: Called Year 2 (or later)
RAN Measure	"rapid naming" OR "naming speed" OR "rapid automat* naming" OR "RAN" OR "rapid serial naming"		If measure was labeled RAN or rapid naming, but tested naming of all 26 letters
Reading Outcome	“reading” OR “dyslexia”		If reading measure was assessed before Grade 1
Language	*Search criteria were not restricted by language*	If sample was English L1 or early/simultaneous bilinguals	If sample was drawn from L2 English immersion school
Study Design	*Search criteria were not restricted by study design*	Longitudinal study, minimum 3 months	If the study was not longitudinal OR conducted for less than 3 months OR If the study was described as a case study

Table 2. Descriptive Statistics for Samples Included in the Full Meta-Analysis for Study 1

Model	**N**	**k**	**n**	**Mean**	**SD**	**Range**
Initial Timepoint						
K	47	295	8552			
Pre-K	13	51	2508			
Mixed K/Pre-K	1	27	139			
Final Timepoint						
Grade 1	27	134	5972			
Grade 2	28	164	3902			
Grade 3	11	51	1621			
Grade 4	8	24	2050			
Time between measures	60	373	10513	27.38	11.16	12-57
RAN Task						
Publication						
Published/Standardized	16	86	4526			
Not Published/Standardized	46	287	6305			
Stimuli						
Alphanumeric	22	109	4425			
Non-Alphanumeric	50	255	9068			
RAN Colors	22	69	4044			
RAN Objects	29	118	5689			
RAN Letters	16	63	3196			
RAN Numbers	12	35	3232			
Composition						
RAN Total Items	48	297	8457	72.28	43.77	24-216
RAN Unique Items	46	288	7136	5.84	2.50	4-20
Sample Risk Proportion						
Low Risk	42	238	8528			
Medium Risk	7	72	1579			
High Risk	12	63	487			
Latent Variable(s) Used						
Yes	5	12	1809			
No	58	361	9879			

Note. N = number of samples/studies; k = number of effect sizes; n = number of participants

The N for some sections may not sum to 10513 as a result of these factors not being mutually exclusive within a study.

Table 3. Main Effects for Study 1: Intercept-only Models

Model	**N**	**k**	**I^2**	**τ^2**	**r**	***t***	**df**	**95% CI**
All Studies/Samples	60	373	74.09	.018	-.38	-22.35	50.21	[-.44 -.37]
RAN Type								
Colors	22	69	66.39	.012	-.32	-11.60	19.69	[-.40 -.27]
Objects	29	118	74.21	.012	-.34	-15.67	25.75	[-.41 -.31]
Letters	16	63	68.11	.017	-.46	-15.01	10.81	[-.57 -.42]
Digits	12	35	76.94	.015	-.45	-11.60	10.42	[-.58 -.39]
Reading Measure Types								
Reading Comprehension	39	87	74.43	.021	-.38	-15.91	31.05	[-.46 -.35]
Reading Fluency	23	54	77.84	.036	-.35	-7.95	17.60	[-.47 -.28]
Single Word Reading	50	193	69.30	.015	-.38	-22.28	40.59	[-.44 -.36]
Reading Measure Splits								
Single Word Reading								
Real Word Reading	45	109	70.24	.015	-.41	-24.43	38.85	[-.46 -.39]
Nonword Reading	38	84	66.59	.013	-.33	-16.05	28.48	[-.38 -.29]
Timing								
Timed Reading	33	137	81.27	.032	-.37	-11.70	26.86	[-.46 -.32]
Untimed Reading	57	223	70.25	.014	-.37	-21.78	48.48	[-.43 -.36]
Efficiency and Accuracy								
Efficiency	22	57	52.52	.009	-.40	-19.15	14.12	[-.47 -.38]
Accuracy	48	155	70.07	.015	-.37	-20.48	41.42	[-.43 -.35]

Note. N = number of samples/studies; k = number of effect sizes
All models were significant at $p < .001$.

Table 4. Primary Moderator Effects for Practical and Theoretical Considerations for Study 1: Pearson Correlations

Model	**N**	**k**	**I^2**	**τ^2**	**r**	***t***	**df**	***p***	**95% CI**	**Power**
Practical Considerations										
Unique RAN Tokens	45	285	68.49	.016						
Intercept					-.34	-3.84	5.94			
Unique Tokens					-.01	-.81	4.52	.46	[-.04 .03]	.11
Total RAN Items	46	290	63.80	.012						
Intercept					-.46	-13.5	25.6			
Total Items					.00	1.19	10.7	.26	[-.00 .00]	.19
Standardized RAN Test	60	373	73.94	.019						
Intercept					-.40	-21.07	37.81			
Published/Std Test					.06	1.37	23.52	.18	[-.03 .15]	.36
Age at Assessments	60	373	73.97	.019						
Intercept					-.31	-3.28	21.25			
Initial (RAN) Age (mos.)					-.01	-1.95	19.52	.07	[-.01 .00]	.31
Final (Reading) Age (mos.)					.00	0.04	19.51	.97	[-.00 .00]	
Theoretical Considerations										
Alphanumeric vs. Non-Alphanumeric	58	364	69.45	.015						
Intercept					-.46	-11.05	14.22			
Non-Alphanumeric					.13	2.78	21.83	**.01**	[.03 .23]	.33
Nonword vs. Real Word Reading	50	193	66.55	.013						
Intercept					-.33	-15.62	28.50			
Real Word Measure					-.09	-3.73	37.09	**<.001**	[-.14 -.04]	.51
Timed vs. Untimed Reading	58	360	72.56	.017						
Intercept					-.37	-23.08	40.97			
Timed Reading					.00	.13	31.74	.90	[-.06 .07]	.45
Efficiency vs. Accuracy	56	212	69.45	.015						
Intercept					-.38	-20.23	39.61			
Efficiency					-.01	-.035	18.83	.73	[-.09 .06]	.30
Sample Risk Proportion	60	373	74.50	.019						
Intercept					-.35	-5.62	7.18			
Low Risk					-.05	-0.78	9.19	.46	[-.20 .10]	.36
Medium Risk					-.01	-0.17	11.78	.87	[-.19 .16]	

Note. N = number of studies; k = number of effect sizes
All intercepts were significant at $p < .01$. Moderator effects indicated in **bold are *p* < .05**.

Table 5. Main and Moderator Effects for Semipartial Correlation Meta-Analysis for Study 1

Model	**N**	**k**	**I^2**	**τ^2**	**r**	***t***	**df**	***p***	**95% CI**	**Power**
All Studies/Samples	32	353	60.94	.007	-.25	-17.7	27	<.001	[-.28 -.22]	.99
Theoretical Considerations										
Alphanumeric vs. Non-Alphanumeric	31	350	50.95	.015						
Intercept					-.30	-9.66	6.71			
Non-Alphanumeric					.07	2.07	9.45	.07	[-.01 .14]	.28
Nonword vs. Real Word Reading	26	203	65.11	.008						
Intercept					-.23	-8.29	15.40			
Real Word Measure					-.04	-1.35	20.80	.19	[-.09 .02]	.56
Timed vs. Untimed Reading	32	347	58.89	.006						
Intercept					-.23	-16.68	22.80			
Timed Reading					-.06	-1.92	18.80	.07	[-.12 .01]	.52
Efficiency vs. Accuracy	31	223	53.53	.005						
Intercept					-.24	-14.49	22.67			
Efficiency					-.08	-3.06	9.37	**.01**	[-.14 .02]	.28
Sample Risk Proportion	32	353	60.78	.007						
Intercept					-.26	-15.86	18.30			
Risk					.04	1.32	15.20	.21	[-.23 .08]	.43

Note. N = number of studies; k = number of effect sizes

All intercepts were significant at $p < .05$. Moderator effects indicated in **bold are *p* < .05**.

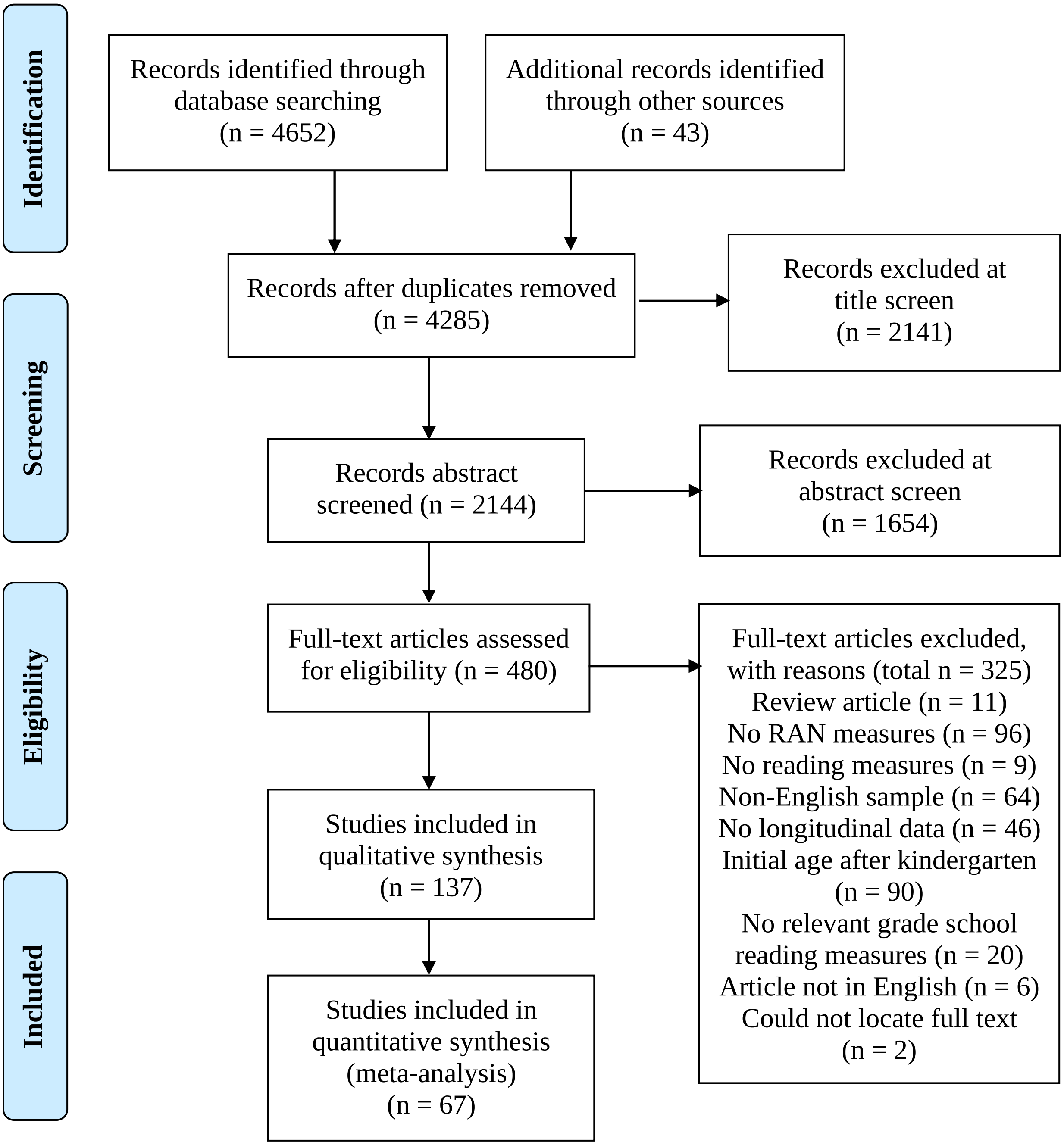
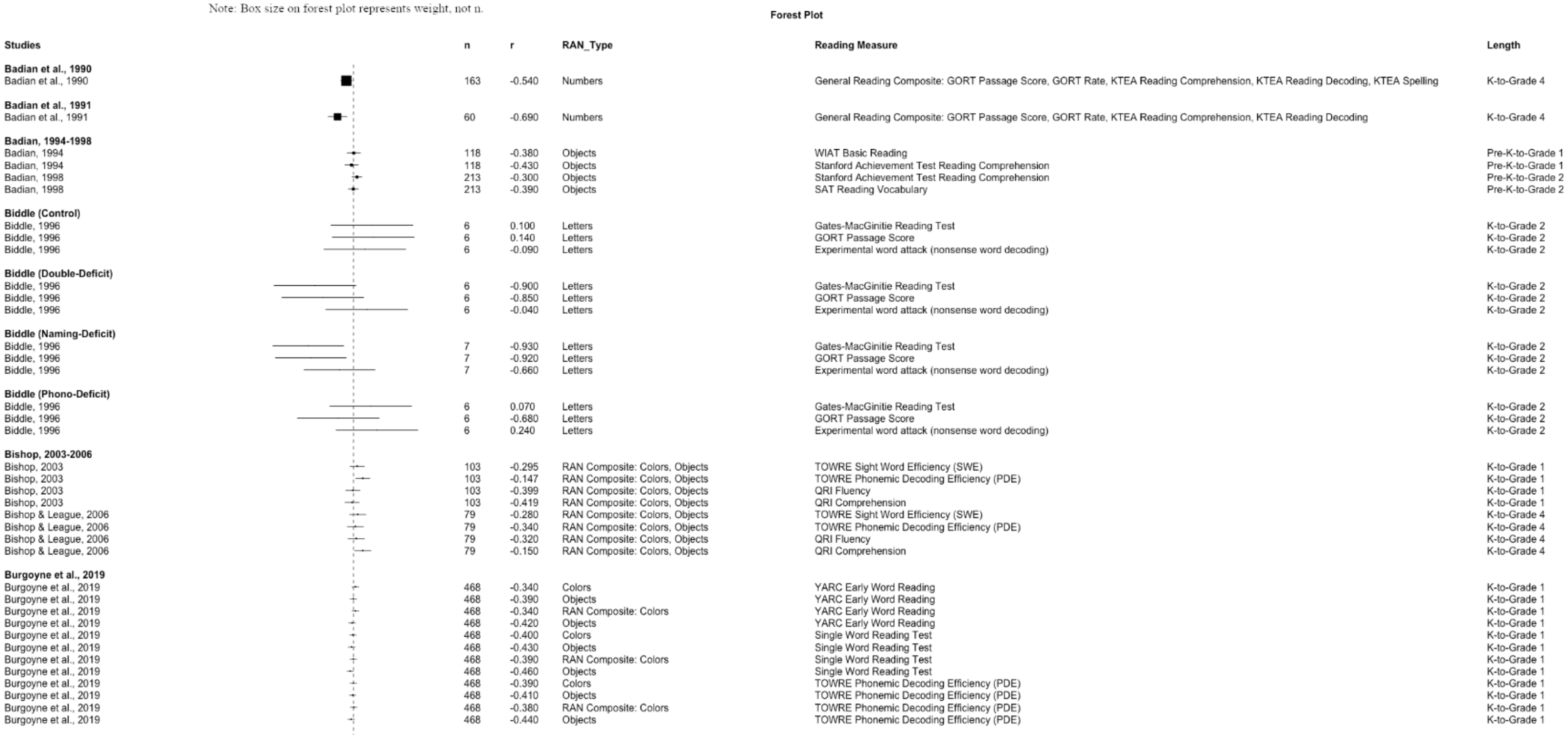


Figure 1. PRISMA Flow Diagram for Study 1

Figure 2. Forest Plot of RAN-Reading Correlations



Note: Box size on forest plot represents weight, not n.

Forest Plot

Studies	n	r	RAN_Type	Reading Measure	Length
Badian et al., 1990					
Badian et al., 1990	163	-0.540	Numbers	General Reading Composite: GORT Passage Score, GORT Rate, KTEA Reading Comprehension, KTEA Reading Decoding, KTEA Spelling	K-to-Grade 4
Badian et al., 1991					
Badian et al., 1991	60	-0.690	Numbers	General Reading Composite: GORT Passage Score, GORT Rate, KTEA Reading Comprehension, KTEA Reading Decoding	K-to-Grade 4
Badian, 1994-1998					
Badian, 1994	118	-0.380	Objects	WIAT Basic Reading	Pre-K-to-Grade 1
Badian, 1994	118	-0.430	Objects	Stanford Achievement Test Reading Comprehension	Pre-K-to-Grade 1
Badian, 1998	213	-0.300	Objects	Stanford Achievement Test Reading Comprehension	Pre-K-to-Grade 2
Badian, 1998	213	-0.390	Objects	SAT Reading Vocabulary	Pre-K-to-Grade 2
Biddle (Control)					
Biddle, 1996	6	0.100	Letters	Gates-MacGinitie Reading Test	K-to-Grade 2
Biddle, 1996	6	0.140	Letters	GORT Passage Score	K-to-Grade 2
Biddle, 1996	6	-0.090	Letters	Experimental word attack (nonsense word decoding)	K-to-Grade 2
Biddle (Double-Deficit)					
Biddle, 1996	6	-0.900	Letters	Gates-MacGinitie Reading Test	K-to-Grade 2
Biddle, 1996	6	-0.850	Letters	GORT Passage Score	K-to-Grade 2
Biddle, 1996	6	-0.040	Letters	Experimental word attack (nonsense word decoding)	K-to-Grade 2
Biddle (Naming-Deficit)					
Biddle, 1996	7	-0.930	Letters	Gates-MacGinitie Reading Test	K-to-Grade 2
Biddle, 1996	7	-0.920	Letters	GORT Passage Score	K-to-Grade 2
Biddle, 1996	7	-0.660	Letters	Experimental word attack (nonsense word decoding)	K-to-Grade 2
Biddle (Phono-Deficit)					
Biddle, 1996	6	0.070	Letters	Gates-MacGinitie Reading Test	K-to-Grade 2
Biddle, 1996	6	-0.680	Letters	GORT Passage Score	K-to-Grade 2
Biddle, 1996	6	0.240	Letters	Experimental word attack (nonsense word decoding)	K-to-Grade 2
Bishop, 2003-2006					
Bishop, 2003	103	-0.295	RAN Composite: Colors, Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Bishop, 2003	103	-0.147	RAN Composite: Colors, Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Bishop, 2003	103	-0.399	RAN Composite: Colors, Objects	QRI Fluency	K-to-Grade 1
Bishop, 2003	103	-0.419	RAN Composite: Colors, Objects	QRI Comprehension	K-to-Grade 1
Bishop & League, 2006	79	-0.280	RAN Composite: Colors, Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 4
Bishop & League, 2006	79	-0.340	RAN Composite: Colors, Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 4
Bishop & League, 2006	79	-0.320	RAN Composite: Colors, Objects	QRI Fluency	K-to-Grade 4
Bishop & League, 2006	79	-0.150	RAN Composite: Colors, Objects	QRI Comprehension	K-to-Grade 4
Burgoyne et al., 2019					
Burgoyne et al., 2019	468	-0.340	Colors	YARC Early Word Reading	K-to-Grade 1
Burgoyne et al., 2019	468	-0.390	Objects	YARC Early Word Reading	K-to-Grade 1
Burgoyne et al., 2019	468	-0.340	RAN Composite: Colors	YARC Early Word Reading	K-to-Grade 1
Burgoyne et al., 2019	468	-0.420	Objects	YARC Early Word Reading	K-to-Grade 1
Burgoyne et al., 2019	468	-0.400	Colors	Single Word Reading Test	K-to-Grade 1
Burgoyne et al., 2019	468	-0.430	Objects	Single Word Reading Test	K-to-Grade 1
Burgoyne et al., 2019	468	-0.390	RAN Composite: Colors	Single Word Reading Test	K-to-Grade 1
Burgoyne et al., 2019	468	-0.460	Objects	Single Word Reading Test	K-to-Grade 1
Burgoyne et al., 2019	468	-0.390	Colors	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Burgoyne et al., 2019	468	-0.410	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Burgoyne et al., 2019	468	-0.380	RAN Composite: Colors	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Burgoyne et al., 2019	468	-0.440	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1

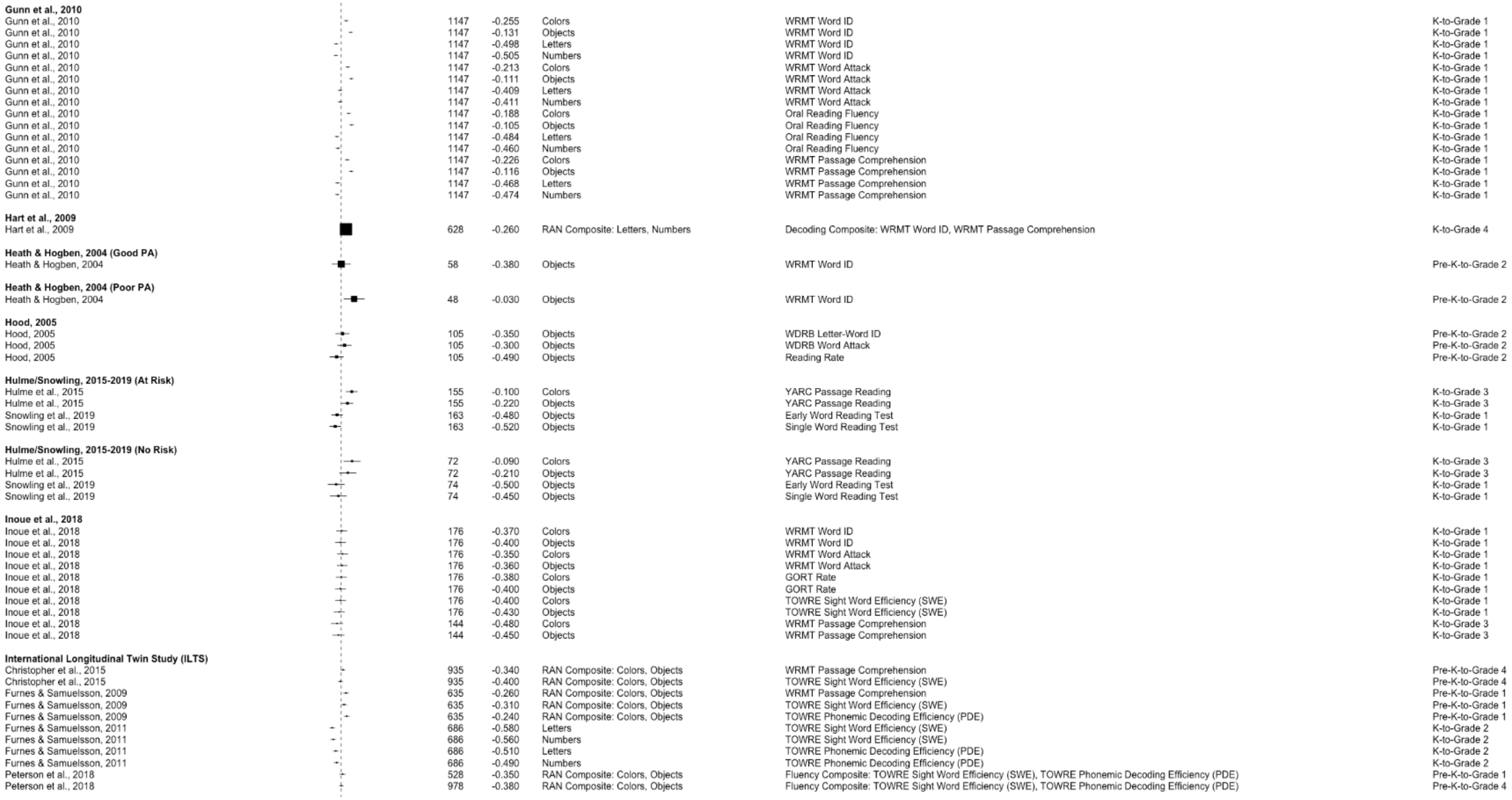
Figure 2 cont. Forest Plot of RAN-Reading Correlations

Caravolas et al., 2012-2019					
Caravolas et al., 2013	185	-0.400	Colors	Unstandardized Picture-Word Matching Task	Pre-K-to-Grade 1
Caravolas et al., 2013	185	-0.440	Objects	Unstandardized Picture-Word Matching Task	Pre-K-to-Grade 1
Caravolas et al., 2019	179	-0.346	Colors	Gates-MacGinitie Reading Test	Pre-K-to-Grade 2
Caravolas et al., 2019	179	-0.383	Objects	Gates-MacGinitie Reading Test	Pre-K-to-Grade 2
Cardoso-Martins & Pennington, 2004 (High Risk)					
Cardoso-Martins & Pennington, 2004	58	-0.540	RAN Composite: Letters, Numbers	WJ Letter-Word ID	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.410	RAN Composite: Colors, Objects	WJ Letter-Word ID	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.490	RAN Composite: Letters, Numbers	WJ Passage Comprehension	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.340	RAN Composite: Colors, Objects	WJ Passage Comprehension	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.460	RAN Composite: Letters, Numbers	WJ Word Attack	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.370	RAN Composite: Colors, Objects	WJ Word Attack	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.600	RAN Composite: Letters, Numbers	GORT: Composite Fluency Score	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	58	-0.390	RAN Composite: Colors, Objects	GORT: Composite Fluency Score	K-to-Grade 2
Cardoso-Martins & Pennington, 2004 (Low Risk)					
Cardoso-Martins & Pennington, 2004	49	-0.470	RAN Composite: Letters, Numbers	WJ Letter-Word ID	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.220	RAN Composite: Colors, Objects	WJ Letter-Word ID	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.370	RAN Composite: Letters, Numbers	WJ Passage Comprehension	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.170	RAN Composite: Colors, Objects	WJ Passage Comprehension	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.460	RAN Composite: Letters, Numbers	WJ Word Attack	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.200	RAN Composite: Colors, Objects	WJ Word Attack	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.530	RAN Composite: Letters, Numbers	GORT: Composite Fluency Score	K-to-Grade 2
Cardoso-Martins & Pennington, 2004	49	-0.300	RAN Composite: Colors, Objects	GORT: Composite Fluency Score	K-to-Grade 2
Carroll et al., 2016					
Carroll et al., 2016	123	-0.337	Objects	British Abilities Scale (BAS)	Pre-K-to-Grade 2
Catts et al., 1999/Adlof et al., 2010					
Catts et al., 1999	593	-0.484	Colored Animals	Reading Comprehension Composite: WRMT Passage Comprehension, GORT Comprehension, Diagnostic Achievement Battery	K-to-Grade 2
Catts et al., 1999	593	-0.444	Colored Animals	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 2
Adlof et al., 2010	433	-0.520	Colored Animals	Reading Comprehension Composite: WRMT Passage Comprehension, GORT Comprehension, Diagnostic Achievement Battery	K-to-Grade 2
Catts, 1991-1993					
Catts, 1991	41	-0.520	Colors	WRMT Word ID	K-to-Grade 1
Catts, 1991	41	-0.570	Objects	WRMT Word ID	K-to-Grade 1
Catts, 1991	41	-0.580	Colored Animals	WRMT Word ID	K-to-Grade 1
Catts, 1991	41	-0.400	Colors	WRMT Word Attack	K-to-Grade 1
Catts, 1991	41	-0.460	Colored Animals	WRMT Word Attack	K-to-Grade 1
Catts, 1993	50	-0.480	Colors	WRMT Word ID	K-to-Grade 2
Catts, 1993	50	-0.540	Objects	WRMT Word ID	K-to-Grade 2
Catts, 1993	50	-0.510	Colored Animals	WRMT Word ID	K-to-Grade 2
Catts, 1993	50	-0.430	Colors	WRMT Word Attack	K-to-Grade 2
Catts, 1993	50	-0.560	Objects	WRMT Word Attack	K-to-Grade 2
Catts, 1993	50	-0.460	Colored Animals	WRMT Word Attack	K-to-Grade 2
Catts, 1993	50	-0.530	Colors	GORT Rate	K-to-Grade 2
Catts, 1993	50	-0.630	Objects	GORT Rate	K-to-Grade 2
Catts, 1993	50	-0.450	Colored Animals	GORT Rate	K-to-Grade 2
Catts, 1993	50	-0.400	Colors	GORT Passage Score	K-to-Grade 2
Catts, 1993	50	-0.460	Objects	GORT Passage Score	K-to-Grade 2
Catts, 1993	50	-0.350	Colored Animals	GORT Passage Score	K-to-Grade 2
Catts, 2015-2017					
Catts et al., 2016	264	-0.408	Objects	Decoding Composite: WRMT Word ID, TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE), WRMT Word Attack	K-to-Grade 2
Catts et al., 2016	264	-0.377	Objects	QRI Comprehension	K-to-Grade 3
Catts et al., 2016	264	-0.437	Objects	MAP	K-to-Grade 3
Catts et al., 2015	313	-0.335	Objects	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1
Catts et al., 2015	313	-0.379	Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Catts et al., 2015	313	-0.372	Objects	Florida Assessment of Instruction for Reading: Oral Reading Fluency	K-to-Grade 1
Catts et al., 2015	292	-0.400	Objects	Decoding Composite: WRMT Word ID, TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE), WRMT Word Attack	K-to-Grade 2
Catts et al., 2015	270	-0.540	Objects	Reading Comprehension Composite: The Reading Comprehension Measure, Measures of Academic Progress: Reading (MAP)	K-to-Grade 3
Catts et al., 2017	292	-0.365	Objects	WRMT Word ID	K-to-Grade 2
Catts et al., 2017	292	-0.330	Objects	WRMT Word ID	K-to-Grade 2
Catts et al., 2017	292	-0.266	Objects	WRMT Word Attack	K-to-Grade 2
Catts et al., 2017	292	-0.268	Objects	WRMT Word Attack	K-to-Grade 2
Catts et al., 2017	292	-0.499	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
Catts et al., 2017	292	-0.442	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
Catts et al., 2017	292	-0.391	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 2
Catts et al., 2017	292	-0.357	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 2

Figure 2 cont. Forest Plot of RAN-Reading Correlations

Cirino et al., 2018					
Cirino et al., 2018	193	-0.690	RAN Composite: Objects, Letters, Numbers	Fluency Composite: TOWRE Sight Word Efficiency (SWE), DIBELS Oral Reading Fluency	K-to-Grade 1
Cirino et al., 2018	193	-0.570	RAN Composite: Objects, Letters, Numbers	Reading Comprehension Composite: AIMSWeb Mazes, WJ Passage Comprehension	K-to-Grade 1
Cirino et al., 2018	193	-0.620	RAN Composite: Objects, Letters, Numbers	Decoding Composite: WRAT Reading, WJ Word Attack, WJ Letter-Word ID	K-to-Grade 1
Cronin & Carver, 1998					
Cronin & Carver, 1998	57	-0.370	Objects	WRMT Word ID	K-to-Grade 1
Cronin & Carver, 1998	57	-0.600	Letters	WRMT Word ID	K-to-Grade 1
Cronin & Carver, 1998	57	-0.470	Numbers	WRMT Word ID	K-to-Grade 1
Cronin & Carver, 1998	57	-0.380	Objects	WRMT Word Attack	K-to-Grade 1
Cronin & Carver, 1998	57	-0.570	Letters	WRMT Word Attack	K-to-Grade 1
Cronin & Carver, 1998	57	-0.510	Numbers	WRMT Word Attack	K-to-Grade 1
Cronin & Carver, 1998	57	-0.310	Objects	WRMT Passage Comprehension	K-to-Grade 1
Cronin & Carver, 1998	57	-0.570	Letters	WRMT Passage Comprehension	K-to-Grade 1
Cronin & Carver, 1998	57	-0.520	Numbers	WRMT Passage Comprehension	K-to-Grade 1
Cronin & Carver, 1998	57	-0.340	Objects	WRMT Word Comprehension	K-to-Grade 1
Cronin & Carver, 1998	57	-0.530	Letters	WRMT Word Comprehension	K-to-Grade 1
Cronin & Carver, 1998	57	-0.430	Numbers	WRMT Word Comprehension	K-to-Grade 1
Cronin, 2013					
Cronin, 2013	108	-0.550	Objects	WRMT Word ID	Pre-K-to-Grade 2
Cronin, 2013	108	-0.510	Objects	WRMT Word Attack	Pre-K-to-Grade 2
Cronin, 2013	108	-0.570	Objects	WRMT Passage Comprehension	Pre-K-to-Grade 2
Dally, 2006					
Dally, 2006	132	-0.450	RAN Composite: Colors, Objects, Numbers	WRMT Word Attack	K-to-Grade 2
Dally, 2006	132	-0.470	RAN Composite: Colors, Objects, Numbers	WRMT Passage Comprehension	K-to-Grade 2
Dally, 2006	132	-0.430	RAN Composite: Colors, Objects, Numbers	Word Recognition: Burns/Roe Informal Reading Inventory	K-to-Grade 2
Dittman, 2016					
Dittman, 2016	116	-0.320	RAN Composite: Colors, Objects	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1
Dittman, 2016	116	-0.440	RAN Composite: Colors, Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Ellis & Large, 1988					
Ellis & Large, 1988	40	-0.480	Colors	Schonell Reading	K-to-Grade 1
Evans et al., 2000					
Evans et al., 2000	64	-0.370	Colors	WRMT Passage Comprehension	K-to-Grade 2
Evans et al., 2006					
Evans et al., 2006	139	-0.390	Colors	WRMT Word ID	K-to-Grade 1
Felton, 1990-1992					
Felton & Brown, 1990	81	-0.300	Colors	WRMT Word ID	K-to-Grade 1
Felton & Brown, 1990	81	-0.270	Objects	WRMT Word ID	K-to-Grade 1
Felton & Brown, 1990	81	-0.370	Letters	WRMT Word ID	K-to-Grade 1
Felton & Brown, 1990	81	-0.390	Numbers	WRMT Word ID	K-to-Grade 1
Felton & Brown, 1990	81	-0.190	Colors	WRMT Word Attack	K-to-Grade 1
Felton & Brown, 1990	81	-0.240	Objects	WRMT Word Attack	K-to-Grade 1
Felton & Brown, 1990	81	-0.190	Letters	WRMT Word Attack	K-to-Grade 1
Felton & Brown, 1990	81	-0.240	Numbers	WRMT Word Attack	K-to-Grade 1
Felton, 1992	221	-0.390	Colors	California Achievement Test - Reading	K-to-Grade 3
Felton, 1992	221	-0.230	Objects	California Achievement Test - Reading	K-to-Grade 3
Felton, 1992	221	-0.360	Letters	California Achievement Test - Reading	K-to-Grade 3
Felton, 1992	221	-0.360	Numbers	California Achievement Test - Reading	K-to-Grade 3
Georgiou et al., 2006					
Georgiou et al., 2006	62	-0.210	Colors	WRMT Word ID	K-to-Grade 1
Georgiou et al., 2006	62	-0.020	Colors	WRMT Word Attack	K-to-Grade 1
Georgiou et al., 2006	62	0.190	Colors	GORT Rate	K-to-Grade 1
Georgiou et al., 2006	62	-0.290	Colors	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Georgiou et al., 2012					
Georgiou et al., 2012	82	-0.440	Colors	WRMT Word Attack	K-to-Grade 2
Georgiou et al., 2012	82	-0.380	Colors	GORT total time	K-to-Grade 2

Figure 2 cont. Forest Plot of RAN-Reading Correlations



Gunn et al., 2010					
Gunn et al., 2010	1147	-0.255	Colors	WRMT Word ID	K-to-Grade 1
Gunn et al., 2010	1147	-0.131	Objects	WRMT Word ID	K-to-Grade 1
Gunn et al., 2010	1147	-0.498	Letters	WRMT Word ID	K-to-Grade 1
Gunn et al., 2010	1147	-0.505	Numbers	WRMT Word ID	K-to-Grade 1
Gunn et al., 2010	1147	-0.213	Colors	WRMT Word Attack	K-to-Grade 1
Gunn et al., 2010	1147	-0.111	Objects	WRMT Word Attack	K-to-Grade 1
Gunn et al., 2010	1147	-0.409	Letters	WRMT Word Attack	K-to-Grade 1
Gunn et al., 2010	1147	-0.411	Numbers	WRMT Word Attack	K-to-Grade 1
Gunn et al., 2010	1147	-0.188	Colors	Oral Reading Fluency	K-to-Grade 1
Gunn et al., 2010	1147	-0.105	Objects	Oral Reading Fluency	K-to-Grade 1
Gunn et al., 2010	1147	-0.484	Letters	Oral Reading Fluency	K-to-Grade 1
Gunn et al., 2010	1147	-0.460	Numbers	Oral Reading Fluency	K-to-Grade 1
Gunn et al., 2010	1147	-0.226	Colors	WRMT Passage Comprehension	K-to-Grade 1
Gunn et al., 2010	1147	-0.116	Objects	WRMT Passage Comprehension	K-to-Grade 1
Gunn et al., 2010	1147	-0.468	Letters	WRMT Passage Comprehension	K-to-Grade 1
Gunn et al., 2010	1147	-0.474	Numbers	WRMT Passage Comprehension	K-to-Grade 1
Hart et al., 2009					
Hart et al., 2009	628	-0.260	RAN Composite: Letters, Numbers	Decoding Composite: WRMT Word ID, WRMT Passage Comprehension	K-to-Grade 4
Heath & Hogben, 2004 (Good PA)					
Heath & Hogben, 2004	58	-0.380	Objects	WRMT Word ID	Pre-K-to-Grade 2
Heath & Hogben, 2004 (Poor PA)					
Heath & Hogben, 2004	48	-0.030	Objects	WRMT Word ID	Pre-K-to-Grade 2
Hood, 2005					
Hood, 2005	105	-0.350	Objects	WDRB Letter-Word ID	Pre-K-to-Grade 2
Hood, 2005	105	-0.300	Objects	WDRB Word Attack	Pre-K-to-Grade 2
Hood, 2005	105	-0.490	Objects	Reading Rate	Pre-K-to-Grade 2
Hulme/Snowling, 2015-2019 (At Risk)					
Hulme et al., 2015	155	-0.100	Colors	YARC Passage Reading	K-to-Grade 3
Hulme et al., 2015	155	-0.220	Objects	YARC Passage Reading	K-to-Grade 3
Snowling et al., 2019	163	-0.480	Objects	Early Word Reading Test	K-to-Grade 1
Snowling et al., 2019	163	-0.520	Objects	Single Word Reading Test	K-to-Grade 1
Hulme/Snowling, 2015-2019 (No Risk)					
Hulme et al., 2015	72	-0.090	Colors	YARC Passage Reading	K-to-Grade 3
Hulme et al., 2015	72	-0.210	Objects	YARC Passage Reading	K-to-Grade 3
Snowling et al., 2019	74	-0.500	Objects	Early Word Reading Test	K-to-Grade 1
Snowling et al., 2019	74	-0.450	Objects	Single Word Reading Test	K-to-Grade 1
Inoue et al., 2018					
Inoue et al., 2018	176	-0.370	Colors	WRMT Word ID	K-to-Grade 1
Inoue et al., 2018	176	-0.400	Objects	WRMT Word ID	K-to-Grade 1
Inoue et al., 2018	176	-0.350	Colors	WRMT Word Attack	K-to-Grade 1
Inoue et al., 2018	176	-0.360	Objects	WRMT Word Attack	K-to-Grade 1
Inoue et al., 2018	176	-0.380	Colors	GORT Rate	K-to-Grade 1
Inoue et al., 2018	176	-0.400	Objects	GORT Rate	K-to-Grade 1
Inoue et al., 2018	176	-0.400	Colors	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Inoue et al., 2018	176	-0.430	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Inoue et al., 2018	144	-0.480	Colors	WRMT Passage Comprehension	K-to-Grade 3
Inoue et al., 2018	144	-0.450	Objects	WRMT Passage Comprehension	K-to-Grade 3
International Longitudinal Twin Study (ILTS)					
Christopher et al., 2015	935	-0.340	RAN Composite: Colors, Objects	WRMT Passage Comprehension	Pre-K-to-Grade 4
Christopher et al., 2015	935	-0.400	RAN Composite: Colors, Objects	TOWRE Sight Word Efficiency (SWE)	Pre-K-to-Grade 4
Furnes & Samuelsson, 2009	635	-0.260	RAN Composite: Colors, Objects	WRMT Passage Comprehension	Pre-K-to-Grade 1
Furnes & Samuelsson, 2009	635	-0.310	RAN Composite: Colors, Objects	TOWRE Sight Word Efficiency (SWE)	Pre-K-to-Grade 1
Furnes & Samuelsson, 2009	635	-0.240	RAN Composite: Colors, Objects	TOWRE Phonemic Decoding Efficiency (PDE)	Pre-K-to-Grade 1
Furnes & Samuelsson, 2011	686	-0.580	Letters	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
Furnes & Samuelsson, 2011	686	-0.560	Numbers	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
Furnes & Samuelsson, 2011	686	-0.510	Letters	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 2
Furnes & Samuelsson, 2011	686	-0.490	Numbers	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 2
Peterson et al., 2018	528	-0.350	RAN Composite: Colors, Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	Pre-K-to-Grade 1
Peterson et al., 2018	978	-0.380	RAN Composite: Colors, Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	Pre-K-to-Grade 4

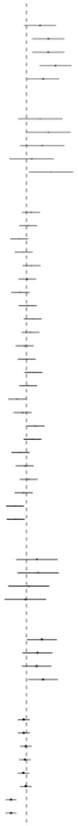
Figure 2 cont. Forest Plot of RAN-Reading Correlations

Kirby et al., 2003					
Kirby et al., 2003	99	-0.444	Colors	WRMT Word Attack	K-to-Grade 3
Kirby et al., 2003	99	-0.325	Objects	WRMT Word Attack	K-to-Grade 3
Kirby et al., 2003	99	-0.531	Colors	WRMT Word ID	K-to-Grade 3
Kirby et al., 2003	99	-0.392	Objects	WRMT Word ID	K-to-Grade 3
Kirby et al., 2003	99	-0.514	Colors	WRMT Passage Comprehension	K-to-Grade 3
Kirby et al., 2003	99	-0.374	Objects	WRMT Passage Comprehension	K-to-Grade 3
Kirby et al., 2011					
Kirby et al., 2011	142	-0.452	Colors	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 3
Kirby et al., 2011	142	-0.443	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 3
Kirby et al., 2011	142	-0.494	Colors	WRMT Passage Comprehension	K-to-Grade 3
Kirby et al., 2011	142	-0.447	Objects	WRMT Passage Comprehension	K-to-Grade 3
Kirby et al., 2011	142	-0.361	Colors	WRMT Word Attack	K-to-Grade 3
Kirby et al., 2011	142	-0.352	Objects	WRMT Word Attack	K-to-Grade 3
Kirby et al., 2011	142	-0.416	Colors	WRMT Word ID	K-to-Grade 3
Kirby et al., 2011	142	-0.405	Objects	WRMT Word ID	K-to-Grade 3
Lachance & Mazzocco, 2006					
Lachance & Mazzocco, 2006	224	-0.350	Colors	WJ Letter-Word ID	K-to-Grade 2
Lachance & Mazzocco, 2006	224	-0.380	Numbers	WJ Letter-Word ID	K-to-Grade 2
Lachance & Mazzocco, 2006	224	-0.240	Colors	WJ Word Attack	K-to-Grade 2
Lachance & Mazzocco, 2006	224	-0.300	Numbers	WJ Word Attack	K-to-Grade 2
Ledesma, 2002					
Ledesma, 2002	58	-0.250	Objects	WRMT Word ID	K-to-Grade 1
Ledesma, 2002	58	-0.670	Letters	WRMT Word ID	K-to-Grade 1
Ledesma, 2002	58	-0.210	Objects	WRMT Word Attack	K-to-Grade 1
Ledesma, 2002	58	-0.560	Letters	WRMT Word Attack	K-to-Grade 1
Ledesma, 2002	58	-0.310	Objects	WRMT Passage Comprehension	K-to-Grade 1
Ledesma, 2002	58	-0.460	Letters	WRMT Passage Comprehension	K-to-Grade 1
Ledesma, 2002	58	-0.300	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Ledesma, 2002	58	-0.700	Letters	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 1
Ledesma, 2002	58	-0.250	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Ledesma, 2002	58	-0.520	Letters	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Ledesma, 2002	58	-0.280	Objects	GORT Passage Score	K-to-Grade 1
Ledesma, 2002	58	-0.490	Letters	GORT Passage Score	K-to-Grade 1
Ledesma, 2002	58	-0.250	Objects	GORT Rate	K-to-Grade 1
Ledesma, 2002	58	-0.440	Letters	GORT Rate	K-to-Grade 1
Ledesma, 2002	58	-0.070	Objects	GORT Comprehension	K-to-Grade 1
Ledesma, 2002	58	-0.280	Letters	GORT Comprehension	K-to-Grade 1
Lewis et al., 2011					
Lewis et al., 2011	105	-0.270	Colors	WRMT Word ID	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.400	Objects	WRMT Word ID	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.560	Letters	WRMT Word ID	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.480	Numbers	WRMT Word ID	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.060	Colors	WRMT Word Attack	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.130	Objects	WRMT Word Attack	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.140	Letters	WRMT Word Attack	Pre-K-to-Grade 4
Lewis et al., 2011	105	-0.010	Numbers	WRMT Word Attack	Pre-K-to-Grade 4
LRRC & Chiu, 2018					
LRRC & Chiu, 2018	277	-0.340	RAN Composite: Colors, Objects	WRMT Passage Comprehension	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.222	RAN Composite: Colors, Objects	QRI Comprehension	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.329	RAN Composite: Colors, Objects	Gates-MacGinitie Reading Test	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.291	RAN Composite: Colors, Objects	WRMT Word ID	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.289	RAN Composite: Colors, Objects	WRMT Word Attack	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.378	RAN Composite: Colors, Objects	TOWRE Phonemic Decoding Efficiency (PDE)	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.322	RAN Composite: Colors, Objects	TOWRE Sight Word Efficiency (SWE)	Pre-K-to-Grade 3
LRRC & Chiu, 2018	277	-0.434	RAN Composite: Colors, Objects	Florida Assessment of Instruction for Reading: Oral Reading Fluency	Pre-K-to-Grade 3
Macdonald et al., 2013					
Macdonald et al., 2013	131	-0.315	RAN Composite: Colors, Objects	WRMT Word ID	K-to-Grade 1
Macdonald et al., 2013	131	-0.409	RAN Composite: Colors, Objects	WRMT Word Attack	K-to-Grade 1

Figure 2 cont. Forest Plot of RAN-Reading Correlations

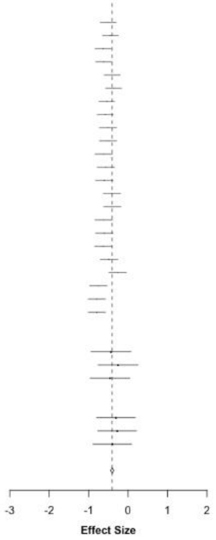
Mazzocco & Grimm, 2013					
Mazzocco & Grimm, 2013	188	-0.263	Colors	WRMT Letter-Word ID	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.324	Objects	WRMT Letter-Word ID	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.330	Letters	WRMT Letter-Word ID	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.333	Numbers	WRMT Letter-Word ID	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.173	Colors	WRMT Word Attack	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.240	Objects	WRMT Word Attack	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.321	Letters	WRMT Word Attack	K-to-Grade 2
Mazzocco & Grimm, 2013	188	-0.263	Numbers	WRMT Word Attack	K-to-Grade 2
McIlraith, 2018					
McIlraith, 2018	264	-0.372	RAN Composite: Colors, Objects	Decoding Composite: WRMT Word ID, WRMT Word Attack	Pre-K-to-Grade 1
McMillen, 2018					
McMillen, 2018	26	-0.240	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
McMillen, 2018	26	-0.120	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 2
Ozernov-Palchik et al., 2017					
Ozernov-Palchik et al., 2017	139	-0.132	Colors	WRMT Word ID	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.204	Objects	WRMT Word ID	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.259	Letters	WRMT Word ID	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.128	Colors	WRMT Word Attack	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.211	Objects	WRMT Word Attack	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.209	Letters	WRMT Word Attack	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.312	Colors	TOWRE Sight Word Efficiency (SWE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.391	Objects	TOWRE Sight Word Efficiency (SWE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.427	Letters	TOWRE Sight Word Efficiency (SWE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.288	Colors	TOWRE Phonemic Decoding Efficiency (PDE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.374	Objects	TOWRE Phonemic Decoding Efficiency (PDE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.377	Letters	TOWRE Phonemic Decoding Efficiency (PDE)	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.295	Colors	GORT Rate	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.383	Objects	GORT Rate	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.357	Letters	GORT Rate	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.324	Colors	GORT Accuracy	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.406	Objects	GORT Accuracy	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.385	Letters	GORT Accuracy	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.345	Colors	GORT Fluency	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.432	Objects	GORT Fluency	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.398	Letters	GORT Fluency	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.283	Colors	GORT Comprehension	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.296	Objects	GORT Comprehension	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.328	Letters	GORT Comprehension	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.345	Colors	GORT ORI	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.400	Objects	GORT ORI	Mixed-to-Grade 2
Ozernov-Palchik et al., 2017	139	-0.399	Letters	GORT ORI	Mixed-to-Grade 2
Parrila et al., 2004					
Parrila et al., 2004	102	-0.460	Colors	WRMT Word ID	K-to-Grade 2
Parrila et al., 2004	102	-0.500	Colors	WRMT Passage Comprehension	K-to-Grade 2
Petersen & Gillam, 2013					
Petersen & Gillam, 2013	16	-0.706	Colored Animals	WRMT Word ID	K-to-Grade 1
Petersen & Gillam, 2013	16	-0.663	Colored Animals	WRMT Passage Comprehension	K-to-Grade 1
Petersen & Gillam, 2013	16	-0.244	Colored Animals	DIBELS NWF	K-to-Grade 1
Petersen & Gillam, 2013	16	-0.382	Colored Animals	DIBELS Oral Reading Fluency	K-to-Grade 1
Scanlon & Vellutino, 1996					
Scanlon & Vellutino, 1996	1000	-0.320	Objects	WRMT Word ID	K-to-Grade 1
Scanlon & Vellutino, 1996	1000	-0.280	Objects	WRMT Word Attack	K-to-Grade 1
Schatschneider et al., 2004					
Schatschneider et al., 2004	189	-0.310	Objects	WJ Passage Comprehension	K-to-Grade 2
Schatschneider et al., 2004	189	-0.360	Letters	WJ Passage Comprehension	K-to-Grade 2
Schatschneider et al., 2004	189	-0.350	Objects	WJ Letter-Word ID	K-to-Grade 2
Schatschneider et al., 2004	189	-0.470	Letters	WJ Letter-Word ID	K-to-Grade 2
Schatschneider et al., 2004	189	-0.440	Objects	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2
Schatschneider et al., 2004	189	-0.550	Letters	TOWRE Sight Word Efficiency (SWE)	K-to-Grade 2

Figure 2 cont. Forest Plot of RAN-Reading Correlations



Smith & Locke, 2008 (Control Group)					
Smith et al., 2008	28	-0.072	RAN Composite: Colors, Objects	WRMT Word ID	Pre-K-to-Grade
Smith et al., 2008	28	0.134	RAN Composite: Colors, Objects	WRMT Word Attack	Pre-K-to-Grade
Smith et al., 2008	28	0.131	RAN Composite: Colors, Objects	WRMT Passage Comprehension	Pre-K-to-Grade
Smith et al., 2008	28	0.304	RAN Composite: Colors, Objects	GORT Passage Score	Pre-K-to-Grade
Smith et al., 2008	28	0.007	RAN Composite: Colors, Objects	GORT Comprehension	Pre-K-to-Grade
Smith & Locke, 2008 (Dyslexia Group)					
Smith et al., 2008	16	-0.063	RAN Composite: Colors, Objects	WRMT Word ID	Pre-K-to-Grade
Smith et al., 2008	16	0.134	RAN Composite: Colors, Objects	WRMT Word Attack	Pre-K-to-Grade
Smith et al., 2008	16	-0.018	RAN Composite: Colors, Objects	WRMT Passage Comprehension	Pre-K-to-Grade
Smith et al., 2008	16	-0.269	RAN Composite: Colors, Objects	GORT Passage Score	Pre-K-to-Grade
Smith et al., 2008	16	0.199	RAN Composite: Colors, Objects	GORT Comprehension	Pre-K-to-Grade
Uhry, 2002					
Uhry, 2002	86	-0.282	Colors	Silent Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.353	Objects	Silent Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.530	Letters	Silent Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.448	Numbers	Silent Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.274	Colors	Oral Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.367	Objects	Oral Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.505	Letters	Oral Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.362	Numbers	Oral Comprehension	K-to-Grade 2
Uhry, 2002	86	-0.249	Colors	Words in Text Passage	K-to-Grade 2
Uhry, 2002	86	-0.300	Objects	Words in Text Passage	K-to-Grade 2
Uhry, 2002	86	-0.421	Letters	Words in Text Passage	K-to-Grade 2
Uhry, 2002	86	-0.380	Numbers	Words in Text Passage	K-to-Grade 2
Uhry, 2002	86	-0.229	Colors	WRMT Word ID	K-to-Grade 2
Uhry, 2002	86	-0.350	Objects	WRMT Word ID	K-to-Grade 2
Uhry, 2002	86	-0.565	Letters	WRMT Word ID	K-to-Grade 2
Uhry, 2002	86	-0.463	Numbers	WRMT Word ID	K-to-Grade 2
Uhry, 2002	86	-0.183	Colors	WRMT Word Attack	K-to-Grade 2
Uhry, 2002	86	-0.253	Objects	WRMT Word Attack	K-to-Grade 2
Uhry, 2002	86	-0.503	Letters	WRMT Word Attack	K-to-Grade 2
Uhry, 2002	86	-0.420	Numbers	WRMT Word Attack	K-to-Grade 2
Uhry, 2002	86	-0.342	Colors	Words per Second	K-to-Grade 2
Uhry, 2002	86	-0.447	Objects	Words per Second	K-to-Grade 2
Uhry, 2002	86	-0.602	Letters	Words per Second	K-to-Grade 2
Uhry, 2002	86	-0.588	Numbers	Words per Second	K-to-Grade 2
Vadasy et al., 2008 (Control)					
Vadasy & Sanders, 2008	18	-0.147	Objects	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1
Vadasy & Sanders, 2008	18	-0.119	Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Vadasy & Sanders, 2008	18	-0.337	Objects	Oral Reading Fluency	K-to-Grade 1
Vadasy & Sanders, 2008	18	-0.411	Objects	WRMT Passage Comprehension	K-to-Grade 1
Vadasy et al., 2008 (Treatment)					
Vadasy & Sanders, 2008	32	-0.026	Objects	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1
Vadasy & Sanders, 2008	32	-0.133	Objects	Fluency Composite: TOWRE Sight Word Efficiency (SWE), TOWRE Phonemic Decoding Efficiency (PDE)	K-to-Grade 1
Vadasy & Sanders, 2008	32	-0.154	Objects	Oral Reading Fluency	K-to-Grade 1
Vadasy & Sanders, 2008	32	0.003	Objects	WRMT Passage Comprehension	K-to-Grade 1
Wagner, Torgesen, Rashotte					
Hecht et al., 2000	197	-0.440	Letters	WRMT Word ID	K-to-Grade 3
Hecht et al., 2000	197	-0.440	Numbers	WRMT Word ID	K-to-Grade 3
Hecht et al., 2000	197	-0.400	Letters	WRMT Word Attack	K-to-Grade 3
Hecht et al., 2000	197	-0.420	Numbers	WRMT Word Attack	K-to-Grade 3
Hecht et al., 2000	197	-0.450	Letters	WRMT Passage Comprehension	K-to-Grade 3
Hecht et al., 2000	197	-0.400	Numbers	WRMT Passage Comprehension	K-to-Grade 3
Torgesen et al., 1994	244	-0.660	RAN Composite: Letters, Numbers	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1
Wagner et al., 1994	244	-0.660	RAN Composite: Letters, Numbers	Decoding Composite: WRMT Word ID, WRMT Word Attack	K-to-Grade 1

Figure 2 cont. Forest Plot of RAN-Reading Correlations



Wolf, 1984-1986					
Wolf, 1984	98	-0.467	Colors	Gates-MacGinitie Reading Test	K-to-Grade 1
Wolf, 1984	98	-0.421	Objects	Gates-MacGinitie Reading Test	K-to-Grade 1
Wolf, 1984	98	-0.562	Letters	Gates-MacGinitie Reading Test	K-to-Grade 1
Wolf, 1984	98	-0.550	Numbers	Gates-MacGinitie Reading Test	K-to-Grade 1
Wolf, 1984	98	-0.385	Colors	GORT Passage Score	K-to-Grade 1
Wolf, 1984	98	-0.353	Objects	GORT Passage Score	K-to-Grade 1
Wolf, 1984	98	-0.497	Letters	GORT Passage Score	K-to-Grade 1
Wolf, 1984	98	-0.526	Numbers	GORT Passage Score	K-to-Grade 1
Wolf et al., 1986	83	-0.470	Colors	Gates-MacGinitie Reading Test	K-to-Grade 2
Wolf et al., 1986	83	-0.470	Objects	Gates-MacGinitie Reading Test	K-to-Grade 2
Wolf et al., 1986	83	-0.550	Letters	Gates-MacGinitie Reading Test	K-to-Grade 2
Wolf et al., 1986	83	-0.510	Numbers	Gates-MacGinitie Reading Test	K-to-Grade 2
Wolf et al., 1986	83	-0.540	RAN Composite: Colors, Objects, Letters, Numbers	Gates-MacGinitie Reading Test	K-to-Grade 2
Wolf et al., 1986	83	-0.390	Colors	GORT Passage Score	K-to-Grade 2
Wolf et al., 1986	83	-0.380	Objects	GORT Passage Score	K-to-Grade 2
Wolf et al., 1986	83	-0.550	Letters	GORT Passage Score	K-to-Grade 2
Wolf et al., 1986	83	-0.540	Numbers	GORT Passage Score	K-to-Grade 2
Wolf et al., 1986	83	-0.560	RAN Composite: Colors, Objects, Letters, Numbers	GORT Passage Score	K-to-Grade 2
Wolf et al., 1986	83	-0.450	Colors	Single Word Reading Task	K-to-Grade 2
Wolf et al., 1986	83	-0.260	Objects	Single Word Reading Task	K-to-Grade 2
Wolf et al., 1986	83	-0.640	Letters	Single Word Reading Task	K-to-Grade 2
Wolf et al., 1986	83	-0.660	Numbers	Single Word Reading Task	K-to-Grade 2
Wolf et al., 1986	83	-0.660	RAN Composite: Colors, Objects, Letters, Numbers	Single Word Reading Task	K-to-Grade 2
Wolter et al., 2011 (LI Group)					
Wolter et al., 2011	18	-0.410	Colored Animals	WRMT Word ID	K-to-Grade 4
Wolter et al., 2011	18	-0.250	Colored Animals	WRMT Word Attack	K-to-Grade 4
Wolter et al., 2011	18	-0.430	Colored Animals	WRMT Passage Comprehension	K-to-Grade 4
Wolter et al., 2011 (TL Group)					
Wolter et al., 2011	19	-0.300	Colored Animals	WRMT Word ID	K-to-Grade 4
Wolter et al., 2011	19	-0.270	Colored Animals	WRMT Word Attack	K-to-Grade 4
Wolter et al., 2011	19	-0.380	Colored Animals	WRMT Passage Comprehension	K-to-Grade 4

2.7 Supplemental Methods

2.7.1 Data Extraction: Whole Sample vs. Subgroupings

We prioritized extracting effect sizes from whole samples if they were available. This design consideration was also intended to minimize variability to focus on the utility of RAN as an early screener. However, there were many cases in which authors reported only data from subsamples. For example, Heath and Hogben (2004) report correlations separately for groups with Good PA and Poor PA.

2.7.2 Data Extraction: Multiple Timepoints

The only cases in which we collected multiple grade school timepoints were where the measures used at each timepoint differed considerably and did not overlap. For example, Badian (1994) collected only decoding measures in fall of Grade 1, and collected comprehension measures in the spring. In this case, two timepoints were extracted, but only so that all possible types of reading measures were included. When multiple outcome points were available, we coded and analyzed the one closest to the end of grade 2, as this is the period in which children typically develop automaticity in word reading (Chall, 1983; Wolf et al., 2000) and when dyslexia is commonly diagnosed in the US.

2.7.3 Data Extraction: Age and Grade

Grade was further specified as fall (July-December) versus spring (January-June) semesters when available. Fall versus spring specifications from Australia were flipped to match the US/UK/Canadian school year. UK samples were coded as kindergarten if they were called Year 1 and as Grade 1 if they were called Year 2; US/AUS/CAN samples were coded as the reported grade, unless otherwise specified in Supplemental Methods. In the cases that age was reported without grade, we used the following guidelines to derive a code for grade: mean age <60 months was coded as preschool fall, and mean age between 60 and 66 months was coded as preschool spring. This 6-month progression was used all the way through grade school. For example, a sample with a mean age of 100 months at follow-up would be coded as Grade 2 spring if grade was not reported. There was one case in which our general coding of grade/age was incongruent. For two Australian samples (Dally, 2006; Dittman, 2016), the initial timepoints both have an initial age of 67 months, but the samples are referred to as kindergarten and 1st grade, respectively. As a result, both are coded as Kindergarten Fall for the initial timepoint, which also corresponds to our age guidelines.

2.7.4 Excluded Articles Examples

Our inclusion and exclusion criteria were applied strictly, which resulted in excluding papers that are topic relevant, but did not quite fit our criteria. For example, Powell & Atkinson (2021) was a study conducted in the UK; its initial timepoint was reception year and its final timepoint was the summer of Year 1 (US Kindergarten). Due to our exclusion criteria being primarily based around grade rather than age (as many studies did not report age), this paper did not meet inclusion for final timepoint being in grade school.

An example relating to our definition of a RAN task is that tasks that measured speeded naming of the alphabet (e.g., Fuchs et al., 2001) were excluded, despite calling their tasks RAN or rapid letter naming. Several English-speaking samples were excluded due to being in French immersion schools where English is not the primary language of instruction (e.g., Jared et al., 2011).

2.7.5 Effect Size Extraction: Ambiguous Cases

There were a number of ambiguous cases in terms of coding effect sizes. For example, Bishop (2003) reports positive correlations using standard scores, as expected. Using the same sample, Bishop & League (2006) report positive correlations with raw time scores. As a large majority of our correlations and those reported in previous were negative when using raw scores, we decided to multiply Bishop & League's (2006) effect sizes by -1 (contrary to Hjetland et al., 2017, which reported a positive value for this study). Six papers did not report their scoring method. Of these six, Peterson et al. (2018), Wolf (1984), Caravolas et al., (2013) and Torgesen et al. (1994) all have other papers with overlapping data included in this meta-analysis and were set to match the data from the other papers; for Peterson (2018) and Torgesen et al. (1994), this meant multiplying the reported effect by -1. Carroll et al. (2016) and Petersen & Gillam (2013) did not report in their papers, but clarified through email correspondence.

In another case, Catts and colleagues (1991, 1993) used absolute values for their reporting even though they used raw time scores; thus, their effect sizes were multiplied by -1. Catts et al. (1999) and Adlof et al. (2010; same data used for these two papers) had positive correlations but with no mention of absolute values. These correlations were in strongly in line with those from Catts et al., (1991, 1993); as RAN was a highly relevant construct to these papers, it would've been highly unusual for the authors to experience a switch from r = -.4 to r = .4 and not mention the direction change. As a result, we decided to multiply effect sizes from Catts et al. (1999) and Adlof et al. (2010) by -1.

Lewis et al., (2011) states that standard scores were used for both RAN and reading, though all eight of their correlations are negative and some of their correlations are strongly negative (e.g., r = -.56 for WRID and RAN Letters). Correlations of this positive magnitude would be highly anomalous; as a result, we decided to use negative correlations.

2.7.6 Outlier Handling

There were 4 samples with outlier effect sizes (above 97.5 or below 2.5%ile); 3 of these 4 were from a thesis with n = 6 or 7 subjects. The only remaining outlier effect size was the control group from Smith & Locke (2008). All these samples were retained, as we used models that are robust to small samples, and there was no valid reason to exclude Smith & Locke's control group, as we retained their dyslexic group.

For total RAN items, two studies (Cirino et al., 2018; Lachance & Mazzocco, 2006) were flagged as outliers, as they used composite RAN measures, which summed to more than 200 items. These were removed from the moderator analysis on total RAN items. Three studies were flagged as outliers for unique RAN items; two of these studies (Kirby et al., 2003; Parrila et al., 2004) had 4 unique items and one had 20 unique items (Hood, 2005). The mean number of unique items was 5.84 (SD = 2.5); as a result, we decided to retain Kirby et al. (2003) and Parrila et al. (2004) but exclude Hood (2005) from this analysis as it was above the 97.5 percentile cutoff described above.

Two studies (Burgoyne et al., 2019; Ledesma, 2002) were flagged as outliers for measuring reading early (only assessed reading in Fall 1st grade); however, as we included other studies and effects with this timepoint, we retained these studies for all analyses. One study (Caroll et al., 2016) was flagged as an outlier for measuring RAN only in the fall of preschool; however, as we included other studies that measured in preschool but did not describe when exactly, we decided to retain this study for all analyses.

3. Auditory Processing and Reading Disability: A Systematic Review and Meta-Analysis

3.1 Introduction

Difficulty in learning to read can have many possible causes, and any given individual may struggle with one or many of the skills necessary to read accurately, fluently, and with comprehension. This multifactorial view of reading ability posits that multiple underlying skills, including phonological awareness (PA), letter knowledge, and rapid automatized naming (RAN) ability, predict reading outcomes, and that weaknesses in one skill may be able to be compensated with facility in another (Compton, 2020; O'Brien & Yeatman, 2020). These skills are mostly linguistic; however, children with reading disability[4] (RD) also struggle with skills that are not purely linguistic, such as auditory processing (Hämäläinen et al., 2013; Rosen, 2003). Auditory processing skills are often tested with sound stimuli that are difficult to discriminate or detect; for example, in a frequency discrimination task, an individual might be asked to indicate whether two similar sounds were the same or different frequencies (i.e., pitch). Other types of auditory processing skills include stream segregation (Helenius et al., 1999), beat detection (Goswami et al., 2002), speech-in-noise (Nittrouer et al., 2018), and rapid temporal order judgment tasks (Tallal, 1980). Any given individual with RD may perform worse (or better) than a typical peer on any constellation of auditory processing tasks, as different peripheral and central mechanisms are responsible for encoding each sound feature.

There is considerable overlap between children who struggle with auditory processing tasks (who may be diagnosed with auditory processing disorder, APD)[5] and children who have

---

[4] Here, we use the term RD to describe a primary deficit in reading accuracy, speed, or comprehension; this is a slightly broader term than developmental dyslexia, which is typically defined by deficits primarily at the word level, though the terms are sometimes used interchangeably (Peterson & Pennington, 2015).

[5] There is considerable debate in the field about the criteria for APD (see Chermak et al., 2018; Vermiglio, 2018). This is orthogonal to the current analyses because isolated, continuous auditory task scores are used, thus this debate will not be discussed here.

impairment in reading or language (Sharma et al., 2009). However, many children with RD have typical auditory processing abilities (Heath & Hogben, 2004; Ramus et al., 2003; Tallal, 1980), further reinforcing the notion that RD is a heterogeneous diagnostic category and leaving unanswered the question of how auditory processing and RD may be related.

Some previous studies suggested that auditory processing is directly causal to RD. Tallal (1980) hypothesized that children with RD cannot “consistently process” rapidly changing sounds, such as formant transitions. This rapid auditory temporal processing deficit model for reading difficulties presented by Tallal (1980; Tallal et al., 1998) spurred a number of studies looking a short vs. long inter-stimulus intervals (ISI), rapid formant transitions, and fast vs. slow amplitude or frequency modulation rates in individuals with RD (Marshall et al., 2001; Ramus et al., 2003; Wright & Conlon, 2009). Key to this hypothesis is that these small acoustic changes could result in cascading difficulties with perceiving the difference between phonemes and thus, the meaning of words. For example, vowel duration differences create a phoneme distinction in some languages (e.g., Finnish), so if a child has trouble consistently discriminating the duration of sounds, then they may have trouble with phonemic awareness for sounds or words that differ in duration. Similar patterns might be observed for speakers of Mandarin, for which frequency (tone/pitch) is phonemic, and so on.

The results of these studies about processing rapidly changing sounds were mixed, with some finding marked accuracy differences between RD and control children at fast intervals (Reed, 1989), but many others found impairment among poorer readers regardless of whether the stimuli were rapidly changing or short in duration (Amitay, Ben-Yehudah, et al., 2002; Marshall et al., 2001; Mody et al., 1997; Ramus et al., 2003). Further, the rapid auditory temporal processing hypothesis fails to explain a few key findings in the literature. First, the deficit in auditory processing in RD is not specific to rapidly changing stimuli, nor is it present in all rapid auditory tasks (Marshall et al., 2001; Protopapas, 2014; Rosen, 2003). For example, Marshall et al. (2001) observed no interaction between typical and RD group performance on the auditory task used by Tallal (1980) and ISI, indicating that RD children's deficits were not specific to rapid stimuli. Second, this hypothesis is challenged by weak clinical intervention results from a popular computerized program for training rapid auditory processing, Fast ForWord. The training showed no meta-analytic effect on remediating reading difficulties in randomized controlled trials (Strong et al., 2011). Finally, several early results that reported large effect sizes supporting the hypothesis have not been replicated. Specifically, according to this hypothesis, the severity of language or reading impairment and the severity of auditory impairment should be highly correlated. Tallal (1980) reported an extremely high Spearman correlation of $r = .81$ between auditory processing skills and score on a phonics test. Later studies didn't find such strong relationships (Witton et al., 1998), and within groups with a disorder, the relationship was essentially non-existent (Rosen, 2003).

Alternative hypotheses to try to explain the relationship between auditory processing deficits and reading deficits have been proposed. The imprecise temporal sampling (Goswami, 2011) and neural noise (Hancock et al., 2017) hypotheses discuss the role of neural oscillations and neuronal timing in affecting both sensory/perceptual and language systems. Other hypotheses implicate decision-making and executive functions as part of a multifactorial model of RD, both of which may impact psychophysical task performance (O'Brien & Yeatman, 2020). These non-causal hypotheses generally share the idea that shared neural architecture between reading and auditory processing skills explain why both may be impaired in RD.

Although the mechanism of auditory processing deficit hypotheses for RD is not clear, evidence points to auditory processing deficits in RD, on average (i.e., when comparing groups with versus without RD). Many qualitative reviews exist regarding auditory processing deficits in individuals with RD, but even the best qualitative reviews, such as the one completed by Hämäläinen and colleagues (2013), can fall short of identifying deficits in RD for various categories of auditory processing (e.g., duration perception, intensity perception). In order to draw conclusions about whether a meaningful effect was present, they relied on the number of significant results in a category. For example, intensity discrimination had mostly non-significant results (only 2 of 16 studies were significant at $p = .05$), and their conclusion was that intensity discrimination is not impaired in RD. The study reported that a mean weighted effect size (weighted by sample size) for that analysis as $d = .5$, which is a medium effect. Many of the original results were near $p = .05$, and the lower bounds of confidence intervals were often extremely close to 0. These facts are not factored into the interpretation, in favor of focus on the proportion of significant *p*-values. Because that review did not perform a true meta-analysis, the true effect and its confidence interval remain unknown. As a result, researchers could draw incorrect conclusions from this review, and a formal meta-analysis may reveal that this effect size is significant. Clinicians and school professionals are also left without clear guidelines on the relationship between auditory processing and RD which could inform decisions such as when to refer children with APD or RD for comprehensive evaluation.

The only extant meta-analysis related to this topic focused only on frequency discrimination impairment in RD, and potential moderators of this relationship such as reading impairment severity and psychophysical task design (Witton et al., 2020). This meta-analysis describes a large frequency discrimination impairment ($d = .76$; $p < .001$) in RD. Moderator analyses revealed that psychophysical task design and performance on phoneme deletion, a phonological awareness task, moderate the relationship between RD and frequency discrimination. This study is an important first step toward understanding the complex mechanisms shared by reading and pitch perception, but its narrow scope does not contextualize the findings alongside other auditory tasks. The present study advances knowledge by expanding scope to a variety of auditory processing measures and implementing needed improvements to the methodology, which will be described below.

The primary goal of the proposed study is to estimate the mean auditory processing impairment in RD as compared to typical reader groups, across multiple auditory domains. We accomplished this by updating and extending the studies by Hämäläinen et al. (2013) and Witton et al. (2020). Specifically, we made five key methodological and statistical improvements that are required in order to produce high-quality evidence. First, we used snowball searching to identify related articles using references and citations, which neither study did. Snowball searching is essential in order to identify all relevant effect sizes in the literature that should be included in analysis (Greenhalgh & Peacock, 2005). As an example, in a recent snowball search for a meta-analysis in our lab related to reading (McWeeny et al., in press), we initially screened 4,000 titles identified during database searching and found 91 relevant articles to include in analysis; we found an additional 39 relevant articles from the snowball search. Second, we used robust variance estimation (RVE) models to estimate effect sizes rather than sample-size based effect size weighting (Hedges et al., 2010). RVE models allow for multiple effect sizes from the same sample to be included in each analysis, so we will not have to take the mean of effect sizes within a given study, as was done in Witton et al. (2020). We also assessed the degree to which each analysis is powered, which is essential for meta-analytic moderator analyses (Hedges &

Pigott, 2004; Schmidt, 2017). Fourth, we described and analyzed measures of study quality and risk of bias, which allows for both descriptive data about the health of the literature and allows inference testing on whether study quality affects effect sizes. For example, we can test whether lower quality studies tend to find larger effect sizes. Each of these design considerations were made in concordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009; Page et al., 2021) and represent the gold-standard in meta-analysis research. These considerations will improve the quality of meta-analysis in our field at all steps of the process, from literature searching through analysis. Finally, as we took the same approach for frequency discrimination, duration discrimination, intensity discrimination, and gap detection, we are able to contextualize summary effect sizes across domains, which provides key advances over the existing single-domain meta-analysis and cross-domain qualitative analysis. This allowed us to address key theoretical questions; for example, many studies suggest that intensity discrimination can serve as a control for psychophysical task demands (e.g., Goswami et al., 2010); comparing these summary effect sizes for each domain will allow us to directly test the size of a frequency discrimination deficit as compared to an intensity discrimination deficit.

A secondary aim of this meta-analysis is to test the hypothesis that auditory processing impairments in RD are in part due to differing demands across psychophysical tasks used to measure auditory processing. This hypothesis predicts that different task designs (e.g., same-different vs. two-alternative forced choice; 2AFC) will yield different estimates for the average auditory processing impairment in RD. A meta-analysis of frequency discrimination in RD (Witton et al., 2020) found evidence of task design moderating effect sizes; specifically, same-different tasks yielded smaller effect sizes (d = 0.40, SD = 0.27, n = 4) than either AXB (d =

1.10, SD = 0.79, n = 11) or 2AFC tasks (*d* = 0.92, SD = 0.39, n = 9). We predicted that this moderation effect will apply to any auditory domain in which discrimination is measured, as tasks that have greater executive function (EF) demands are likely to result in lower scores in RD samples, due to the high co-occurrence of EF difficulties and ADHD with RD (Germanò et al., 2010; Lonergan et al., 2019). This analysis should provide additional information related to the mechanism by which RD is associated with poorer auditory processing.

3.2 Methods

3.2.1 Overview

We examined the relationship between reading impairment and auditory processing; as a result, we encountered designs that analyzed reading from a disorder vs. typical development perspective (categorial) and ones that measure reading ability as a continuum within typical readers or across varied reading ability (correlational). Based on the literature review from Hämäläinen et al. (2013) and pilot searches, most studies treated reading categorically by comparing groups; as a result, the primary focus of the analyses are categorical. The heterogeneity present in correlational analyses, such as whether the whole sample is analyzed together, or subgroups are analyzed separately, as well as the variety within each sample's composition (e.g., typical sample, language disorder, or APD sample), precludes meta-analyzing these correlations.

3.2.2 Study Inclusion Criteria

Studies included in the quantitative analysis had four characteristics: 1) a group of children with RD (see below for definition), 2) a group of control/typically developing children, 3) a relevant behavioral auditory processing task (see details in Auditory Processing Tasks, below), and 4) calculable standardized mean difference. No neural data is included in this meta-analysis (for a meta-analysis of the mismatch negativity (MMN), see Gu & Bi, 2020). No study was excluded because its effect size is an outlier, as was done in Hämäläinen et al. (2013); unless the study should be excluded for poor study quality, then using these effect sizes will most accurately represent the true effect size in the population. In order to assess whether extreme outliers exert effects on our main questions of interests, models with extreme outliers (≥2 SD) both included and excluded are presented.

For RD definition, we included studies that use alternate terms such as dyslexia or “poor readers,” given the heterogeneity of definitions used in both research and practice that use different types of reading measures and scores for inclusion and for formal diagnosis. We did not exclude a study if the sample has comorbidities such as ADHD, or other language impairment, but we did exclude the study if the sample is designed for autism spectrum disorder, schizophrenia, or chromosomal disorder (e.g., Fragile X, Down Syndrome). If the study does not list their participant exclusion list, they may thus include individuals with ASD incidentally; we included these studies. As mentioned above, we chose to use the broader term reading disability, as the term dyslexia is sometimes understood as a deficit only at the single word level with no other co-occurring deficits and typical IQ. Participants’ reading skills need not be reported in the paper; however, study quality measures (see Appendix 1) serve to mark studies that only use participant report or purported diagnosis for reading ability distinctions as lower quality. We also required that children have begun formal reading instruction and thus would be able to be diagnosed with RD, which diverges from the Hämäläinen et al. (2013) study approach. Inclusion of pre-readers dilutes the analysis of the relationship between auditory processing and RD, as approximately 50% children who are identified as at-risk for reading problems based on family history of reading problems will not go on to develop RD (Puolakanaho et al., 2007).

3.2.3 Auditory Processing Tasks

Auditory processing tasks that are analyzed here comprise four primary domains: frequency discrimination, duration discrimination, intensity discrimination, and gap detection. We chose these domains because they represent basic acoustic features (i.e., frequency, duration, and intensity), with gap detection being an additional way to test duration processing. We also restricted our tasks only to thresholds in the relevant domains, as was done in Witton et al. (2020). We also further restricted our stimuli to only tones/tone pips or noise.

3.2.3.1 Auditory Processing Task Descriptions and Analytic Considerations

Definitions use to guide inclusion by domain are described below.

3.2.3.1.1 Frequency. Any task that assesses an individual's ability to tell the difference between two frequencies and for which a threshold (in Hz) was considered a frequency discrimination task. We included all frequency discrimination thresholds in the main analysis, regardless of the frequency measured. This decision is motivated by comparison to previous literature (Witton et al., 2020), despite pitch being encoded by primarily different mechanisms (e.g., cochlear place, timing) above and below ~4kHz. Thresholds must be reported in Hz; some authors refer to auditory repetition tasks with varying inter-stimulus intervals (ISIs; e.g., Tallal, 1980) as frequency discrimination, but thresholds from these tasks are reported in ms.

3.2.3.1.2 Intensity. Any task that assesses an individual's ability to tell the difference between two intensities and for which a threshold (in dB) is derived was considered a duration discrimination task.

3.2.3.1.3 Duration. Any task that assesses an individual's ability to tell the difference between two durations and for which a threshold (in ms) is derived, was considered a duration discrimination task.

3.2.3.1.4 Gap Detection. Any task in which a participant is asked whether they hear a period of silence (i.e., a gap) in a period of noise or a tone and a threshold (in ms) is derived from it, was considered a gap detection task.

3.2.3.2 Task Designs

There are several possible methods that can be used to estimate behavioral discrimination thresholds. Most use adaptive thresholds to estimate a percent detectable change. We categorized each task according to the guidelines used by Witton et al. (2020). *Two-alternative forced choice* (2AFC) tasks require participants to choose between two stimuli on the domain of interest (e.g., which tone was higher/lower, longer/shorter, or louder/quieter). *Same-different* tasks also use two stimuli, but only require the participant to respond whether the stimuli were the same or different (similar to the yes-no method). *Three-alternative forced choice* (3AFC) tasks require participants to select the “odd one out” from a group of three sounds. *AXB* tasks also use three sounds, but the middle sound is a fixed reference, and the listener chooses between the first and last sound on the domain of interest. *ABABA/AAAAA* tasks use 10 stimuli over two intervals and participants are asked which interval had two different sounds. A similar paradigm will play four sounds in two groups and participants will be asked which interval (1 or 2) had different tones; authors differ in their description of this task, which we called two interval, two-alternative forced choice (2I-2AFC). *Gap detection* tasks have an additional design, in which the participant is asked if they heard two sounds or one. These “fusion” tasks only exist for gap detection thresholds. Finally, oddball paradigms with fixed “levels” are occasionally used to estimate a threshold. These tasks were included as long as they had a sufficient description of their methodology used to estimate the threshold.

3.2.4 Procedure

3.2.4.1 Data Collection

To begin, we conducted a snowball search (i.e., a backward and forward search, and subsequent searches from results) using references and citations of the 61 included studies from the Hämäläinen et al. (2013) study. To conduct this snowball search, we used Microsoft Academic Graph (Wang et al., 2019), which is a database that tracks connections between publications (peer-reviewed papers and other scientific products such as dissertations and theses), such that every backward reference is also a forward citation. Microsoft Academic has higher citation counts than Scopus or Web of Science, with a high number of unique items (Harzing & Alakangas, 2017). This allowed us to efficiently identify studies that have been published since the search for their review was completed in 2010. Newly identified articles were also snowball searched.

3.2.4.2 Abstract and Title Screening

Titles of papers from the search were reviewed individually by the first author, who has expertise in the relevant constructs and meta-analysis methods; titles that were deemed to be clearly irrelevant were screened out. Potentially relevant abstracts were then each reviewed by two different screeners using a checklist of inclusion criteria; consensus was reached in all cases of conflict. Articles with relevant abstracts were then full text screened by two independent coders. Agreement for full-text inclusion was 86.9%; consensus was reached in all cases of conflict. A PRISMA flow diagram describing the search and screening procedure is presented in **Figure 3.**

3.2.4.3 Data Extraction

For each of the studies meeting the inclusion criteria listed above, two independent coders extracted information from each paper in four domains: 1) sample characteristics, 2) reading measures or groups, 3) auditory processing tasks, and 4) study quality. Sample characteristics include age, demographics (i.e., SES, race, language, etc.), and how the RD category was defined (e.g., were they all individuals with a formal diagnosis, or were they tested and scored below a cutoff?). We included studies of all languages and demographics. The specific reading measure(s) used were extracted and categorized as single word or connected text reading that is rate-based, accuracy-based, both rate and accuracy-based, or a task of reading comprehension, as these constructs provide good coverage of the types of reading measures used in the literature across countries/orthographies. Information relating to the auditory processing tasks include what auditory domain was being tested (frequency, duration, and intensity discrimination, as well as gap detection) and the stimuli characteristics.

Study quality was assessed for each study using the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies[6]. Although many meta-analyses use a sum score of different quality metrics, these types of scores are not recommended for meta-regression (Shamliyan et al., 2010; Whiting et al., 2005). Instead, each study was categorized as good, fair, or poor according to the NIH tool independently by two coders and consensus was met in all cases of conflict. These categories were then used as categorical variables in meta-regression.

For extracting effect sizes, we coded standardized mean difference (SMD) measures, whether they are reported as separate sample means and SDs or if they are reported as an effect size, typically Cohen's *d*. For a study to be included, it needed to have either a reported SMD and corresponding variance, or group means and SDs.

3.2.5 Analytic Plan

---

[6] The study quality tool was tailored from its original version to better fit the study designs and constructs present in the literature. The original tool and the modified tool are presented in Supplemental Materials.

Main analyses include descriptive statistics (e.g., demographics, stimuli characteristics), main effects, and within- and between-study bias analyses for each of the four categories of auditory processing skills. Only select moderator analyses (i.e., subgroup analyses and meta-regression) with large anticipated effect sizes are included in the main planned analyses, as moderator analyses have much lower power than main meta-analyses because they depend on the number of studies in each category (Hedges & Pigott, 2004; Schmidt, 2017). Questions such as developmental trajectories and differences across languages have strong theoretical implications but cannot be tested with adequate power given the size of the extant literature.

All standard mean difference (SMD) effect sizes were transformed to Hedges' *g*, which corrects for small sample sizes. All statistical analyses were conducted in R, using *robumeta* for statistical modeling and *metafor* for effect size calculation and auxiliary functions (e.g., generating funnel plots) (Fisher et al., 2017; Hedges et al., 2010; R Core Team, 2013). *Robumeta* uses robust variance estimation (RVE) models, which allow for correlated effects within a study, maximizing data retention. Intercept-only models were generated for each of the auditory domain categories listed above.

To test for study quality bias, we used the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies as described above. This allowed us to test whether high quality studies have systematically smaller or larger effect sizes. We tested this across all studies rather than within each auditory category, due to low power. To test for funnel plot asymmetry, which is indicative of publication or reporting bias, we used a technique that allows for multiple effect sizes per category. Traditional methods for examining funnel plot asymmetry, such as Egger's Regression or trim-and-fill analyses, only accommodate one effect size per study or category. Recently, these traditional methods have been expanded to correlated effects models with “sandwich” estimators (Rodgers & Pustejovsky, 2020). We therefore used an “Egger’s Sandwich Regression” to test for funnel plot asymmetry.

3.2.6 Power Analyses

Using the SMD effect size estimates reported by Hämäläinen et al., 2013, we calculated power at $\alpha = .05$ for each of the categories they analyzed. This provided us with a strong estimate of the likely effect size, making our power analyses more accurate. With a mean $n = 22$ per group across all studies, assuming moderate heterogeneity in a random effects model, we calculated power for effect sizes, listed in **Table 6**, using the method described in Valentine et al. (2010). Power for all analyses were extremely high (≥.99), given the medium and large effect sizes present and the number of studies found during pilot searching.

At present, there is no clear consensus on calculating power for moderator analyses for RVE models. To calculate power for task design as a moderator, we used the *metapower* package in R (Griffin, 2020) assuming moderate heterogeneity ($I^2 = 50\%$), a minimum k = 20 per discrimination category, a mean $n = 22$ per group, and the effect sizes present in Witton et al. (2020; $d_1 = .4$, $d_2 = 1.1$). Power for this analysis was .94; however, the 20 studies will not fit into two categories as there are many possible task designs (e.g., 2AFC, AXB, 3AFC, same-different, etc.). Power may decrease as a result of the 20 studies not being evenly distributed between two categories. We approached this in two ways; we only included a task design category in the moderator analyses if it had at least $N = 5$ samples, and we reported the *a priori* power alongside the analysis in the Stage 2 manuscript, as has been done in other meta-analyses (e.g., Araújo et al., 2015).

For bias analyses, we pooled studies across all categories, as the primary question is about bias rather than auditory processing. Assuming a moderate effect of study quality (d =

±.35; that is, .35 between poor and average quality and .35 between average and high quality) and moderate heterogeneity (I^2 = 50%) in a random-effects model, we are powered at .96 assuming we have k = 77 studies, as estimated from pilot searching. For Egger's Regression, we used a standard regression power calculation. There would need to be 70 studies included to be adequately (0.90) powered to find a moderate effect (f^2 = .15). We believe that this is highly feasible given the snowball search that will be conducted. If the analysis is not powered at .9 or greater, we will report this fact in the manuscript.

3.3 Results

3.3.1 Main Analyses

3.3.1.1 Descriptive Statistics

The analytic sample (n = 3,545) was drawn from 63 independent samples across 65 reports that assessed at least one of the four auditory task categories. A majority of the participants (n = 2,206) were children under 12 years old, whereas adults (>18 years old; n = 1,045), adolescents ($12 \leq x \leq 18$ years; n = 253), and combined child/adolescent samples (n = 82) comprised the remainder of the sample[7]. A majority of participants (n = 2,003) spoke English, with Hebrew (n = 558), Chinese (n = 230), German (n = 182), Greek (n = 181), Dutch (n = 166) participants, and other languages (Portuguese, Spanish, Finnish, and French; combined n = 225) comprising a sizable minority.

3.3.1.2 Intercept-Only Models

To estimate the average impairment in each of the four auditory task categories measured, we ran intercept-only RVE models. Significant deficits were found in each auditory

---

[7] The *n*s of each age group do not add to the full sample because Goswami et al., 2010 and Thomson & Goswami, 2008 are part of the same longitudinal study, in which the subjects were children at the initial timepoint and are adolescents by the final reported timepoint. These participants are counted twice only in descriptive statistics and are considered correlated effects in the meta-analyses.

category at an alpha level of $p = .01$. The forest plots for each auditory task category are presented in **Figures 4-7**. Full model results are present in **Table 7**, and model results with and without outliers are presented in **Table 10.** Heterogeneity was moderate to high for each category (all $I^2 \geq .48$.) with and without outliers excluded. Model results for frequency discrimination ($g = .79$), duration discrimination ($g = .80$), and intensity discrimination ($g = .60$) changed minimally with the exclusion of outliers. However, the model for gap detection with a high-end outlier excluded yielded a smaller effect size estimate ($g = .80$ as opposed to $g = 1.05$), less heterogeneity ($I^2 = 75.20$ and $\tau^2 = .35$ as opposed to $I^2 = 89.13$ and $\tau^2 = .97$), and a larger t-value despite a smaller effect size estimate ($t = 4.89$ as opposed to $t = 3.65$). Because we consider the data excluding the outlier to be a more accurate representation of the relationship, we will primarily use the model with outliers excluded in the interpretation of our gap detection analyses.

3.3.1.3 Task Design Analyses

In order to test whether specific task designs (e.g., 2AFC, AXB) yielded larger effects than others, we ran RVE meta-regression models for each auditory task category. In order to be analyzed, a task design needed to be present in 5 or more samples. The number of task designs analyzed thus varied for the different auditory task categories; 3 task designs were included in the frequency discrimination analysis, two task designs were included in each of the duration discrimination analysis, the intensity discrimination analysis, and the gap detection analysis. No significant moderating effect of task design was found for any task. *A priori* power as registered was .94 for each analysis; however, many of our assumptions were incorrect, and these power values were inflated. Specifically, one of the key conditions, same-different designs did not reach the N = 5 studies threshold to be included and calculating power with its corresponding effect size of $g = .4$ would inflate power greatly. To address this concern, we created an adjusted power calculation that used our data's *N*, *n* (*n* = 29 per group for frequency and duration discrimination, *n* = 26 per group for intensity discrimination, and *n* = 20 per group for gap detection), and I^2 values (rounded to the nearest quartile to reflect the benchmarks present in *metapower*), with the effect sizes from Witton et al. (2020). For example, for duration discrimination, we calculated adjusted power using the corresponding effect sizes from Witton et al. (2020) for the two designs that met the *N* = 5 studies threshold ($g_{2AFC} = .9$, $g_{AXB} = 1.1$). As gap detection designs (i.e., fusion and the gaps-in-noise test) were not present in the Witton et al. meta-analysis, we chose a moderate difference ($\Delta g = .3$) between the two conditions to match the intensity discrimination power calculation ($g_{ABABA} = .6$ and $g_{2AFC} = .9$). Full model results are present in **Table 8.**

3.3.1.4 Bias Analyses

3.3.1.4.1 Within-Study Bias, Study Quality.

Of the 63 studies, 4 were good quality, 43 were fair quality, and 16 were poor quality. The criteria that prevented most fair quality studies from being coded as good was the lack of reported reliability for both auditory processing tasks and for reading tasks and the lack of reported power analyses. To test whether study quality systematically biased effect sizes, we collapsed across auditory domains to increase power. No significant effect of study quality was found, though good quality studies had marginally smaller effect sizes ($\Delta g = -.48$; $p = .07$) than fair studies. Poor studies had slightly larger effect sizes than fair studies, but the effect was not significant ($\Delta g = .28$; $p = .17$). Full model results are presented in **Table 9**.

3.3.1.4.2 Between-Study Bias, Publication Bias.

To test whether publication bias exists in the literature, we performed a one-tailed Egger's sandwich regression across all auditory task categories (Rodgers & Pustejovsky, 2020).

This analysis revealed a significant effect of sampling bias (i.e., standard error) on effect sizes, indicating that studies with larger sampling bias had systematically larger effect sizes, which is characteristic of publication bias. The degrees of freedom for the regressor, sampling variance, was notably low ($df < 4$) due to a few high-end outliers (i.e., large sampling variance due to small sample size). We offer caution in the interpretation of significant results in the presence of $df < 4$, as type I error is inflated in the t-distribution at low degrees of freedom. Full model results are presented in **Table 4**.

3.4 Discussion

The results of these meta-analyses mark an important step forward for the field of auditory processing and RD, documenting a large, non-linguistic, multiple-domain auditory processing impairment for the first time. Analyses revealed a significant impairment for individuals with RD as compared with typical readers in auditory domains of frequency (g = .79), duration (g = .80), and intensity discrimination (g = .60), as well as gap detection (g = .80). These results are broadly consistent with previous reviews that documented deficits in frequency discrimination (Hämäläinen et al., 2013; Witton et al., 2020), as well as duration discrimination and gap detection (Hämäläinen et al., 2013). However, in contrast to past reviews, we also found a significant impairment in intensity discrimination, similar in magnitude to the other domains analyzed.

The effect sizes presented here can be reasonably compared to the weighted effect sizes from Hämäläinen et al. (2013). Despite meaningful differences in inclusion criteria, namely the restriction to behavioral thresholds from non-linguistic stimuli in only four auditory task categories, our effect sizes for each auditory task category were highly similar. Effect sizes were within .1 SMD for frequency discrimination (g = .79 in this analysis, g = .7 in Hämäläinen et al. and g = .76 in Witton et al.), duration discrimination (g = .80 in this analysis and g = .9 in

Hämäläinen et al.), and intensity discrimination (*g* = .60 in this analysis and *g* = .5 in Hämäläinen et al.), and within .2 SMDs for gap detection (*g* = .80 in this analysis and *g* = .6 in Hämäläinen et al.). The present analysis has a distinct advantage in also reporting confidence intervals for each category, describing not only a mean effect but a spread of effect sizes around each auditory task category mean. These confidence intervals ranged considerably in size among the auditory task categories, partly due to how many subjects and studies were included in each analysis, and partly due to the spread of effects among the included studies. The smallest confidence interval among the four categories was intensity discrimination [.44 .76] and the largest was gap detection [.45 1.15]. In sum, these results suggest that individuals with RD are on average, poorer than age-matched typical-reader peers in multiple domains of non-linguistic auditory processing, reflecting even broader deficits than previous reviews have suggested.

The range of effects for auditory task category discussed above reflects considerable heterogeneity in the literature. To illustrate some potential sources of this heterogeneity, we present an example of how much effect sizes can vary within the same study and sample. In this example, Thomson et al. (2013) tested intensity discrimination in the first year of a longitudinal study when children were age 9;8 (years;months). In that assessment, the RD children (n = 33) performed extremely poorly (*g* = 1.88) relative to controls (n = 11) in a 2AFC task (29.25 dB standard). Thomson and Goswami (2008) used a subset of these participants' data from one year later (age 10;8) to test intensity discrimination in an AAAAA/ABABA task. There, the RD group (n = 25) had no intensity discrimination deficit (75dB standard) and outperformed the controls (n = 23; *g* = -.14). There are a number of possibilities for this large effect size change between the two studies: differences in task design, stimulus properties, developmental changes, selection and subsetting bias, selective reporting, low reliability of intensity discrimination thresholds in children, or any combination of these.

In an attempt to explain some of this heterogeneity across studies, we tested whether specific task designs yielded larger effects across domains, as was found for frequency discrimination in Witton et al. (2020). We found no differences for any of the task designs (e.g., whether task designs that required different processing, such as higher working memory demands, led to greater group differences) in any of the auditory task categories. Due to our criteria of needing 5 samples to include a given design, neither the "same-different" design nor 2I-2AFC were not included in any of our moderator analysis. We initially calculated moderator power under the assumption that same-different designs would be included and accordingly chose large *a priori* effect sizes (d_1 = .4, d_2 = 1.1), as was justifiable given the large differences for same-different and 2I-2AFC tasks as opposed to 2AFC and AXB tasks present in Witton et al. (2020). The *adjusted* power calculation present in **Table 3**, which we believe to be the most accurate representation of the analyses' power, is considerably lower than the *a priori* power calculation. Accordingly, the interpretation of these findings is limited in that the null effects described are unsurprising given the adjusted powered.

3.4.1 Risk of Bias Within and Between Studies

Study quality is a key consideration in assessing a literature's risk of bias, with more standardized, transparent, and replicable methodology decreasing the risk of bias within each study. Using the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, we marked only 6% of studies as "good," 68% as "fair," and 25% as "poor." The primary study quality dimensions on which studies differed were the clarity and strength of the RD diagnostic inclusion criteria, whether any framework for handling comorbid disorders such as ADHD and DLD/SLI was given, and the description of the relevant auditory processing tasks. Though there were varying strengths and weakness in these papers, no study reported *a priori* power analyses, and only three studies (Georgiou et al., 2010a; Heath et al., 2006a; Papadopoulos et al., 2012) measured or reported reliability for non-standardized behavioral tasks, thus reducing the overall variability of study quality. No systematic bias for low quality or high-quality articles were found, though the lack of variability likely contributed to finding no effect of study quality.

Finally, we found an effect of publication bias in the included studies, contrary to the findings of Hämäläinen et al. (2013). The comparison between the publication bias described here and the literature summarized in Hämäläinen et al. (2013) is difficult given the differences in inclusion criteria described above. We do offer some caution in the interpretation of our publication bias analysis, as the degrees of freedom were below 4, at which point the *t*-distribution inflates type I error. Going forward, studies of auditory processing and RD should report *a priori* power and pre-register analysis plans to combat difficulties in replication (Ansari & Gervain, 2018)

3.4.2 Comparative Discussion of Cross-Domain Auditory Processing Deficits

As was a primary goal of this study, the shared methodology among our four meta-analyses allows us to compare these effect sizes to each other and discuss the context of each described deficit. As noted in other reviews, the presence of large and broad auditory processing undermines the specificity of hypotheses relating to temporal sampling and rapid auditory temporal processing (Goswami, 2011; Protopapas, 2014). Though the present meta-analysis does not include rise-time discrimination (often described as beat detection) tasks nor auditory repetition tasks similar to those from Tallal (1980), the presence of large cross-domain auditory processing deficits necessitates a shift toward hypotheses that can explain the broad impairments described here.

The magnitude of the deficits among the four auditory task categories was strikingly similar, and quite large. The deficits measured here are also similar, though slightly smaller in magnitude, to several meta-analyses looking at deficits for RD in PA (*g* = 1.37; Melby-Lervåg et al., 2012), RAN (*g* = 1.19; Araújo & Faísca, 2019), and orthographic knowledge (*g* = 1.17; Georgiou et al., 2021). The similarity in magnitude does not necessarily suggest that auditory processing plays similarly import roles in reading development as PA and RAN, but rather that the role of auditory processing in reading development needs further testing so that it can properly be factored into theoretical models or included in universal screening.

Though deficits in frequency discrimination, duration discrimination, and gap detection have been described before, our review is the first to conclude that there is a significant weakness in intensity processing in RD. Though intensity discrimination is often included as a "control task" to ensure that other effects of interest are not due to task demands (e.g., Goswami et al., 2010), its magnitude was highly similar to frequency and duration discrimination as well as gap detection. The reason that many individual studies and a major systematic review (Hämäläinen et al., 2013) described no intensity discrimination deficits in RD is unclear. However, a potential explanation is that low power in individual studies and the comparison of their respective *p*-values across studies led researchers to believe that there was no intensity discrimination deficit.

Our cross-domain analyses offer two primary interpretations for the intensity discrimination deficit. One is that intensity discrimination is truly impaired in RD and should be considered as a key auditory processing impairment alongside spectral and temporal processing deficits. Another possible explanation is that there is no true effect of intensity discrimination, and that as a control task, the magnitude of its effect should be subtracted from the other auditory task categories to quantify their respective effects. This interpretation likely necessitates that the magnitude of the intensity discrimination deficit correlates with the magnitude of frequency or duration discrimination deficit. Only Tong et al. (2018) reports correlations between intensity thresholds and other psychoacoustic tasks ($r = .29$ and $r = .37$ for intensity discrimination's correlation with one-rise and rise rove task, respectively). Few studies have even published the correlations between the tasks described in this meta-analysis with other psychoacoustic tasks more broadly (Gibson et al., 2006; Halliday & Bishop, 2006b). For example, Gibson et al. (2006) finds moderate to low correlations ($r = .29$ and $r = .39$ for the control and RD groups, respectively) for the relationship between thresholds of frequency discrimination and frequency modulation detection. In sum, we find it unlikely that the intensity discrimination deficit is highly correlated with a frequency or duration discrimination deficit, though further research is needed to understand the relationship between auditory processing measures.

In sum, these results broaden our understanding of auditory processing deficits in RD. This broadening was necessary, given the previous interpretation of intensity discrimination as a control task. However, we are not left with any clearer picture for how these deficits fit with the prevailing multifactorial models of RD (Compton, 2020), as evidence for a cascading causal path remains weak. Models such as the neural noise hypothesis (Hancock et al., 2017) remain viable under these results, but need updating to reflect the relative lack of time-specific requirements for intensity processing.

3.4.3 Underlying Distributions of Auditory Processing and Reading Ability

Some papers have described bimodal distributions for auditory processing ability, such that some individuals with RD have no auditory processing impairment whereas others have a large impairment (Banai & Ahissar, 2004, 2005; Ben-Yehudah et al., 2004). If this is indeed the case, it raises the question of what utility describing a “mean impairment” serves when it is possible that no individual child falls at the mean of the auditory processing ability distribution. The mean impairments described here still have utility for creating clinical benchmarks for children diagnosed with comorbid APD and RD, such that children performing worse than the estimated effect sizes for a given auditory task category likely fall into the impaired group of a bimodal distribution, regardless of how many children fall at the mean.

Another factor that may influence the underlying distribution of auditory processing skills in the general population, or within RD as a category, is whether children with dyslexia have increased variability in auditory task performance has been noted by a number of studies (e.g., Banai & Ahissar, 2005; King et al., 2003). Using a rule of thumb for comparing variance ratios ($s^2max/s^2min < 3$; Dean & Voss, 1999), 58 of the 135 effect sizes analyzed here had unequal variance (43%), with RD having the larger variance in 55 of these 58 effect sizes (94.8%). It was quite common for studies to have unequal variances for one effect size (e.g., frequency discrimination), but equal variance for others (e.g., intensity discrimination; N = 17). Among those analyzed here, 30 studies had effect sizes with equal variance, whereas only 16 studies had unilaterally unequal variances. Moreover, this type of variance ratio depends strongly on sample size, in which small samples are more likely to yield unequal variance in their sample. In sum, there is some evidence that children with RD are more variable in their performance on auditory processing tasks, but the effects are not unilateral, nor are they particularly damaging to the interpretation of mean impairments.

3.4.4 Limitations

The results of our meta-regression are primarily limited by the lack of variety of task designs, as discussed above. The task design differences in Witton et al. (2020) appear to be driven by the differences between same-different designs and traditional psychophysical designs (e.g., 2AFC, AXB, etc.). Unfortunately, the same-different design did not meet the $N = 5$ threshold for inclusion in our meta-regression as specified in the Stage 1 manuscript for any of the auditory processing categories. The remaining evidence that auditory processing deficits are in part due to psychophysical task demands is thus limited to whether the intensity discrimination deficits described above reflect a true deficit, an artifact of task demands, or some combination of the two.

It is also worth noting that our sample, much like most studies on RD, overrepresents English-speaking and reading participants (Share, 2021). Though we were sensitive to the differing diagnostic cutoff criteria across orthographies, such as using non-word or fluency reading measures in transparent languages where reading accuracy plateaus after first grade (Papadopoulos et al., 2021), we are unable to remove the English-speaking bias from our meta-analyses.

Finally, it is worthwhile to discuss in detail the limitations of meta-analysis approaches more broadly, and the critiques that may be levied against the current meta-analyses. In pooling data across studies and samples, there is a potential to cancel out or erase theoretically motivated and potentially meaningful variability between effect sizes. For example, Ahissar et al. (2006) tests the hypothesis that individuals with RD are particularly poor performers in the presence of a perceptual anchor or reference tone, and they find starkly different effect sizes in a “no reference” condition ($g = 1.78$) versus a “reference” condition ($g = -.07$). The inclusion criteria we chose in the Stage 1 manuscript necessitate the pooling of these effect sizes and thus wash away potentially meaningful variability. There are many such critiques to be levied against data pooling, ranging from concern over pooling frequency perception data above and below ~4kHz, to including adults alongside children for skills that we know change over the course of development, and to concerns about including RD samples who may have comorbid ADHD. We recognize the validity of these critiques and invite researchers to use the data and code from the Open Science Framework (https://osf.io/nwctx/) to test specific hypotheses from extant data, if there is sufficient power to do so and if there is not, to use this meta-analysis as justification for funding/support to generate high quality data in future studies.

3.4.5 Conclusions and Future Directions

The meta-analytic methods described here could reasonably be applied to other psychoacoustic tasks, such as others presented in Hämäläinen et al. (2013), as well as to stimuli from non-linguistic to include linguistic stimuli. Further specifying these impairments is paramount to being able to formulate detailed hypotheses surrounding why these deficits exist in RD. We strongly encourage these lines of research, as the primary limitation in completing a high-quality meta-analysis is time, as opposed to funding.

Despite the seemingly elusive mechanisms that link auditory processing and RD, an independent effort toward translation to clinical impact should continue. Namely, the size of these behavioral impairments warrants further investigation as potential screening measures for RD. Comparatively few prospective longitudinal studies (e.g., Law et al., 2017) include measures of auditory processing rather than only more traditional measures of letter and letter-sound knowledge, PA, and RAN, as the evidence base surrounding auditory processing in RD is comparatively weaker (for a prospective, longitudinal meta-analysis on RAN and PA, see

McWeeny et al., 2022). We hope that this meta-analysis can be used as justification for creating standardized auditory processing measures with published reliability, which, if satisfactory, can be followed by their inclusion in prospective longitudinal studies.

Taken together, the findings of this systematic review and meta-analysis advance the field forward in multiple ways. The addition of an intensity discrimination deficit to meta-analytically documented deficits in frequency discrimination (Witton et al., 2020), suggests that auditory impairments in RD are broader, and perhaps larger, than previous thought, precluding the use of domain-specific causal hypotheses (Protopapas, 2014). Study quality measures reveal the need to make major updates to reliability reporting and *a priori* power calculations. Improving reliability and registering adequately powered analysis plans may help reduce the extant publication bias. Finally, we hope that the present study can be used as motivation for increased research activity in the area of auditory processing and RD, with the goal of improving our understanding of why these deficits exist, and whether we can leverage them in identifying RD earlier.

3.5 References

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. https://doi.org/10.1038/nn1800

Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory processing deficits in reading disabled adults. *Journal of the Association for Research in Otolaryngology*, *3*(3), 302–320. https://doi.org/10.1007/s101620010093

Amitay, S., Ben-Yehudah, G., Banai, K., & Ahissar, M. (2002). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain*, *125*(10), 2272–2285. https://doi.org/10.1093/brain/awf231

Ansari, D., & Gervain, J. (2018). Registered Reports: Introducing a new article format in Developmental Science. *Developmental Science*, *21*(1), e12650. https://doi.org/10.1111/desc.12650

Araújo, S., & Faísca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading*, *23*(5), 349–368. https://doi.org/10.1080/10888438.2019.1572758

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Banai, K., & Ahissar, M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiology & Neurotology*, *9*(6), 328–340. https://doi.org/10.1159/000081282

Banai, K., & Ahissar, M. (2005). Psychoacoustics and working memory in dyslexia. In J. Syka & M. M. Merzenich (Eds.), *Plasticity and signal representation in the auditory system.*

(2005-03154-021; pp. 233–242). Springer Publishing Co; https://doi.org/10.1007/0-387-23181-1_21

Beattie, R. L., & Manis, F. R. (2013). Rise time perception in children with reading and combined reading and language difficulties. *Journal of Learning Disabilities*, *46*(3), 200–209. https://doi.org/10.1177/0022219412449421

Ben-Yehudah, G., Banai, K., & Ahissar, M. (2004). Patterns of deficit in auditory temporal processing among dyslexic adults. *NeuroReport: For Rapid Communication of Neuroscience Research*, *15*(4), 627–631. https://doi.org/10.1097/00001756-200403220-00011

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020

Breier, J. I., Fletcher, J. M., Foorman, B. R., Klaas, P., & Gray, L. C. (2003). Auditory temporal processing in children with specific reading disability with and without attention deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research*, *46*(1), 31–42. eric.

Cacace, A. T., McFarland, D. J., Ouimet, J. R., Schrieber, E. J., & Marro, P. (2000). Temporal processing deficits in remediation-resistant reading-impaired children. *Audiology and Neurotology*, *5*(2), 83–97. https://doi.org/10.1159/000013871

Chaubet, J., Pereira, L., & Perez, A. P. (2014). Temporal resolution ability in students with dyslexia and reading and writing disorders. *International Archives of Otorhinolaryngology*, *18,* 146–149. https://doi.org/10.1055/s-0033-1363465

Chermak, G. D., Vivian, I., Doris-Eva, B., & E., M. F. (2018). Letter to the editor: Response to Vermiglio, 2018, “The gold standard and auditory processing disorder.” *Perspectives of the ASHA Special Interest Groups*, *3*(6), 77–82. https://doi.org/10.1044/persp3.SIG6.77

Colling, L. J., Noble, H. L., & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, *11*. https://www.frontiersin.org/article/10.3389/fnins.2017.00360

Compton, D. L. (2020). Focusing our view of dyslexia through a multifactorial lens: A commentary. *Learning Disability Quarterly*, 0731948720939009. https://doi.org/10.1177/0731948720939009

Dean, A., & Voss, D. (Eds.). (1999). Checking model assumptions. In *Design and analysis of experiments* (pp. 103–134). Springer. https://doi.org/10.1007/0-387-22634-6_5

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: An R-package for robust variance estimation in meta-analysis*.

Fostick, L., Bar-El, S., & Ram-Tsur, R. (2012). Auditory temporal processing as a specific deficit among dyslexic readers. *Journal of Psychology Research*, *2*(2). https://doi.org/10.17265/2159-5542/2012.02.001

Georgiou, G. K., Martinez, D., Vieira, A. P. A., & Guo, K. (2021). Is orthographic knowledge a strength or a weakness in individuals with dyslexia? Evidence from a meta-analysis. *Annals of Dyslexia*, *71*(1), 5–27. https://doi.org/10.1007/s11881-021-00220-6

Georgiou, G. K., Protopapas, A., Papadopoulos, T. C., Skaloumbakas, C., & Parrila, R. (2010a). Auditory temporal processing and dyslexia in an orthographically consistent language.

Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, *46*(10), 1330–1344. https://doi.org/10.1016/j.cortex.2010.06.006

Georgiou, G. K., Protopapas, A., Papadopoulos, T. C., Skaloumbakas, C., & Parrila, R. (2010b). Auditory temporal processing and dyslexia in an orthographically consistent language. *Cortex*, *46*(10), 1330–1344. https://doi.org/10.1016/j.cortex.2010.06.006

Germanò, E., Gagliano, A., & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental Neuropsychology*, *35*(5), 475–493. https://doi.org/10.1080/87565641.2010.494748

Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, *23*(4), 621–642. https://doi.org/10.1080/02643290500412545

Gokula, R., Sharma, M., Cupples, L., & Valderrama, J. T. (2019). Comorbidity of auditory processing, attention, and memory in children with word reading difficulties. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02383

Gooch, D., Snowling, M., & Hulme, C. (2011). Time perception, phonological skills and executive function in children with dyslexia and/or ADHD symptoms. *Journal of Child Psychology and Psychiatry*, *52*(2), 195–203. https://doi.org/10.1111/j.1469-7610.2010.02312.x

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

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. https://doi.org/10.1111/j.1467-7687.2010.00955.x

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019. https://doi.org/10.1007/s11145-009-9186-6

Goswami, U., Huss, M., Mead, N., & Fosker, T. (2021). Auditory sensory processing and phonological development in high IQ and exceptional readers, typically developing readers, and children with dyslexia: A longitudinal study. *Child Development*, *92*(3), 1083–1098. https://doi.org/10.1111/cdev.13459

Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, *49*(5), 1363–1376. https://doi.org/10.1016/j.cortex.2012.05.005

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. https://doi.org/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. https://doi.org/10.1073/pnas.122368599

Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese.

Journal of Cognitive Neuroscience, *23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ : British Medical Journal*, *331*(7524), 1064–1065. https://doi.org/10.1136/bmj.38636.593461.68

Griffin, J. W. (2020). *metapower: Power analysis for meta-analysis* (0.2.1). https://CRAN.R-project.org/package=metapower

Gu, C., & Bi, H.-Y. (2020). Auditory processing deficit in individuals with dyslexia: A meta-analysis of mismatch negativity. *Neuroscience & Biobehavioral Reviews*, *116*, 396–405. https://doi.org/10.1016/j.neubiorev.2020.06.032

Halliday, L. F., & Bishop, D. V. M. (2006a). Auditory frequency discrimination in children with dyslexia. *Journal of Research in Reading*, *29*(2), 213–228. https://doi.org/10.1111/j.1467-9817.2006.00286.x

Halliday, L. F., & Bishop, D. V. M. (2006b). Is poor frequency modulation detection linked to literacy problems? A comparison of specific reading disability and mild to moderate sensorineural hearing loss. *Brain and Language*, *97*(2), 200–213. https://doi.org/10.1016/j.bandl.2005.10.007

Hämäläinen, J. A., Fosker, T., Szücs, D., & Goswami, U. (2011). N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. *International Journal of Psychophysiology*, *81*(1), 51–59. https://doi.org/10.1016/j.ijpsycho.2011.04.005

Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. T. (2013). Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/field evidence. *Journal of Learning Disabilities*, *46*(5), 413–427. https://doi.org/10.1177/0022219411436213

Hancock, R., Pugh, K. R., & Hoeft, F. (2017). Neural noise hypothesis of developmental dyslexia. *Trends in Cognitive Sciences*, *21*(6), 434–448. https://doi.org/10.1016/j.tics.2017.03.008

Hari, R., Sääskilahti, A., Helenius, P., & Uutela, K. (1999). Non-impaired auditory phase locking in dyslexic adults. *NeuroReport*, *10*(11), 2347–2348.

Harzing, A.-W., & Alakangas, S. (2017). Microsoft Academic is one year old: The Phoenix is ready to leave the nest. *Scientometrics*, *112*(3), 1887–1894. https://doi.org/10.1007/s11192-017-2454-3

Heath, S. M., Bishop, D. V. M., Hogben, J. H., & Roach, N. W. (2006a). Psychophysical indices of perceptual functioning in dyslexia: A psychometric analysis. *Cognitive Neuropsychology*, *23*(6), 905–929. https://doi.org/10.1080/02643290500538398

Heath, S. M., Bishop, D. V. M., Hogben, J. H., & Roach, N. W. (2006b). Psychophysical indices of perceptual functioning in dyslexia: A psychometric analysis. *Cognitive Neuropsychology*, *23*(6), 905–929. https://doi.org/10.1080/02643290500538398

Heath, S. M., & Hogben, J. H. (2004). The reliability and validity of tasks measuring perception of rapid sequences in children with dyslexia. *Journal of Child Psychology and Psychiatry*, *45*(7), 1275–1287. https://doi.org/10.1111/j.1469-7610.2004.00313.x

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, *9*(4), 426–445. https://doi.org/10.1037/1082-989X.9.4.426

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1,* 39–65. https://doi.org/10.1002/jrsm.5

Helenius, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain*, *122*(5), 907–913. https://doi.org/10.1093/brain/122.5.907

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674–689. https://doi.org/10.1016/j.cortex.2010.07.010

Ingelghem, M. V., Boets, B., van Wieringen, A., Onghena, P., Ghesquière, P., Vandenbussche, E., & Wouters, J. (2005). *An auditory temporal processing deficit in children with dyslexia?* Unpublished manuscript.

King, W. M., Lombardino, L. J., Crandell, C. C., & Leonard, C. M. (2003). Comorbid auditory processing disorder in developmental dyslexia. *Ear and Hearing*, *24*(5), 448–456. https://doi.org/10.1097/01.AUD.0000090437.10978.1A

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00482

Law, J. M., Vandermosten, M., Ghesquière, P., & Wouters, J. (2017). Predicting future reading problems based on pre-reading auditory measures: A longitudinal study of children with a familial risk of dyslexia. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.00124

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. https://doi.org/10.1016/j.jml.2010.09.003

Lonergan, A., Doyle, C., Cassidy, C., MacSweeney Mahon, S., Roche, R. A. P., Boran, L., & Bramham, J. (2019). A meta-analysis of executive functioning in dyslexia with consideration of the impact of comorbid ADHD. *Journal of Cognitive Psychology*, *31*(7), 725–749. https://doi.org/10.1080/20445911.2019.1669609

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech, Language & Hearing Research*, *44*(4), 925–940. https://doi.org/10.1044/1092-4388(2001/073)

Matthews, N. H. (2013). *Musical beat perception in children with developmental dyslexia—Effects of pitch and training*. University of Cambridge.

McArthur, G. M., Atkinson, C. M., & Ellis, D. (2010). Can training normalize atypical passive auditory ERPs in children with SRD or SLI? *Developmental Neuropsychology*, *35*(6), 656–678. https://doi.org/10.1080/87565641.2010.508548

McArthur, G. M., & Hogben, J. H. (2001). Auditory backward recognition masking in children with a specific language impairment and children with a specific reading disability. *The Journal of the Acoustical Society of America*, *109*(3), 1092–1100. https://doi.org/10.1121/1.1338559

McCroskey, R. L., & Kidder, H. C. (1980). Auditory fusion among learning disabled, reading disabled, and normal children. *Journal of Learning Disabilities*, *13*(2), 69–76. https://doi.org/10.1177/002221948001300205

Melby-Lervåg, M., Lyster, S.-A. H., & 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

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, *64*(2), 199–231. https://doi.org/10.1006/jecp.1996.2343

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting Items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Nittrouer, S., Krieg, L. M., & Lowenstein, J. H. (2018). Speech recognition in noise by children with and without dyslexia: How is it related to reading? *Research in Developmental Disabilities*, *77*, 98–113. https://doi.org/10.1016/j.ridd.2018.04.014

O'Brien, G., & Yeatman, J. (2020). Bridging sensory and language theories of dyslexia: Towards a multifactorial model. *Developmental Science*, *n/a*(n/a), e13039. https://doi.org/10.1111/desc.13039

Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia*, *50*(8), 1895–1905. https://doi.org/10.1016/j.neuropsychologia.2012.04.014

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Papadopoulos, T. C., Csépe, V., Aro, M., Caravolas, M., Diakidoy, I.-A., & Olive, T. (2021). Methodological issues in literacy research across languages: Evidence from alphabetic orthographies. *Reading Research Quarterly*, *56*(S1), S351–S370. https://doi.org/10.1002/rrq.407

Papadopoulos, T. C., Georgiou, G. K., & Parrila, R. K. (2012). Low-level deficits in beat perception: Neither necessary nor sufficient for explaining developmental dyslexia in a consistent orthography. *Research in Developmental Disabilities*, *33*(6), 1841–1856. https://doi.org/10.1016/j.ridd.2012.04.009

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory processing of amplitude envelope rise time in adults diagnosed with developmental dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286. https://doi.org/10.1080/10888430701344280

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*(1), 283–307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., & Wouters, J. (2011). Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities*, *32*(6), 2810–2819. https://doi.org/10.1016/j.ridd.2011.05.025

Protopapas, A. (2014). From temporal processing to developmental language disorders: Mind the gap. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20130090. https://doi.org/10.1098/rstb.2013.0090

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A.-M., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, *48*(9), 923–931. https://doi.org/10.1111/j.1469-7610.2007.01763.x

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865. https://doi.org/10.1093/brain/awg076

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, *48*(2), 270–292. https://doi.org/10.1016/0022-0965(89)90006-4

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, *10*(3), 215–233. https://doi.org/10.1002/dys.276

Rocheron, I., Lorenzi, C., Füllgrabe, C., & Dumont, A. (2002). Temporal envelope perception in dyslexic children. *NeuroReport*, *13*(13), 1683–1687.

Rodgers, M., & Pustejovsky, J. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *MetaArXiv*, 1–25. https://doi.org/10.31222/osf.io/vqp8u

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, *31*(3), 509–527. https://doi.org/10.1016/S0095-4470(03)00046-9

Santurette, S., Poelmans, H., Luts, H., Ghesquiére, P., Wouters, J., & Dau, T. (2010). Detection and identification of monaural and binaural pitch contours in dyslexic listeners. *Journal* *of the Association for Research in Otolaryngology*, *11*(3), 515–524. https://doi.org/10.1007/s10162-010-0216-5

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *Career Development International*, *22*(5), 469–476. https://doi.org/10.1108/CDI-08-2017-0136

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Role of auditory temporal processing for reading and spelling disability. *Perceptual and Motor Skills*, *86*(3, Pt 1), 1043–1047. https://doi.org/10.2466/pms.1998.86.3.1043

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1999). The role of phonological awareness, speech perception, and auditory temporal processing for dyslexia. *European Child & Adolescent Psychiatry*, *8*(Suppl 3), III/28-III/34. https://doi.org/10.1007/PL00010690

Shamliyan, T., Kane, R. L., & Dickinson, S. (2010). A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, *63*(10), 1061–1070. https://doi.org/10.1016/j.jclinepi.2010.04.014

Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly*, *56*(S1), S391–S402. https://doi.org/10.1002/rrq.401

Sharma, M., Purdy, S. C., & Kelly, A. S. (2009). Comorbidity of auditory processing, language, and reading disorders. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(3), 706–722. https://doi.org/10.1044/1092-4388(2008/07-0226)

Sharma, M., Purdy, S. C., Newall, P., Wheldall, K., Beaman, R., & Dillon, H. (2006). Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder. *Clinical Neurophysiology*, *117*(5), 1130–1144. https://doi.org/10.1016/j.clinph.2006.02.001

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the 'Fast ForWord' language intervention program. *Journal of Child Psychology and Psychiatry*, *52*(3), 224–235. https://doi.org/10.1111/j.1469-7610.2010.02329.x

Talcott, J. B., Witton, C., Hebb, G. S., Stoodley, C. J., Westwood, E. A., France, S. J., Hansen, P. C., & Stein, J. F. (2002). On the relationship between dynamic visual and auditory processing and literacy skills; results from a large primary-school study. *Dyslexia*, *8*(4), 204–225. https://doi.org/10.1002/dys.224

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, *9*(2), 182–198. https://doi.org/10.1016/0093-934X(80)90139-X

Tallal, P., Merzenich, M. M., Miller, S., & Jenkins, W. (1998). Language learning impairments: Integrating basic science, technology, and remediation. *Experimental Brain Research*, *123*(1), 210–219. https://doi.org/10.1007/s002210050563

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, *29*(3), 334–348. https://doi.org/10.1111/j.1467-9817.2006.00312.x

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, *102*(1), 120–129. https://doi.org/10.1016/j.jphysparis.2008.03.007

Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, *26*(2), 139–161. https://doi.org/10.1007/s11145-012-9359-6

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. https://doi.org/10.1177/0022219417712018

Toplak, M. e., Rucklidge, J. j., Hetherington, R., John, S. c. f., & Tannock, R. (2003). Time perception deficits in attention-deficit/ hyperactivity disorder and comorbid reading difficulties in child and adolescent samples. *Journal of Child Psychology and Psychiatry*, *44*(6), 888–903. https://doi.org/10.1111/1469-7610.00173

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities*, *33*(2), 635–644. https://doi.org/10.1016/j.ridd.2011.11.005

Vermiglio, A. J. (2018). The gold standard and auditory processing disorder. *Perspectives of the ASHA Special Interest Groups*, *3*(6), 6–17. https://doi.org/10.1044/persp3.SIG6.6

Walker, M. M., Shinn, J. B., Cranford, J. L., Givens, G. D., & Holbert, D. (2002). Auditory temporal processing performance of young adults with reading disorders. *Journal of Speech, Language, and Hearing Research*, *45*(3), 598–605. https://doi.org/10.1044/1092-4388(2002/048)

Wang, H.-L. S., Huss, M., Hämäläinen, J. A., & Goswami, U. (2012). Basic auditory processing and developmental dyslexia in Chinese. *Reading and Writing*, *25*(2), 509–536. https://doi.org/10.1007/s11145-010-9284-5

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. https://doi.org/10.1016/j.ridd.2019.01.006

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., & Rogahn, R. (2019). A review of Microsoft Academic services for science of science Studies. *Frontiers in Big Data*, *2*. https://doi.org/10.3389/fdata.2019.00045

Whiting, P., Harbord, R., & Kleijnen, J. (2005). No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodology*, *5*(1), 19. https://doi.org/10.1186/1471-2288-5-19

Willburger, E., & Landerl, K. (2010). Anchoring the deficit of the anchor deficit: Dyslexia or attention? *Dyslexia*, *16*(2), 175–182. https://doi.org/10.1002/dys.404

Witton, C., Swoboda, K., Shapiro, L. R., & Talcott, J. B. (2020). Auditory frequency discrimination in developmental dyslexia: A meta-analysis. *Dyslexia: An International Journal of Research and Practice*, *26*(1), 36–51. https://doi.org/10.1002/dys.1645

Witton, C., Talcott, J. B., Hansen, P. C., Richardson, A. J., Griffiths, T. D., Rees, A., Stein, J. F., & Green, G. G. R. (1998). Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Current Biology*, *8*(14), 791–797. https://doi.org/10.1016/S0960-9822(98)70320-3

Wright, C. M., & Conlon, E. G. (2009). Auditory and Visual Processing in Children With Dyslexia. *Developmental Neuropsychology*, *34*(3), 330–355. https://doi.org/10.1080/87565640902801882

Zaidan, E. (2009). *An investigation of temporal resolution abilities in school-aged children with and without dyslexia* (2010-99051-126; Issues 9-A) [ProQuest Information & Learning].

http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-99051-126&site=ehost-live

Zaidan, E., & Baran, J. A. (2013). Gaps-in-noise (GIN©) test results in children with and without reading disabilities and phonological processing deficits. *International Journal of Audiology*, *52*(2), 113–123. https://doi.org/10.3109/14992027.2012.733421

Zhang, M., Xie, W., Xu, Y., & Meng, X. (2018). Auditory temporal perceptual learning and transfer in Chinese-speaking children with developmental dyslexia. *Research in Developmental Disabilities*, *74*, 146–159. https://doi.org/10.1016/j.ridd.2018.01.005

3.6 Tables and Figures

Table 6. Power estimates from known extant literature for categorical studies

Analysis	Estimate (Hedges' *g*)	*k*	Power
Frequency	.7	46	≥.99
Duration	.9	23	≥.99
Intensity	.5	21	≥.99
Gap Detection	.6	19	≥.99

Table 7. Main Effects: Intercept-only Models for Study 2

Model	**N**	**k**	**I^2**	**τ^2**	**g**	***t***	**SE**	**df**	**95% CI**
Task Category									
Frequency Discrimination	30	55	76.12	.23	.79	7.56	.11	27.9	[.58 1.01]
Duration Discrimination	14	22	66.49	.15	.80	6.12	.13	12.5	[.52 1.08]
Intensity Discrimination	24	34	52.39	.10	.60	7.65	.08	21.3	[.44 .76]
Gap Detection [a]	16	23	75.20	.347	.80	4.89	.16	14.8	[.45 1.15]

Note. N = number of samples/studies; k = number of effect sizes

All models were significant at $p < .001$.

a – model is presented with an extreme outlier excluded

Table 8. Moderator Effects: Meta-Regression Models for Study 2

Model	***N***	***k***	**I^2**	**τ^2**	***A priori* Power**	**Adjusted Power**	***g***	***t***	**SE**	**df**	***p***	**95% CI**
Frequency	24	38	77.40	.26	.94	.65[b]						
Intercept							.73	3.20	.21	10.89		[.22 1.24]
ABABA							-.03	-.13	.26	6.51	.90	[-.66 .60]
AXB							.25	.94	.27	16.75	.36	[-.32 .82]
Duration	13	19	70.01	.19	.94	.10[b]						
Intercept							.78	4.30	.18	5.89		[.33 1.22]
AXB							.08	.23	.32	10.10	.81	[-.64 .79]
Intensity	15	21	31.36	.04	.94	.42[c]						
Intercept							.48	6.55	.07	11.47		[.32 .64]
ABABA							.26	1.13	.23	3.33	.33	[-.44 .97]
Gap Detection [a]	11	17	62.85	.21	.94	.12 [b]						
Intercept							.57	6.44	.09	6.99		[.36 .78]
GIN							1.04	1.82	.57	3.81	.15	[-.58 2.66]

Note. N = number of samples/studies; k = number of effect sizes

All intercepts were significant at $p < .01$.

a – model is presented with an extreme outlier excluded

b – I^2 used in power calculation was 75%

c – I^2 used in power calculation was 25%

Table 9. Between- and Within-Study Bias Analyses for Study 2

Model	**N**	**k**	**I^2**	**τ^2**	**g**	***SE***	***t***	**df**	***p***	**95% CI**
Within-Study Bias	63	135	76.81	.28						
Intercept					.73	.08	8.77	40.79		[.56 .89]
Good Quality					-.48	.19	-2.52	3.59	.07	[-1.03 .07]
Poor Quality					.28	.20	1.38	26.24	.18	[-.14 .69]
Between-Study Bias	63	135	75.28	.26						
Intercept					.45	.10	4.45	9.61		
Standard Error					2.84	.10	2.85	2.84	.03*	[-.44 6.11]

Note. N = number of studies; k = number of effect sizes
All intercepts were significant at $p < .05$.
* One-tailed test was used, as is customary for Egger's Regression

Table 10. Main Effects: Intercept-only Models with and Without Outliers for Study 2

Model	**N**	**k**	**I^2**	**τ^2**	**g**	***t***	**SE**	**df**	**95% CI**
Outliers Included									
Frequency Discrimination	30	55	76.12	.23	.79	7.56	.11	27.9	[.58 1.01]
Duration Discrimination	14	22	66.49	.15	.80	6.12	.13	12.5	[.52 1.08]
Intensity Discrimination	24	34	52.39	.10	.60	7.65	.08	21.3	[.44 .76]
Gap Detection	17	24	89.13	.97	1.05	3.65	.29	16	[.44 1.66]
Outliers Excluded									
Frequency Discrimination	30	52	74.51	.21	.78	7.75	.10	28	[.58 .99]
Duration Discrimination	13	21	53.75	.08	.73	6.87	.11	11.6	[.50 .96]
Intensity Discrimination	24	31	48.84	.09	.59	7.65	.08	21.3	[.43 .75]
Gap Detection	16	23	75.20	.35	.80	4.89	.16	14.8	[.45 1.15]

Note. N = number of samples/studies; k = number of effect sizes
All models were significant at $p < .001$ except gap detection with outliers included, which was significant at $p < .003$.
a – model is presented with an extreme outlier excluded

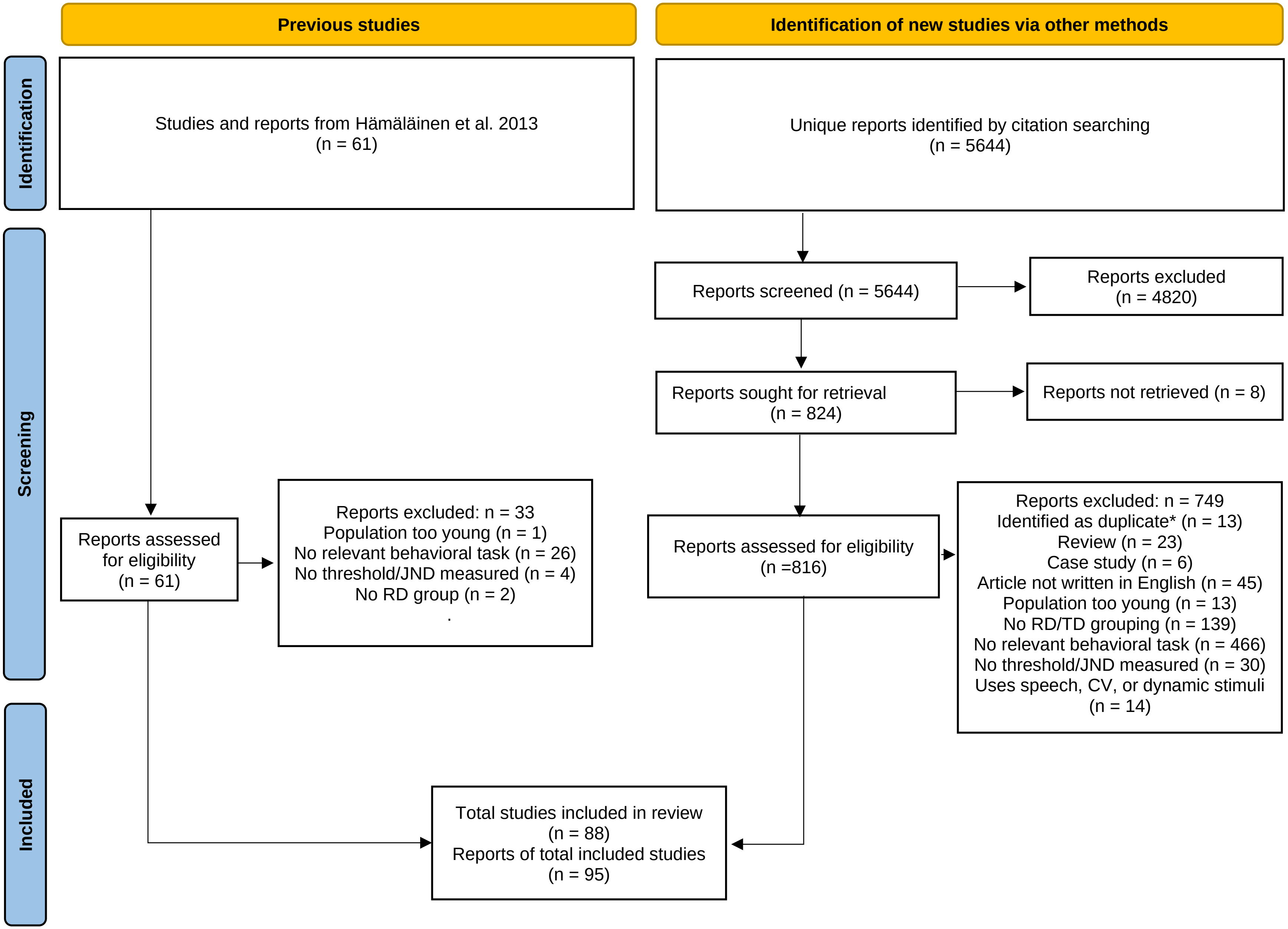
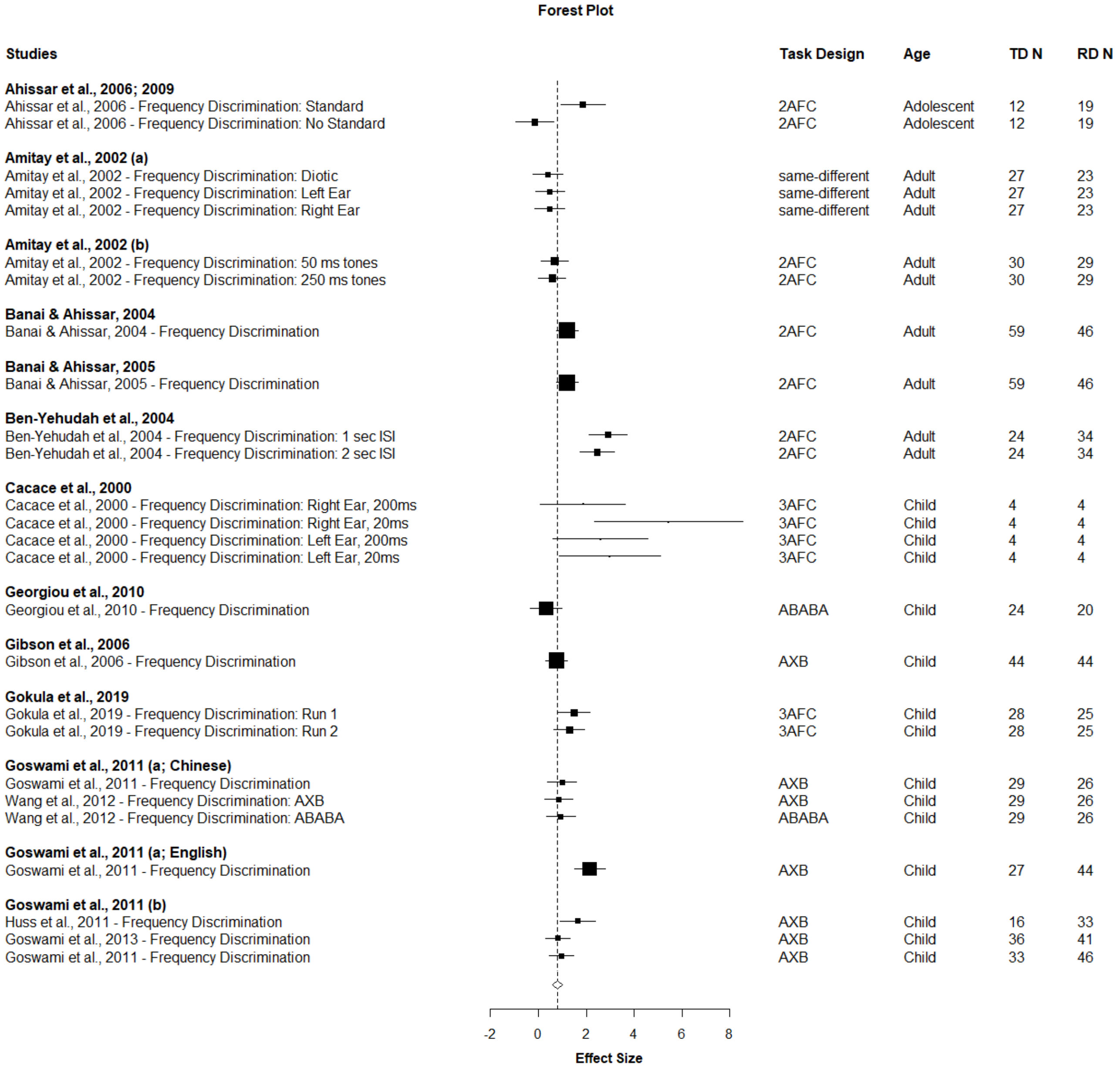
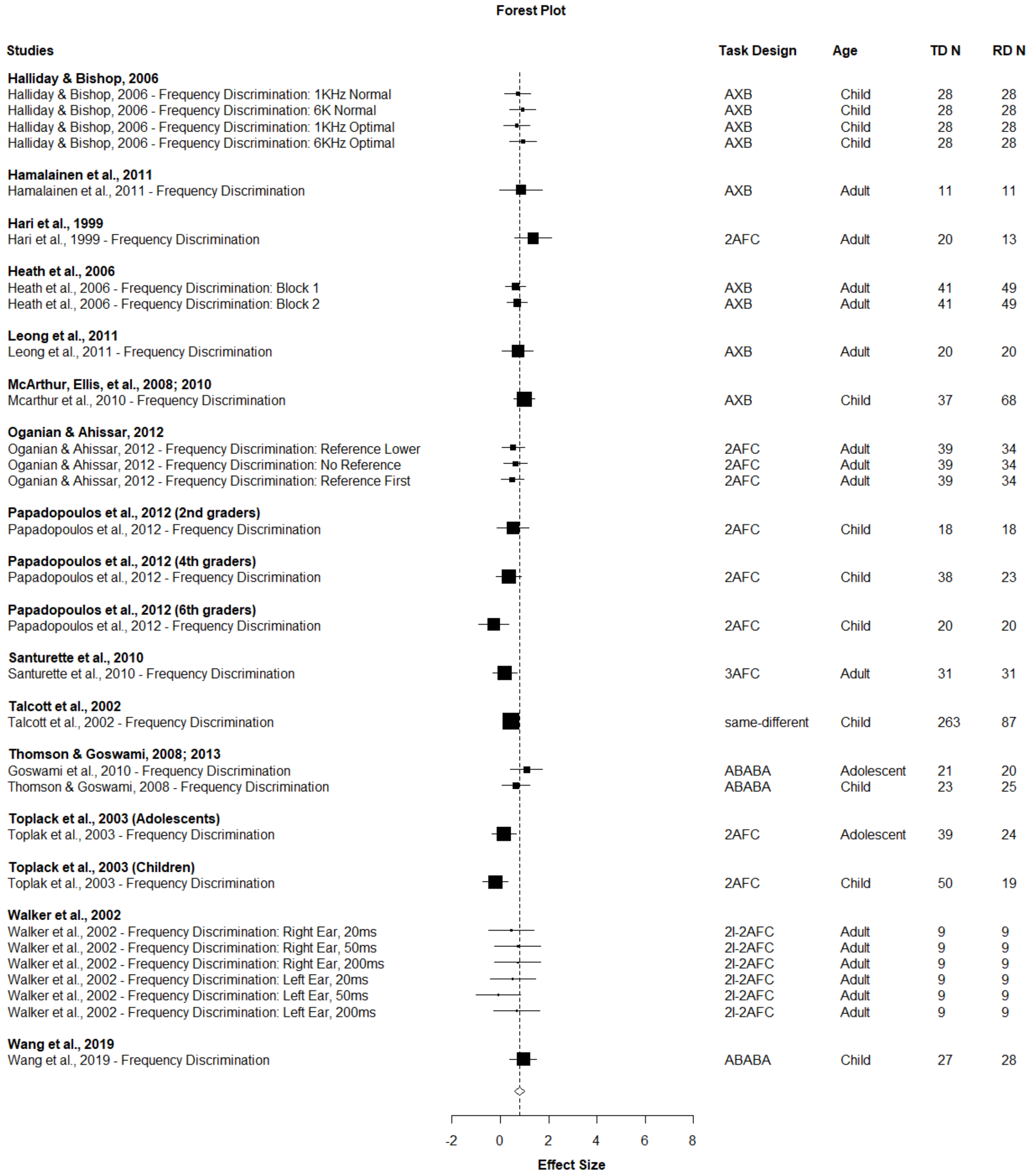

Previous studies
Identification of new studies via other methods
Identification
Screening
Included
Studies and reports from Hämäläinen et al. 2013
(n = 61)
Unique reports identified by citation searching
(n = 5644)
Reports screened (n = 5644)
Reports excluded
(n = 4820)
Reports sought for retrieval
(n = 824)
Reports not retrieved (n = 8)
Reports assessed
for eligibility
(n = 61)
Reports excluded: n = 33
Population too young (n = 1)
No relevant behavioral task (n = 26)
No threshold/JND measured (n = 4)
No RD group (n = 2)
Reports assessed for eligibility
(n =816)
Reports excluded: n = 749
Identified as duplicate* (n = 13)
Review (n = 23)
Case study (n = 6)
Article not written in English (n = 45)
Population too young (n = 13)
No RD/TD grouping (n = 139)
No relevant behavioral task (n = 466)
No threshold/JND measured (n = 30)
Uses speech, CV, or dynamic stimuli
(n = 14)
Total studies included in review
(n = 88)
Reports of total included studies
(n = 95)

Figure 3. PRISMA Flow Chart for Study 2



Studies	Task Design	Age	TD N	RD N
Ahissar et al., 2006; 2009				
Ahissar et al., 2006 - Frequency Discrimination: Standard	2AFC	Adolescent	12	19
Ahissar et al., 2006 - Frequency Discrimination: No Standard	2AFC	Adolescent	12	19
Amitay et al., 2002 (a)				
Amitay et al., 2002 - Frequency Discrimination: Diotic	same-different	Adult	27	23
Amitay et al., 2002 - Frequency Discrimination: Left Ear	same-different	Adult	27	23
Amitay et al., 2002 - Frequency Discrimination: Right Ear	same-different	Adult	27	23
Amitay et al., 2002 (b)				
Amitay et al., 2002 - Frequency Discrimination: 50 ms tones	2AFC	Adult	30	29
Amitay et al., 2002 - Frequency Discrimination: 250 ms tones	2AFC	Adult	30	29
Banai & Ahissar, 2004				
Banai & Ahissar, 2004 - Frequency Discrimination	2AFC	Adult	59	46
Banai & Ahissar, 2005				
Banai & Ahissar, 2005 - Frequency Discrimination	2AFC	Adult	59	46
Ben-Yehudah et al., 2004				
Ben-Yehudah et al., 2004 - Frequency Discrimination: 1 sec ISI	2AFC	Adult	24	34
Ben-Yehudah et al., 2004 - Frequency Discrimination: 2 sec ISI	2AFC	Adult	24	34
Cacace et al., 2000				
Cacace et al., 2000 - Frequency Discrimination: Right Ear, 200ms	3AFC	Child	4	4
Cacace et al., 2000 - Frequency Discrimination: Right Ear, 20ms	3AFC	Child	4	4
Cacace et al., 2000 - Frequency Discrimination: Left Ear, 200ms	3AFC	Child	4	4
Cacace et al., 2000 - Frequency Discrimination: Left Ear, 20ms	3AFC	Child	4	4
Georgiou et al., 2010				
Georgiou et al., 2010 - Frequency Discrimination	ABABA	Child	24	20
Gibson et al., 2006				
Gibson et al., 2006 - Frequency Discrimination	AXB	Child	44	44
Gokula et al., 2019				
Gokula et al., 2019 - Frequency Discrimination: Run 1	3AFC	Child	28	25
Gokula et al., 2019 - Frequency Discrimination: Run 2	3AFC	Child	28	25
Goswami et al., 2011 (a; Chinese)				
Goswami et al., 2011 - Frequency Discrimination	AXB	Child	29	26
Wang et al., 2012 - Frequency Discrimination: AXB	AXB	Child	29	26
Wang et al., 2012 - Frequency Discrimination: ABABA	ABABA	Child	29	26
Goswami et al., 2011 (a; English)				
Goswami et al., 2011 - Frequency Discrimination	AXB	Child	27	44
Goswami et al., 2011 (b)				
Huss et al., 2011 - Frequency Discrimination	AXB	Child	16	33
Goswami et al., 2013 - Frequency Discrimination	AXB	Child	36	41
Goswami et al., 2011 - Frequency Discrimination	AXB	Child	33	46

Figure 4. Forest Plot of Frequency Discrimination Effect Sizes



Studies	Task Design	Age	TD N	RD N
Halliday & Bishop, 2006				
Halliday & Bishop, 2006 - Frequency Discrimination: 1KHz Normal	AXB	Child	28	28
Halliday & Bishop, 2006 - Frequency Discrimination: 6K Normal	AXB	Child	28	28
Halliday & Bishop, 2006 - Frequency Discrimination: 1KHz Optimal	AXB	Child	28	28
Halliday & Bishop, 2006 - Frequency Discrimination: 6KHz Optimal	AXB	Child	28	28
Hamalainen et al., 2011				
Hamalainen et al., 2011 - Frequency Discrimination	AXB	Adult	11	11
Hari et al., 1999				
Hari et al., 1999 - Frequency Discrimination	2AFC	Adult	20	13
Heath et al., 2006				
Heath et al., 2006 - Frequency Discrimination: Block 1	AXB	Adult	41	49
Heath et al., 2006 - Frequency Discrimination: Block 2	AXB	Adult	41	49
Leong et al., 2011				
Leong et al., 2011 - Frequency Discrimination	AXB	Adult	20	20
McArthur, Ellis, et al., 2008; 2010				
Mcarthur et al., 2010 - Frequency Discrimination	AXB	Child	37	68
Oganian & Ahissar, 2012				
Oganian & Ahissar, 2012 - Frequency Discrimination: Reference Lower	2AFC	Adult	39	34
Oganian & Ahissar, 2012 - Frequency Discrimination: No Reference	2AFC	Adult	39	34
Oganian & Ahissar, 2012 - Frequency Discrimination: Reference First	2AFC	Adult	39	34
Papadopoulos et al., 2012 (2nd graders)				
Papadopoulos et al., 2012 - Frequency Discrimination	2AFC	Child	18	18
Papadopoulos et al., 2012 (4th graders)				
Papadopoulos et al., 2012 - Frequency Discrimination	2AFC	Child	38	23
Papadopoulos et al., 2012 (6th graders)				
Papadopoulos et al., 2012 - Frequency Discrimination	2AFC	Child	20	20
Santurette et al., 2010				
Santurette et al., 2010 - Frequency Discrimination	3AFC	Adult	31	31
Talcott et al., 2002				
Talcott et al., 2002 - Frequency Discrimination	same-different	Child	263	87
Thomson & Goswami, 2008; 2013				
Goswami et al., 2010 - Frequency Discrimination	ABABA	Adolescent	21	20
Thomson & Goswami, 2008 - Frequency Discrimination	ABABA	Child	23	25
Toplack et al., 2003 (Adolescents)				
Toplak et al., 2003 - Frequency Discrimination	2AFC	Adolescent	39	24
Toplack et al., 2003 (Children)				
Toplak et al., 2003 - Frequency Discrimination	2AFC	Child	50	19
Walker et al., 2002				
Walker et al., 2002 - Frequency Discrimination: Right Ear, 20ms	2I-2AFC	Adult	9	9
Walker et al., 2002 - Frequency Discrimination: Right Ear, 50ms	2I-2AFC	Adult	9	9
Walker et al., 2002 - Frequency Discrimination: Right Ear, 200ms	2I-2AFC	Adult	9	9
Walker et al., 2002 - Frequency Discrimination: Left Ear, 20ms	2I-2AFC	Adult	9	9
Walker et al., 2002 - Frequency Discrimination: Left Ear, 50ms	2I-2AFC	Adult	9	9
Walker et al., 2002 - Frequency Discrimination: Left Ear, 200ms	2I-2AFC	Adult	9	9
Wang et al., 2019				
Wang et al., 2019 - Frequency Discrimination	ABABA	Child	27	28

Figure 4 cont. Forest Plot of Frequency Discrimination Effect Sizes

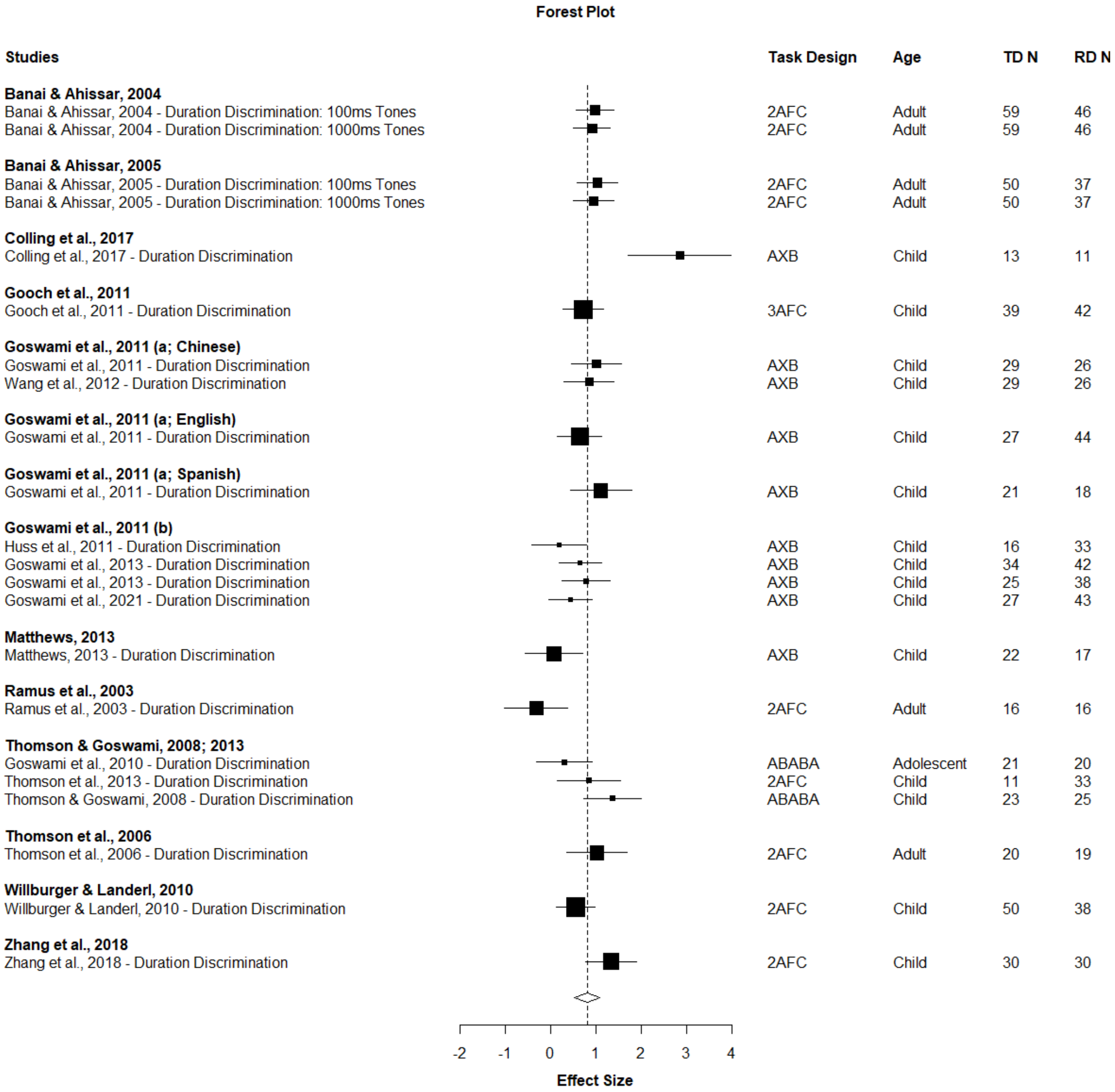
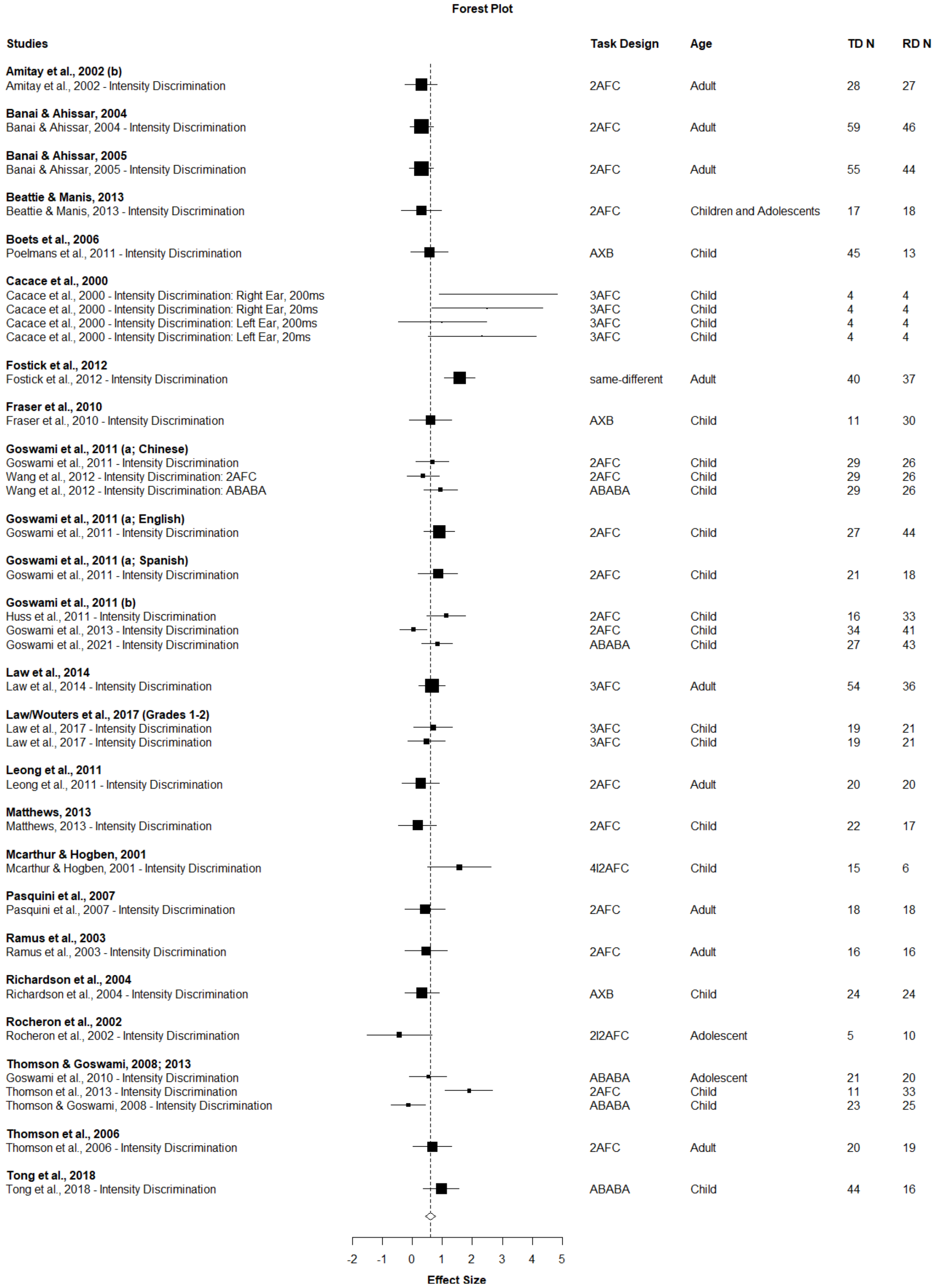
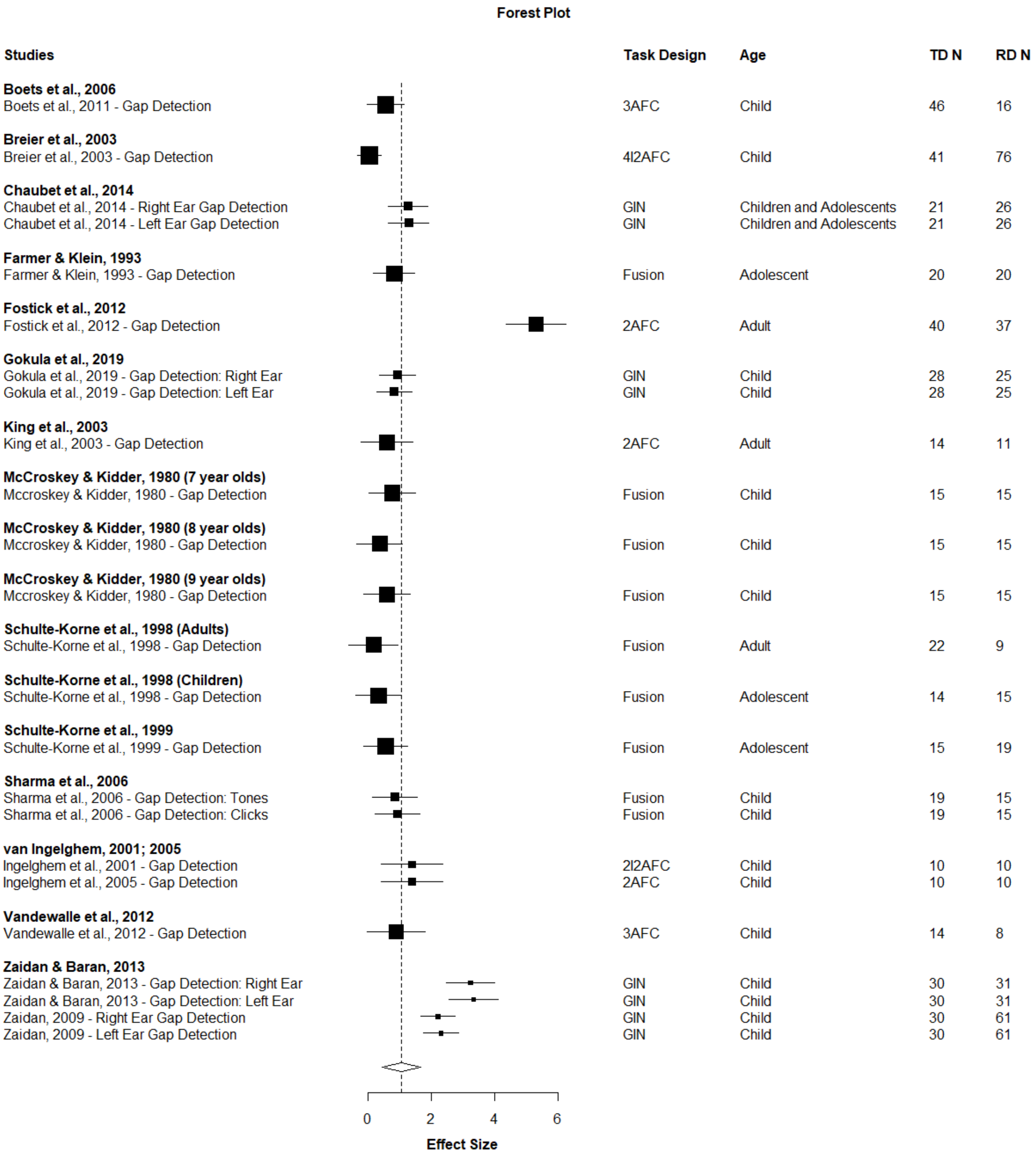


Figure 5. Forest Plot of Duration Discrimination Effect Sizes



Studies	Task Design	Age	TD N	RD N
Amitay et al., 2002 (b)				
Amitay et al., 2002 - Intensity Discrimination	2AFC	Adult	28	27
Banai & Ahissar, 2004				
Banai & Ahissar, 2004 - Intensity Discrimination	2AFC	Adult	59	46
Banai & Ahissar, 2005				
Banai & Ahissar, 2005 - Intensity Discrimination	2AFC	Adult	55	44
Beattie & Manis, 2013				
Beattie & Manis, 2013 - Intensity Discrimination	2AFC	Children and Adolescents	17	18
Boets et al., 2006				
Poelmans et al., 2011 - Intensity Discrimination	AXB	Child	45	13
Cacace et al., 2000				
Cacace et al., 2000 - Intensity Discrimination: Right Ear, 200ms	3AFC	Child	4	4
Cacace et al., 2000 - Intensity Discrimination: Right Ear, 20ms	3AFC	Child	4	4
Cacace et al., 2000 - Intensity Discrimination: Left Ear, 200ms	3AFC	Child	4	4
Cacace et al., 2000 - Intensity Discrimination: Left Ear, 20ms	3AFC	Child	4	4
Fostick et al., 2012				
Fostick et al., 2012 - Intensity Discrimination	same-different	Adult	40	37
Fraser et al., 2010				
Fraser et al., 2010 - Intensity Discrimination	AXB	Child	11	30
Goswami et al., 2011 (a; Chinese)				
Goswami et al., 2011 - Intensity Discrimination	2AFC	Child	29	26
Wang et al., 2012 - Intensity Discrimination: 2AFC	2AFC	Child	29	26
Wang et al., 2012 - Intensity Discrimination: ABABA	ABABA	Child	29	26
Goswami et al., 2011 (a; English)				
Goswami et al., 2011 - Intensity Discrimination	2AFC	Child	27	44
Goswami et al., 2011 (a; Spanish)				
Goswami et al., 2011 - Intensity Discrimination	2AFC	Child	21	18
Goswami et al., 2011 (b)				
Huss et al., 2011 - Intensity Discrimination	2AFC	Child	16	33
Goswami et al., 2013 - Intensity Discrimination	2AFC	Child	34	41
Goswami et al., 2021 - Intensity Discrimination	ABABA	Child	27	43
Law et al., 2014				
Law et al., 2014 - Intensity Discrimination	3AFC	Adult	54	36
Law/Wouters et al., 2017 (Grades 1-2)				
Law et al., 2017 - Intensity Discrimination	3AFC	Child	19	21
Law et al., 2017 - Intensity Discrimination	3AFC	Child	19	21
Leong et al., 2011				
Leong et al., 2011 - Intensity Discrimination	2AFC	Adult	20	20
Matthews, 2013				
Matthews, 2013 - Intensity Discrimination	2AFC	Child	22	17
Mcarthur & Hogben, 2001				
Mcarthur & Hogben, 2001 - Intensity Discrimination	4I2AFC	Child	15	6
Pasquini et al., 2007				
Pasquini et al., 2007 - Intensity Discrimination	2AFC	Adult	18	18
Ramus et al., 2003				
Ramus et al., 2003 - Intensity Discrimination	2AFC	Adult	16	16
Richardson et al., 2004				
Richardson et al., 2004 - Intensity Discrimination	AXB	Child	24	24
Rocheron et al., 2002				
Rocheron et al., 2002 - Intensity Discrimination	2I2AFC	Adolescent	5	10
Thomson & Goswami, 2008; 2013				
Goswami et al., 2010 - Intensity Discrimination	ABABA	Adolescent	21	20
Thomson et al., 2013 - Intensity Discrimination	2AFC	Child	11	33
Thomson & Goswami, 2008 - Intensity Discrimination	ABABA	Child	23	25
Thomson et al., 2006				
Thomson et al., 2006 - Intensity Discrimination	2AFC	Adult	20	19
Tong et al., 2018				
Tong et al., 2018 - Intensity Discrimination	ABABA	Child	44	16

Figure 6. Forest Plot of Intensity Discrimination Effect Sizes



Studies	Task Design	Age	TD N	RD N
Boets et al., 2006				
Boets et al., 2011 - Gap Detection	3AFC	Child	46	16
Breier et al., 2003				
Breier et al., 2003 - Gap Detection	4I2AFC	Child	41	76
Chaubet et al., 2014				
Chaubet et al., 2014 - Right Ear Gap Detection	GIN	Children and Adolescents	21	26
Chaubet et al., 2014 - Left Ear Gap Detection	GIN	Children and Adolescents	21	26
Farmer & Klein, 1993				
Farmer & Klein, 1993 - Gap Detection	Fusion	Adolescent	20	20
Fostick et al., 2012				
Fostick et al., 2012 - Gap Detection	2AFC	Adult	40	37
Gokula et al., 2019				
Gokula et al., 2019 - Gap Detection: Right Ear	GIN	Child	28	25
Gokula et al., 2019 - Gap Detection: Left Ear	GIN	Child	28	25
King et al., 2003				
King et al., 2003 - Gap Detection	2AFC	Adult	14	11
McCroskey & Kidder, 1980 (7 year olds)				
Mccroskey & Kidder, 1980 - Gap Detection	Fusion	Child	15	15
McCroskey & Kidder, 1980 (8 year olds)				
Mccroskey & Kidder, 1980 - Gap Detection	Fusion	Child	15	15
McCroskey & Kidder, 1980 (9 year olds)				
Mccroskey & Kidder, 1980 - Gap Detection	Fusion	Child	15	15
Schulte-Korne et al., 1998 (Adults)				
Schulte-Korne et al., 1998 - Gap Detection	Fusion	Adult	22	9
Schulte-Korne et al., 1998 (Children)				
Schulte-Korne et al., 1998 - Gap Detection	Fusion	Adolescent	14	15
Schulte-Korne et al., 1999				
Schulte-Korne et al., 1999 - Gap Detection	Fusion	Adolescent	15	19
Sharma et al., 2006				
Sharma et al., 2006 - Gap Detection: Tones	Fusion	Child	19	15
Sharma et al., 2006 - Gap Detection: Clicks	Fusion	Child	19	15
van Ingelghem, 2001; 2005				
Ingelghem et al., 2001 - Gap Detection	2I2AFC	Child	10	10
Ingelghem et al., 2005 - Gap Detection	2AFC	Child	10	10
Vandewalle et al., 2012				
Vandewalle et al., 2012 - Gap Detection	3AFC	Child	14	8
Zaidan & Baran, 2013				
Zaidan & Baran, 2013 - Gap Detection: Right Ear	GIN	Child	30	31
Zaidan & Baran, 2013 - Gap Detection: Left Ear	GIN	Child	30	31
Zaidan, 2009 - Right Ear Gap Detection	GIN	Child	30	61
Zaidan, 2009 - Left Ear Gap Detection	GIN	Child	30	61

Figure 7. Forest Plot of Gap Detection Effect Sizes

4. Conclusion

The present meta-analyses are important steps forward in summarizing the relationship between reading ability and some of its key correlates, namely RAN, PA, and auditory processing. The respective literatures are at different scientific stages; RAN and PA are considerably more prominent in modern theories of RD and have a stronger evidence base, allowing for increased specificity in research questions with regards to theoretical path models (e.g., Georgiou et al., 2016), RD subtypes over time (e.g., Ozernov-Palchik et al., 2017), different orthographies (Caravolas et al., 2019; Furnes et al., 2019), and more. Nonetheless, the longitudinal relationship between kindergarten RAN and later reading is highly similar in magnitude to the relationship between auditory processing tasks and RD ($d = .8$ from the auditory processing meta-analysis converts to $r = .37$; $r = .38$ from the RAN meta-analysis converts to $d = .82$). The similarity in magnitude does not necessarily suggest that RAN and auditory processing play similarly important roles in reading development, but rather that the role of auditory processing in reading development needs further testing so that it can properly be factored into theoretical models or included in universal screening. In this conclusory chapter, I discuss the broader paradigms that underlie the auditory processing and RAN literatures and the future directions for each literature.

The end goal of research with RD, or any clinical population, is to improve outcomes for those affected. The route for how RAN might have the desired clinical impact is clear: it can be reliably measured before reading age, it is a significant predictor of reading outcomes many years into the future, and it explains unique variance as part of a larger pre-reading screening battery. These larger screening batteries are currently insufficient to identify children with RD, and additional behavioral or neural measures are needed to improve prediction (Norton & Wolf, 2012). Though research with behavioral auditory processing measures has traditionally focused on theoretical perceptual underpinnings of RD, the possibility of using these measures longitudinally from pre-reading age has only been explored in a handful of papers (Boets et al., 2011; Law et al., 2017). These papers show some promise in that some auditory processing measures (rise-time discrimination, speech-in-noise were not measured in our meta-analysis) predicts reading years into the future, but critical information, such as the reliability of these tasks in a clinical population of young children has been minimally explored in the published literature. Psychometric information, particularly in the form of standardized administration and measurement, age-norming, and published reliability are essential if these measures are going to contribute to a universal screening battery, especially considering the lack of published reliability of these tasks described by the study quality measures. Further, correlations among the auditory processing tasks and traditional pre-reading measures should be tested for multicollinearity in order to ensure that these tasks explain *unique* variance in reading outcomes. Finally, cut-off scores for both RAN and auditory processing tasks would need to be established by cross-validated classification models in order to make these models applicable to clinicians and educators.

Before the aforementioned future directions in the auditory processing and RD literature proceed, there will likely be considerable debate surrounding the legitimacy of our findings. These critiques are reasonable at first glance; in collapsing biological mechanisms, such as pitch perception above and below ~4kHz, or paradigms, such as “reference-free” discrimination (Ahissar et al., 2006; Oganian & Ahissar, 2012), potentially meaningful variability is washed away. Arguments may be made specifically about the large amounts of heterogeneity in the auditory processing literature, in which a comparison to the RAN meta-analysis reveals similar levels of heterogeneity in each literature. Despite this heterogeneity, the present data make it difficult to argue anything beyond a model of general frequency, duration, and intensity discrimination deficits of unknown cause. This is not to suggest that there is no effect of these potentially meaningful moderators, but that there is insufficient high-level evidence beyond the general deficits described in our meta-analysis.

In contrast to the many directions that auditory processing research could go, the future directions of RAN research are relatively clear. Expanding our findings to other languages is perhaps the most important of these directions, given the overrepresentation of English orthography in reading research (Share, 2008, 2021). The timing of reading acquisition, and thus how RAN relates to it, is considerably different in transparent orthographies and non-alphabetic orthographies (Araújo et al., 2015; Gottardo et al., 2021; Seymour et al., 2003). Accordingly, continued investigation of how the relationship between RAN and reading changes across time and orthography will be able to further disentangle the literature's reliance on English-language research for RAN theory, while also creating practical universal screening guidelines for children with RD in languages other than English. The advent of RVE meta-analysis models may be particularly useful in analyzing these nested, longitudinal data not only in the number of RAN and reading tasks, but also in those tasks measured at multiple time points. To my knowledge, this type of meta-analytic longitudinal path model has not been investigated in the RAN-reading literature, nor in RVE models more broadly and warrants further investigation.

In sum, RAN and the auditory processing tasks described in this dissertation are clear correlates of reading ability at distinct meta-scientific stages. They share a paradigm in which theoretical advances need to be accompanied by advancement toward earlier diagnosis and subsequent treatment of RD. Prospective diagnostics, and screening to a lesser extent, need measures that are reliable, stable, and valid; without well-developed and tested measures, these efforts will undoubtedly fail to improve outcomes for those with RD.

4.1 References

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. psyh. https://doi.org/10.1038/nn1800

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*. psyh. https://doi.org/10.1080/10888438.2019.1580284

Furnes, B., Elwér, Å., Samuelsson, S., Olson, R. K., & Byrne, B. (2019). Investigating the double-deficit hypothesis in more and less transparent orthographies: A longitudinal study from preschool to grade 2. *Scientific Studies of Reading*, *23*(6), 478–493. psyh. https://doi.org/10.1080/10888438.2019.1610410

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2016). The anatomy of the RAN-reading relationship. *Reading and Writing*, *29*(9), 1793–1815. psyh. https://doi.org/10.1007/s11145-016-9653-9

Gottardo, A., Chen, X., & Huo, M. R. Y. (2021). Understanding Within- and Cross-Language Relations Among Language, Preliteracy Skills, and Word Reading in Bilingual Learners:

Evidence From the Science of Reading. *Reading Research Quarterly, 56*(S1), S371–S390. https://doi.org/10.1002/rrq.410

Law, J. M., Vandermosten, M., Ghesquière, P., & Wouters, J. (2017). Predicting Future Reading Problems Based on Pre-reading Auditory Measures: A Longitudinal Study of Children with a Familial Risk of Dyslexia. *Frontiers in Psychology, 8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.00124

Norton, E. S., & Wolf, M. (2012). Rapid Automatized Naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427–452. psyh. https://doi.org/10.1146/annurev-psych-120710-100431

Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia, 50*(8), 1895–1905. https://doi.org/10.1016/j.neuropsychologia.2012.04.014

Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental Science*, *20*(5). eric. http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1152907&site=ehost-live

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*(4), 584–615. pdh. https://doi.org/10.1037/0033-2909.134.4.584

Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly*, *56*(S1), S391–S402. https://doi.org/10.1002/rrq.401

Appendix 1. NIH Study Quality Checklist

1. Was the research question or objective in this paper clearly stated?
2. Was the study population clearly specified and defined?
    a. Was dyslexia/reading disability (and "typical development") clearly defined using cutoff scores and appropriate measures? Was the control group free of RD subjects?
3. Was the participation rate of eligible persons at least 50%?
    a. You will usually answer "Not Applicable" here.
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?
    a. Does the dyslexia group have the same inclusion/exclusion criteria other than reading performance?
    b. Were dyslexics recruited from the same source as typical participants (e.g., from a classroom)?
5. Was a sample size justification, power description, or variance and effect estimates provided?
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?
    a. Not applicable
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?
    a. check "not applicable" in all cases
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
    a. Were the tests used to measure dyslexia (and other subgroupings, such as ADHD) specifically defined in detail?
10. Was the exposure(s) assessed more than once over time?
    a. Were the participants' reading skills (or dyslexia Yes status) assessed more than once over time?
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
    a. Were the auditory processing tasks clearly defined, valid, reliable, and implemented consistently across all study participants?
12. Were the outcome assessors blinded to the exposure status of participants?
    a. Were the assessors blind to the participants' diagnosis?
13. Was loss to follow-up after baseline 20% or less?
    a. (If the study was not longitudinal, select Not Applicable)
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?

a. Were key potential confounding variables (e.g., attention or language impairments, age) measured and adjusted statistically for their impact on the relationship dyslexia and auditory processing?

Guidance for Assessing the Quality of Observational Cohort and Cross-Sectional Studies

The guidance document below is organized by question number from the tool for quality assessment of observational cohort and cross-sectional studies.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher quality scientific research explicitly defines a research question.

Questions 2 and 3. Study population

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

You may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list.

If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias.

Question 4. Groups recruited from the same population and uniform eligibility criteria

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies–which is when data are obtained from the past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a "no."

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a "yes."

Question 5. Sample size justification

Did the authors present their reasons for selecting or recruiting the number of people included or analyzed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the methods section of the article may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be "yes."

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be "no." This is not a "fatal flaw." It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a prespecified question–i.e., it may have been an exploratory, hypothesis-generating study.

Question 6. Exposure assessed prior to outcome measurement

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to nonexposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioral factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be "yes," since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principal applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and nonexposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be "no."

Question 7. Sufficient timeframe to see an effect

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a "no" response.

Question 8. Different levels of the exposure of interest

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes–e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an "NA," and it should not count negatively towards the quality rating.

Question 9. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable–for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a "no" and the latter a "yes."

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers more frequently than those without elevated BP (nonexposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion.

Question 10. Repeated exposure assessment

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the followup period, compared to those who started out high then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design.

Question 11. Outcome measures

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable–for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death–the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a "yes." An example of a "no" would be self-report by subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented.

Question 12. Blinding of outcome assessors

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called "masking." The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As you assess this criterion, think about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants' exposure status. The committee would then be provided with copies of participants' medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for prespecified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark "NA" and explain the potential for bias.

Question 13. Followup rate

Higher overall followup rates are always better than lower followup rates, even though higher rates are expected in shorter studies, whereas lower overall followup rates are often seen in studies of longer duration. Usually, an acceptable overall followup rate is considered 80 percent or more of participants whose exposures were measured at baseline. However, this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent followup, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent followup rate.

Question 14. Statistical analyses

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome–that are not of interest to the research question–should be controlled for in the analyses.

For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders.

Some general guidance for determining the overall quality rating of observational cohort and cross-sectional studies

The questions on the form are designed to help you focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that you simply tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study–in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality. (Thus, the greater the risk of bias, the lower the quality rating of the study.)

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher quality the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding–all concepts reflected in the tool.

Generally, when you evaluate a study, you will not see a "fatal flaw," but you will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, you should ask yourself about the potential for bias in the study you are critically appraising. For any box where you check "no" you should ask, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, does this factor cause you to doubt the results that are reported in the study or doubt the ability of the study to accurately assess an association between exposure and outcome?

The best approach is to think about the questions in the tool and how each one tells you something about the potential for bias in a study. The more you familiarize yourself with the key concepts, the more comfortable you will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias.

Appendix 2. Table of Selected RAN-Reading Meta-Analysis References

Meta-Analysis Reference	Effect Size	Sample Size	English-only?	Longitudinal?	Key Details
Swanson et al., 2003	*r* = -.41	*n* = 2,257	Yes	No	First meta-analysis of RAN and reading, English-only
Swanson et al., 2003	*r* = -.45	*n* = 2,257	No	No	First meta-analysis of RAN and reading, English-only
Araújo et al., 2015	*r* = -.43	*n* = 28,826	No	No	Across all ages, languages, RAN, and reading measures
Hjetland et al., 2017	*r* = -.37	*n* = 3,285	No	Yes	Longitudinal from 1st grade or before, only including studies with a comprehension measure
Hjetland et al., 2017	*r* = -.34	*n* = 3,285	No	Yes	Longitudinal from 1st grade or before, only including studies with a comprehension measure
Kudo et al., 2015	*d* = -.89 (*r* = -.41 equivalent)	NR	No	No	RD vs. Average Achievement
Araújo et al., 2019	*d* = -1.19 (*r* = -.51 equivalent)	*n* = 22,418	No	No	Dyslexia vs. Controls

Appendix 3. List of References Included in Study 1

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. https://doi.org/10.1177/0022219410369067

Badian, N. A. (1994). Preschool prediction: Orthographic and phonological skills, and reading. *Annals of Dyslexia, 44*(1), 1-25.

Badian, N. A. (1998). A validation of the role of preschool phonological and orthographic skills in the prediction of reading. *Journal of Learning Disabilities, 31*(5), 472-481.

Badian, N. A., Duffy, F. H., Als, H., & McAnulty, G. B. (1991). Linguistic profiles of dyslexic and good readers. *Annals of Dyslexia, 41*, 221–245. https://doi.org/10.1007/BF02648088

Badian, N. A., McAnulty, G. B., Duffy, F. H., & Als, H. (1990). Prediction of dyslexia in kindergarten boys. *Annals of Dyslexia*, *40*, 152–169. https://doi.org/10.1007/BF02648146

Biddle, K. R. (1996). *Timing deficits in impaired readers: An investigation of visual naming speed and verbal fluency*. Thesis, Tufts University.

Bishop, A. G. (2003). Prediction of first-grade reading achievement: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly*, 26(3), 189-200.

Bishop, A. G., & League, M. B. (2006). Identifying a multivariate screening model to predict reading difficulties at the onset of kindergarten: A longitudinal analysis. *Learning Disability Quarterly*, 29(4), 235-252.

Burgoyne, K., Malone, S., Lervag, A., & Hulme, C. (2019). Pattern understanding is a predictor of early reading and arithmetic skills. *Early Childhood Research Quarterly*, *49*, 69–80. https://doi.org/10.1016/j.ecresq.2019.06.006

Caravolas, M., Lervag, A., Defior, S., Seidlova Malkova, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*(8), 1398–1407. https://doi.org/10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, *23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

Cardoso-Martins, C., & Pennington, B. F. (2004). The relationship between phoneme awareness and rapid serial naming skills and literacy acquisition: The role of developmental period and reading ability. *Scientific Studies of Reading*, *8*(1), 27–52. https://doi.org/10.1207/s1532799xssr0801_3

Carroll, J. M., Solity, J., & Shapiro, L. R. (2016). Predicting dyslexia using prereading skills: the role of sensorimotor and cognitive abilities. *Journal of Child Psychology and Psychiatry,* 57(6), 750-758.

Catts, H. W. (1991). Early identification of dyslexia: Evidence from a follow-up study of speech-language impaired children. *Annals of Dyslexia*, *41*(1), 163–177. https://doi.org/10.1007/BF02648084

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, *36*(5), 948–958. https://doi.org/10.1044/jshr.3605.948

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*(4), 331–361. https://doi.org/10.1207/s1532799xssr0304_2

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407–1425. https://doi.org/10.1007/s11145-015-9576-x

Catts, H. W., McIlraith, A., Bridges, M. S., & Nielsen, D. C. (2017). Viewing a phonological deficit within a multifactorial model of dyslexia. *Reading and Writing*, *30*(3), 613–629. https://doi.org/10.1007/s11145-016-9692-2

Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, *48*(3), 281–297. https://doi.org/10.1177/0022219413498115

Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education*, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B. F., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2015). Genetic and environmental etiologies of the longitudinal relations between prereading skills and reading. *Child Development*, *86*(2), 342–361. https://doi.org/10.1111/cdev.12295

Cirino, P. T., Child, A. E., & Macdonald, K. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002

Cronin, V., & Carver, P. (1998). Phonological sensitivity, rapid naming, and beginning reading.

Applied Psycholinguistics, *19*(3), 447–461. https://doi.org/10.1017/S0142716400010262

Cronin, V. S. (2013). RAN and double-deficit theory. *Journal of Learning Disabilities*, *46*(2), 182–190. https://doi.org/10.1177/0022219411413544

Dally, K. (2006). The influence of phonological processing and inattentive behavior on reading acquisition. *Journal of Educational Psychology*, *98*(2), 420–437. https://doi.org/10.1037/0022-0663.98.2.420

Dittman, C. K. (2016). The impact of early classroom inattention on phonological processing and word-reading development. *Journal of Attention Disorders*, *20*(8), 653–664. https://doi.org/10.1177/1087054713478979

Ellis, N., & Large, B. (1988). The early stages of reading: A longitudinal study. *Applied Cognitive Psychology*, *2*(1), 47–76. https://doi.org/10.1002/acp.2350020106

Evans, M. A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds and phonological awareness: An examination of kindergarten children across letters and of letters across children. *Reading and Writing: An Interdisciplinary Journal*, *19*(9), 959–989.

Evans, M. A., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *54*(2), 65–75.

Felton, R. H. (1992). Early identification of children at risk for reading disabilities. *Topics in Early Childhood Special Education*, *12*(2), 212-229.

Felton, R. H., & Brown, I. S. (1990). Phonological processes as predictors of specific reading skills in children at risk for reading failure. *Reading and Writing*, *2*(1), 39–59. https://doi.org/10.1007/BF00383373

Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading*, *32*(3), 275–292. https://doi.org/10.1111/j.1467-9817.2009.01393.x

Furnes, B., & Samuelsson, S. (2011). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, *21*(1), 85–95. https://doi.org/10.1016/j.lindif.2010.10.005

Georgiou, G. K., Parrila, R., & Kirby, J. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading*, *10*(2), 199–220. https://doi.org/10.1207/s1532799xssr1002_4

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing: An Interdisciplinary Journal*, *25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Gunn, B., Smolkowski, K., & Vadasy, P. (2010). Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, *4*(1), 53–86. https://doi.org/10.1080/19345747.2010.488716

Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology*, *101*(2), 388–402. https://doi.org/10.1037/a0015115

Heath, S. M., & Hogben, J. H. (2004). Cost-effective prediction of reading difficulties. *Journal of Speech, Language, and Hearing Research*, *47*(4), 751–765. https://doi.org/10.1044/1092-4388(2004/057)

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing*, *12*(1), 99–128. https://doi.org/10.1023/A:1008033824385

Hood, M. H. (2005). *The role of temporal and phonological processing in early reading development: A longitudinal study*. Thesis, Griffith University. https://doi.org/10.25904/1912/2673

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Inoue, T., Georgiou, G. K., Parrila, R., & Kirby, J. R. (2018). Examining an extended home literacy model: The mediating roles of emergent literacy skills and reading fluency. *Scientific Studies of Reading, 22*(4), 273-288. doi.org/10.1080/10888438.2018.1435663

Kirby, J. R., Ball, A., Geier, B. K., Parrila, R., & Wade-Woolley, L. (2011). The development of reading interest and its relation to reading ability. *Journal of Research in Reading, 34*(3), 263-280.

Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P. G., & Landerl, K. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, *45*(3), 341–362. https://doi.org/10.1598/RRQ.45.3.4

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, *95*(3), 453–464. https://doi.org/10.1037/0022-0663.95.3.453

Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences*, *16*(3), 195–216. https://doi.org/10.1016/j.lindif.2005.12.001

Ledesma, H. M. L. (2002). *Language factors influencing early reading development in bilingual (Filipino-English) boys*. Thesis, Georgia State University.

Lewis, B. A., Avrich, A. A., Freebairn, L. A., Hansen, A. J., Sucheston, L. E., Kuo, I., Taylor, H., Iyengar, S., & Steina, C. (2011). Literacy outcomes of children with early childhood speech sound disorders: Impact of endophenotypes. *Journal of Speech, Language, and Hearing Research*, *54*(6), 1628–1643. https://doi.org/10.1044/1092-4388(2011/10-0124)

Macdonald, H. H., Sullivan, A. L., & Watkins, M. W. (2013). Multivariate screening model for later word reading achievement: Predictive utility of prereading skills and cognitive ability. *Journal of Applied School Psychology*, *29*(1), 52–71. https://doi.org/10.1080/15377903.2013.751476

Mazzocco, M. M. M., & Grimm, K. J. (2013). Growth in rapid automatized naming from grades K to 8 in children with math or reading disabilities. *Journal of Learning Disabilities*, *46*(6), 517–533. https://doi.org/10.1177/0022219413477475

McIlraith, A. L. and Language and Reading Research Consortium (2018). Predicting word reading ability: A quantile regression study. *Journal of Research in Reading*, *41*(1), 79–96. https://doi.org/10.1111/1467-9817.12089

McMillen, S. M. (2018). Pescado or fish? *Rapid automatic naming performance for young Spanish-speaking English language learners*. Thesis, University of Memphis.

Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental*

Science, *20*(5). https://doi.org/10.1111/desc.12471

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading*, *8*(1), 3–26. https://doi.org/10.1207/s1532799xssr0801_2

Petersen, D. B., & Gillam, R. B. (2013). Accurately predicting future reading difficulty for bilingual Latino children at risk for language impairment. *Learning Disabilities Research & Practice*, 28(3), 113–128. https://doi.org/10.1111/ldrp.12014

Peterson, R. L., Arnett, A. B., Pennington, B. F., Byrne, B., Samuelsson, S., & Olson, R. K. (2018). Literacy acquisition influences children's rapid automatized naming. *Developmental Science*, *21*(3), 1–9. https://doi.org/10.1111/desc.12589

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities Research Reviews*, *2*(1), 54–63. https://doi.org/10.1002/(SICI)1098-2779(1996)2:1<54::AID-MRDD9>3.0.CO;2-X

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Smith, S. L., Scott, K. A., Roberts, J., & Locke, J. L. (2008). Disabled readers' performance on tasks of phonological processing, rapid naming, and letter knowledge before and after kindergarten. *Learning Disabilities Research & Practice*, *23*(3), 113–124. https://doi.org/10.1111/j.1540-5826.2008.00269.x

Snowling, M. J., Lervåg, A., Nash, H. M., & Hulme, C. (2019). Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Developmental Science, 22*(1), e12723.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5), 276–286. https://doi.org/10.1177/002221949402700503

Uhry, J. K. (2002). Kindergarten phonological awareness and rapid serial naming as predictors of Grade 2 reading and spelling. In E. Witruk, A. D. Friederici, & T. Lachmann (Eds.), *Basic functions of language, reading and reading disability.* (pp. 299–313). Kluwer Academic. https://doi.org/10.1007/978-1-4615-1011-6_18

Vadasy, P. F., & Sanders, E. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading and Writing*, *21*(9), 929–963. https://doi.org/10.1007/s11145-008-9119-9

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*(1), 73–87. https://doi.org/10.1037/0012-1649.30.1.73

Wolf, M. (1984). Naming, reading, and the dyslexias: A longitudinal overview. *Annals of Dyslexia*, *34*(1), 87–115. https://doi.org/10.1007/BF02663615

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. *Child Development, 57*(4), 988–1000. https://doi.org/10.1111/j.1467-8624.1986.tb00260.x

Wolter, J. A, Self, T., & Apel, K. (2011). Initial mental graphemic representation acquisition and later literacy achievement in children with language impairment: A longitudinal study.

Journal of Learning Disabilities, *44*(6), 543–555.

https://doi.org/10.1177/0022219410392042

Appendix 4. List of References Included in Study 2

*Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. https://doi.org/10.1038/nn1800

*Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory processing deficits in reading disabled adults. *Journal of the Association for Research in Otolaryngology*, *3*(3), 302–320. https://doi.org/10.1007/s101620010093

*Amitay, S., Ben-Yehudah, G., Banai, K., & Ahissar, M. (2002). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain*, *125*(10), 2272–2285. https://doi.org/10.1093/brain/awf231

*Banai, K., & Ahissar, M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiology & Neurotology*, *9*(6), 328–340. https://doi.org/10.1159/000081282

*Banai, K., & Ahissar, M. (2005). Psychoacoustics and working memory in dyslexia. In J. Syka & M. M. Merzenich (Eds.), *Plasticity and signal representation in the auditory system.* (2005-03154-021; pp. 233–242). Springer Publishing Co; https://doi.org/10.1007/0-387-23181-1_21

*Beattie, R. L., & Manis, F. R. (2013). Rise time perception in children with reading and combined reading and language difficulties. *Journal of Learning Disabilities*, *46*(3), 200–209. https://doi.org/10.1177/0022219412449421

*Ben-Yehudah, G., Banai, K., & Ahissar, M. (2004). Patterns of deficit in auditory temporal processing among dyslexic adults. *NeuroReport: For Rapid Communication of*

Neuroscience Research, *15*(4), 627–631. https://doi.org/10.1097/00001756-200403220-00011

*Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020

*Breier, J. I., Fletcher, J. M., Foorman, B. R., Klaas, P., & Gray, L. C. (2003). Auditory temporal processing in children with specific reading disability with and without attention deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research*, *46*(1), 31–42. eric.

*Cacace, A. T., McFarland, D. J., Ouimet, J. R., Schrieber, E. J., & Marro, P. (2000). Temporal processing deficits in remediation-resistant reading-impaired children. *Audiology and Neurotology*, *5*(2), 83–97. https://doi.org/10.1159/000013871

*Chaubet, J., Pereira, L., & Perez, A. P. (2014). Temporal resolution ability in students with dyslexia and reading and writing disorders. *International Archives of Otorhinolaryngology*, *18*, 146–149. https://doi.org/10.1055/s-0033-1363465

*Colling, L. J., Noble, H. L., & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, *11*. https://www.frontiersin.org/article/10.3389/fnins.2017.00360

*Farmer, M. E., & Klein, R. (1993). Auditory and visual temporal processing in dyslexic and normal readers. *Annals of the New York Academy of Sciences*, *682*(1), 339-341. https://doi.org/10.1111/j.1749-6632.1993.tb22987.x

*Fostick, L., Bar-El, S., & Ram-Tsur, R. (2012). Auditory temporal processing as a specific deficit among dyslexic readers. *Journal of Psychology Research*, *2*(2). https://doi.org/10.17265/2159-5542/2012.02.001

*Fraser, J., Goswami, U., & Conti-Ramsden, G. (2010). Dyslexia and specific language impairment: The role of phonology and auditory processing. *Scientific Studies of Reading*, *14*(1), 8-29.

*Georgiou, G. K., Protopapas, A., Papadopoulos, T. C., Skaloumbakas, C., & Parrila, R. (2010). Auditory temporal processing and dyslexia in an orthographically consistent language. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *46*(10), 1330–1344. https://doi.org/10.1016/j.cortex.2010.06.006

*Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, *23*(4), 621–642. https://doi.org/10.1080/02643290500412545

*Gokula, R., Sharma, M., Cupples, L., & Valderrama, J. T. (2019). Comorbidity of auditory processing, attention, and memory in children with word reading difficulties. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02383

*Gooch, D., Snowling, M., & Hulme, C. (2011). Time perception, phonological skills and executive function in children with dyslexia and/or ADHD symptoms. *Journal of Child Psychology and Psychiatry*, *52*(2), 195–203. https://doi.org/10.1111/j.1469-7610.2010.02312.x

*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. https://doi.org/10.1111/j.1467-7687.2010.00955.x

*Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019. https://doi.org/10.1007/s11145-009-9186-6

*Goswami, U., Huss, M., Mead, N., & Fosker, T. (2021). Auditory sensory processing and phonological development in high IQ and exceptional readers, typically developing readers, and children with dyslexia: A longitudinal study. *Child Development*, *92*(3), 1083–1098. https://doi.org/10.1111/cdev.13459

*Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, *49*(5), 1363–1376. https://doi.org/10.1016/j.cortex.2012.05.005

*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. https://doi.org/10.1016/j.jml.2013.03.001

*Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

*Halliday, L. F., & Bishop, D. V. M. (2006). Auditory frequency discrimination in children with dyslexia. *Journal of Research in Reading*, *29*(2), 213–228. https://doi.org/10.1111/j.1467-9817.2006.00286.x

*Hämäläinen, J. A., Fosker, T., Szücs, D., & Goswami, U. (2011). N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. *International Journal of Psychophysiology*, *81*(1), 51–59. https://doi.org/10.1016/j.ijpsycho.2011.04.005

*Hari, R., Sääskilahti, A., Helenius, P., & Uutela, K. (1999). Non-impaired auditory phase locking in dyslexic adults. *NeuroReport*, *10*(11), 2347–2348.

*Heath, S. M., Bishop, D. V. M., Hogben, J. H., & Roach, N. W. (2006). Psychophysical indices of perceptual functioning in dyslexia: A psychometric analysis. *Cognitive Neuropsychology*, *23*(6), 905–929. https://doi.org/10.1080/02643290500538398

*Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674–689. https://doi.org/10.1016/j.cortex.2010.07.010

*Ingelghem, M. V., Van Wieringen, A., Wouters, J., Vandenbussche, E., Onghena, P., & Ghesquière, P. (2001). Psychophysical evidence for a general temporal processing deficit in children with dyslexia. *Neuroreport*, *12*(16), 3603-3607.

*Ingelghem, M. V., Boets, B., van Wieringen, A., Onghena, P., Ghesquière, P., Vandenbussche, E., & Wouters, J. (2005). *An auditory temporal processing deficit in children with dyslexia?* Unpublished manuscript.

*King, W. M., Lombardino, L. J., Crandell, C. C., & Leonard, C. M. (2003). Comorbid auditory processing disorder in developmental dyslexia. *Ear and Hearing*, *24*(5), 448–456. https://doi.org/10.1097/01.AUD.0000090437.10978.1A

*Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia.

Frontiers in Human Neuroscience, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00482

*Law, J. M., Vandermosten, M., Ghesquière, P., & Wouters, J. (2017). Predicting future reading problems based on pre-reading auditory measures: A longitudinal study of children with a familial risk of dyslexia. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.00124

*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. https://doi.org/10.1016/j.jml.2010.09.003

*Matthews, N. H. (2013). *Musical beat perception in children with developmental dyslexia—Effects of pitch and training*. University of Cambridge.

*McArthur, G. M., Atkinson, C. M., & Ellis, D. (2010). Can training normalize atypical passive auditory ERPs in children with SRD or SLI? *Developmental Neuropsychology*, *35*(6), 656–678. https://doi.org/10.1080/87565641.2010.508548

*McArthur, G. M., & Hogben, J. H. (2001). Auditory backward recognition masking in children with a specific language impairment and children with a specific reading disability. *The Journal of the Acoustical Society of America*, *109*(3), 1092–1100. https://doi.org/10.1121/1.1338559

*McCroskey, R. L., & Kidder, H. C. (1980). Auditory fusion among learning disabled, reading disabled, and normal children. *Journal of Learning Disabilities*, *13*(2), 69–76. https://doi.org/10.1177/002221948001300205

*Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia*, *50*(8), 1895–1905. https://doi.org/10.1016/j.neuropsychologia.2012.04.014

*Papadopoulos, T. C., Georgiou, G. K., & Parrila, R. K. (2012). Low-level deficits in beat perception: Neither necessary nor sufficient for explaining developmental dyslexia in a consistent orthography. *Research in Developmental Disabilities*, *33*(6), 1841–1856. https://doi.org/10.1016/j.ridd.2012.04.009

*Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory processing of amplitude envelope rise time in adults diagnosed with developmental dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286. https://doi.org/10.1080/10888430701344280

*Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., & Wouters, J. (2011). Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities*, *32*(6), 2810–2819. https://doi.org/10.1016/j.ridd.2011.05.025

*Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865. https://doi.org/10.1093/brain/awg076

*Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, *10*(3), 215–233. https://doi.org/10.1002/dys.276

*Rocheron, I., Lorenzi, C., Füllgrabe, C., & Dumont, A. (2002). Temporal envelope perception in dyslexic children. *NeuroReport*, *13*(13), 1683–1687.

*Santurette, S., Poelmans, H., Luts, H., Ghesquiére, P., Wouters, J., & Dau, T. (2010). Detection and identification of monaural and binaural pitch contours in dyslexic listeners. *Journal of the Association for Research in Otolaryngology*, *11*(3), 515–524. https://doi.org/10.1007/s10162-010-0216-5

*Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Role of auditory temporal processing for reading and spelling disability. *Perceptual and Motor Skills*, *86*(3, Pt 1), 1043–1047. https://doi.org/10.2466/pms.1998.86.3.1043

*Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1999). The role of phonological awareness, speech perception, and auditory temporal processing for dyslexia. *European Child & Adolescent Psychiatry*, *8*(Suppl 3), III/28-III/34. https://doi.org/10.1007/PL00010690

*Sharma, M., Purdy, S. C., & Kelly, A. S. (2009). Comorbidity of auditory processing, language, and reading disorders. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(3), 706–722. https://doi.org/10.1044/1092-4388(2008/07-0226)

*Sharma, M., Purdy, S. C., Newall, P., Wheldall, K., Beaman, R., & Dillon, H. (2006). Electrophysiological and behavioral evidence of auditory processing deficits in children with reading disorder. *Clinical Neurophysiology*, *117*(5), 1130–1144. https://doi.org/10.1016/j.clinph.2006.02.001

*Talcott, J. B., Witton, C., Hebb, G. S., Stoodley, C. J., Westwood, E. A., France, S. J., Hansen, P. C., & Stein, J. F. (2002). On the relationship between dynamic visual and auditory processing and literacy skills; results from a large primary-school study. *Dyslexia*, *8*(4), 204–225. https://doi.org/10.1002/dys.224

*Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, *29*(3), 334–348. https://doi.org/10.1111/j.1467-9817.2006.00312.x

*Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, *102*(1), 120–129. https://doi.org/10.1016/j.jphysparis.2008.03.007

*Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, *26*(2), 139–161. https://doi.org/10.1007/s11145-012-9359-6

*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. https://doi.org/10.1177/0022219417712018

*Toplak, M. E., Rucklidge, J. J., Hetherington, R., John, S. C. F., & Tannock, R. (2003). Time perception deficits in attention-deficit/ hyperactivity disorder and comorbid reading difficulties in child and adolescent samples. *Journal of Child Psychology and Psychiatry*, *44*(6), 888–903. https://doi.org/10.1111/1469-7610.00173

*Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities*, *33*(2), 635–644. https://doi.org/10.1016/j.ridd.2011.11.005

*Walker, M. M., Shinn, J. B., Cranford, J. L., Givens, G. D., & Holbert, D. (2002). Auditory temporal processing performance of young adults with reading disorders. *Journal of*

Speech, Language, and Hearing Research, *45*(3), 598–605. https://doi.org/10.1044/1092-4388(2002/048)

*Wang, H.-L. S., Huss, M., Hämäläinen, J. A., & Goswami, U. (2012). Basic auditory processing and developmental dyslexia in Chinese. *Reading and Writing*, *25*(2), 509–536. https://doi.org/10.1007/s11145-010-9284-5

*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. https://doi.org/10.1016/j.ridd.2019.01.006

*Willburger, E., & Landerl, K. (2010). Anchoring the deficit of the anchor deficit: Dyslexia or attention? *Dyslexia*, *16*(2), 175–182. https://doi.org/10.1002/dys.404

*Zaidan, E. (2009). *An investigation of temporal resolution abilities in school-aged children with and without dyslexia* (2010-99051-126; Issues 9-A) [ProQuest Information & Learning]. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-99051-126&site=ehost-live

*Zaidan, E., & Baran, J. A. (2013). Gaps-in-noise (GIN©) test results in children with and without reading disabilities and phonological processing deficits. *International Journal of Audiology*, *52*(2), 113–123. https://doi.org/10.3109/14992027.2012.733421

*Zhang, M., Xie, W., Xu, Y., & Meng, X. (2018). Auditory temporal perceptual learning and transfer in Chinese-speaking children with developmental dyslexia. *Research in Developmental Disabilities*, *74*, 146–159. https://doi.org/10.1016/j.ridd.2018.01.005