Abstract

Title of dissertation: EXAMINING THE DISPROPORTIONATE REPRESENTATION OF BILINGUAL CHILDREN IN SPECIAL EDUCATION

José A. Ortiz, Doctor of Philosophy, 2022

Dissertation directed by: Associate Professor Kelli Cummings
Department of Counseling, Higher, and Special Education

Chapter II

Nonword repetition has been endorsed as a less biased method of assessment for children from culturally and linguistically diverse backgrounds, but there are currently no systematic reviews or meta-analyses on its use with bilingual children. The purpose of this study is to evaluate diagnostic accuracy of nonword repetition in the identification of language impairment (LI) in bilingual children. Using a keyword search of peer-reviewed literature from several large electronic databases, as well as ancestral and forward searches, 13 studies were identified that met the eligibility criteria. Studies were evaluated on the basis of quality of evidence, design characteristics, and reported diagnostic accuracy. A meta-regression analysis, based on study results, was conducted to identify task characteristics that may be associated with better classification accuracy. Diagnostic accuracy across studies ranged from poor to good. Bilingual children with LI performed with more difficulty on nonword repetition tasks than those with typical language. Quasiuniversal tasks, which account for the phonotactic constraints of multiple languages, exhibited better diagnostic accuracy and resulted in less misidentification of children with typical language than language-specific tasks. Evidence suggests that nonword repetition may be a useful tool in the assessment and screening of LI in bilingual children, though it should be used in conjunction with other measures. Quasiuniversal tasks demonstrate the potential to further reduce assessment bias, but extant research is limited.

Chapter III

The disproportionate identification of language-related disorders in schools, including communication disorders and specific learning disability, is an ongoing problem for bilingual children, with evidence of both over- and underrepresentation. Previous research has uncovered distinct identification patterns for emergent and English-proficient bilinguals, as well as differences in identification rates across grades. However, there is limited information about disability identification for different groups of bilinguals across grades. Thus, the purpose of this study is to examine the prevalence and incidence of language-related disorders in emergent and English-proficient bilinguals in elementary school. Using a nationally representative, individual-level, longitudinal data set, this study examined representation in language-related disorder categories, as well as identification rates by year. This study also examined individual- and school-level predictors of disability identification for bilingual children.

Results indicate that emergent and English-proficient bilinguals exhibit distinct patterns of language-related disorder identification. Emergent bilinguals experienced a disproportionate increase in disability identification rates in third grade, resulting in significant overrepresentation in subsequent grades. By fifth grade, emergent bilinguals experienced approximately twice the odds of being identified with a language-related disorder, compared to monolinguals. English-proficient bilinguals, on the other hand, were underrepresented in language-related disorder categories in early elementary school grades, but experienced identification rates similar to monolinguals by fifth grade. Outcomes from this study provide insight into patterns of language-related disorder identification for bilinguals that have not been addressed in previous research. The implications for education practice and policy are discussed.

Chapter IV

The disproportionate representation of bilingual children in special education is an ongoing issue in US schools, with evidence of both over- and underrepresentation. Identification rates of language-related disorders, including communication disorders and specific learning disability, are particularly relevant for bilingual children given the challenges associated with differentiating language difference from disorder and the possibility of misidentification. School-based speech-language pathologists are well positioned to address the issue, but many do not engage in practices that may reduce disproportionate disability identification.

The purpose of this practitioner paper is to provide school-based clinicians with an evidence-based model for addressing disproportionality in bilingual children, with a focus on prevention. This paper provides a review of the literature on the topic and integrates information from relevant studies to provide a clear depiction of the nature of the problem. In addition, this paper describes a model of disproportionality prevention, and provides a set of evidence-based methods that clinicians can employ. Topics include, pre-referral intervention, early identification, parent engagement, and collaboration. By adopting the methods described in this paper, school-based speech-language pathologists can strengthen their ability to meaningfully address many of the issues that contribute to over- and underrepresentation of bilingual children in special education.

EXAMINING THE DISPROPORTIONATE REPRESENTATION OF BILINGUAL CHILDREN IN SPECIAL EDUCATION

By

José A. Ortiz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Fall 2022

Advisory Committee:
Associate Professor Kelli Cummings
Associate Professor Jason Chow
Professor Susan De La Paz
Associate Professor Yi Ting Huang
Professor Ana Taboada Barber

Acknowledgements

I would like to thank all who supported me in the writing of this dissertation. Thank you to Kelli Cummings for serving as my mentor in this program. I am incredibly appreciative of the guidance, feedback, and insight that she has offered, and I am very fortunate for having had the opportunity to work with her. A special thank you to my wife, Mónica L. Caudillo, who has been a constant source of support throughout my program.

Table of Contents

List of Tables

Chapter II

Chapter III

List of Figures

Chapter I: Introduction

English language learners (ELLs) comprise approximately 10% of the public school population in the US (National Center for Education Statistics, 2021). Despite their large and growing numbers, children with limited English proficiency children exhibit poorer outcomes than their peers in reading, math, and science (National Center for Education Statistics, 2021). These children also frequently experience a lack of equitable access to special education services. This is particularly evident in the disproportionate identification of language-related disorders—speech or language impairments, hereafter referred to as communication disorders (CD), and specific learning disability (SLD) in this population—(e.g., P. L. Morgan, Farkas, et al., 2017; Sullivan, 2011; Umansky et al., 2017). Previous research into disproportionality has identified evidence of both overidentification (Samson & Lesaux, 2009; Yamasaki & Luk, 2018), as well as underidentification (P. L. Morgan, Farkas, et al., 2017; Umansky et al., 2017). Both of these patterns may result in problematic outcomes for students. Although underidentification may result in the delay or lack of receipt of needed services, overidentification suggests that many children may be receiving unnecessarily services, which is concerning given the potential for long-term negative outcomes associated with being placed in special education. Children in special education exhibit higher drop-out rates (Kemp, 2006), are less likely to enroll in post-secondary education (Wagner et al., 2006), and are at an increased risk of incarceration (Bell, 2016). Given the risks associated with misidentification, the precision with which children are identified with disabilities is of critical importance.

Despite having been identified over 50 years ago (Dunn, 1968), the disproportionate representation of children from racial-, ethnic-, and language-minority groups continues to be a pervasive issue in special education. Although federal policy has largely focused on addressing overrepresentation (US Department of Education, 2016a, 2016b), a growing research base has increasingly revealed a more complex issue. Although our understanding of disproportionality has

substantially improved in the past decades, the apparently conflicting evidence demonstrating both underrepresentation and overrepresentation has resulted in a lack of consensus about the best way to address the issue (Skiba et al., 2016). The importance of a clear understanding of representation patterns cannot be understated given the implications for practice as well as policy (Shakir, 2018). Regardless of the form that it takes, disproportionality represents a manifestation of a lack of equitable access to education for students from minority backgrounds. A better understanding of the problem, as well as its underlying mechanisms, will provide insight into possible solutions. What follows is a description of the relevant factors that contribute to disproportionality, a discussion of policy and practice issues, followed by an overview of the methods used to further investigate the topic.

Patterns of Disproportionality

Disproportionality has commonly been argued to predominantly be a problem of overrepresentation (Artiles & Trent, 1994; Coutinho & Oswald, 2000; Dunn, 1968), reflecting a scenario in which the proportion of minority students identified with a disability differs from the expected proportion given the relative size of the minority student population. Although there is evidence of overrepresentation of bilingual children in special education (Artiles et al., 2005; Sullivan, 2011), several studies have demonstrated just the opposite (P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017). A number of explanatory factors could account for these apparent discrepancies, such as the language background of students, grade level, and disability category.

Language Background

The variability in disability identification rates for different bilingual groups, such as English-proficient bilinguals and emergent bilinguals, illustrates the importance of considering language background in the examination of disproportionality (Yamasaki & Luk, 2018). One of the challenges in bilingual research is the lack of a shared definition of the population under investigation. Studies of

disproportionality have used a variety of different identifiers for children who speak multiple languages, such as *ELL, language-minority,* and *bilingual.* Although there is some overlap in these classifications, they each include unique groups of children not shared by the other categories. For this reason, in the subsequent sections I will use each of these terms when appropriate for the context being described. The term *ELL* refers specifically to the educational classification used to designate children with limited English proficiency. *Language-minority* refers to children who were not born in an English-speaking country or whose dominant home language is not English (Artiles et al., 2005). This term is useful when describing the contrasting experiences and outcomes of minority- and majority-language speakers. The term *bilingual,* which captures the broadest range of students and is inclusive of both language-minority and ELL children, refers to those children who have learned or are learning multiple languages; this term is synonymous with the *dual language learner* (Paradis et al., 2011). Although majority-language speakers learning a minority language can be considered to be bilingual, for the purposes of the present work the term bilingual will refer specifically to children from language minority backgrounds. This term bilingual accounts for a spectrum of ability across languages, comprising emergent bilinguals as well as English-proficient bilinguals (García et al., 2008). Although emergent bilinguals are analogous to ELLs, English-proficient bilinguals include children who are simultaneous bilinguals from birth, as well as sequential bilinguals whose English proficiency exceeds the criterion to be considered an ELL.

Although bilinguals are commonly treated as a monolithic group, there is an emerging trend toward accounting for a greater level of specificity of language background in disproportionality research (Artiles et al., 2005; Umansky et al., 2017; Yamasaki & Luk, 2018). This is an especially important consideration given how an individual's unique language background may influence their performance on language assessment, a fundamental component of disability evaluations (Bedore et al., 2018; Gillam et al., 2013; Peña et al., 2016). English-proficient bilinguals may exhibit an entirely different pattern of performance from emergent bilinguals on both academic tasks as well as language

assessment. Because of these potential performance differences, English-proficient bilinguals may also exhibit a different level of risk for disability identification than emergent bilinguals.

Several studies have addressed bilingual heterogeneity by accounting for language background in their analytic approaches. Umansky et al., (2017), for example, applied an ever-ELL framework in their study of disproportionality in special education, using a longitudinal sample comprising 13 years of data from elementary, middle and high school. Using this framework, the authors distinguished between current-ELLs and ever-ELLs. Current-ELLs were those students who were classified as ELLs at the time of analysis, whereas ever-ELLs were those students who had ever been classified as ELLs at any point in time. This approach accounts for the lack of stability in the ELL population, as children may be reclassified as their English proficiency improves. The authors identified differences in the patterns of special education representation for ever-ELLs compared to current-ELLs, observing that current-ELLs were overrepresented at the secondary level, but ever-ELLs were consistently underrepresented across grades. Yamasaki and Luk (2018) similarly identified differences in special education representation for emergent bilinguals and English-proficient bilinguals in their study of elementary school students in a school district in Massachusetts. Although English-proficient bilinguals were consistently underrepresented in special education, emergent bilinguals exhibited a shift from under- to overrepresentation as they progressed through grades.

The relationship between language background and family immigration status may also play a role in the identification of disabilities at different grade levels, given the likelihood that children with limited English proficiency are from immigrant families. Hibel and Jasper (2012) examined the risk of being identified with SLD for children of immigrants, from kindergarten to fifth grade. Using a nationally representative, individual-level sample, the authors found that children of immigrants initially exhibited a lower likelihood of being identified with SLD, but identification rates increased over time. The authors also reported that, after controlling for family immigration status, language-minority status, defined as

participation in English as a second language services, was not a significant predictor of SLD identification. Results of this study suggest that family immigration status may play a key role in predicting risk for disability identification among some bilingual populations.

Grade Level

One particularly important dimension to consider is the grade level at which disability identification rates are measured. Relative rates of disability identification may vary from grade to grade, making outcomes derived from a single grade difficult to generalize to other grades. Although most studies have focused on elementary school, rates of identification for bilinguals may differ substantially even from early elementary grades to later ones. The potential for change in rates of special education identification as students progress through school illustrates the importance of considering grade level in the examination of disability identification patterns.

Several studies have used longitudinal samples to demonstrate how the likelihood of disability identification may increase as students progress through school. Samson and Lesaux (2009), in their evaluation of disproportionality of bilingual students from kindergarten to third grade, identified growth in relative rates of identification, such that bilingual children were underrepresented in special education in kindergarten, but overrepresented by third grade. This change from under- to overrepresentation for bilingual students has also been demonstrated in several other studies (Artiles et al., 2005; Umansky et al., 2017; Yamasaki & Luk, 2018). Results from studies such as these demonstrate the value of accounting for changes in identification patterns across grades.

Disability Category

The likelihood of being identified with a disability differs depending on the disability category, and different minority groups may also experience a systematically greater likelihood of identification for certain disabilities (P. L. Morgan et al., 2015; Sullivan, 2011; Yamasaki & Luk, 2018). For bilingual

children, language-related disorders (CD and SLD) are of particular relevance given the challenge of differentiating between language differences and language disorders (Bedore & Peña, 2008; Castilla-Earls et al., 2020). Children with limited English proficiency are often disproportionately identified with these disabilities (P. L. Morgan, Farkas, et al., 2017; Sullivan, 2011; Umansky et al., 2017). The identification of both CD and SLD relies on the measurement of skills related to language ability, including verbal expression, comprehension, reading, and writing, as well as associated cognitive areas such as executive functioning and working memory. Identification also depends on the expertise of team members from a number of different disciplines, including general education teachers, special education teachers, school psychologists, and speech-language pathologists. For bilingual children, it is possible that common characteristics associated with typical dual language development may be mistaken as symptoms of an underlying disability, particularly when standardized assessment is used (Anaya et al., 2018; Barragan et al., 2018; Bedore et al., 2005). For these reasons, any discussion of disproportionate representation of bilingual children in special education should specifically consider the identification of CD and SLD, and the evaluation process used to identify these disabilities.

The evidence for disproportionality of bilinguals in these categories is somewhat mixed with respect to the pattern of representation and rates of identification. Morgan et al. (2017) examined the prevalence of CD in two cross-sectional cohorts of language-minority kindergarten students, and found evidence of underrepresentation in both cohorts. Across the two cohorts, language-minority students exhibited 43-50% lower odds of being identified with a CD compared to language-majority students. Similarly, Morgan et al. (2015), using discrete-time modeling to examine the incidence of language-related disorders from kindergarten to eighth grade, also identified evidence of lower rates of identification for language-minority students in the categories of both CD and SLD. In this study, language-minority students exhibited 40% lower odds of being identified with CD and 28% lower odds of being identified with SLD, compared to language-majority students. Although these nationally

representative studies do provide evidence for underrepresentation of bilinguals in these disability categories, other studies using district- and state-level data have provided evidence of increasing patterns of risk for identification over time. Sullivan (2011), for example, found that while emergent bilinguals were consistently overrepresented in the SLD category over an eight year period in the school districts studied, CD identification grew over time. Although emergent bilinguals were initially underrepresented in the CD category, school districts were increasingly likely to overidentify these students over the years studied. Umansky et al. (2017), in their examination of disproportionality in two states over an eight year period, identified evidence of underidentification of ever-ELLs, relative to non-ELLs, in one state but not the other. The authors also found that the incidence of SLD identification increased over time, such that ever-ELLs were overidentified by middle school. Despite evidence of underrepresentation from nationally representative studies, the examination of local contexts suggests a more nuanced pattern of representation with substantial variation across district and state lines.

Mechanisms of Disproportionate Representation

Several potential mechanisms have been theorized to explain how bilingual children are disproportionately represented in special education. The empirical evidence for the existence of both overrepresentation and underrepresentation suggests that there are likely multiple related casual mechanisms at play. Overrepresentation in special education is often argued to be the result of cultural and linguistic bias, the presence of which systematically drives up disability identification for bilinguals (Abedi, 2009; Sanatullova-Allison & Robison-Young, 2016; Skiba, 2002). Overrepresentation, resulting from misidentification, may stem from the use of inadequate or poorly designed assessment tools (Barragan et al., 2018; Macswan & Rolstad, 2006), a misapplication of standardized assessment (Laing & Kamhi, 2003; Restrepo & Silverman, 2001), a reliance on assessment of knowledge rather than learning capacity (O'Connor et al., 2013; Orellana et al., 2019; Peña et al., 2001), or a lack culturally responsive assessment practices (Skiba, 2002). In bilingual children, exposure to a language other than English is a

distinguishing characteristic and common aspects of typical bilingual development, such as cross-linguistic influence, may be misinterpreted as symptoms of a language-related disorder. For this reason, overrepresentation is often viewed as a direct consequence of misidentification, due to the incorrect attribution of observed areas of difficulty to an underlying disability, rather than considering the role that language differences may play. An emergent-bilingual student, for example, may be incorrectly identified as having a language-related disorder when they are simply in the process of acquiring English. The misidentification of disabilities based on language differences also extends to children who speak non-mainstream dialects of English, such as African American English. The argument that cultural or linguistic bias in disability identification explains overrepresentation may be especially compelling given such well-documented court cases as *Larry P v. Riles* (*Larry P. v. Riles*, 1979) and the *Ann Arbor Decision* (*Martin Luther King Jr., ETC. v. Ann Arbor Sch. Dist.*, 1979). Both of these incidences exemplify the degree to which biased assessment methods, or those rooted in linguistic prescriptivism, may result in real harm for children from language-minority backgrounds.

The quality of instruction may also contribute to overrepresentation of bilinguals in special education. Children from language-minority communities, particularly children of immigrants, may be more likely to attend schools with limited resources (Buckley et al., 2004). Schools in areas of high poverty are likely to have larger percentages of inexperienced or uncertified teachers (Artiles et al., 2002). These poorly resourced schools may be unable to meet the specific educational needs of students with limited English proficiency, potentially increasing the risk for academic difficulties (Donovan & Cross, 2002; Hibel et al., 2010). Although not all bilingual students will face poor quality instruction, the link between instructional quality and academic outcomes is clear. Empirical evidence from the response to intervention literature, for example, suggests the quality of instruction has a substantial impact on academic outcomes, as well as future disability identification, for bilingual students. When provided with increased frequency and more individualized instruction, within a

response to intervention model, bilingual students at risk of academic difficulties are able to catch up to their peers (e.g., McMaster et al., 2008; O'Connor et al., 2013; Vaughn, Mathes, et al., 2006). Furthermore, there is evidence to suggest that response to intervention may result in lower levels of disproportionate identification of reading difficulties in bilingual students (O'Connor et al., 2013, 2014). The outcomes of these studies suggest that the quality of instruction has a direct impact on rates of identification for bilinguals, and that modification to instruction has the potential to result in substantially improved outcomes.

Although assessment bias and instructional quality may provide convincing explanations for overrepresentation, they do not adequately account for the existence of underrepresentation, which may be more closely tied to socioeconomic, cultural, or linguistic barriers. The obstacles faced by language-minority children may lead to underrepresentation in special education or poorer academic outcomes broadly (Donovan & Cross, 2002). These barriers may also result in lower parent participation and consequently reduced access to services. Families may not feel sufficiently informed about the processes involved in disability identification, an issue which is then compounded by communication barriers due to language differences, resulting in an inability to engage effectively (Wolfe & Durán, 2013). Lack of familiarity with the processes involved in disability identification may be partly explained by fewer interactions with healthcare professionals due to limited access to these services broadly, as well as fewer positive interactions with providers (Calvo & Hawkins, 2015). Cross-cultural differences in perceptions and beliefs about disability may also impact the decision to seek out services (K. P. Cummings & Hardin, 2017). Families may also avoid school-based services, due to less-than-ideal past experiences with the disability identification process (Hardin et al., 2009; Klingner & Harry, 2006; Wolfe & Durán, 2013). The "frog pond effect," in which minority students attending schools with lower levels of academic achievement experience a decreased risk of disability identification, may also play a role in underrepresentation (Hibel et al., 2010; P. L. Morgan et al., 2015).

Although there are unique factors that contribute to each under- and overrepresentation, it is equally important to consider the relationship between these two patterns. In many instances both trends may exist simultaneously within the same district or school, in the form of a shift from under- to overrepresentation. As previously mentioned, several studies have identified patterns of underrepresentation leading to overrepresentation for emergent bilinguals (Artiles et al., 2005; Samson & Lesaux, 2009; Yamasaki & Luk, 2018). One possible explanation for this pattern may be hesitation on the part of teachers to refer emergent bilinguals for special education assessment until they exhibit a sufficiently high level of English proficiency. This may be driven by the belief that observed areas of difficulty are related limited English proficiency, and that academic ability will improve as a function of English ability (Limbos & Geva, 2001). Teachers may also falsely believe that emergent bilinguals must demonstrate a sufficient level of English proficiency to benefit from special education services (Hibel & Jasper, 2012). This delay in referring for evaluation creates a situation in which emergent bilinguals receive delayed access to special education compared to their monolingual peers. To further compound the issue, children classified as ELLs who receive special education services are also less likely to be reclassified as English-proficient than their ELL peers outside of special education (Slama, 2014; Umansky et al., 2017). This creates a "reclassification bottleneck", inflating the numbers of children classified as ELLs in special education in later grades (Umansky et al., 2017).

Policy and Practice

As a policy issue, disproportionality is mainly centered on race and ethnicity. A focus on bilingual representation in special education is largely absent from current federal policy, as exemplified in the reporting requirements for the Individuals with Disabilities Education Act (IDEA; Individuals with Disabilities Education Act, 2004). Although states are required to report the degree to which students across racial and ethnic groups are represented in special education, reporting of representation based on bilingual status (i.e., limited English proficiency) is not required and is considered out of the scope of

the law (US Department of Education, 2016a). Because of the absence of reporting requirements based on language background, this information is often not publicly available, making it a far more opaque issue to examine.

Policy has also tended toward the presumption that overrepresentation is the predominant issue, as exemplified in the recently enacted Equity in IDEA regulations (US Department of Education, 2016a). This set of regulations, which took effect in March 2019, is intended to strengthen federal special education policy by ensuring that children from minority racial and ethnic backgrounds are provided with equal access to special education. Equity in IDEA emphasizes the need to ensure that minority students are not overrepresented in special education, by addressing the "the well-documented and detrimental overidentification of certain students for special education services" (US Department of Education, 2016a, p. 1). Although efforts such as Equity in IDEA are ostensibly intended to improve access to special education services, the singular focus on overrepresentation may not address the needs of local education agencies in which overrepresentation does not exist, or where patterns of special education representation differ depending on the grade level. Although these regulations do not diminish a local education agency's ability to address underrepresentation, they also do not provide the tools required to mitigate it.

The narrative regarding the responsibility of practitioners in addressing disproportionality has centered on reducing assessment bias in order to lessen overidentification, reflecting the emphasis expressed in federal policy. For those practitioners involved in disability evaluations—school psychologists and speech-language pathologists—a great deal of attention has been paid to nonbiased assessment methods, as exemplified in the practitioner literature (e.g., Campbell et al., 1997; Roseberry-McKibbin & O'Hanlon, 2005; Saenz & Huer, 2003). With respect to language assessment many have advocated for the adoption of alternative assessment methods, such as dynamic assessment (Orellana et al., 2019; Peña et al., 2001) or processing dependent measures (Dollaghan & Campbell, 1998; J. A.

Ortiz, 2021a). Although these methods can reduce assessment bias, many clinicians continue to rely on standardized assessment (Arias & Friberg, 2017). It is also unclear to what degree reducing assessment bias results in a reduction in overrepresentation in real world settings. Outcomes from studies examining the diagnostic accuracy of less-biased assessment methods for bilinguals do suggest that they lead to more accurate diagnostic outcomes (Lazewnik et al., 2019; J. A. Ortiz, 2021a; Paradis et al., 2013; Petersen et al., 2017). Despite the supporting evidence, it is important to acknowledge that assessment occurs as the last step along a student's path to special education, and the myriad of other factors that cumulatively contribute to a referral for evaluation may dwarf the contribution that nonbiased assessment makes in reducing overrepresentation.

Response to intervention, another potentially useful tool in the mitigation of disproportionate representation, has also received substantial attention in the practitioner literature (e.g., Carreker & Joshi, 2010; Justice, 2006; Ruiz, 2020). Because it is a prevention model, the aims of response to intervention are to identify students who need support early and provide small-group intervention in general education to accelerate progress. Thus, response to intervention may contribute more to reducing disproportionality than nonbiased assessment alone. There are several different ways that practitioners, including school psychologists and speech-language pathologists, could be included in the response to intervention process, such as through universal screening or by participating in secondary and tertiary interventions (Ehren & Nelson, 2005). One major advantage for response to intervention is that it may be equally useful in addressing both overrepresentation and underrepresentation. In cases of overrepresentation, participation in secondary or tertiary tier interventions may be sufficient for at-risk students to meet academic standards, thereby avoiding a special education referral. In cases where underrepresentation exists, for example, universal screening may improve early identification of children who may have otherwise been missed, potentially increasing the number of children referred for evaluation. Although there is some evidence of its use by school psychologists and speech-language

pathologists (Silva et al., 2021; Sullivan & Long, 2010), widespread adoption remains limited (American Speech-Language-Hearing Association, 2020b; McKenna et al., 2021).

Overview of the Dissertation

The present chapter provides an overview of the theoretical framework and major issues associated with disproportionality. The subsequent chapters will address several important topics that contribute to our understanding of the issue. Chapter 2 comprises a meta-analysis of one specific alternative language assessment method for bilingual children: nonword repetition. This study, which was published in *American Journal of Speech-Language Pathology*, examines the diagnostic accuracy of nonword repetition in the identification of language impairment in bilingual children based on an analysis of outcomes from thirteen studies. This study addressed the following questions: (a) What is the quality of the extant research into nonword repetition in bilingual children? (b) What is the diagnostic accuracy of nonword repetition when predicting language impairment in bilingual children? (c) Are different types of nonword repetition tasks or scoring methods associated with better classification accuracy of language impairment in bilingual children? Results of this study suggest that nonword repetition, a processing-dependent measure that is considered to be less-biased than other forms of assessment (Dollaghan & Campbell, 1998), may be a useful tool in the identification of language impairment when paired with other assessment methods. The use of alternative assessment tools, such as nonword repetition, in the disability evaluation process may help to reduce misidentification of bilingual children with language-related disorders.

Chapter 3 consists of a study focusing on the disproportionate identification of language-related disorders, including CD and SLD. This study addresses the following research questions: (a) Are English proficient bilinguals and emergent bilinguals disproportionately represented in language-related disorder categories from kindergarten to fifth grade? (b) How do the rates of identification of language-

related disorders differ for emergent bilinguals, English-proficient bilinguals, and monolinguals, from kindergarten to fifth grade? (c) What individual- and school-level variables predict disability identification for emergent bilinguals, English-proficient bilinguals, and monolinguals? This study utilizes data from the Early Childhood Longitudinal Survey Kindergarten class of 2010-11 (ECLS-K:2011), an individual-level, nationally representative data set with observations for over 18,000 children from grades K to 5. Results from this study provide insight into the degree to which over- and underrepresentation affect English-proficient and emergent bilinguals, as well as the variables that predict these classification outcomes.

Although most studies of disproportionality have examined racial and ethnic representation in special education (Cruz & Rodl, 2018), the are relatively few studies focusing specifically on bilinguals. Because the factors that drive racial and ethnic disproportionality may differ from those related to bilingualism, it is important to examine the unique set of circumstances that influence disability identification for bilinguals. With respect to methodology, many studies have used district- or school-level samples, limiting the ability to derive detailed inferences from these results (Sullivan, 2011; Yamasaki & Luk, 2018). A small, but growing, number of studies of bilinguals have used individual-level samples, which provide for a far more robust analysis than samples that use aggregated data (Hibel & Jasper, 2012; P. L. Morgan, Farkas, et al., 2017; Samson & Lesaux, 2009; Umansky et al., 2017). Although some studies have utilized analyses that do not account for relevant variables that may influence observed outcomes (Robinson & Norton, 2019; Yamasaki & Luk, 2018), a growing number of studies utilize techniques that more accurately account for these variables, providing a strong foundation for further research into bilingual disproportionality (Hibel & Jasper, 2012; P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017; Umansky et al., 2017). This study builds on existing research by examining the large sample provided by the ECLS-K:2011, which allows for a robust analysis of outcomes for individual students across the country.

Chapter 4 comprises a paper intended for practitioners, drawing from the outcomes of the study described in Chapter 3, as well as recent research into bilingual disproportionality. The purpose of this paper will be to provide those school-based speech-language pathologists involved in disability assessment with insight into aspects of disproportionality that have not been addressed in the practitioner literature. This paper will have two main areas of focus: (a) the underlying nature of disproportionality and how risk for disability identification may change across grades, and (b) a model of preventing disproportionate identification for bilinguals. Although a number of practitioner papers have been written on the topic of disproportionality in bilingual children, most do not consider more recent research in their recommendations, particularly with respect to the dynamic nature of disproportionality across grades. In addition, most published practitioner papers portray English language learners as a monolithic group, and do not directly address the role that diversity of language background plays in identification. Lastly, many practitioners focus on methods of disproportionality mitigation rather than prevention, particularly those in the communication disorders literature. This paper will provide clinicians with an up-to-date overview of the current issues surrounding bilingual disproportionality by addressing several key considerations not previously discussed in the practitioner literature.

Chapter II: Using Nonword Repetition to Identify Language Impairment in Bilingual Children: A Meta-Analysis of Diagnostic Accuracy

This chapter comprises a meta-analysis that was published in the American Journal of Speech-Language in September 2021, the full text of which is included in the following pages.

Ortiz, J. A. (2021). Using nonword repetition to identify language impairment in bilingual children: A meta-analysis of diagnostic accuracy. *American Journal of Speech-Language Pathology*, *30*(5), 2275–2295.

Using Nonword Repetition to Identify Language Impairment in Bilingual Children: A Meta-Analysis of Diagnostic Accuracy

Bias in language assessment is a pervasive issue which adversely affects children from culturally and linguistically diverse backgrounds, often resulting in misidentification. Because of the limited number of appropriate diagnostic tools available, arriving at accurate conclusions about a child's true communication ability can be a complex task. For children learning more than one language, it is particularly difficult to determine the degree to which apparent difficulties may be related to language impairment (LI), or whether they reflect typical bilingual development. The obstacles to appropriate identification are exemplified by the disproportionate number of children with limited English proficiency in special education classrooms in the US. English learners, who comprise about 5 million students in public schools in the US (Hussar et al., 2020), are at a higher risk for being misidentified in special education evaluations (Samson & Lesaux, 2009; Sullivan, 2011; Yamasaki & Luk, 2018). These children after often over- or underidentified across a range of different disability categories, with individual language profiles playing a major role in misidentification (Yamasaki & Luk, 2018). Although attempts to improve assessment methods have resulted in the development of an array of different clinical tools, disproportionate identification continues to persist.

Misidentification is often attributed to factors such as incorrect application of standardized testing (e.g., Bedore & Peña, 2008), confirmatory bias (e.g., Knotek, 2003), or over-referral of linguistically diverse students (e.g., Ysseldyke et al., 1997). Speech-language pathologists play a particularly important role in special education eligibility as essential team members in the disability evaluation process. Although accurate and nonbiased assessment is critical in ensuring appropriate identification of LI, many of the diagnostic tools available to clinicians are inadequate. Some of the most commonly used tools have been shown to be inaccurate when used with bilingual children, due to content or linguistic bias (Barragan et al., 2018; Restrepo & Silverman, 2001). Tests that are reliant on

prior knowledge, contain culturally irrelevant stimuli, or do not account for cross-linguistic skills may result in inaccurate conclusions about a child's language abilities. Furthermore, the inappropriate use of standardized assessment with populations that are not represented in a test's normative sample, as is frequently the case for bilinguals, can also result in flawed clinical conclusions. In spite of the concerns around standardized assessment, its use continues to be common among clinicians who work with bilingual clients (Arias & Friberg, 2017).

There have been numerous methods recommended to reduce bias in assessment, such as dynamic assessment (e.g., Dollaghan & Campbell, 1998; Peña et al., 2001), conceptual vocabulary scoring (e.g., Bedore et al., 2005), parent report (e.g., Paradis et al., 2010; Restrepo, 1998), and cross-linguistic language sample analysis (e.g., Castilla-Earls et al., 2020; Gutiérrez-Clellen & Simon-Cereijido, 2009). All of these efforts reflect an attempt to lessen the degree to which bilingual children are disadvantaged by traditional forms of assessment. With the goal of evaluating the quality of such assessment tools, Dollaghan and Horner (2011) published a meta-analysis of diagnostic accuracy which focused specifically on tools intended to be used with bilingual Spanish-English speakers. In their study, the authors summarized the outcomes of 17 measures from nine studies. Of the included studies, the vast majority focused on measures of morphosyntax, while one study investigated the use of nonword repetition. The authors concluded that no single method of assessment demonstrated particularly strong classification accuracy, but that the included measures may be useful when used alongside other methods as part of an assessment battery. In their analysis of the quality of evidence, the authors also identified a number of areas of improvement for future diagnostic accuracy studies. Results of this meta-analysis indicated a clear need for continued research into appropriate assessment methods for the identification of LI in bilinguals.

Although nonword repetition was represented by a single study in the meta-analysis conducted by Dollaghan and Horner (2011), it has long been endorsed for its potential as a less biased assessment

method (Dollaghan & Campbell, 1998; Laing & Kamhi, 2003). In a nonword repetition task, an individual is asked to repeat a set of novel word-like stimuli that adhere to the phonotactic properties of a specific language. These tasks typically include between 20 to 40 nonwords of increasing length, ranging from one to five syllables. Because their sound sequences are permissible for a given language, nonwords sound like plausible words. Nonword repetition is thought to be less biased than other forms of assessment because all children, regardless of cultural or linguistic background, are presented with a set of stimuli that they have never heard before, resulting in a somewhat more level playing field. Since the listener has presumably not been previously exposed to these words, performance should not be associated with prior linguistic knowledge, unlike many commonly used assessment methods that may be partially or entirely reliant on prior knowledge.

Nonword repetition, as a type of processing-dependent measure, relies on an individual's ability to perceive, briefly retain, and then repeat a novel word (e.g., Gathercole & Baddeley, 1990, 1993). This has commonly been thought of as a measure of phonological working memory (e.g., Bishop et al., 1996; Botting & Conti-Ramsden, 2001; Lely & Howard, 1993). As noted by Graf Estes et al. (2007), a number of different skills have also been proposed as underlying nonword repetition ability, including phonological encoding (e.g., Kamhi & Catts, 1986), phonological awareness or sensitivity (e.g., Metsala, 1999), general phonological processing ability (e.g., Bowey, 2001), or multiple cognitive processes (e.g., Edwards & Lahey, 1998).

Regardless of the specific skill recruited in nonword repetition, there is evidence of its clinical utility in the identification of LI. Several early studies found an association between the presence of LI and poor performance on the task (Gathercole & Baddeley, 1990; Kamhi & Catts, 1986). This effect has since been replicated numerous times, as exemplified in the meta-analysis conducted by Graf Estes et al. (2007). Across the 23 studies of nonword repetition included, the authors of this meta-analysis found that children with LI exhibited substantially more difficulty in their ability to repeat nonwords than

children with TL. Performance on nonword tasks across studies was 1.27 standard deviations lower for children with LI than for those with TL, with more difficulty on nonwords of increased length. The effect size was considered to be very large and did not significantly differ across ages. The authors also identified associations between the various types of tasks used across studies and variation in effect sizes, indicating that not all tasks are equivalent. Results of this meta-analysis suggest that nonword repetition may be a useful tool in the identification of LI, and that certain characteristics of the stimuli may have an impact on performance.

Although extant evidence supports the use of nonword repetition in the identification of LI in English speakers, there is less research into its utility with children from culturally and linguistically diverse backgrounds. Some researchers have addressed this issue by evaluating the degree to which performance on nonword repetition may differ for children from different racial, ethnic, or linguistic backgrounds. Dollaghan and Campbell (1998), in their study of nonword repetition on a sample that included both African American and white participants, found clear differences in task performance for groups of children with TL and those with LI. Reported positive and negative likelihood ratios, which represent how well the task distinguished between the two groups, were 25.15 and .03, respectively, indicating that the nonword measure was diagnostically informative for both ruling in and ruling out the presence of LI. Importantly, the authors did not identify an association between nonword repetition ability and race, suggesting that their task did not disadvantage children from different racial backgrounds.

Given the concern regarding misidentification of bilinguals, Kohnert et al. (2006) sought to determine how well nonword repetition could distinguish between monolinguals and bilinguals with TL, and monolinguals with LI. In their study, the authors used the English nonword task developed by Dollaghan and Campbell (1998) to evaluate differences in performance across these groups. They found that, although the task could be used to successfully distinguish between monolinguals with and without

LI, outcomes for bilinguals were less favorable. For bilinguals with TL, overall performance fell somewhere between monolinguals with TL and those with LI. Particularly concerning was that, for one-to-two syllable nonwords, the performance of bilinguals with TL was even lower than that of monolinguals with LI, suggesting that the task may in fact be biased. For monolinguals and bilinguals, the task was found to be diagnostically informative for ruling out LI, with negative likelihood ratios of .08 for both groups. The ability of the task to rule in LI, on the other hand, was less favorable for bilinguals than it was for monolinguals. Reported positive likelihood ratios of 10.7 and 5.07 for monolinguals and bilinguals, respectively, indicate that the task was diagnostically informative for ruling in LI in monolinguals, but only suggestive of LI when including bilinguals. Although the results of this study do not support the use of nonword repetition with linguistically diverse children, given that bilinguals with TL could not be reliably distinguished from monolinguals with LI, the use of a task based on English phonology may have affected the performance of bilinguals. All of the bilinguals in the sample were sequential Spanish-English speakers, with four to eight years of experience learning English. Although these participants had significant exposure to English, one cannot rule out the potential influence of language background as a factor in difficulty repeating nonwords that adhered to the phonotactic constraints of their second language, rather than to the constraints of their native language. Given this linguistic profile, it is plausible that the language of the task may have played a role in observed differences in performance.

As a method of language assessment for bilingual children, nonword repetition may be a useful addition to a clinical tool kit, given some of its particularly appealing qualities. Administration can be completed in about five minutes, making it very convenient to include in an assessment battery without significantly increasing the length of an evaluation. The task itself is also relatively straightforward for both the clinician and the client, in contrast to many assessment methods that require a substantial amount of training. With respect to bias, nonword repetition is also less reliant on prior knowledge than

some other commonly used assessment methods. Despite its potential advantages, nonword repetition is not without limitations. The task itself lacks face validity and may feel somewhat contrived for examinees. In addition, although the use of stimuli that are devoid of cultural or linguistic relevance may result in a less-biased task, it may also result in a task that is less engaging. Although it may be clinically useful, there are a number of unanswered questions regarding its utility with bilingual children. Given the extant evidence of its effectiveness with monolinguals, and its potential as a nonbiased assessment method, the present study was developed with the goal of evaluating the efficacy of nonword repetition in the identification of LI in bilingual children. Although numerous studies of nonword repetition in bilinguals have been published, there are presently no systematic reviews or meta-analyses examining its diagnostic accuracy with this population.

The purpose of the present meta-analysis is to determine the diagnostic accuracy of nonword repetition in the identification of LI in bilingual children, and to evaluate the quality of the associated evidence. The following sections provide a description of the methods used to identify articles and extract and analyze the relevant data, a detailed review of both descriptive and quantitative outcomes, and finally a discussion of these findings. Results of this meta-analysis are intended to offer insight into the effectiveness of nonword repletion as a method of bilingual assessment and any task characteristics that may improve diagnostic outcomes. Thus, this meta-analysis will address the following research questions:

1. What is the quality of the extant research into nonword repetition in bilingual children?
2. What is the diagnostic accuracy of nonword repetition when predicting language impairment in bilingual children?
3. Are different types of nonword repetition tasks or scoring methods associated with better classification accuracy of language impairment in bilingual children?

Method

Using the Preferred Reporting Items *for Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies* checklist (PRISMA-DTA; McInnes et al., 2018), the following section provides an overview of the search procedure used to identify relevant articles for the present meta-analysis, how variables from each study were selected and recorded, how the quality of each study was evaluated, and finally how quantitative outcomes were analyzed across studies. The *Quality of Studies* section describes the criteria used to evaluate the quality of each study to address the first research question. The *Quantitative Analysis* section describes the procedures used to address the second and third research questions, which examined the classification accuracy of nonword repetition tasks. Much of the methodology used in this study, including the search procedures, eligibility criteria, and quantitative analysis, was guided by the framework for meta-analytic research described in the in-progress *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Deeks et al., 2013).

Search Procedure

Before beginning the literature search, a review protocol was developed to ensure that the search strategy effectively identified all relevant articles. This protocol included the intended purpose and research questions, as well as the search terms, databases, and eligibility criteria described below. Potential studies were identified through a search of the following electronic databases: ERIC, EBSCO Academic Search Ultimate, Medline, and PsycInfo. These databases were searched for studies published in peer-reviewed journals in English, using the following terms: (nonword repetition OR nonword OR repetition OR processing dependent) AND (language impair* OR language disorder* OR language disab*) AND bilingual. The date range covered by the database search was from inception until September 2019. No other restrictions were applied to the search of databases.

Using a study selection procedure similar to that detailed by Orellana et al. (2019), potentially relevant articles were identified by taking the following steps. First, after conducting the initial search, duplicate records were removed. Unique records were then screened on the basis of their titles and abstracts. A full text review of the remaining records was then completed to identify those that met the eligibility criteria. From this set of identified articles, an ancestral search was conducted, which consisted of identifying potentially relevant studies from the reference lists of articles selected in the initial search. Lastly, a forward search was completed using the "Cite By" function in Google Scholar for each paper selected in the initial search. Both the ancestral and forward searches followed the same procedures as the initial search to screen out irrelevant results on the basis of their abstracts or full texts. The ancestral and forward searches were completed in October 2019.

Eligibility Criteria

A set of eligibility criteria was established before reviewing potential studies, in order to identify relevant papers to be included. First, studies were required to evaluate a measure of nonword repetition. Second, studies needed to report quantitative results for nonword repetition measures separate from other outcomes. Because the focus of the present meta-analysis was specifically on nonword repetition, studies that reported these outcomes as part of a composite measure, and did not disaggregate these data, were excluded. Third, studies were required to report the diagnostic accuracy of the nonword measure they were evaluating. Although a number of quantitative methods can be used to examine the efficacy of a tool, specific metrics, such as sensitivity and specificity, are required to be able to evaluate its diagnostic accuracy. Studies that provided mean differences in the absence of diagnostic accuracy, for example, were excluded. Fourth, articles needed to include both bilingual children with TL and those with LI in their samples. Studies that only focused on monolinguals, or did not include bilingual children with LI, were excluded. Fifth, studies were required to include children under

the age of 18, in their samples. Finally, studies had to be published in a peer-reviewed journal and written in English in order to be included.

Coding Procedure

In order to identify relevant characteristics of each study, a coding matrix was developed based on the variables described by Dollaghan & Horner (2011) and Orellana et al. (2019), with the addition of several elements specific to nonword repetition. Coding was primarily completed by the author of the present study, along with a doctoral student who was trained on the coding procedure. To ensure reliability, coders independently double-coded four of the included articles (30%), which were selected at random. Coders reviewed the full texts of each study and recorded data for each of the variables in the coding matrix. Inter-rater agreement was 98.17%, which was calculated by comparing responses on each of the coding variables. After independently recording these variables, the two coders discussed any discrepancies and achieved consensus. If necessary, coders contacted authors of primary studies regarding any data that required clarification.

Coders recorded general information for each study, including paper title, authors, journal, and year of publication. Study sample variables included the number of individuals with LI and, age mean/range, gender, SES, country, and languages spoken. For studies that included multiple countries or languages, data for each country/language were recorded if they were reported separately. With respect to the methodology of each study, coders recorded the following variables: nonword repetition task characteristics, source of the nonword repetition stimuli, scoring method, reliability, participant selection, reference and index test administration characteristics, and bilingual status determination. In a study of diagnostic accuracy, the reference standard is the method by which researchers establish the presence of the condition under investigation. The index measure refers to the assessment method being evaluated which, for the present study, was nonword repetition. For both the reference and index

measures, coders recorded the specific tests/tasks administered, cut scores, language of administration, the number and duration of test sessions for each group. Because the determination of who is bilingual can be made in a number of ways, such as through parent report or standardized assessment, coders also recorded the manner by which studies classified participants as bilingual and, if reported, how they established language dominance. Lastly, coders recorded both descriptive and quantitative outcomes from each study. This included a summary of major findings, in addition to all metrics of diagnostic accuracy reported. A description of specific quantitative outcomes, and how these data were analyzed, is provided in the *Quantitative Analysis* section below.

Quality of Evidence

The quality of evidence for each study was evaluated based on the Critical Appraisal of Diagnostic Evidence (Dollaghan, 2007), as well as the quality indicators used by Dollaghan & Horner (2011) and Orellana et al (2019). The eight quality indicators were as follows: (a) sample size, (b) representativeness of sample, (c) gate design, (d) independent testing, (e) blinded testing, (f) reliability, (g) index measure administration, (h) validity and reliability of reference standard. The evaluation of study quality used an 8-point scale, in which each study was assigned 1 point for design features that met the criteria of each of the quality indicators.

Sample Characteristics

For each study, the respective number of participants with TL and LI was recorded. A sufficient number of participants is necessary to ensure that reported diagnostic accuracy is representative of the larger population from which a sample was drawn. Consistent with previously published meta-analytic research, a total sample of greater than 30 participants across these groups was preferred (Orellana et al., 2019). In addition to the combined sample size, the ratio of children with TL to those with LI was also considered. In the US, the prevalence of language impairment in individuals aged 3-17 is approximately

3.3% (Black et al., 2015). It is preferable to have a sample that reflects the prevalence of LI in the general population, as nonrepresentative ratios may result in inflated or otherwise imprecise diagnostic accuracy (Dollaghan & Horner, 2011).

Gate Design

The gate design of each study was also evaluated, distinguishing between one- and two- gate diagnostic designs. The distinction between these two designs relates to the participant selection process (for an overview, see Rutjes et al., 2005). In order to evaluate the ability of an assessment method to distinguish between those with and without a particular condition (e.g., LI), the sample must include participants from both groups. A one-gate design includes a large representative sample of the population on which to administer the method of assessment being evaluated. A sample that is sufficiently large would presumably include some individuals with LI, as well as those with TL, but the presence of LI is unknown during participant selection. In a two-gate design, a sample of individuals known to have LI is selected alongside a sample of individuals known to have TL. In effect there a two-gate design utilizes two sets of inclusion criteria; participants with LI are selected on the basis of low language ability, whereas those with TL are selected because they are known to have typical language ability. When evaluating the diagnostic accuracy of a specific measure, a one-gate design may not be practical or feasible given the greater sample size required. Despite this, two-gate designs may lead to spectrum bias, or the inclusion of individuals from extreme ends of the language ability spectrum, which has the potential to result in inaccurate and inflated diagnostic accuracy (Dollaghan & Horner, 2011; Rutjes et al., 2005). For this reason, one-gate designs are preferable to two-gate designs.

Testing Procedures

Three key elements related to testing procedures were included in the evaluation of quality of evidence. The first was whether the same index measure was administered to all participants. Studies

that administered the same measure to participants with LI as they did to those with TL were preferred, given the potential for verification bias (Dollaghan, 2007). The next element that was considered was whether testing for the reference standard was conducted by an independent examiner, rather than by the research team itself. Studies that received referrals from schools or clinics, but subsequently administered follow-up testing to confirm the diagnosis of LI were not considered to have utilized independent testing. The last element that was evaluated was blinding. Studies were classified as blinded when they clearly stated that examiners were unaware of relevant diagnostic information about the participants, such as the presence of LI. The absence of independent testing or blinding may lead to clinical review bias, due to the potential influence of subjective knowledge or the expectations of examiners on test outcomes (Dollaghan, 2007).

Reliability

In addition to testing procedures, the reliability of the scoring procedures was also evaluated by coding for whether studies used multiple raters and reported inter-rater agreement. Nonword repetition tasks require an examiner to judge a client's production and subsequentially make a determination regarding its accuracy. The reliance on examiner judgement introduces a possible threat to the reliability of the task, as there is some potential for variation in the measured accuracy of repetition, depending on the scorer. Phoneme-level scoring is particularly susceptible, as there is an increased potential for human error when judging the accuracy of individual phonemes. Variability in scoring accuracy may be further exacerbated by the presence of developmental phoneme substitutions, distortions, or cross-linguistic influence, for which a number of scoring schemes may be implemented (e.g., Dollaghan & Campbell, 1998; Guiberson & Rodríguez, 2016). Because of these potential sources of error, the use of two or more raters and was preferred.

Validity and Reliability of Reference Standard

One of the most critical features of a study of diagnostic accuracy is the reference standard that was used identify the disorder in question. For the purposes of the present meta-analysis, a reference standard was considered to be valid and reliable if it met two key criteria. First, the measure must have included assessment in both languages, consistent with current best practice (e.g., Kohnert, 2010; Peña et al., 2016). Second, the reference standard was required to include converging evidence from multiple sources, both direct (e.g., clinician-administered tasks) and indirect (e.g., parent report) in order ensure a high level of certainty regarding the presence of LI (e.g., Barragan et al., 2018; Castilla-Earls et al., 2020; Restrepo, 1998). Parent report was considered to be acceptable when it was provided in the form of a questionnaire or a structured interview, as there is evidence of their diagnostic utility when used along with other assessment methods (Paradis et al., 2010, 2013; Restrepo, 1998). Previous diagnosis, if used as the sole component of a reference standard, was considered to be a valid and reliable identifier of LI only if the specific measures used by the diagnostician were reported and they met the criteria specified above. This decision was made due to the fact that bilingual children may be misdiagnosed with LI when using commonly available assessment tools (Barragan et al., 2018; Restrepo & Silverman, 2001) and consequently, in the absence of other sources of information, a prior diagnosis may not be a reliable indicator of LI. If the specific assessment methods used to make a previous diagnosis were not reported, it would not be possible to know how they were validated for use with the population on which they were administered.

Quantitative Analysis

In the evaluation of diagnostic accuracy, studies may report one or more metrics, including sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-, respectively), diagnostic odds ratio (DOR), and area under the curve (AUC). Sensitivity and specificity are two of the most commonly reported metrics. Sensitivity, or the true positive rate, represents the proportion of individuals with LI that were correctly identified, expressed as a decimal or percentage. Specificity, or the true negative

rate, represents the proportion of individuals with TL that were correctly identified. Sensitivity and specificity values of greater than 90% suggest that a test has good discriminant accuracy, values from 80% to 89% suggest fair accuracy, and values below 80% suggest poor accuracy (Plante & Vance, 1994).

LR+ represents the probability that someone with LI will be correctly diagnosed (i.e., true positive rate or sensitivity) versus the probability that someone with TL will be diagnosed as having LI (i.e., false positive rate or 1 – specificity): LR+ = $\frac{\text{true positive rate}}{\text{false positive rate}}$. LR-, on the other hand, represents the false negative rate (i.e., 1 – sensitivity) versus the true negative rate (i.e., specificity): LR– = $\frac{\text{false negative rate}}{\text{true negative rate}}$. Higher values of LR+ and lower values LR- indicate stronger diagnostic accuracy. Values of LR+ ≥ 10 and LR- < .1 are diagnostically informative, while LR+ and LR- values around 3 and .3, respectively, are only clinically suggestive of LI (Dollaghan, 2007; Dollaghan & Horner, 2011).

AUC and DOR are two additional metrics that are particularly useful for summarizing the overall classification accuracy of a test. DOR, in the present study, represents the odds of a positive test result (i.e., scoring below the cut score) coming from someone with LI, relative to the odds of a positive result coming from someone with TL. DOR is calculated as $DOR = \frac{true\ positives}{false\ negatives} / \frac{false\ positives}{true\ negatives}$, with values ranging from 0 to infinity. Higher values of DOR indicate better discriminant accuracy, with values close to 1 suggesting that a test is diagnostically uninformative (Dollaghan & Horner, 2011). AUC is derived from the receiver operating characteristic (ROC) analysis, which determines a diagnostic test's ability to detect a "signal" (K. D. Cummings & Smolkowski, 2015). As a summary measure, this metric represents overall classification accuracy, with values ranging from 0 to 1, and provides information about how well a test performs at discriminating between LI and TL in a given sample. A value of 1 indicates perfect diagnostic accuracy, and a value of .5 suggests that classification accuracy is no better than chance, and thus is diagnostically uninformative (Deeks et al., 2013). AUC is unique in that it does not just provide information about the classification ability of a measure for a single criterion level and can also be used

to select the optimal cut score of a measure, by simultaneously accounting for both sensitivity and specificity (K. D. Cummings & Smolkowski, 2015).

Because studies of diagnostic accuracy may report any of the previously mentioned metrics, 2x2 contingency tables of true positives, false positives, false negatives, and true negatives, were derived in order to compare outcomes across studies. Sensitivity and specificity represent proportions of correctly identified participants with the condition and those without it, and as such can be used to derive contingency tables of true and false positives and negatives (for an overview, see Dollaghan, 2007). If raw numbers of individuals correctly and incorrectly identified were not reported, these values were calculated based on reported sensitivity and specificity and sample sizes of individuals with LI and those with TL, consistent with the convention used in meta-analytical research (e.g., Orellana et al., 2019). A continuity correction of 0.5 was added to all cells of 2x2 tables that contained at least one value of zero, to enable calculation of further metrics (Schwarzer et al., 2007). All quantitative analyses were based on these estimates, which resulted in slight discrepancies in estimated versus reported results but did not meaningfully alter any of the studies' outcomes. Statistical analyses were completed using the software program *R* (R Core Team, 2020) in conjunction with the *mada* (Doebler & Holling, 2015), *meta* (Schwarzer, 2007), and *metafor* (Viechtbauer, 2010) packages. Several studies reported multiple values for diagnostic accuracy metrics for different cut scores. If studies reported an optimal cut score, only the diagnostic accuracy metrics for that cut score were included, in order to derive estimates using values from a single cut score (Deeks et al., 2013). In the case that the authors did not specify an optimal cut score, the value that resulted in the highest DOR was used. For studies that contrasted different types of nonword repetition tasks and scoring methods, or reported separate results for different groups, each was included as a separate observation.

Pooled values of sensitivity, specificity, likelihood ratios, and DOR were estimated using the bivariate method described by Reitsma et al. (2005). This method accounts for a possible negative

correlation between sensitivity and specificity, which is not typically considered when estimating pooled diagnostic accuracy metrics. A summary receiver operating characteristic analysis was subsequently conducted from estimates of sensitivity and specificity across studies to estimate AUC, as a summary of the overall effect size.

Variability in effect size beyond that which can be attributed to sampling error may be due to heterogeneity between studies. A lack of homogeneity in diagnostic accuracy outcomes may reflect fundamental differences in the study population or methods. For this reason, the presence of heterogeneity was evaluated using Cochran's *Q*. Because of its low power to identify heterogeneity in small samples, this was further quantified using the I^2 statistic, which provides an estimate of the percentage of variability across measures (Higgins et al., 2003). To further investigate sources of heterogeneity, a random effects meta-regression model was estimated using DOR as the outcome variable. In the presence of established heterogeneity, a random effects model is preferable to a fixed effects model, due to the increased risk of type I error when using fixed effects models (Higgins & Thompson, 2004). This model was estimated using the DerSimonian and Laird (1986) method of inverse variance weighting and included a Knapp-Harting adjustment (Hartung & Knapp, 2001; Sidik & Jonkman, 2002) which results in wider confidence intervals, reflecting the uncertainty of estimates of between study heterogeneity (Higgins et al., 2019). Because many studies reported results for multiple measures, robust variance estimation was applied using clusters at the group level. Finally, the presence of publication bias was evaluated using the method described by Deeks et al. (2005).

Results

Study Selection

Figure 1 shows a diagram of the search and study selection process. The initial database search yielded 288 results, of which 166 were unique, and covered years of publication from 1983 to 2019. Of

these results, 62 were removed based on their title, and 82 on the basis of their abstract, leaving 22 potentially relevant studies. Twelve of these studies were excluded because they did not meet the eligibility criteria following a full text review, leaving 10 studies that were included. The ancestral and forward searches yielded three additional results that met the eligibility criteria. The final pool consisted of 13 studies with years of publication ranging from 2007 to 2018.

Potential candidate studies identified in the database, forward, and ancestral searches were excluded for several reasons, with many not meeting multiple eligibility criteria. Twelve studies did not include children with LI in their samples, and eight included only monolingual participants. With respect to quantitative outcomes, nine studies did not provide disaggregated outcomes for nonword repetition and 12 did not report the data required to calculate diagnostic accuracy.

Insert Figure 2 about here

Study Characteristics

Participant Characteristics

See Table 1 for an overview of study characteristics. Sample characteristics varied across studies. The total number of participants across studies was 1,366 (916 with TL and 450 LI), with ages ranging from 3;0 to 11;6, and an estimated mean of 6;8. Data were collected in a variety of countries including Australia, Canada, France, Germany, Israel, Russia, Spain, the Netherlands, and the United States. Spanish was the most common first or minority language, while English was the most common second or majority language. Regarding participant grouping, there were two predominant types of designs: those that included bilinguals only, and those that included monolingual comparison groups. Five studies used samples that included bilinguals with LI and bilinguals with TL in the absence of monolinguals. Eight of the studies included four or more groups, adding monolinguals with LI and monolinguals with TL. Armon-Lotem & Meir (2016) used a design consisting of six groups; in addition to

bilinguals, this study included monolinguals with TL and monolinguals with LI, who were speakers of either Hebrew or Russian. Tuller et al. (2018) included eight groups consisting of monolinguals with TL and those with LI who were speakers of either French or German, in addition to French-speaking bilinguals with TL and those with LI, and German-speaking bilinguals with TL and those with LI. One study (Girbau & Schwartz, 2007) labeled participants as Spanish-speakers who exhibited comprehension of Catalan due to a high degree of diglossic bilingualism in the area of Spain in which the data were collected. These participants were considered to be bilingual, as the profile described by the authors is consistent with bilinguals who exhibit differential abilities across expressive and receptive language domains, or those who have experienced language attrition (Anderson, 2012).

Insert Table 1 about here

Information related to socioeconomic status (SES) was not consistently reported across studies. All but three studies reported information about the SES of their participants (de Almeida et al., 2017; dos Santos & Ferré, 2018; Guiberson & Rodríguez, 2013). Those studies that did report SES generally used free or reduced school lunch and/or parental education level as a proxy. Five studies reported the specific SES level of their participants or provided indices of SES (Girbau & Schwartz, 2007, 2008; Gutiérrez-Clellen & Simon-Cereijido, 2010; Li'el et al., 2018; Tuller et al., 2018). One study reported significant differences in maternal education between the TL and LI groups (Thordardottir & Brandeker, 2013). The remaining studies reported matching their participants by SES.

Nonword Repetition Stimuli

A variety of nonword repetition stimuli were used across studies, differing in both the number and complexity of nonwords used in each task. For an overview see Table A.1 in the online appendix. A total of 18 nonword repetition tasks were included across studies, ranging from 14 to 71 nonwords of one to five syllables in length.

Quality of Evidence

The first research question sought to appraise the quality of the evidence of nonword repetition research in bilingual children. The methodological quality of each study was evaluated using the quality indicators shown in Table 2. One point was assigned for each of the eight indicators listed. Eight studies received a score of 4 points (Boerma et al., 2015; dos Santos & Ferré, 2018; Guiberson & Rodríguez, 2013; Gutiérrez-Clellen & Simon-Cereijido, 2010; Hamann & Abed Ibrahim, 2017; Li'el et al., 2018; Tuller et al., 2018; Windsor et al., 2010), while another four studies received 3 points (Armon-Lotem & Meir, 2016; de Almeida et al., 2017; Girbau & Schwartz, 2008; Thordardottir & Brandeker, 2013). The remaining study received 2 points (Girbau & Schwartz, 2007).

Insert Table 2 about here

The first quality indicators to be evaluated were sample size and representative ratio. Samples ranged from 22 to 144 participants. Ten studies met the criterion of at having least 30 participants. The two studies conducted by Girbau & Schwartz (2007, 2008) each had samples of 22 participants, while Thordardottir & Brandeker (2013) nearly met the criterion with 28 participants. With respect to the ratio of participants with LI to those with TL, none of the studies used samples that were representative of the prevalence of LI in the general population.

Gate design was the next quality indicator that that was evaluated. All studies utilized a two-gate design. One of the key distinctions between a one- and two-gate design is pre-selection of individuals with the condition being evaluated. In a one-gate design, the same set of inclusion criteria exists for those with and those without the condition being evaluated. A one-gate design, for example, may include administration of an index measure to an entire school without pre-selecting children diagnosed with LI. Two studies (Gutiérrez-Clellen & Simon-Cereijido, 2010; Windsor et al., 2010) reported recruiting participants from schools, but not whether they were pre-selected based on

language ability. If sampling had been conducted without pre-selection of individuals with and without LI, these may qualify as one-gate designs.

The next set of quality indicators considered the testing procedures. All studies gave the same index measures to all participants, regardless of group. Two studies utilized independent testing for their reference standard (Boerma et al., 2015; Li'el et al., 2018). The remainder of the studies recruited children with suspected or diagnosed LI, and then administered follow-up testing to confirm the presence of LI. None of the studies reported conducting blinded administration of the nonword repetition measure being evaluated. One remarkable aspect reported in three studies was the degree to which misdiagnosis was identified following testing by the researchers (de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018). All of these studies conducted confirmatory testing on potential participants and found that large proportions of bilingual children who had been previously diagnosed with LI actually exhibited TL.

The next quality indicator examined the inter-rater reliability of the scoring procedure. Eleven studies reported having used some method of inter-rater reliability, which was typically accomplished by recording participant repetitions of nonwords and having a second or third rater score some portion of the productions. Two studies did not report whether multiple raters were used (Armon-Lotem & Meir, 2016; de Almeida et al., 2017).

The last quality indicator to be evaluated was the validity and reliability of the reference standard used to establish the presence of LI in each study. Recall that assessment of LI in bilinguals should include multiple converging sources of direct and indirect evidence and assessment in both languages. Ten studies used a reference standard that was valid and reliable according to these criteria. The three most commonly used methods of identifying LI were standardized assessment, previous diagnosis, and parent report. For the three studies that did not meet the validity and reliability criteria,

the studies' designs may have dictated the choice of reference standards. Li'el et al. (2018) used previous diagnosis based on assessment in English and parent report as the method of identification, but their study also included more than 20 languages, potentially making cross-linguistic assessment unfeasible. Boerma et al. (2015) included a similarly diverse set of language profiles, with 14 languages represented, and relied on independent testing by clinicians who all used the same set of standardized assessment tools. Lastly, Girbau & Schwartz (2007) described their participants as Spanish speakers, not as bilingual Spanish-Catalan speakers. The chosen reference standard was consistent with the defined linguistic profile of the participants.

Diagnostic Accuracy of Nonword Repetition

The second research question examined the diagnostic accuracy of nonword repetition in bilingual children. To evaluate this question, diagnostic accuracy metrics were derived for each measure in the included set of studies, comprising 37 measures, as shown in Table 3. There were 25 measures for bilingual groups, and 12 for monolingual groups. Studies that included monolingual comparison groups, but did not report disaggregated outcomes for bilinguals, were not included in the estimation of pooled metrics or the meta-regression model, as these data do not reflect the diagnostic accuracy of their task for bilinguals separate from monolinguals.

Insert Table 3 about here

Figure 2 shows forest plots of LR+, and LR- for bilingual measures. The estimated LR+ mean of 3.91 ($95\%\ C.I.\ [3.25, 4.68]$) exceeded a value of 3, considered to be suggestive of LI, but did not meet or exceed the diagnostically informative value of 10 (Dollaghan & Horner, 2011). Across measures, LR+ ranged from uninformative to informative and included values from 2.19 to 12.5, with most items falling in the suggestive range. Fifteen of the bilingual measures from 11 studies exhibited diagnostically suggestive LR+ values, of greater than 3 and less than 10. Two measures from as many studies had

diagnostically informative LR+ values, of greater than 10. The estimated mean LR- of .33 ($95\%\ C.I.[.26, .41]$) was similarly only diagnostically suggestive but did not meet the criteria considered to be informative of <.1. Across measures, LR- values ranged from informative to uninformative and included values from .05 to .67, with many items falling in the suggestive or uninformative ranges. Eight measures, from seven studies, exhibited LR- values between .3 and .1 and thus were considered to be only suggestive of LI, while four measures from as many studies exhibited values that were informative, with values less than .1. The confidence intervals for LR+ and LR- revealed mixed outcomes. For LR+, the confidence interval contained only diagnostically suggestive values, whereas the confidence interval for LR- contained uninformative values. The upper confidence bound for LR- fell outside of the range considered to be either informative or suggestive, indicating that the true mean may be uninformative. Finally, for nearly all measures, the LR+ and LR- confidence intervals included uninformative values.

Insert Figure 2 about here

The estimated mean sensitivity was 74% ($95\%\ C.I.[66\%, 80\%]$), with values ranging from 41% to 96%. Eight measures from seven studies were found to have fair discriminant accuracy, meeting the 80% criterion. Only four measures, from as many studies, surpassed the 90% criterion for good discriminant accuracy (Plante & Vance, 1994). Specificity values exhibited a more restricted range, from 57% to 96%, with an estimated mean of 81% ($95\%\ C.I.[77\%, 85\%]$). Seven measures, from six studies, exhibited fair specificity, while only five measures from four studies had good specificity. As with likelihood ratios, confidence intervals for sensitivity and specificity estimates included diagnostically uninformative values.

The pooled DOR was 12.1 ($95\%\ C.I.[8.71, 16.3]$), indicating that a positive result on a nonword repetition task is about 12 times more likely to have come from someone with LI, than from someone

with TL. The ROC analysis, as shown in Figure 3, resulted in an estimated AUC of .87 $(95\%\ C.I.[.84, .91])$. The AUC, which provides an estimate of overall classification accuracy across studies, indicates that the identification ability of nonword repetition is substantially better than chance (.5) but well below perfect classification (1.0).

Insert Figure 3 about here

Classification Accuracy and Task Characteristics

The third research question examined whether specific types of nonword tasks or scoring methods were associated with higher diagnostic accuracy. To answer this question, three relevant task characteristics were identified: (a) the complexity of the stimuli, (b) the scoring method used, and (c) whether a task was specific to a given language. In addition to these task characteristics, several studies reported outcomes for monolingual comparison groups.

Complexity of nonword stimuli. The complexity of nonword stimuli, in terms of minimum of and maximum syllable length, varied among the tasks used in each study. Across studies, seven nonword tasks had a maximum nonword length of five syllables, five tasks used four syllables as their maximum, and another five used a maximum of three syllables. Minimum syllable lengths exhibited a more restricted range, with seven tasks using one syllable nonwords as their minimum, and six tasks using two syllable nonwords. Four studies reported optimal diagnostic accuracy outcomes when using a subset of nonword stimuli for scoring, omitting shorter length nonwords. Girbau & Schwartz (2007, 2008) and Guiberson et al. (2013) omitted one and two syllable nonword and included only nonwords of three, four, and five syllables. In the study conducted by Windsor et al. (2010), scoring for the Spanish task was based on five syllable nonwords only, whereas scoring for the English task was based solely on nonwords of four syllables.

Phoneme- and item- level scoring. Across studies, there were two methods of scoring: percent of phonemes correct (PPC) and percent of items correct (PIC), often referred to as phoneme- and item-level scoring, respectively. Eight studies used PIC scoring exclusively, three studies used PPC scoring exclusively, and two studies compared the two scoring methods. Boerma et al. (2015), using a sample of Dutch speakers from diverse linguistic backgrounds, found that PPC scoring exhibited better classification accuracy than PIC scoring. Guiberson & Rodriguez (2013), on the other hand, concluded that although PPC did differentiate between bilinguals with TL and those with LI, PIC scoring was far more accurate.

Language-specific and quasi-universal tasks. With respect to the language on which tasks were based, two types of measures were represented across studies: those that were purely language-specific and those that included language-independent elements. Language-specific (LS) tasks are constructed based on the phonology of a single language. A LS task that is based on English phonology, for example, asks a participant to repeat words that are plausible in English, but not necessarily in other languages. Language-independent tasks, on the other hand, are designed to consider the phonemes and phonotactic constraints of a broad range of languages. These tasks are also referred to as quasi-universal (QU), because they contain nonword stimuli that may be plausible words in a number of languages, though they still retain some language-specific elements (for an overview, see Chiat, 2015).

Eight studies exclusively used tasks that were LS, while five studies used tasks that included both QU and LS nonwords. One study directly compared the performance of bilingual children on LS and QU nonword repetition, to determine which exhibited better classification accuracy (Boerma et al., 2015). In their comparison of the two types, the authors found that children with LI performed with more difficulty than children with TL on both types of tasks, but that the QU measure resulted in higher classification accuracy. Three studies sought to determine whether performance on tasks that included

QU items was moderated by language exposure, and did not find a significant association (de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018).

Three studies evaluated the classification accuracy of LS tasks administered in two languages. Armon-Lotem & Meir (2016) administered LS tasks in both Russian and Hebrew, and found poor sensitivity (65%) but good specificity (94%), when using failing scores on both tasks as their criterion. Gutiérrez-Clellen & Simon-Cereijido (2010) also used a failing score on both Spanish and English tasks, and found similarly inadequate sensitivity (41%) but good specificity (96%). Windsor et al. (2010) administered LS tasks in English and Spanish to bilinguals and found that the English task exhibited good sensitivity (95%) but poor specificity (57%). Results of the Spanish task, on the other hand, revealed poor sensitivity (58%) and fair specificity (82%). The authors did not report classification accuracy when considering the results of both tasks together.

Monolingual comparisons. Several studies included monolingual comparison groups. In the evaluation of QU and LS tasks, a single study directly compared the performance of monolinguals and bilinguals. Boerma et al. (2015) found a significant difference in performance on a Dutch LS task, but not on a QU one. The authors concluded that the LS task disadvantaged bilingual children due to limited language-specific knowledge of Dutch to support working memory for the purposes of repeating novel words. Windsor et al. (2010), using solely LS tasks, reported similar performance patterns. Applying a novel approach, the researchers administered tasks in both Spanish and English to bilinguals, as well as to monolingual English speakers. They found that, on the English task, bilinguals with TL exhibited performance below that of monolingual children with TL, but above that of all groups of children with LI. Performance on the Spanish task, though, revealed just the opposite; bilingual children with TL performed better than monolinguals, suggesting that the language of the task affected performance.

In comparing bilinguals and monolinguals, studies also varied in the ways that they applied cut scores to each group. Some studies used the same cut score for both groups, while others derived optimal cut scores for each group. Those studies that used the same cut score for both groups reported lower classification accuracy with bilinguals (de Almeida et al., 2017; Thordardottir & Brandeker, 2013; Windsor et al., 2010). Studies that used separate cut scores for each group found that the optimal cut score varied for bilinguals compared to monolinguals. In their study, Armon-Lotem & Meir (2016) found that monolingual cut scores resulted in inadequate diagnostic accuracy when used with bilinguals, and classification accuracy improved when cut scores were derived specifically for the bilingual group. In their study, Tuller et al. (2018) found that the optimal cut score varied not only by bilingual status, but also by the language of the task, and the age of participants. In comparing bilinguals to monolinguals with optimal cut scores applied to each group, the authors reported lower specificity for bilinguals in both their German and French samples, but lower sensitivity in their French sample only. The sensitivity of the task in their German sample was slightly higher for bilinguals. Hamann & Abed Ibrahim (2017) also derived separate cut scores for each group, and similarly reported that the optimal cut score resulted in lower specificity for bilinguals compared to monolinguals, with a corresponding increase in sensitivity.

Meta-regression analysis. The contribution of specific task characteristics to variability in diagnostic accuracy was further evaluated using a meta-regression analysis. First, in order to determine whether differences in reported diagnostic accuracy were due to sampling error, or genuine differences between studies, the potential presence of heterogeneity was examined for all measures used with bilingual groups. Using a random effects model, a test of between groups heterogeneity was nonsignificant at the $\alpha = 0.05$ level, $Q(21) = 31.44, p = .07, I^2 = 33.22$. Although the p-value was not significant, Cochran's *Q*-test has low power to detect heterogeneity, especially in studies with relatively small samples. The I^2 value of 33.22% indicated that there was a low-moderate heterogeneity in DOR

between studies (Higgins et al., 2003). A model that included monolingual groups, in addition to bilingual groups, exhibited somewhat more heterogeneity, $Q(33) = 60.68,\ p = .002, I^2 = 45.62$.

To identify potential sources of heterogeneity, a random effects meta-regression model was estimated, which included all index measures across studies, for both monolingual and bilingual groups. Because many studies reported results for multiple measures, this model comprised 23 clusters with adjusted standard errors for a total of 34 measures. The model contained three categorical variables, each with two levels: (a) scoring method: PIC or PPC, (b) language status: monolingual or bilingual, and (c) nonword repetition task type: LS only or tasks with both QU and LS elements. The single study that included a fully QU measure (Boerma et al., 2015) was included in the latter group. Results of the model are shown in Table 4.

In the meta-regression model, the coefficient representing the difference between QU and LS tasks was statistically significant at the $\alpha = 0.05$ level ($DOR = 2.47, 95\%\ C.I[1.17, 5.21], p = .02$), using LS tasks as the reference group. Across studies, tasks that included QU elements exhibited better classification accuracy than purely LS tasks. The relative DOR of 2.47 indicates that tasks with QU stimuli were about two and a half times more likely to result in correct identification of LI compared to LS tasks, for both monolingual and bilingual groups. The coefficient representing the difference between monolingual and bilingual groups, was not statistically significant ($DOR = 1.84, 95\%\ C.I[0.65, 5.18], p = .23$), suggesting that the diagnostic accuracy of nonword repetition was no different for monolinguals or bilinguals. The coefficient for scoring type, using PPC scoring as the reference group, was similarly nonsignificant ($DOR = 1.61, 95\%\ C.I[0.87, 3], p = .12$), which suggests that there was no difference between the two scoring methods in the included studies. The effect for task type (QU vs. LS) was robust to the inclusion of mean participant age as well as minimum and maximum syllable length as control variables.

Insert Table 4 about here

To identify whether diagnostic accuracy was the same for each group, regardless of task type, an additional model was estimated, which included a covariate representing the interaction between task type and language status. The resulting interaction term was nonsignificant ($DOR = 1.48$, $95\%\ C.I[0.29, 7.44], p = .62$), indicating that the association between diagnostic accuracy and nonword repetition task type was the same for both monolinguals and bilinguals. The corrected Akaike information criterion (AICc) values for this model and the reduced model without the interaction term, were 103.56 and 100.81, respectively. The lower AICc value for the reduced model indicates that excluding the interaction resulted in a better fit, making it the preferred model in this analysis.

Publication Bias

An examination of potential publication bias was conducted by comparing the average DOR for bilingual measures in each study with the effective sample size (Deeks et al., 2005). Results of the test revealed a significant relationship between sample size and diagnostic accuracy, $F = 7.2\ (1, 10), p = 0.023$, such that larger samples were associated with lower diagnostic accuracy. This suggests possible publication bias due to the fact that smaller studies with lower diagnostic accuracy were underrepresented among the included articles.

Discussion

This meta-analysis presented an overview of studies of the diagnostic accuracy of nonword repetition in the identification LI in bilingual children. A systematic review of the literature identified 13 articles that met the eligibility criteria. Although the initial search yielded a greater number of studies, a relatively small proportion examined diagnostic accuracy in bilinguals. The three aims of this study were to: (a) evaluate the quality of evidence of extant research, (b) examine diagnostic accuracy, and (c) identify task characteristics that may be associated with higher diagnostic accuracy. The overarching

goal of this study was to determine the degree to which current research supports the use of nonword repetition as a means of identifying LI in bilingual children.

Quality of Evidence

The evaluation of the quality of evidence revealed a great deal of similarity in research methodology across studies, though there were notable differences. One key piece of information in diagnostic test research is who truly has the condition that the index measure is intended to identify, which is accomplished by way of the reference standard. Studies utilized a variety of different methods to establish the presence of LI, with standardized testing, parent report, and previous diagnosis being the most common. Consistent with current best practice, all studies used multiple converging sources of information. In spite of the fact that the inclusion of language sampling as part of an assessment battery may yield better classification accuracy (Lazewnik et al., 2019), only two studies included this as part of their reference standard (Gutiérrez-Clellen & Simon-Cereijido, 2010; Thordardottir & Brandeker, 2013). It is also promising that all but three studies used a reference standard that included both languages, given the importance of considering cross-linguistic abilities in the diagnostic process. Lastly, although independent testing was included as a quality indicator, this may be unrealistic for studies of bilinguals to implement, given the lack of an agreed upon reference standard. Legitimate concerns about the possibility of misdiagnosis may drive researchers to conduct their own assessment. The fact that several studies found a number of misdiagnosed participants (de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018) highlights the value of adhering to a stringent reference standard. These children would have been misclassified as having LI, if not for the assessment administered directly by the researchers.

Diagnostic Accuracy of Nonword Repetition

Results from the analysis of diagnostic accuracy suggest that, in the identification of LI in bilinguals, nonword repetition does not demonstrate sufficient precision when used on its own, though it may be a useful addition to an assessment battery. Classification accuracy varied greatly across studies, ranging from poor to good. Despite some studies reporting strong classification accuracy, pooled averages of sensitivity and specificity were poor and fair, respectively. The pooled specificity of 81% was higher than the sensitivity value of 74%, indicating that individuals with TL were more accurately classified than those with LI. Pooled positive and negative likelihood ratios of 3.91 and .33, respectively, only met the criteria to be considered clinically suggestive. The estimated AUC of .87 suggests that overall classification accuracy was far better than chance, but this value should be interpreted alongside other reported metrics. No specific measure of nonword repetition stood out as objectively more accurate based on the evidence reviewed, though the range of values shows that some measures did exhibit better performance than others. An inspection of Figure 2 shows nine measures whose LR+ values exceed the confidence interval for the pooled mean, but only a single measure whose confidence interval does not overlap with that of the mean. LR- values followed a similar trend, with no measure exhibiting a confidence interval that is exclusive of the mean. In addition, the 95% confidence intervals for the majority of the studies included values that are considered to be diagnostically uninformative, for both LR+ and LR-. This was also observed in the confidence interval for the pooled average of LR-. Taken as a whole, these results indicate the nonword repetition lacks sufficient diagnostic accuracy to be used in the absence of other assessment tools.

Although diagnostic accuracy outcomes across studies do not support the use of nonword repetition as the sole method identification of LI in bilinguals, in most clinical applications it is unlikely to be used in the absence of other diagnostic tools and is best-suited for use as a part of an assessment battery. Previous diagnostic research into LI in bilinguals has demonstrated the effectiveness of nonword repetition as part of a composite measure when used alongside parent report (Paradis et al., 2013),

narrative production (Boerma & Blom, 2017), or standardized assessment (Thordardottir & Brandeker, 2013).

Task Characteristics and Classification Accuracy

In the examination of task characteristics that may affect diagnostic accuracy, three potentially salient features were identified: (a) nonword complexity, (b) scoring method, and (c) task type (i.e., LS vs. QU). With respect to complexity, nonword tasks varied substantially in their minimum and maximum syllable lengths. Because of the variety of nonword stimuli used across studies, the ideal range of complexity for optimal discriminant accuracy remains unclear. The four studies that restricted the range of nonword complexity reported better diagnostic accuracy when omitting shorter nonwords. Although these outcomes provide some support for the omission of shorter nonwords, results cannot be generalized beyond the specific nonword tasks that were used in these studies.

The use of two different scoring methods, PIC and PPC, was distributed roughly evenly across all measures, with conflicting outcomes for those studies that directly compared the two methods (Boerma et al., 2015; Guiberson & Rodríguez, 2013). Furthermore, results of the meta-regression model did not reveal a significant difference in the diagnostic accuracy between the two scoring methods. Given the differences in study design, it is not possible to definitively conclude which of the two scoring methods resulted in better classification accuracy. Further investigation, controlling for nonword repetition stimuli, task type, and language background may provide more insight into the differential classification ability of the two scoring methods, when used in bilingual assessment. Because studies utilized a number of different sources for their nonword repetition stimuli, there was little direct evidence to be able compare the outcomes for any specific stimuli.

One very important issue to consider is whether nonword repetition tasks that are designed for a single language disadvantage bilinguals, given the potential risk for misidentification when both

languages are not considered, as identified by Gutiérrez-Clellen & Simon-Cereijido (2010). Evidence from other studies also gives rise to concerns about the utility of nonword repetition administered in a single language. Kohnert et al (2006), in their investigation of English nonword repetition for monolinguals and bilinguals with TL and monolinguals with LI, found that bilinguals with TL performed with more difficulty than monolinguals with TL, resulting in potential over-identification of bilinguals. In the present meta-analysis there was evidence of a difference between LS and QU nonword repetition. Results of Boerma et al. (2015) provide support for QU measures, though no other studies directly compared the two types of tasks. Outcomes from the meta-regression model also showed that tasks that included QU elements were associated with significantly higher classification accuracy, providing additional support for their use. In comparing bilingual and monolingual performance, several of the included studies found that classification accuracy was lower for bilinguals (Armon-Lotem & Meir, 2016; Thordardottir & Brandeker, 2013; Windsor et al., 2010). Importantly, LS tasks were used in all of these studies. The small number of studies that utilized QU tasks generally found that the performance difference between monolinguals and bilinguals diminished when this source of potential bias was reduced. Importantly, tasks that included QU elements comprised only 10 measures across all studies. Although results presented here suggest a potential advantage for QU tasks, further investigation is warranted regarding any difference in classification accuracy compared to LS nonword repetition, given the small number of QU tasks included.

The way that cut scores are applied also has a direct impact on classification accuracy. In deriving cut scores, it important to consider that misclassification may occur if monolingual cut scores are used for bilinguals, as demonstrated by Armon-Lotem & Meir (2016). The use of monolingual cut scores for bilinguals has been shown to result in poorer classification accuracy in other assessment methods as well (Gillam et al., 2013). Tuller et al. (2018) also showed that optimal cut scores varied by

bilingual status. These results suggest that the optimal cut score for any given nonword task cannot be presumed to be the same for monolinguals and bilinguals.

Limitations and Future Directions

There were several imitations of the present meta-analysis. First, coders were not blinded to the authors of studies, resulting in potential bias. Second, it is possible that articles were missed during the search and study selection process. Although all efforts were made to conduct a comprehensive literature review, the possibility of having missed certain relevant articles cannot be ruled out. Third, the small sample size, though common in meta-analytical research, limits the ability to draw clear and definitive conclusions from observed outcomes.

With respect to quantitative results, it is important to keep in mind that the individual outcomes do not tell the whole story. Although measures of sensitivity, specificity, likelihood ratios, DOR, and AUC serve as useful objective metrics to compare the diagnostic accuracy of different measures, they should not be interpreted as the sole indicator of a specific measure's utility. The associated confidence intervals for each of these values provides some additional insight into the accuracy of the index measures, and it is telling that most confidence intervals included diagnostically uninformative values. The presence of heterogeneity between measures must also be considered. Less heterogeneity in the classification accuracy values would suggest more certainty in the interpretation of the outcomes for each index measure. In addition, the meta-regression, although informative, was constrained by a relatively small number of studies. For this reason, outcomes must be tempered by the knowledge that the sample used in this analysis may not be representative of the population. In addition, results of this model only revealed associations between the included variables and should not be interpreted as causal. Lastly, the presence of publication bias may have contributed to inflated diagnostic accuracy overall, given the negative association between diagnostic accuracy and sample size. If no bias were

present, we would have expected the primary study outcomes to be normally distributed, which was not the case. Although other factors may have played a role in this asymmetric distribution (Deeks et al., 2005), potential publication bias needs to be considered when interpreting diagnostic accuracy outcomes.

The characteristics of the participants across studies is another relevant aspect when considering how nonword repetition may be used for individuals of different ages or linguistic backgrounds. The included studies represented a diverse set of languages from several different countries, and the analysis in the present study included all languages as an aggregate. This is a limitation in that the diversity of language backgrounds included may have obscured diagnostic accuracy outcomes for any given language. Similarly, studies did not universally report dimensions of bilingual language acquisition such as age of exposure to and use of a second language. Because of the heterogeneity inherent in bilingualism, factors such as these may affect performance on a nonword task presented in a given language. Despite this limitation, because results were not limited to a single language pair or acquisitional environment, outcomes are potentially more generalizable to larger range of cultural and linguistic contexts than results from any single study. Another limitation was the variability of ages across studies, which ranged from 3;0 to 11;6. Age was not evaluated as a predictor of interest in the current meta-analysis, due to its lack of an identified effect in previous nonword repetition research (Graf Estes et al., 2007). Despite this, it is plausible that diagnostic accuracy may vary as a function of age. This should certainly be explored in future studies that include sample sizes sufficiently large to identify any possible effect. In addition, although most studies did not include participants under the age of 5;0, some studies did use preschool-age children in their samples. Given the potential of nonword repetition for use with three to five year-olds, and the evidence of its utility with even younger children (Guiberson & Rodríguez, 2016; Hoff et al., 2008), future studies may focus on a broader age range to further generalize outcomes to children of differing ages.

Finally, in the examination of nonword task characteristics, the meta-regression model included a set of predictors that were identified as particularly relevant for bilingual assessment. Though other elements of nonword tasks may play a role in diagnostic accuracy, the small sample of the present study size restricted the ability to include additional predictors. For this reason, future studies should assess the potential effect of other task characteristics on diagnostic accuracy, such as the length and number of stimuli.

Clinical Implications

As a diagnostic tool nonword repetition may be a useful addition to a clinical assessment battery given its straightforward administration and the unique information that it provides. It is particularly appealing for use with bilingual children, due to the minimal reliance on prior linguistic knowledge, potentially reducing assessment bias. Although it exhibits a number of advantages over traditional forms of assessment, nonword repetition lacks sufficient diagnostic accuracy to be used without the aid of additional measures of LI. Average diagnostic accuracy for the set of measures evaluated fell in the poor to fair range, suggesting that the task may be of limited value when used on its own. Despite its variable diagnostic accuracy, nonword repetition may prove useful as a supplemental assessment tool, and it may offer information above and beyond that which other methods are able to provide. Given its relative ease of administration, nonword repetition may be particularly useful as a screening tool in the early identification of LI in bilinguals when used in conjunction with other screening techniques, such as parent report. Lastly, although the focus of this study was specifically on diagnostic accuracy, nonword repetition performance may also be useful for informing clinical decisions for both assessment and treatment, given its associations with working memory, vocabulary development, and phonological awareness.

When including nonword repetition in their diagnostic toolkit, clinicians should consider the phonotactic similarity of the task to both of the client's languages. The limited evidence supporting the use of QU tasks suggests that they are a potentially promising alternative to LS tasks. Clinicians should take particular caution when interpreting results for LS tasks administered only in one language. Clinicians should also be cautious in generalizing results to any language for which the use of nonword repetition has not been sufficiently examined, or when administering tasks that have not been validated for use in a given language. Furthermore, there is only limited evidence supporting the use of nonword repetition with preschool-age bilingual children. In this age range, bilingual phonological development may be heavily influenced by changes in language exposure and use, potentially resulting in inconsistent nonword repetition performance over time. Clinicians should be cautious when interpreting results in children under the age of five, for diagnostic purposes. Limitations notwithstanding, nonword repetition demonstrates some potential in the identification of LI in bilingual children when used alongside other assessment methods, though a fuller understanding of its clinical application is necessary.

Tables and Figures

Table 1

Study Characteristics

Study	Group	n TL [a]	n LI [a]	Age Range	L1 [b]	L2 [c]
Armon-Lotem & Meir (2016)	Bi	117	27	5;5-6;8 [d]	Russian	Hebrew
	Mo	38	14		Hebrew	N/A
	Mo	20	14		Russian	N/A
Boerma et al. (2015)	Bi	30	30	4;6-7;2	Chinese, Danish, Dari, Pashto, Egyptian Arabic, Frisian, Kirundi, Moroccan Arabic, Portuguese, Russian, Suryoyo, Tarifit-Berber, Turkish	Dutch
	Mo	30	30	4;11-7;3	Dutch	N/A
de Almeida et al. (2017)	Bi	61	21	5;4-8;11	Arabic, European Portuguese, Turkish	French
	Mo	12	17	5;6-8;7	French	N/A
dos Santos & Ferré (2018)	Bi	30	13	5;4-8;2	English, Arabic	French
	Mo	14	10	5;4-8;5	French	N/A
Girbau & Schwartz (2007)	Bi	11	11	8;3-10;11	Catalan	Spanish
Girbau & Schwartz (2008)	Bi	11	11	7;6-10;10	Spanish	English
Guiberson & Rodríguez (2013)	Bi	23	21	3;0-5;10	Spanish	English
Gutiérrez-Clellen & Simon-Cereijido (2010)	Bi	95	49	3;11-7;10	Spanish	English
Hamann & Abed Ibrahim (2017)	Bi	46	8	5;5-9;3	Arabic, Portuguese, Turkish	German
	Mo	10	12	5;6-9;4	German	N/A
Li'el et al. (2018)	Bi	42	19	5;0-6;11 [e]	African Arabic, Cantonese, Chin, Dutch, Dzongkha, Farsi, Filipino, French, German, Hebrew, Hindi, Indonesian, Italian, Japanese, Kirundi, Kurdish, Macedonian, Malayalam, Mandarin, Middle Eastern Arabic, Serbian, Shilluk, Spanish, Tamil, Telegu, Turkish, Vietnamese	English
Thordardottir & Brandeker (2013)	Bi	14	14	5;0 [f]	Arabic, Dutch, Japanese, Punjabi, Russian, Singhalese, Spanish, Tamil, Urdu	French
	Mo	14	14		French	N/A
Tuller et al. (2018)	Bi	69	26	5;4-8;11	Arabic, Portuguese, Turkish	French
	Bi	48	8	5;1-9;11	Arabic, Portuguese, Turkish	German
	Mo	37	17	5;7-8;8	French	N/A
	Mo	10	12	5;6-9;4	German	N/A
Windsor et al. (2010)	Bi	65	19	6;0-11;6 [g]	Spanish	English
	Mo	69	34		English	N/A

Table 1

Study Characteristics

Note. Mo = monolingual, Bi = bilingual

(table continues)

[a] The number of participants on which nonword repetition data was collected. [b] Languages listed represent first, home, or minority languages. [c] Languages listed reflect second or majority languages. [d] The aggregated age range across groups. [e] Age range reported as between five and six years. [f] Participants were reported to be equivalent in age with a mean of five years across groups. [g] The age range used for participant recruitment.

Table 2

Methodological Quality Indicators

Study	Sample >30	Representative ratio	Gate design	Independent testing	Blinded testing	Reliability	Same index measure given to all	Valid & reliable reference standard	Total
Armon-Lotem & Meir (2016)	yes [a]	no	2	no	no	no	yes [a]	yes [a]	3
Boerma et al. (2015)	yes [a]	no	2	yes [a]	no	yes [a]	yes [a]	no	4
de Almeida et al. (2017)	yes [a]	no	2	no	no	no	yes [a]	yes [a]	3
dos Santos & Ferré (2018)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4
Girbau & Schwartz (2007)	no	no	2	no	no	yes [a]	yes [a]	no	2
Girbau & Schwartz (2008)	no	no	2	no	no	yes [a]	yes [a]	yes [a]	3
Guiberson & Rodríguez (2013)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4
Gutiérrez-Clellen & Simon-Cereijido (2010)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4
Hamann & Abed Ibrahim (2017)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4
Li'el et al. (2018)	yes [a]	no	2	yes [a]	no	yes [a]	yes [a]	no	4
Thordardottir & Brandeker (2013)	no	no	2	no	no	yes [a]	yes [a]	yes [a]	3
Tuller et al. (2018)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4
Windsor et al. (2010)	yes [a]	no	2	no	no	yes [a]	yes [a]	yes [a]	4

Note. Scoring: yes = 1, no = 0

[a] Preferred

Table 3

Diagnostic Accuracy Metrics For all Nonword Repetition Measures Across

Study	Group	Language [a]	NWR Task	Scoring	Cut score	Sensitivity	Specificity	LR+	LR-	AUC [95% C.I.] [b]
Armon-Lotem & Meir (2016)	Bi	Hebrew	LS	PIC	71%	81%	79%	3.78	.24	.89
	Bi	Russian	LS	PIC	79%	70%	76%	2.94	.39	.78
	Bi	Hebrew & Russian	LS [c]	PIC	71%/79% [d]	65%	94%	10.83 [e]	.37 [e]	
	Mo	Hebrew	LS	PIC	86%	93%	66%	2.71	.11	.84
	Mo	Russian	LS	PIC	71%	86%	90%	8.57	.16	.94
Boerma et al. (2015)	Bi	Dutch	QU	PPC	78.1%	83%	93%	11.9	.14	.9 [.81, .99]
	Bi	Dutch	LS	PPC	63.8%	63%	93%	9	.4	.79 [.68, .91]
	Bi	Dutch	QU	PIC	NR	87%	83%	5.12 [e]	.16 [e]	.89 [.79, .98]
	Bi	Dutch	LS	PIC	NR	77%	73%	2.85 [e]	.32 [e]	.76 [.63, .89]
	Mo	Dutch	QU	PPC	77.7%	83%	90%	8.3	.19	.94 [.89, 1]
	Mo	Dutch	LS	PPC	72.7%	93%	93%	13.3	.08	.95 [.91, 1]
de Almeida et al. (2017)	Bi	French	LS + QU	PIC	80%	81%	79%	3.86 [e]	.24 [e]	
	Mo	French	LS + QU	PIC	80%	88%	83%	5.18 [e]	.14 [e]	
dos Santos & Ferré (2018)	Bi + Mo [f]	French	LS + QU	PIC	-2 SD	59%	95%	13	.43	
	Bi + Mo [f]	French	QU	PIC	-1.65 SD	59%	91%	6.5	.45	
	Bi + Mo [f]	French	LS	PIC	-1.28 SD	77%	93%	11.33	.24	
	Mo	French	LS + QU	PIC	-2 SD	90%	92%	11.7	.11	
Girbau & Schwartz (2007)	Bi	Spanish	LS	PIC	50% [g]	100%	91%	11	0	
Girbau & Schwartz (2008)	Bi	Spanish	LS	PIC	33% [g]	82%	91%	9	.2	
Guiberson & Rodríguez (2013)	Bi	Spanish	LS	PIC	DF [g]	71%	74%	2.74	.39	
	Bi	Spanish	LS	PPC	DF [g]	48%	78%	2.19	.67	
Gutiérrez-Clellen & Simon-Cereijido (2010)	Bi	English	LS	PPC	70%	55.1%	82.1%	3.08 [e]	.55 [e]	.76 [.67, .84]
	Bi	Spanish	LS	PPC	70%	61.2%	82.1%	3.42 [e]	.47 [e]	.76 [.68, .84]
	Bi	English & Spanish	LS [c]	PPC	70%	40.8%	95.8%	9.71	.62	
Hamann & Abed Ibrahim (2017)	Bi	German	LS + QU	PIC	63.5%	100%	73.91%	3.83	0	

(table continues)

Table 3

Diagnostic Accuracy Metrics For all Nonword Repetition Measures Across

	Mo	German	LS + QU	PIC	59.9%	91.7%	90%	9.17	.09	.95
Li'el et al. (2018)	Bi	English	LS	PIC	≤7	89.5%	81%	4.7	.13	.9 [.81, .97]
Thordardottir & Brandeker (2013)	Bi	French	LS	PPC	82%	85%	79%	4.05 [e]	.19 [e]	
	Mo	French	LS	PPC	82%	92%	100%	26.76 [e]	.08 [e]	
Tuller et al. (2018)	Bi	French	LS + QU	PIC	79.6%	84%	77%	4.8	.28	.86
	Bi	German	LS + QU	PIC	58.3%	100%	88%	7.39 [e]	0	.94
	Mo	French	LS + QU	PIC	77.5%	88%	92%	10.63	.13	.97
	Mo	German	LS + QU	PIC	59.9%	92%	90%	9.17	.09	.95
Windsor et al. (2010)	Bi	English	LS [c]	PPC	78% [h]	94%	57%	2.2	.09	
	Bi	Spanish	LS [c]	PPC	80% [i]	58%	82%	3.14	.52	
	Mo	English	LS [c]	PPC	78% [h]	76%	73%	2.78	.32	
	Mo	Spanish	LS [c]	PPC	80% [i]	77%	64%	2.11	.37	

Note. Cut scores, sensitivity, specificity, are listed as reported by primary studies. AUC values are rounded to two decimals. Bi = bilingual; Mo = monolingual; PPC = percent phonemes correct; PIC = percent items correct; LS = language-specific; QU = quasi-universal; NR = not reported; SD = standard deviation; DF = discriminant function

[a] The phonology of the language(s) to which the task conforms. [b] Missing AUC values and associated confidence intervals indicate that the study did not report these data. [c] Language-specific tasks administered in both languages. [d] The cut scores for Hebrew and Russian, respectively. [e] Studies did not report LR+ or LR-; values were estimated from sensitivity and specificity using a continuity correction of .5 added to all cells of a study's 2x2 table if zero values were present. [f] These data represent the aggregated accuracy of bilinguals and monolinguals and were not included in the calculation of pooled metrics, or the meta-regression model. [g] Cut score using 3, 4, and 5 syllable nonwords. [h] Cut score using 4 syllable nonwords. [i] Cut score using 5 syllable nonwords.

Table 4

Meta-regression Analysis

Coefficient	DOR	95% C.I.	β	p
Intercept	8.53*	[5.5, 13.24]	2.14	0
Monolingual [a]	1.84	[0.65, 5.18]	0.61	.23
LS + QU NWR [b]	2.47*	[1.17, 5.21]	0.9	.02
PIC scoring [c]	1.61	[0.87, 3]	0.48	.12

Note. DOR = Relative diagnostic odds ratio, derived by exponentiating the regression coefficient (β).

[a] The difference between the monolingual and bilingual groups, using bilingual as the reference group. [b] The difference between fully language-specific (LS) tasks and those with quasi-universal (QU) elements, using LS as the reference group. [c] The difference between item-level (PIC) and phoneme-level (PPC) scoring, using PPC as the reference group.

*Significant at the <.05 level.

Figure 1. Flowchart of Search Results and Study Selection

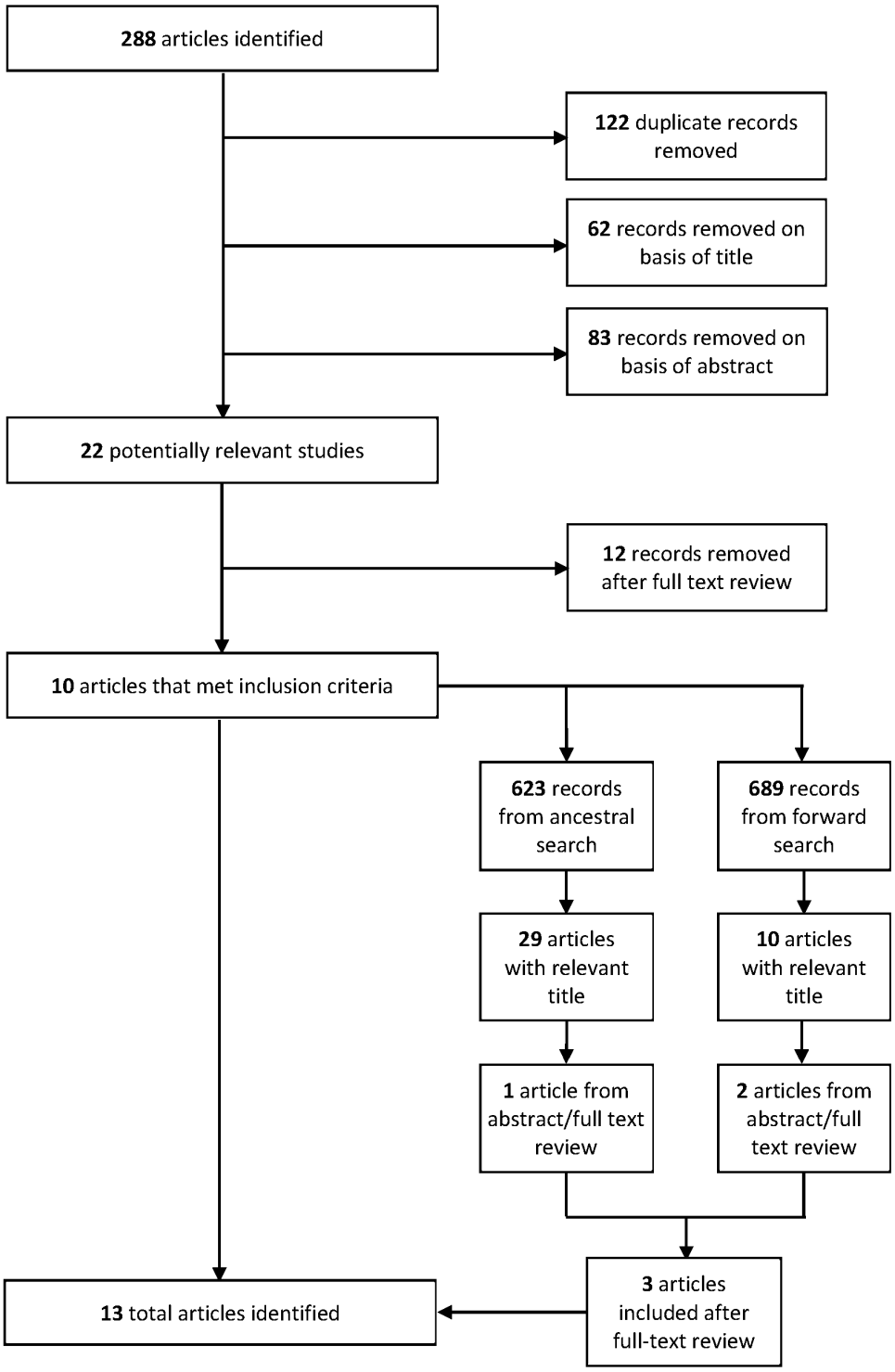


Figure 2. Forest Plot of LR+ (top) and LR- (bottom) for Bilingual Measures

Note. He = Hebrew; Ru = Russian; En = English; Sp = Spanish; Fr = French; Ge = German; QU = quasiuniversal; LS = language-specific; PPC = percent phonemes correct; PIC = percent items correct Squares represent LR+ for each study; diamond represents the pooled LR+ for all studies; horizontal lines represent the 95% confidence intervals for each nonword repetition measure. A continuity correction of .5 was added to all cells of a study's 2x2 table if zero values were present.

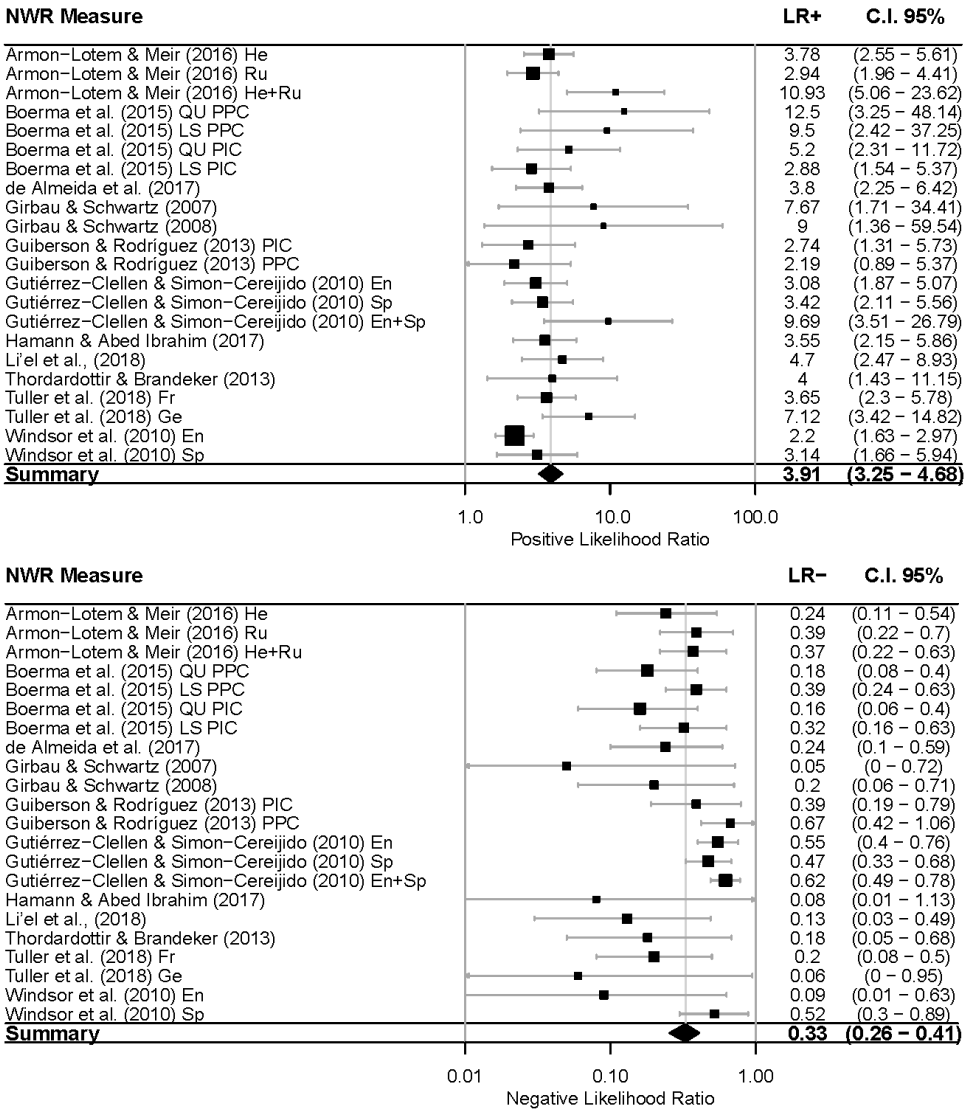
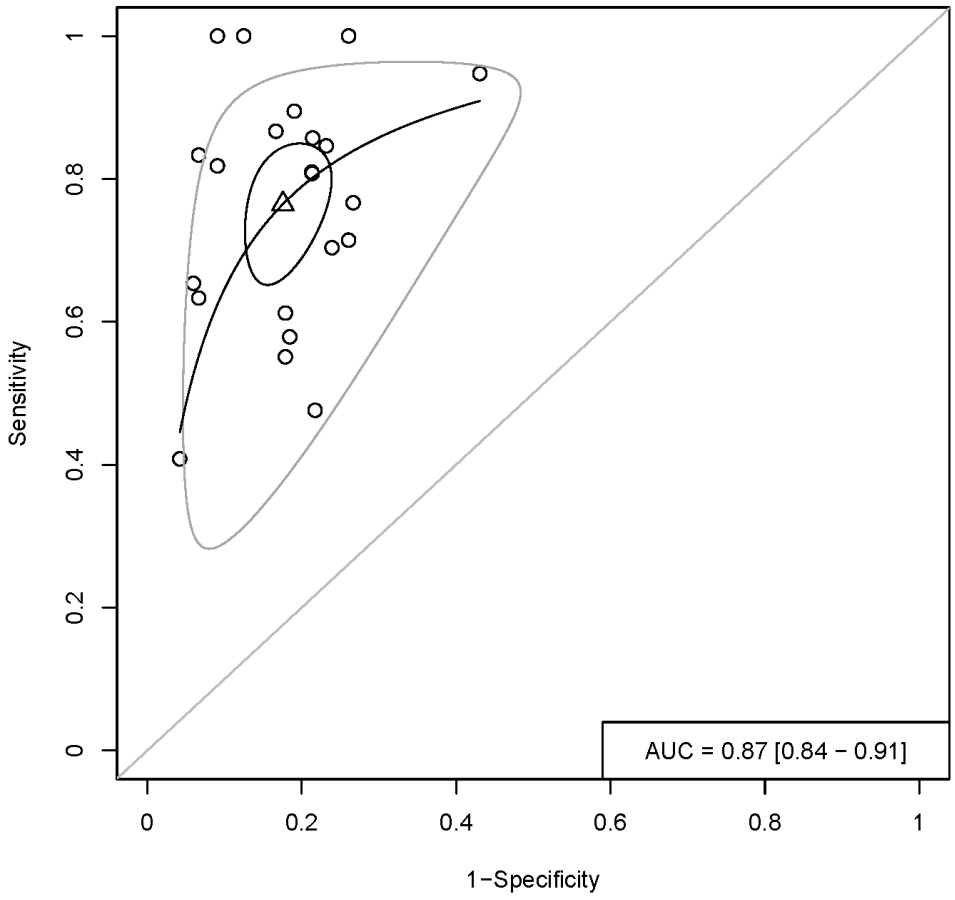


Figure 3. Summary ROC Analysis for Bilingual Measures

Note. Small circles represent each nonword repetition measure; triangle represents the summary estimate across studies; curved line represents estimated summary ROC curve for all studies; small ellipse represents the 95% confidence interval; large ellipse represents the 95% prediction interval.



Chapter III: The Disproportionate Identification of Language-Related Disorders in Bilingual Children

The lack of equitable access to education services for bilingual children is a long-standing and pervasive issue in US schools. Despite making up approximately 10% of the public school population, children from language-minority backgrounds are at a higher risk for academic difficulties (National Center for Education Statistics, 2021). Bilingual children are also disproportionately represented in special education, with evidence of both over- and underrepresentation (Artiles et al., 2005; Cruz & Firestone, 2022; De Valenzuela et al., 2006; Hibel & Jasper, 2012; Linn & Hemmer, 2011; P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017; Samson & Lesaux, 2009; Shifrer et al., 2011; Sullivan, 2011; Umansky et al., 2017; Yamasaki & Luk, 2018). The presence of disproportionality suggests unequal access to special education services, and both over- and underrepresentation have potential consequences for those children affected by it. Overrepresentation represents a scenario in which bilingual children are more likely to receive special education than their monolingual peers, suggesting that some children may be unnecessarily receiving services. This is concerning given that children with identified disabilities are more likely to drop out of school (Kemp, 2006), are less likely to enroll in post-secondary education (Wagner et al., 2006), and are at an increased risk of incarceration (Bell, 2016). Underrepresentation, on the other hand, may represent a delay or denial of access to needed services, the consequences of which are likely to be detrimental to achievement in future grades.

The identification of language-related disorders, including specific learning disabilities (SLD) and speech or language impairments, hereafter referred to as communication disorders (CD), is an area of particular relevance for bilingual children, given the challenge of differentiating between language differences and disorders (Case & Taylor, 2005; Castilla-Earls et al., 2020; Shenoy, 2014). Several studies have identified evidence of disproportionate identification of CD and SLD in bilingual students (Artiles et al., 2005; P. L. Morgan, Farkas, et al., 2017; Sullivan, 2011; Umansky et al., 2017; Yamasaki & Luk, 2018).

The identification of both CD and SLD relies on the measurement of skills related to language ability, including verbal expression, comprehension, reading, writing, as well as associated cognitive areas, such as working memory and executive functioning. Typical characteristics of dual language development may be misattributed to symptoms of an underlying language-related disorder if not accounted for in assessment (Bedore et al., 2005; Bialystok et al., 2010). Although the difficulties associated with CD and SLD may be manifested differently, the well-established link between oral and written language disorders (Boudreau & Hedberg, 1999; Peterson et al., 2009) reinforces the importance of examining these two disabilities together.

Studies of children from linguistically diverse backgrounds have used a variety of terms to define the population of interest, such as *dual language learner, English language learner (ELL), language-minority, or bilingual.* Although these identifiers are frequently used interchangeably, there are some key differences with respect to who is being described. In the following sections, *bilingual* will be the preferred term to refer to children who are exposed to and have had the opportunity to learn multiple languages, and is synonymous with *dual language learner* (Paradis et al., 2011). This term is an acknowledgement that children with limited English proficiency are, in fact, emergent bilinguals and that many bilingual children do not exhibit limited English proficiency, but rather are English-proficient bilinguals (García et al., 2008).

Patterns of Disproportionate Representation

Disproportionality has commonly been regarded primarily as a problem of overrepresentation (Artiles & Trent, 1994; Coutinho & Oswald, 2000; Dunn, 1968), a perspective that is mirrored in federal education policy (US Department of Education, 2016a). Despite this focus, research has increasingly highlighted a more complex issue, providing evidence for both overrepresentation (Artiles et al., 2005; De Valenzuela et al., 2006; Sullivan, 2011) and underrepresentation (P. L. Morgan et al., 2015; P. L.

Morgan, Farkas, et al., 2017; Sullivan & Bal, 2013) of bilingual children in special education. Although these two patterns may appear to be at odds with one another, empirical evidence suggests that both exist to differing degrees depending on factors such as grade level and language background. Several studies point to a dynamic pattern of disproportionality, demonstrating that the likelihood of disability identification may differ as students progress through school. Samson and Lesaux (2009), in their evaluation of disproportionality of bilingual students from kindergarten to third grade, identified an increased likelihood of disability identification over time. The authors found increasing levels of overrepresentation, such that bilingual children were underrepresented in kindergarten, but overrepresented by third grade. Several other studies have found similar evidence of an increasing likelihood of disability identification over time for bilinguals (Artiles et al., 2005; Artiles & Trent, 1994; Hibel & Jasper, 2012; Samson & Lesaux, 2009).

Linguistic differences in the population under investigation may also account for the various outcomes of previous studies. Although bilingual children are often treated as a monolithic group, differences in language background—proficiency across languages—may impact the likelihood of being identified with a disability. English-proficient bilinguals may experience an entirely different pattern of special education representation than emergent bilinguals. This distinction is a particularly important consideration in the area of language assessment, given the potential performance differences for children with different linguistic profiles (Bedore et al., 2018; Gillam et al., 2013; Peña et al., 2016). A small but growing number of studies of disability identification in bilinguals have acknowledged the importance of accounting for differences in language background. Yamasaki and Luk (2018), using parent- and school-reported data in their examination of disability identification in a school district in Massachusetts, found differences in disability identification rates for emergent and English-proficient bilinguals. In the sample studied, these two groups exhibited significant differences in their likelihood of

being identified with language-related disorders. Although English-proficient bilinguals experienced consistent underrepresentation in the categories of CD and SLD, emergent bilinguals experienced a shift from under- to overrepresentation. Artiles et al. (2005), in a study of 11 school districts in California, similarly found that emergent bilinguals were overrepresented in special education at the secondary level, but English-proficient bilinguals were underrepresented. Umansky et al. (2017), using a longitudinal sample that included students from elementary, middle, and high school, identified differences in the incidence of identification of SLD for two distinct groups: current-ELLs and ever-ELLs. The former group comprised students who were classified as ELLs at the point in time being analyzed, and the latter group were those students who had been classified as ELLs at any point in time. The label ever-ELL included both emerging bilinguals, as well as those students who were reclassified after having exhibited a sufficient level of English proficiency. The authors found different patterns of SLD identification for current- and ever-ELLs, noting that current-ELLs were overrepresented in special education only at the secondary level, but ever-ELLs were consistently underrepresented across grades. The findings of these studies underscore the importance of considering the role that language background plays in the likelihood of disability identification.

Although there is no clear consensus regarding the root cause of disproportionate disability identification in bilingual children, there are likely several underlying mechanisms at play. One of the major factors commonly argued to contribute to overrepresentation is the misidentification of language-related disorders due to biased assessment (Abedi, 2009; Chu & Flores, 2011; Sanatullova-Allison & Robison-Young, 2016). Bilingual children may have an increased risk of being misidentified due to poorly designed or inadequate tests (Barragan et al., 2018; Macswan & Rolstad, 2006; Restrepo & Silverman, 2001), an overreliance on standardized assessment (De Lamo White & Jin, 2011; Laing & Kamhi, 2003), the measurement of achievement rather than learning capacity (Camilleri & Law, 2007; Hall-Mills, 2019;

Peña et al., 2001), or a lack of culturally responsive assessment practices more broadly (Laher & Cockcroft, 2017; Skiba, 2002). For children learning two languages, common features of typical bilingual development may be attributed to symptoms of an underlying language-related disorder. Emergent bilinguals, for example, are likely perform poorly on vocabulary testing when considering ability in only one of their languages (Anaya et al., 2018; Bedore et al., 2005). The importance of accounting for bilingualism in assessment also extends to other linguistic and cognitive domains, such as phonology (Goldstein & Fabiano, 2007; Yavas & Goldstein, 1998), morphosyntax (Bedore et al., 2018; G. P. Morgan et al., 2013), and phonological working memory (J. A. Ortiz, 2021a), making the identification of language-related disorders more challenging in bilinguals than in monolinguals. The misidentification of children from language-minority backgrounds, stemming from cultural or linguistic bias, is not unique to bilingual children. It is also a concern for children who speak non-mainstream dialects of English, as exemplified in the *Ann Arbor Decision*, in which speakers of African-American English were disproportionately identified with disabilities (*Martin Luther King Jr., ETC. v. Ann Arbor Sch. Dist.*, 1979). Clearly the effects of assessment bias are an important consideration in disability evaluations, given the documented risk of misidentification for language-minority students and the potential consequences.

Poor quality of instruction is another factor that may contribute to overrepresentation. Many of the academic difficulties exhibited by bilingual students may be attributable to instruction that is insufficient to meet their educational needs. A lack of access to high quality instruction may result in a greater number of bilingual students identified as at-risk for academic difficulties, and less of an opportunity to make gains in identified areas of difficulty (Connor et al., 2007). If provided with access to higher quality instruction, bilingual students may be less likely to need the additional support of special education services, as exemplified by evidence from studies of response to intervention. Response to intervention provides a useful method of evaluating learning capacity, by measuring the degree to which

students exhibit change in performance following a series of interventions of increasing intensity, most commonly in the area of reading. When provided with this additional instructional support, at-risk bilinguals are able to catch-up to their peers (e.g., Gerber et al., 2004; McMaster et al., 2008; Vaughn, Mathes, et al., 2006). Furthermore, response to intervention may result in an overall reduction in disproportionate identification of students with language-related disorders (O'Connor et al., 2013, 2014). Outcomes from studies such as these demonstrate the effect of instructional quality on academic achievement, as well as the association with reduced levels of disproportionate representation.

Although assessment bias and instructional quality may offer a plausible explanation for overrepresentation, these factors do not sufficiently account for the existence of underrepresentation. Underrepresentation may be linked more closely with the barriers imposed by socioeconomic, cultural, or linguistic factors (Donovan & Cross, 2002). The challenges faced by language-minority families may lead to lower levels of participation and consequently reduced access to special education services. Families may feel ill-informed about special education eligibility procedures, resulting in an inability to engage effectively with educators (Wolfe & Durán, 2013). Immigrant families, in particular, may be generally less familiar with the processes involved in disability identification due to fewer interactions with healthcare providers (Calvo & Hawkins, 2015). A lack of access to healthcare services may also decrease the likelihood of a referral generated by a provider (Flores & The Committee on Pediatric Research, 2010). Cross-cultural differences in beliefs and perceptions of disability may also impact participation (K. P. Cummings & Hardin, 2017). Families may also exhibit limited engagement following previous negative experiences with the disability identification process (Hardin et al., 2009; Klingner & Harry, 2006; McHatton & Correa, 2005). Another possible explanation for underrepresentation is the "frog pond effect," in which children in poor performing schools experience a decreased likelihood of being identified with a disability (Hibel et al., 2010; P. L. Morgan et al., 2015). Because children of

immigrants are more likely to attend schools with fewer resources, they may face a corresponding reduction in the likelihood that they will receive special education services.

Another important pattern to consider is the aforementioned increasing likelihood of disability identification as bilingual students progress though school (Artiles et al., 2005; Samson & Lesaux, 2009; Umansky et al., 2017; Yamasaki & Luk, 2018). One possible explanation for this pattern of identification may be related to a hesitancy to refer emergent bilinguals for disability evaluations until they exhibit a sufficiently high level of English proficiency. Teachers may attribute observed areas of difficulty to limited English proficiency, and subsequently delay a referral for special education (Limbos & Geva, 2001). This inaction may be also rooted in a false belief that emergent bilingual students will be unable to benefit from special education services if they do not demonstrate a mastery of English (Hibel & Jasper, 2012). The delayed referral of emergent bilinguals results in a barrier to early access to services that other students, such as English-proficient bilinguals, do not face. The existence of a "reclassification bottleneck," may exacerbate the problem; children classified as ELLs who also receive special education services experience a decreased likelihood of being reclassified as English-proficient (Slama, 2014; Umansky et al., 2017). This bottleneck results in inflated numbers of ELLs in special education in higher grades.

Regardless of its manifestation as over- or underrepresentation, disproportionality represents a fundamental lack of equity in access to special education services for bilingual students. To better address the problem, researchers and educators need a clearer understanding of the nature of disproportionate disability identification for bilinguals, as well as its underlying mechanisms. Disproportionality policy and research has primarily focused on disparities related to race and ethnicity, but rates of disability identification for bilingual students have received less attention. Although many studies do include bilinguals in their analyses, few consider the role that differences in language

background may play. In addition, although some studies include parent report of language exposure, a reliable measure language ability (Paradis, 2017), many rely solely on the educational classification of ELL, a label which is not inclusive of all bilingual students and whose classification criteria may vary widely across schools (National Research Council, 2011). With respect to methodology, many studies use aggregated school-level or district-level data that do not provide information about individual-level predictors of disability identification risk. Although conclusions drawn from these data are valuable in examining the unique factors that contribute to disproportionality at a local level, they are difficult to generalize to the US as a whole. Additionally, many studies have used cross-sectional samples representing a single grade or data aggregated from multiple grades. Neither of these analyses provide a sufficient level of detail to examine how disproportionate identification may change over grades. Lastly, few published studies draw a distinction between the prevalence and incidence of disability identification. While prevalence tells us about the number of children currently represented within a disability category, incidence provides information related to the rate at which children are identified. The distinction between representation and identification is critical, as neither of these metrics is able to provide a comprehensive portrayal of disproportionality across grades in the absence of the other.

Purpose of the Present Study

The purpose of the present study is to determine the degree to which bilingual children are disproportionately identified with language-related disorders, and to identify predictors associated with disability identification. Using a large, nationally representative data set, this study will examine the incidence and prevalence of CD and SLD for bilingual students in elementary school, and associated predictors. To address the lack of focus on language background in the extant research, this study will include an analysis of the differences in outcomes for emergent- and English-proficient bilinguals. This study will address the following research questions:

1. Are emergent bilinguals and English-proficient bilinguals disproportionately represented in language-related disorder categories from kindergarten to fifth grade, compared to monolinguals?
2. How do the rates of identification of language-related disorders differ for emergent bilinguals, English-proficient bilinguals, and monolinguals, from kindergarten to fifth grade?
3. What individual- and school-level variables predict the identification of language-related disorders for emergent bilinguals, English-proficient bilinguals, and monolinguals?

Method

Data

The sample used in this study was drawn from the Early Childhood Longitudinal Survey, Kindergarten Class of 2010-11 (ECLS-K:2011), an individual-level, nationally representative data set. This data set includes observations from children, parents, and teachers from kindergarten ($n = 18{,}170$) to fifth grade ($n = 12{,}350$). The data were collected in the fall and spring from kindergarten to second grade, and in the spring only from third to fifth grade. The ECLS-K:2011 includes a diverse composition of participants with respect to race, ethnicity, and language background.

Measures

A variety of outcome measures from the ECLS-K:2011 were selected for the present study, including both time-invariant and time-varying variables. The following sections provide a description of these variables, and the rationale for each. These variables draw from those used in previous studies of disproportionality (P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017; Umansky et al., 2017). Table 1 provides an overview of descriptive statistics for each of these variables.

Disability identification. Variables for disability identification comprised two time-varying measures: receipt of special education services and disability category. For each participant, special education teachers reported whether they received special education services, as well as the disability category for those children receiving services. These variables were used to determine which children were identified with CD and SLD. In addition, a third aggregate category representing disability categories other than CD or SLD, was included as a comparator. Importantly, the disability categories used in the analysis are those that were reported by the schools themselves, and therefore some degree of variability in the accuracy of these determinations is likely present.

Linguistic characteristics. Two variables were used to determine the language background of participants: parent report of exposure to a language other than English at home and the results of an English language screener, the Preschool Language Assessment Scale English (preLAS; Duncan & De Avila, 1998). Bilingual status was determined from the time-invariant parent report of language exposure at home, which was collected in fall of kindergarten. Estimating language exposure through the use of parent report has been used in previous in studies of disproportionality (P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017; Samson & Lesaux, 2009). Parent report of language exposure has also been used extensively in studies of CD in bilingual children, as a means of estimating language dominance (Anaya et al., 2018; Bedore et al., 2018; Gillam et al., 2013), as well as screening for language impairment (Guiberson et al., 2011; Paradis et al., 2010; Pua et al., 2017). This method of identifying participants as bilingual is inclusive of both students with and without limited English proficiency. The time-varying variable for a passing score on the English language screener, administered in fall and spring of kindergarten, was used to identify students with limited English proficiency. The screener comprised two subtests from the preLAS, the Simon Says and Art Show tasks, which were used to evaluate receptive and expressive English ability, respectively. The cut score used in the ECLS-K:2011 for

this screener was 16 out of 20 possible points; children who received a score of 16 or greater were considered to have passed the screener. The combination of parent report and English screener variables was used to distinguish between emergent bilinguals and English-proficient bilinguals. Bilingual children, as identified by parent report, who did not pass the English screener in either fall or spring of kindergarten were classified as emergent bilinguals, whereas bilingual children who passed the screener were classified as English-proficient bilinguals.

Due to the lack of variation in English proficiency after kindergarten, English screener outcomes were included as a time-invariant variable. By spring of first grade, 99.9% of children passed the screener, and it was subsequently not administered in the following grades, making the inclusion of language background (i.e., emergent vs. English-proficient bilingual) of limited utility in subsequent grades. For this reason, the analysis included language background in kindergarten as a time-invariant variable. Use of the variable in this manner provides insight into the predictive significance of having entered school as an emergent or English-proficient bilingual. It is important to acknowledge the English screener results do not necessarily align with the classification of students as ELLs. There are a variety of different methods of identifying students in need of English as second language support (National Research Council, 2011) and the included schools may not have used the results from the English screener as their primary means of identification. For this reason, results of the present study reflect the distinction between emergent and English-proficient bilinguals as determined by a common measure, rather than by their classification as ELLs.

Sociodemographic characteristics. The analysis included several time-invariant sociodemographic variables, including race-ethnicity, geographic region, country, socioeconomic status, family immigration status, and marital status. Variables for race-ethnicity included White, Black, Hispanic, and other. Family socioeconomic status consisted of a composite variable measured in the

spring of kindergarten, which included family income, parent occupation, and education level, as used in previous work with the ECLS-K:2011 data (P. L. Morgan, Farkas, et al., 2017). Family immigration status was included to account for possible differences in disability identification risk for children of immigrants. Participants were identified as children of immigrants if at least one parent was born outside of the US (Hibel & Jasper, 2012). Lastly, marital status, as reported by parents in the spring of kindergarten, was also included to account for differences related to family composition.

Developmental and health characteristics. The analysis included several time-invariant variables related to the developmental and health characteristics of participants. The reported gender of each participant was included to account for the higher rates of disability identification for boys than for girls (National Center for Education Statistics, 2021). Prematurity at birth and low birth weight (< 5.5 pounds) were included as they are associated with a greater risk of language impairment (Sansavini et al., 2010). Maternal age at birth was also considered, due to the possible differences in the risk for language-related disorders for children born to younger or older mothers (Harrison & McLeod, 2010). Lastly, the presence or absence of health insurance coverage for participants was considered due to the possibility that families without insurance may have more restricted access to healthcare providers, who play an important role in identifying developmental disabilities.

Direct child assessment. The analysis included several variables derived from direct child assessment, including academic achievement, behavioral functioning, and working memory. Academic achievement is strongly associated with variation in disabilities; children with lower levels of achievement are more likely to be identified with language-related disorders (P. L. Morgan et al., 2015). For this reason, two variables, representing academic achievement measures from reading and mathematics tests administered in kindergarten, were included in the analysis. The reading test contained items related to phonological awareness, expressive/receptive vocabulary, and

comprehension. The mathematics test included items related to number properties and operations, measurement, geometry, data analysis and probability, and algebra. These direct measures were administered individually to participants, and scores were calculated using item-response theory to ensure the comparability of results within and across cohorts (Najarian et al., 2020). Theta reliabilities from kindergarten to fifth ranged from .86 to .95 for the reading test, and from .91 to .94 for the mathematics test. Both assessments exhibited low levels of differential item functioning, indicating that participant subgroups exhibited similar performance on test items provided (Tourangeau et al., 2019). Regarding language of administration, the majority of the assessments were conducted in English, though portions were administered in Spanish, for Spanish-speaking bilingual students. The administration of these tests using translations for languages other than Spanish were not reported. In the present study, the analysis included standardized scores for these tests to facilitate ease of interpretation.

Differences in child behavior characteristics are associated with the incidence of language-related disorders (Harrison & McLeod, 2010). To account for behavioral functioning, variables were derived from two rating scales administered in kindergarten: a modified version of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) and the Approaches to Learning Scales. The SSRS can be used to measure problem behaviors that may negatively impact social skill development, using the Externalizing Problem Behaviors and Internalizing Problem Behaviors scales. Internal consistency reliability ranged from .73 to .79 for the Internalizing Problem Behaviors scale and .86 to .89 for the Externalizing Problem Behaviors scale. The teacher reported Approaches to Learning Scales was used to measure self-regulatory behaviors, such as demonstrating an eagerness to learn, keeping items organized, and persistence in completing tasks. The internal consistency reliability for this scale ranged from .91 to .92 (Tourangeau et al., 2019).

Working memory is strongly associated with language ability. Children who perform poorly on working memory tasks, such as nonword and number repetition, frequently exhibit language difficulties, and performance on these tasks can be used to aid the identification of CD (J. A. Ortiz, 2021a) and SLD (Schuchardt et al., 2008). For this reason the analysis included working memory, as measured with the Numbers Reversed task form the Woodcock Johnson Psychoeducational Battery, Third Edition (Woodcock et al., 2001). In this task, participants are required to repeat a series of digits in the reverse order that they were presented. Standard scores from this task were included in the analysis. The task was administered in English and Spanish; administration in other languages was not reported.

For measures of academic achievement and working memory, the analysis only included outcomes from kindergarten to avoid biased parameter estimates due to endogeneity, or reciprocal causation (Singer & Willett, 2003). While outcomes were collected in all grades, the inclusion of these variables as time-varying (i.e., measured at each grade level) is likely to introduce endogeneity. The inclusion of these variables as time-invariant represents outcomes for children from different language backgrounds across grades, controlling for academic achievement in kindergarten. This approach has been used in previous studies of disproportionality (Hibel & Jasper, 2012; Umansky et al., 2017). Measures of behavioral functioning were included as time-varying, due to the inherent subjectivity of the method of measurement—teacher-rating. The inclusion of multiple raters across grades may help to reduce the effect of potential assessment bias stemming from measurements from a single rater for each child.

School-level variables. The analysis included two school-level variables: percent free/reduced price lunch and percent minority enrollment. The percentage of students receiving free or reduced lunch was used as a proxy for socioeconomic status. These variables were included in order to control for potential differences in rates of identification for schools serving larger proportions of students from

low socioeconomic status or minority backgrounds. In addition, geographic region (Northeast, Midwest, South, and West) was included to account for potential differences in disability identification practices in the different areas of the country.

Analytic Method

To answer the first research question, regarding the degree to which bilinguals are disproportionately represented in language-related disorders categories, the analysis examined the proportion of children identified with disabilities by language background. Using person-period data sets, separate full and reduced logistic regression models were estimated for the categories of CD, SLD, and other disabilities as the outcome measures. All models included a categorical variable representing language background with three levels: emergent bilinguals, English-proficient bilinguals, and monolinguals. In addition, all models included a term representing the interaction between grade level and language background, to account for potential change in the likelihood of disability identification for bilinguals as they progress through school (Artiles et al., 2005; Samson & Lesaux, 2009; Yamasaki & Luk, 2018). Full models included all previously described time-varying and time-invariant predictors, to provide insight to the prevalence of disabilities for students with similar sociodemographic, health/developmental, academic, behavior, and school characteristics. Marginal contrasts were derived from each model for each grade to examine the relative likelihood of being represented in each disability category by grade.

To answer the second research question, regarding the rate at which bilinguals and monolinguals are identified with disabilities, discrete time logit models were estimated as a means of examining the incidence of disability identification from kindergarten to fifth grade. Discrete time models provide information about the risk of a specific event, contingent upon that event not having previously occurred (Singer & Willett, 2003) and have been used in the investigation of

disproportionality in several previous studies (Hibel & Jasper, 2012; P. L. Morgan et al., 2015; Umansky et al., 2017). These discrete time models were estimated in the same manner previously described for logistic regression models but used a person-period data set in which individuals who were identified with a disability in a given grade were excluded from analysis in future grades. In this manner, discrete time models provide information about the incidence of disability identification for individuals who were not previously identified.

The third research question, regarding which student- and school-level characteristics predict disability identification, was addressed through the examination of the parameter estimates from the full logistic and discrete time models. These estimates provide insight into the role that each predictor plays in the likelihood of disability identification.

To ensure that parameter estimates were representative, all models included survey weights provided by the ECLS-K:2011 for each year included, which account for nonresponse to survey items. Models included those weights that accounted for items that had the highest rates of nonresponse, to avoid biased estimates. The sample comprised participants clustered within schools and for this reason all models utilized robust variance estimation to account for dependent outcomes for participants in the same schools. Because of the nested structure of the data, all models were estimated using Taylor series linearization, which provides unbiased parameter and standard error estimates while accounting for clustering in complex survey data. The number of clusters (i.e., schools) included in the analysis was 860 and results estimated using Taylor series linearization approximate those of multilevel models when there are greater than 30 clusters (Huang, 2016). The original sample comprised 18,170 participants in kindergarten but 12,350 in fifth grade, resulting in missing data due to attrition. To ensure that outcomes were not biased due to missing data, outcomes from unimputed models were compared to models that included multiple imputation from five complete data sets. There was no meaningful

difference between the estimates of imputed and unimputed models. The results below represent models estimated from unimputed models. All analyses were conducted using the software program R (R Core Team, 2021) along with the EdSurvey (Bailey et al., 2021) and survey (Lumley, 2020) packages.

Results

Representation in Language-Related Disorder Categories

Table 2 shows the results of the marginal contrasts from the full and reduced logistic regression models, representing the difference in the odds of receiving services for CD, SLD, or any other disability, for emergent bilinguals and English-proficient bilinguals compared to monolinguals. Figure 1 shows the predicted probabilities of receiving special education services in each disability category for emergent bilinguals, English-proficient bilinguals, and monolinguals.

In the CD category, results from the reduced model show a gradual decline in prevalence rates for monolinguals from kindergarten to fifth grade. Compared to monolinguals, emergent bilinguals exhibited significantly lower rates of CD in kindergarten, with 54% ($OR = 0.46$) lower odds of being represented. This was followed by a sharp increase in third grade, resulting in significant overrepresentation by fifth grade, with 119% ($OR = 2.19$) greater odds of being represented in the CD category than monolinguals. English-proficient bilinguals, on the other hand, exhibited significantly lower rates of CD representation than monolinguals from kindergarten through third grade, with 32% ($OR = 0.68$) to 56% ($OR = 0.44$) lower odds of being represented than monolinguals during this period. The addition of the covariates included in the full model resulted in significant levels of underrepresentation for emergent bilinguals in the CD category in most grades, compared to monolinguals with similar individual- and school-level characteristics (i.e., sociodemographic, developmental, health, academic, behavioral, cognitive characteristics and composition of school population). The odds of being represented in the CD category from kindergarten to fourth grade for

emergent bilinguals ranged from 73% ($OR = 0.27$) to 96% ($OR = 0.04$) lower than for otherwise similar monolinguals. In the full model, the sharp increase in CD prevalence for emergent bilinguals in third grade was no longer present, suggesting that factors other than language background may have played a role in mediating this effect. For English-proficient bilinguals, results from the full model showed significant underrepresentation compared to otherwise similar monolinguals in first, third, and fifth grade, with 78% ($OR = 0.22$), 53% ($OR = 0.47$), and 67% ($OR = 0.33$) lower odds of being represented, respectively.

In the SLD category, results from the reduced model show a gradual increase in prevalence from kindergarten to fifth grade, for both English-proficient bilinguals and monolinguals. English-proficient bilinguals generally experienced similar rates of SLD prevalence to monolinguals, but were significantly underrepresented in first and second grade, with 61% ($OR = 0.39$) and 34% ($OR = 0.66$) lower odds of being represented, respectively. Although emergent bilinguals exhibited similar SLD prevalence rates to monolinguals in early elementary grades, they showed a marked increase starting in third grade, resulting in significant overrepresentation by fifth grade. Compared to monolinguals, the odds of being represented in the SLD category for emergent bilinguals in fifth grade were 96% ($OR = 1.96$) greater than for monolinguals. When controlling for the covariates included in full model, emergent bilinguals experienced significant underrepresentation, compared to otherwise similar monolinguals, in first and second grade with 100% ($OR = 0$) and 74% ($OR = 0.26$) lower odds of being represented, respectively. Like with CD, the growth in SLD prevalence for emergent bilinguals diminished with the inclusion of the additional covariates in the full model. In the full model, English-proficient bilinguals exhibited no significant difference in rates of SLD representation across grades, compared to otherwise similar monolinguals.

The prevalence of other, non-language-related, disabilities followed a pattern distinct from that of CD or SLD, as shown in the reduced model. For both monolinguals and bilinguals, the overall levels of prevalence of non-language-related disabilities were much lower than CD or SLD, and generally did not exhibit the same degree of variability across grades. For emergent bilinguals, there was only evidence of underrepresentation in fourth grade, with 72% ($OR = 0.28$) lower odds of being represented than monolinguals. Unlike prevalence rates for CD and SLD, emergent bilinguals did not experience a shift from under- to overrepresentation in non-language-related disability categories as they progressed through school. English-proficient bilinguals, on the other hand, exhibited significant levels of underrepresentation across all grades, ranging from 45% ($OR = 0.55$) to 63% ($OR = 0.37$) lower odds of being represented in non-language-related disability categories than monolinguals. When controlling for the covariates included in the full model, emergent bilinguals exhibited significant levels of underrepresentation in non-language-related disability categories in first, third, fourth, and fifth grade, compared to otherwise similar monolinguals, with 94% ($OR = 0.06$) to 100% ($OR = 0$) lower odds of being represented across these grades. In contrast, there was no evidence of under- or overrepresentation for English-proficient bilinguals in non-language-related disability categories, compared to otherwise similar monolinguals.

Identification Rates of Language-Related Disorders

Table 2 shows the results of the marginal contrasts for the full and reduced discrete time models, representing the difference in the odds of being identified with CD, SLD, or any other disability, for emergent bilinguals and English-proficient bilinguals compared to monolinguals. Figure 2 shows the predicted probabilities of being identified within each of the disability categories for emergent bilinguals, English-proficient bilinguals, and monolinguals.

The incidence of CD in the reduced model shows higher rates of identification for monolinguals in kindergarten, followed by a gradual decrease over time. Compared to monolinguals, emergent bilinguals experienced significantly lower rates of identification in kindergarten, with 54% ($OR = 0.46$) lower odds of being identified. This was followed by a dramatic increase in identification rates after second grade, resulting in significant overidentification in subsequent grades, such that the odds of being identified with CD for emergent bilinguals compared to monolinguals were 239% ($OR = 3.39$) greater in third grade and 557% ($OR = 6.57$) greater in fifth grade. English-proficient bilinguals, on the other hand, exhibited significant underidentification of CD in early elementary grades, compared to monolinguals, with 46% ($OR = 0.54$) and 61% ($OR = 0.39$) lower odds being identified in kindergarten and first grade, respectively. When controlling for the covariates included in the full model, rates of identification for emergent bilinguals were significantly lower than for otherwise similar monolinguals in kindergarten and fourth grade, with 86% ($OR = 0.14$) and 100% ($OR = 0$) lower odds of being identified, respectively. Like with the CD prevalence model, the inclusion of the additional covariates in the full model reduced the sharp increase in identification rates for emergent bilinguals in third grade. For English-proficient bilinguals, results from the full model showed significantly lower CD incidence rates than otherwise similar monolinguals in kindergarten, first, and fourth grade, with 51% ($OR = 0.49$), 95% ($OR = 0.05$), and 66% ($OR = 0.34$) lower odds of being identified, respectively.

The incidence of SLD shown in the reduced model revealed relatively low rates of initial identification, with less than 1% of children in kindergarten identified and no significant difference between groups. Like with CD, emergent bilinguals experienced a sharp increase in identification rates after second grade, resulting in significant levels of overidentification in subsequent grades. Compared to monolinguals, emergent bilinguals experienced 163% ($OR = 2.63$) greater odds of being identified with SLD in third grade and 302% ($OR = 4.02$) greater odds in fifth grade. Although English-proficient

bilinguals followed a similar trend to monolinguals, they experienced a gradual shift from under- to overidentification of SLD, such that the odds of being identified, compared to monolinguals, were 57% $(OR = 0.43)$ lower in first grade, but 118% $(OR = 2.18)$ greater in fifth grade. When controlling for the covariates included in the full model, the odds of being identified with SLD for emergent bilinguals, compared to otherwise similar monolinguals, were 100% $(OR = 0)$ lower in first grade. As observed in the previously described models, the controls included in the full model accounted for the spike in identification rates for emergent bilinguals in third grade. For English-proficient bilinguals, the inclusion of the covariates in the full model resulted in significant underidentification of SLD in fourth grade, with 77% $(OR = 0.23)$ lower odds of being identified than otherwise similar monolinguals, and overidentification in fifth grade, with 201% $(OR = 3.01)$ greater odds of being identified.

With respect to other, non-language-related, disability categories, monolinguals experienced the highest initial rates of identification in kindergarten, but not significantly different than those for emergent bilinguals across grades. English-proficient bilinguals, however, experienced 45% $(OR = 0.55)$, 51% $(OR = 0.49)$, 64% $(OR = 0.36)$ lower odds of being identified than monolinguals in kindergarten, first, and fifth grades, respectively. When controlling for the additional covariates in the full model, rates of identification of non-language-related disabilities were substantially lower for all groups. Emergent bilinguals experienced significant underidentification of non-language-related disabilities in multiple grades, with 100% $(OR = 0)$ lower odds of being identified in first, third, fourth, and fifth grade, than otherwise similar monolinguals. For English-proficient bilinguals, on the other hand, the inclusion of the additional covariates resulted in non-language-related disability identification rates that were not significantly different than otherwise similar monolinguals.

Predictors of Language-Related Disorders

With respect to the specific variables associated with disability identification, results from all models revealed several predictors that accounted for significant variation in identification rates. Results were similar across the prevalence (logistic regression) and incidence (discrete time) models, such that most predictors associated with prevalence were also associated with incidence, with a few exceptions, as shown in Table 3. All predictors found to be significant in the incidence models were also significant in the prevalence models. For this reason, the section below describes those predictors identified as significant in the prevalence models.

Regarding sociodemographic variables, there was a significant negative association between race-ethnicity and CD classification, such that rates of prevalence were lower for Black students ($OR = 0.7$) and Hispanic students ($OR = 0.6$) compared to White students. Hispanic students also exhibited significantly lower rates of non-language-related disabilities than White students ($OR = 0.45$). Socioeconomic status was also negatively associated with SLD classification ($OR = 0.79$); students with higher levels of socioeconomic status exhibited lower rates of identification. There was no significant effect for family immigration status.

In the area of developmental and health characteristics, there was a significant positive association between gender and the likelihood of being represented in any disability category. Estimates from all models showed that males experienced significantly higher rates of CD ($OR = 1.77$), SLD ($OR = 1.57$), and non-language-related disabilities ($OR = 1.81$) compared to females. The age of a child's mother at birth was also significantly associated with the prevalence of CD as well non-language-related disabilities. Compared to children born to mothers between 18 and 38 years old, rates of CD were significantly higher for children born to older mothers, over 38 years old ($OR = 1.51$). Prevalence rates of non-language-related disabilities were significantly lower for younger mothers, under 18 years old ($OR = 0.46$). No other developmental or health variables exhibited a significant effect.

In the domain of academic achievement, reading ability was found to have a significant negative association with the prevalence of CD ($OR = 0.63$) and SLD ($OR = 0.41$), such that higher scores were associated with lower rates of representation. Math ability was similarly negatively associated with CD ($OR = 0.34$), SLD ($OR = 0.42$), and non-language-related disabilities ($OR = 0.22$). With respect to behavioral functioning, self-regulation was negatively associated with both CD ($OR = 0.75$) and non-language-related disabilities ($OR = 0.49$), such that higher teacher-rated self-regulation ability was associated with lower rates of representation. Externalizing problem behaviors were similarly negatively associated only with the prevalence of CD ($OR = 0.79$). Internalizing problem behaviors were positively associated with the prevalence CD ($OR = 1.28$), SLD ($OR = 1.48$), and non-language-related disabilities ($OR = 1.36$), indicating that higher levels of teacher-rated internalizing problem behaviors were associated with higher rates of representation. Working memory was negatively associated with the prevalence CD ($OR = 0.97$), SLD ($OR = 0.97$), and non-language-related disabilities ($OR = 0.97$), suggesting that stronger working memory ability was associated with lower rates of representation. Lastly, age at school entry was positively associated with the prevalence of CD ($OR = 1.07$), SLD ($OR = 1.1$), and non-language-related disabilities ($OR = 1.1$), such that children who entered school at an older age were more likely to receive special education services than those who entered at a younger age.

Of the school level predictors included in the model, geographic region exhibited a significant association with the prevalence of language-related disorders. Rates of SLD were lower for schools in the Midwest ($OR = 0.61$) and South ($OR = 0.33$) compared to schools in the Northeast. Schools in the West, on the other hand, exhibited lower rates of CD ($OR = 0.54$) and non-language-related disabilities ($OR = 0.45$) than schools in the Northeast. The variable representing the percentage of students receiving free/reduced price lunch was nonsignificant for all disability categories. The proportion of

minority students was not associated with CD or SLD representation but was positively associated with the prevalence of non-language-related disabilities ($OR = 1.58$). Students in schools with minority populations of between 25% and 50% were more likely to be represented in a non-language-related disability category than those with minority populations of between 0% and 25%

Discussion

The results of this study contribute to the growing evidence that bilingual children are disproportionately represented in special education in elementary school. The examination of disability identification for different groups of bilinguals remains a relatively underexplored area of investigation, and this study exemplifies the differences in identification and representation trends across language backgrounds. Although previous research has shown differences in risk associated with language background (Artiles et al., 2010; Umansky et al., 2017; Yamasaki & Luk, 2018), few studies have simultaneously considered both the prevalence and incidence of language-related disorders across different bilingual groups. In addition, few studies have examined the role that language background at school entry plays in the likelihood of disability identification in later grades. Outcomes from this study extend beyond those in the extant disproportionality literature by providing insight into the unique patterns of disability prevalence and incidence experienced by monolinguals, emergent bilinguals, and English-proficient bilinguals. The first research question asked about the degree to which bilingual children are disproportionality represented in language-related disorders categories. The second research question asked about the rate of identification of language-related disorders. Lastly, the third research question asked about those student- and school-level characteristics that predict disability identification. What follows is a discussion of the results from this study and how they relate to each of these questions.

Representation in Language-Related Disorder Categories and Rates of Identification

The question of whether bilingual children are over- or underrepresented in language-related disorder categories presumes that disproportionality follows a single direction, common to all bilinguals. Despite much of narrative surrounding the issue, disproportionality is a problem whose complexity cannot be reduced to a one-dimensional characterization, as over- or underrepresentation. Results from this study exemplify the dynamic nature of disproportionality and how it affects different groups of bilinguals in distinct ways. There is substantial variability across disability categories, language background, and grade level. Because of this complexity, the investigation of disproportionality requires a careful examination of how prevalence and incidence change over time, and the role that within group diversity plays. Results from this study provide further evidence that bilinguals are a heterogeneous group with a variety of characteristics that may influence the likelihood of being identified with a disability, and whose identification patterns vary over time.

In the domain of language background, there were some important distinctions between emergent and English-proficient bilinguals, with respect to their likelihood of being identified with a language-related disorder. One of the most obvious differences between these two groups was the growth in prevalence rates for emergent bilinguals, when not controlling for individual- and school-level covariates. Emergent bilinguals exhibited an increase in their likelihood of language-related disorder identification, such that they were more likely to be represented in both the CD and SLD categories after second grade, consistent with previous studies (Artiles et al., 2005; Samson & Lesaux, 2009; Umansky et al., 2017; Yamasaki & Luk, 2018). The prevalence rates of CD and SLD for English-proficient bilinguals, on the other hand, were generally lower than for monolinguals in early elementary school grades, with similar rates in later grades.

The present study builds on previous research by examining the relationship between prevalence and incidence for each bilingual group. The results highlight the importance of examining

both incidence and prevalence to gain insight into how rates of identification drive representation in subsequent grades. Incidence models revealed a sharp increase in identification rates in third grade for both CD and SLD, resulting in an increased prevalence of these disorders in subsequent grades. In other words, the uptick in identification rates in third grade resulted in overrepresentation in later grades. The substantial increase in the incidence of language-related disorders for emergent bilinguals in third grade did decrease in subsequent grades, but identification rates remained higher than for monolinguals by fifth grade. This same disproportionate growth in disability prevalence was absent in non-language-related disability categories, suggesting a unique association between language background and language-related disorder identification, not shared with other disability categories.

Outcomes from this study also showed the effect of English proficiency at school entry on identification rates in subsequent grades. In the present study, language background was defined using two pieces of information: parent report of language exposure and English screener outcomes in kindergarten. Results from the English screener were used to distinguish between emergent and English-proficient bilinguals. The impact of entering school with limited English proficiency had far-reaching effects on the likelihood being identified with a language-related disorder, with greater than expected numbers of emergent bilingual children at-risk of academic difficulties and subsequent receipt of special education services by third grade. The outcomes of this study underscore the importance of considering the long-term effects of limited English proficiency at school entry, and align the with findings of previous research in areas such as reading growth trajectories (Kieffer, 2008). The growth in the prevalence of language-related disorders in emergent bilingual children is consistent with the notion that many teachers prefer to wait until children exhibit a certain level of English of proficiency before referring to special education (Hibel & Jasper, 2012). In the sample used for the present study, the caveat is that the overwhelming majority of bilingual children in the sample (99.9%) received a passing

score on the screener by spring of second grade. This suggests that, despite improved English-proficiency by second grade, children who were emergent bilinguals at school entry continued to be at risk of delayed referral in comparison to their monolingual peers.

The inclusion of the additional covariates included in the full models resulted in outcomes that diverged substantially from those of the reduced models, with generally lower rates of language-related disorders, though not in all cases. Results from these models showed underrepresentation of both emergent and English-proficient bilinguals in the CD category in multiple grades, with no instances of overrepresentation. In the SLD category, the representation of emergent bilinguals in the full models was lower than for otherwise similar monolinguals in all grades, but only significantly lower in first and second grade. In contrast, outcomes from these models showed that English-proficient bilinguals were not disproportionately represented compared to otherwise similar monolinguals. One of the most salient differences between the reduced and full models was the diminished growth of disability identification rates for emergent bilinguals. When controlling for the additional covariates in the full models, the sharp increase in incidence that emergent bilinguals experienced in third grade was no longer present. This suggests that the increase in prevalence and incidence of disabilities for emergent bilinguals may be influenced by variables beyond language background. In other words, although emergent bilinguals exhibited growth in their likelihood of being identified with a language-related disorder, language background alone may not fully account for this growth.

It is important to distinguish between the conclusions drawn from the models with and without additional covariates. Although the full, covariate-adjusted, models are informative, relying solely on the analyses derived from such models is likely to lead to flawed conclusions about the nature of disability identification rates. Because these covariate-adjusted models represent the likelihood of disability identification for children who are equated on a variety of different dimensions, they provide insight

into the possible outcomes for children who share the individual- and school-level characteristics in question. Despite their utility, such models only provide outcomes for hypothetical children who may not actually represent the population being studied—a critique of previous disproportionality studies that have similarly made use of extensive statistical controls (Collins et al., 2016; Skiba et al., 2016). Controlling for factors such as reading ability and socioeconomic status, for example, may not result in a model that is representative of the profiles of many emergent bilingual children, given the link between poverty and academic achievement (Cruz & Firestone, 2022; Skiba et al., 2008).

Particular caution should be taken when interpreting outcomes from covariate-adjusted models with radically different effects than those in the corresponding reduced models, without such covariates. In the present study, the inclusion of additional covariates resulted in the disappearance of a shift from under- to over-representation in CD and SLD prevalence for emergent bilinguals after second grade. It would, however, be inaccurate to conclude that emergent bilinguals were not overrepresented in language-related disorder categories after second grade, given the evidence of growth in CD and SLD prevalence previously discussed. Rather, results from the covariate-adjusted models provide evidence that student- and school-level characteristics other than language background may play a role in mediating this effect. This, however, does not make the existence of growth in disability identification for emergent bilinguals any less real or meaningful. It is critical to acknowledge that results from models with extensive statistical controls represent an abstraction and may not reflect the experiences or profiles of children in real-world settings (Collins et al., 2016), and interpreting these results in a prescriptive manner may have serious implications with respect to education policy and practice. The results of the covariate-adjusted models are valuable for the insight they provide into outcomes for children from similar backgrounds, but they should be interpreted with an understanding of their limitations, and how they relate to the real-world context of the population under investigation.

Predictors of Disability Identification

Regarding the specific student- and school-level variables associated with variation in disability identification, a number of relevant predictors were identified. Consistent with previous research, academic achievement, behavioral functioning, and age-at school entry were all significantly associated with disability identification (P. L. Morgan et al., 2015). Reading ability was unique, in that it was associated with the prevalence and incidence of CD and SLD but not with other disabilities. Although difficulties with literacy may be more commonly considered in the diagnosis of SLD, there is substantial evidence demonstrating the association of literacy difficulties with both speech sound disorders (e.g., Peterson et al., 2009) and language impairment (e.g., Boudreau & Hedberg, 1999). Results presented here provide further support for the need to consider literacy ability in the identification of CD. Stronger working memory ability, as measured on a reversed number repetition task, was also associated with lower prevalence of language-related disorders. The utility of several types of working memory tasks has received much attention, due to their potential as less-biased methods of assessment for bilingual children (Armon-Lotem & Meir, 2016; J. A. Ortiz, 2021a). Although only one specific type of working memory task was included in the present study, results are consistent with previous research into the association between working memory and language ability.

Two of the developmental and health variables exhibited significant associations with the identification of language-related disorders: mother's age at birth and gender. The age of the child's mother at birth was significantly associated with CD. Children born to younger mothers showed a lower likelihood of being identified with CD, while those born to older mothers experience an increased likelihood, consistent with the findings of previous studies (P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017). Regarding gender, the prevalence of all disabilities was higher for boys than for girls. This is not unexpected, given that the majority of children identified with disabilities are male

(National Center for Education Statistics, 2021), but this disparity does raise questions about the equitable receipt of special education services for boys and girls (McGregor, 2020). In the present study, boys were 57% to 77% more likely to receive services for a language-related disorder than girls. Considering that some studies have demonstrated evidence of comparable rates of language disorders in in boy and girls (Norbury et al., 2016; Tomblin et al., 1997), the gender discrepancy in receipt of services in schools warrants further investigation.

Race-ethnicity and immigration status have previously been found to be associated with receipt of special education services, with varying outcomes regarding the direction of the effect and the disability being examined (Cruz & Rodl, 2018; Hibel & Jasper, 2012), but were not consistently associated with language-related disorders across models in the present study. Black students and Hispanic students experienced significantly lower rates of representation than White students in the CD category, but not SLD. There was no significant effect for other racial-ethnic groups in either CD or SLD. There was also no significant effect for family immigration status in any model. The lack of effects across disability categories for race-ethnicity and family immigration status in the present study may partly be attributable to their inclusion as control variables, rather than as the main predictors of interest. The primary focus of this study was on language background and the analysis did not include models focusing specifically on race-ethnicity or family immigration status, in the absence of additional control variables. A more nuanced examination of these characteristics, with outcomes from comparable reduced and full models as well as additional racial-ethnic categories, may reveal effects not observed here. The lack of effects across disability categories may also be related to differences in the rates of identification across grades. Like language background, the rates of disability identification fluctuate over time for children from different racial-ethnic backgrounds (Cruz & Firestone, 2022) and immigrant families (Hibel & Jasper, 2012). Thus, it is plausible that significant levels of over- or underidentification

were present for students from different racial-ethnic backgrounds during certain grades not observed in the analysis.

With respect to school-level variables, neither the proportion of students from minority backgrounds nor the proportion of students receiving free/reduced price school lunch was associated with the identification of language-related disorders. Future research examining a greater range of school-level variables that may play a role in disability identification, such as demographic the composition of school personnel, teacher training, and school resources, is warranted. Another school-level variable, geographic region, was significantly associated with the prevalence of CD and SLD. Although no single region exhibited higher rates of identification across language-related disorder categories, results demonstrated the degree to which disability identification rates can vary by region. The goal of this study was to provide a depiction of trends for the nation as a whole, but it is clear that regional differences play a major role in mediating disability identification rates. Students in the Southern US, for example, have a significantly lower likelihood of being represented in the SLD category than those in the Northeast. These differences in regional-, state-, and district-level identification rates are also exemplified in the varied outcomes shown in previous studies of disproportionality (Robinson & Norton, 2019; Sullivan & Bal, 2013; Yamasaki & Luk, 2018). While there are clear national-level trends, as demonstrated here, in many ways disproportionality is a local issue.

Limitations and Future Research

There were several limitations in the present study, several of which were related to characteristics of the data set itself. The ECLS-K:2011 is a nationally representative data set, but outcomes may not be representative of all schools. Given the regional differences identified, continued investigation into disproportionality in local contexts, and the policy solutions that best meet the needs of each community, is needed. Regarding the time range included in the analysis, the ECLS-K:2011 only

includes outcomes from kindergarten to fifth grade, which limited the ability to draw conclusions beyond elementary school. In addition, within the SLD category there was no distinction between children with different types of learning difficulties and included children with both literacy and numeracy difficulties (i.e., dyscalculia). Given the similar prevalence rates of dyslexia and dyscalculia, as well as the likelihood of their co-occurrence (Butterworth et al., 2011), outcomes in the present study should be interpreted with the consideration that children identified with SLD may have exhibited difficulties with literacy, numeracy, or both.

The absence of late arrivals in the ECLS-K:2011 presented another limitation in this study, as children who entered school after kindergarten were not included in the data set. The small emergent bilingual sample size (2.38%) may be partly attributable to the absence of emergent bilinguals who arrived after kindergarten, such as recent immigrants. Immigrant children who start school in the US after kindergarten likely exhibit a unique risk profile, given the association between age of arrival and academic ability (Heath & Kilpi-Jakonen, 2012) and may also start learning English at a later age, resulting in more difficulty with acquisition (Basu, 2018). Emergent bilinguals who begin acquiring English at a later age likely face a longer period of acquisition. This extended time as an emergent bilingual may result in more difficulty distinguishing between language differences and language-related disorders. Later age of exposure to English may also result in a higher likelihood of delayed disability identification, and subsequently greater numbers of emergent bilinguals in special education in later grades. The absence of late arrivals also likely resulted in the limited variability in English proficiency after kindergarten. Language background, as measured in kindergarten, was included as a time-invariant variable, and a greater degree of variability in English proficiency across grades would allow for the inclusion of language background as time-varying.

Another limitation was related to method of distinguishing between emergent and English-proficient bilinguals. In the present study, emergent bilinguals were identified as students who did not receive a passing score on an English screener. Because ELL classification decisions were not uniformly consistent with screener results across schools, not all individuals classified as emergent bilinguals in the present study received English as second language instruction, while some English-proficient bilinguals did. By using a common measure to distinguish between emergent and English-proficient bilinguals—the English screener—results of the present study provide insight into the relationship between language proficiency and disability identification likelihood. These results, however, do not provide information about the relationship between being identified as an ELL and disability identification likelihood. Future studies may wish to further explore the effect of ELL classification in early elementary on disability identification in later grades.

Within the analysis, there were several limitations. The focus of this study was on disability identification, and not on exit from special education. It is possible that rates of exiting special education differ for monolinguals compared to bilinguals, and this should be a consideration in future studies. In addition, results provide insight into differential rates of disability representation and identification, but not information about misidentification, a limitation not unique to the present study. The data used in this study, and in previous studies of disproportionality, can only provide information about differences in identification rates for children from different groups. Although misidentification is a plausible driver of disproportionality, it would be inappropriate to conclude that differences in rates of representation are entirely attributable to misidentification. The data used in this study do not provide information about the veracity of disability classifications, and whether there are greater rates of misidentification in one group or another. Examination of misidentification would require knowledge relating to who was correctly diagnosed with a specific disability, and who was incorrectly diagnosed. Although it is telling

that bilingual children are identified with disabilities at different rates than their monolingual peers, this finding does not provide information about rates of correct or incorrect identification in either group. Indeed, high levels of misidentification may have been present for monolinguals, but the data do not contain observations that would provide insight into this issue. Outcomes from the present study should not be taken as direct evidence of misidentification, but rather evidence of inequitable access to special education services. Future research considering the accuracy of disability evaluations for different groups would provide more insight into the nature of disproportionality.

Conclusion

The disproportionate identification of language-related disorders in bilingual children is a complex issue, given the evidence of both over- and underrepresentation in different grades for different disability categories. As a heterogeneous group, bilinguals experienced differences in their likelihood of being identified with a language-related disorder related to their language background, during the time period studied. Emergent bilingual children experienced a growth in their likelihood of being identified with CD and SLD, such that they were overrepresented by third grade, when they experienced a sharp increase in the identification rates. English-proficient bilingual children, on the other hand, were generally underrepresented in the CD and SLD categories in early elementary school but exhibited rates of identification similar to monolinguals in later grades. A number of predictors were associated with identification of language-related disorders, including gender, academic achievement, behavioral functioning, working memory, and age at school entry. Regional differences in identification reflect the variability in disproportionality across the country. Results of this study highlight the importance of considering the role that language background plays in disability identification for bilingual children.

Tables and Figures

Table 1

Descriptive Statistics for ECLS-K:2011 Data Set

	Emergent bilingual ($n = 390$)	English-proficient bilingual ($n = 2{,}650$)	Monolingual ($n = 12{,}670$)	Total sample ($n = 15{,}700$)
	Proportion (%)			
Weighted proportion of total sample	2.38	14.98	82.63	100
Race-ethnicity				
White	3.8	6.94	61.09	51.61
Black	1.16	3.98	15.65	13.55
Hispanic	89.17	70.25	14.39	24.55
Other	5.87	18.18	8.44	9.84
Child of immigrant	78.93	79.83	10.02	22.12
Developmental/health characteristics				
Male	52.78	50.3	51.79	51.59
Low birth weight	9.87	6.01	7.71	7.5
Premature birth	14.41	10.67	17.66	16.53
Age of mother at birth ≤ 18	5	2.85	2.93	2.97
Age of mother at birth ≥ 38	3.19	3.75	4.94	4.72
No health insurance	11.37	8.51	3.24	4.23
Parents unmarried	42.6	34.11	34.12	34.32
School region				
Northeast	12.45	16.97	15.9	15.98
Midwest	6.76	12.15	23.2	21.15
South	43.83	30.2	38.1	37.05
West	35.43	39.41	20.7	23.85
	Mean (SD)			
Socioeconomic status	-0.96 (0.53)	-0.53 (0.8)	0.03 (0.77)	-0.08 (0.81)
Academic achievement [b]				
Reading	-1.33 (1.09)	-0.42 (0.68)	-0.22 (0.61)	-0.27 (0.66)
Mathematics	-0.91 (0.69)	-0.56 (0.61)	-0.33 (0.61)	-0.38 (0.62)
Behavioral functioning [c]				
Self-regulation	3.17 (0.64)	3.21 (0.61)	3.17 (0.64)	3.17 (0.64)
Externalizing problem behaviors	1.57 (0.63)	1.55 (0.57)	1.66 (0.65)	1.64 (0.64)
Internalizing problem behaviors	1.57 (0.54)	1.47 (0.46)	1.52 (0.5)	1.51 (0.5)
Working memory spring kindergarten	82.99 (16.3)	91.02 (17.46)	96.09 (16.75)	95.02 (17.04)
Age at school entry (months)	66.26 (4.28)	66.79 (4.35)	67.71 (4.54)	67.55 (4.53)
School characteristics				
% free/reduced lunch [d]	74.25 (24.01)	59.7 (30.48)	39.78 (30.25)	43.72 (31.34)
% minority students [d]	83.58 (20.71)	72.67 (27.54)	40.54 (31.77)	46.63 (33.57)

Note. Proportions, means, and standard deviations from kindergarten and are weighted for nonresponse from parents and teachers. Samples sizes by group reflect the number of participants for whom observations about language background were available. Sample sizes rounded to the nearest ten per security rules for data set.

[a] Composite variable, derived from fall or spring ELL status and home language use. [b] Values represent z-scores from Spring kindergarten. [c] Values represent raw scores from teacher-rating scale from Spring kindergarten. [d] Values represent standard scores. [e] Values represent means of proportions at the school-level.

Note. Proportions, means, and standard deviations from kindergarten and are weighted for nonresponse from parents and teachers.

[a] Composite variable, derived from fall or spring ELL status and home language use. [b] Values represent z-scores from Spring kindergarten. [c] Values represent raw scores from teacher-rating scale from Spring kindergarten. [d] Value represent standard scores. [e] Values represent means of proportions at the school-level.

Table 2

Marginal Contrasts from Logistic Regression and Discrete Time Models for Each Disability Category by Grade (Odds Ratios)

	Prevalence (logistic regression models)											
	CD				SLD				Other			
	Reduced model		Full model		Reduced model		Full model		Reduced model		Full model	
	EB	EPB	EB	EPB	EB	EPB	EB	EPB	EB	EPB	EB	EPB
Kindergarten	0.46*	0.54**	0.13**	0.49	0.46	0.58	0.5	0.8	0.43	0.55*	0.29	0.54
Grade 1	0.52	0.44***	0.04**	0.22**	0.89	0.39*	0***	0.28	0.67	0.54**	0***	0.37
Grade 2	0.79	0.62**	0.27*	0.54	0.68	0.66*	0.26*	1.01	0.73	0.37**	0.31	0.37
Grade 3	1.31	0.68*	0.23*	0.47*	1.64	0.98	0.34	1.15	0.73	0.42***	0.06*	0.46
Grade 4	1.31	0.74	0.26*	0.53	1.38	0.87	0.34	0.67	0.28*	0.44***	0***	0.59
Grade 5	2.19**	0.75	0.36	0.33*	1.96**	1.21	0.58	1.63	0.85	0.4***	0***	0.43
	Incidence (discrete time models)											
	CD				SLD				Other			
	Reduced model		Full model		Reduced model		Full model		Reduced model		Full model	
	EB	EPB	EB	EPB	EB	EPB	EB	EPB	EB	EPB	EB	EPB
Kindergarten	0.46*	0.54**	0.14**	0.49*	0.46	0.58	0.48	0.67	0.43	0.55*	0.34	0.62
Grade 1	1.15	0.39**	0.18	0.05**	1.11	0.43*	0***	0.11	1.27	0.49*	0***	0.33
Grade 2	1.29	1.09	0.67	0.92	0.6	0.7	0.28	0.91	1.4	0.53	0.75	1.33
Grade 3	3.39**	0.76	0.35	0.25	2.63**	1.13	0.47	0.91	1.5	0.46	0***	0.72
Grade 4	1.04	1.31	0***	0.34*	1.74	0.73	0.4	0.23*	0.6	0.77	0***	1.71
Grade 5	6.57***	1.55	2.32	0.74	4.02***	2.18***	1.97	3.01*	2.47	0.36*	0***	0.53

Note. Monolinguals used as the reference group. CD = communication disorders; SLD = specific learning disability; Other = other disability categories; EB = emergent bilingual; EPB = English-proficient bilingual.

* p < .05, ** p < .01, *** p < .001

Table 3

Predictors of Disability Prevalence and Incidence Across Grades (Odds Ratios)

	Prevalence (logistic regression models)			Incidence (discrete time models)		
	CD	SLD	Other	CD	SLD	Other
Race (ref: White)						
Black	0.7*	0.58	0.81	0.75	0.71	0.88
Hispanic	0.6*	0.67	0.45**	0.71	0.77	0.6*
Other	1.13	1.01	1.21	1.12	1.19	1.17
Child of immigrant	0.74	2.07	0.74	0.79	1.64	1.03
Socioeconomic status	1.1	0.79*	1.13	1.07	0.82	1.17
Developmental/health characteristics						
Male	1.77***	1.57**	1.81***	1.65***	1.53**	1.83***
Low birth weight	1.06	1.14	0.89	1.03	0.89	0.92
Premature birth	1.17	1.33	1.24	1.18	1.4	1.22
Age of mother at birth ≤ 18 (ref: age 19-37)	0.75	0.58	0.46*	0.83	0.63	0.5
Age of mother at birth ≥ 38 (ref: age 19-37)	1.51**	1.15	1.46	1.35*	0.99	1.52
No health insurance	0.55	0.82	0.35	0.71	1.09	0.61
Parents unmarried	0.92	1.01	1.01	0.88	0.98	1.02
Academic achievement						
Reading	0.63***	0.41***	0.79	0.62***	0.52***	0.78
Math	0.34***	0.42***	0.22***	0.42***	0.36***	0.26***
Behavioral functioning						
Self-regulation	0.75***	0.93	0.49***	0.7***	0.95	0.7**
Externalizing problem behaviors	0.79**	1	1.06	0.83	1.07	1.27
Internalizing problem behaviors	1.28***	1.48***	1.36**	1.15	1.53***	1.38**
Working memory	0.97***	0.97***	0.97***	0.97***	0.97***	0.97***
Age at school entry	1.07***	1.1***	1.1***	1.06***	1.09***	1.09***
School-level variables						
Free/reduced lunch 25-50% (ref: 0-25%)	1.01	0.76	1.3	0.94	0.81	1.07
Free/reduced lunch 50-75% (ref: 0-25%)	1.11	1.05	1.1	1.32	1.11	1.12
Free/reduced lunch 75-100% (ref: 0-25%)	0.98	0.87	0.87	0.92	0.75	0.66
Minority students 25-50% (ref: 0-25%)	1.33	1.04	1.58*	1.31	1.2	1.38*
Minority students 50-75% (ref: 0-25%)	0.98	1.05	1.28	0.84	0.98	1.29
Minority students 75-100% (ref: 0-25%)	0.84	0.6	1.21	0.76	0.72	1.35
Region Midwest (ref: Northeast)	0.85	0.61*	0.71	0.94	0.61*	0.75
Region South (ref: Northeast)	0.75	0.33***	0.61	0.72	0.4***	0.59
Region West (ref: Northeast)	0.54*	0.69	0.45**	0.58*	0.69	0.43**

Note. Models include interactions between language background and grade (not shown). CD = communication disorders; SLD = specific learning disability; Other = other disability categories; ref = Reference groups for categorical variables with more than two levels.

* $p < .05$, ** $p < .01$, *** $p < .001$

Figure 1

Prevalence of Identified Disabilities from Reduced and Full (Covariate-adjusted) Models

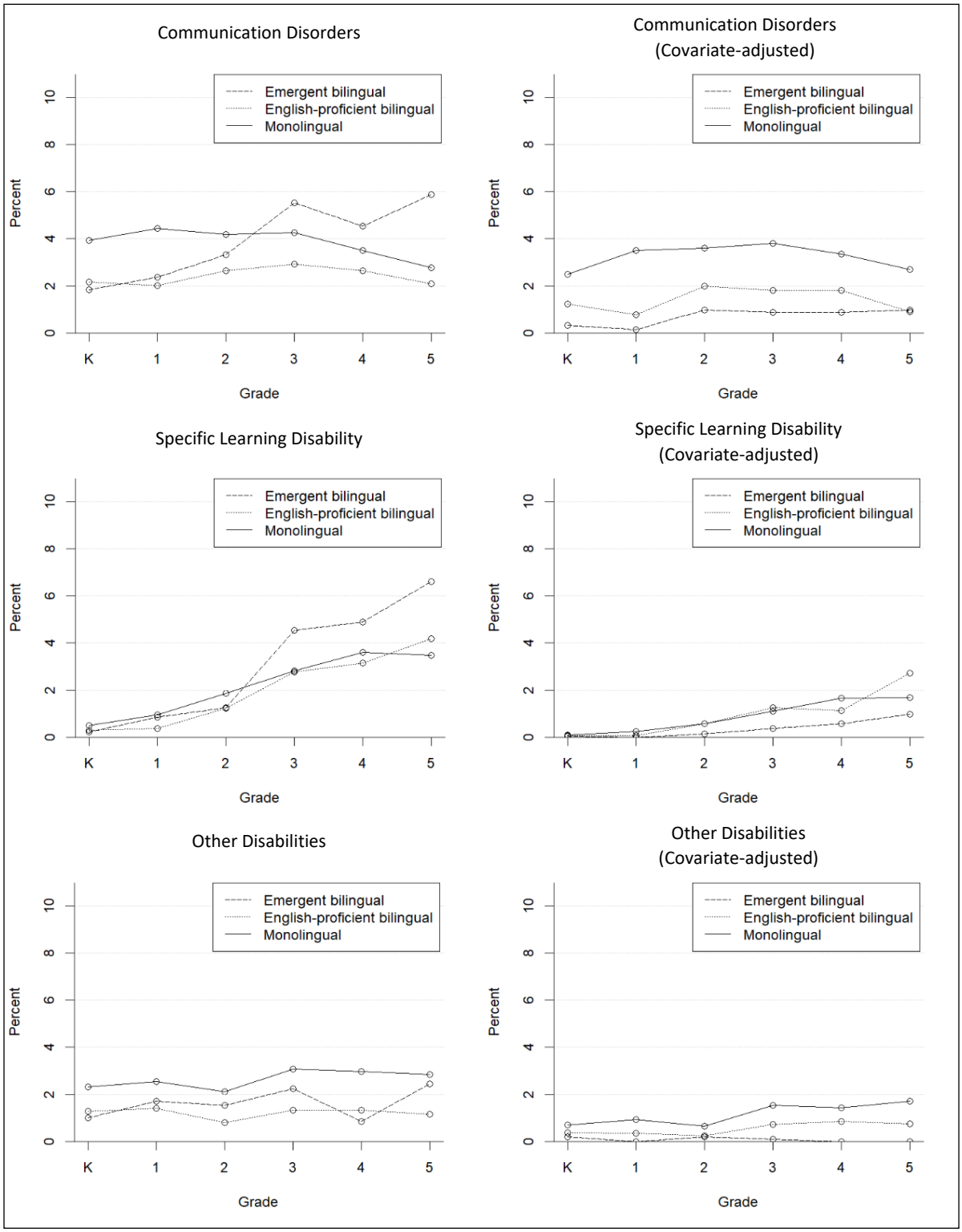
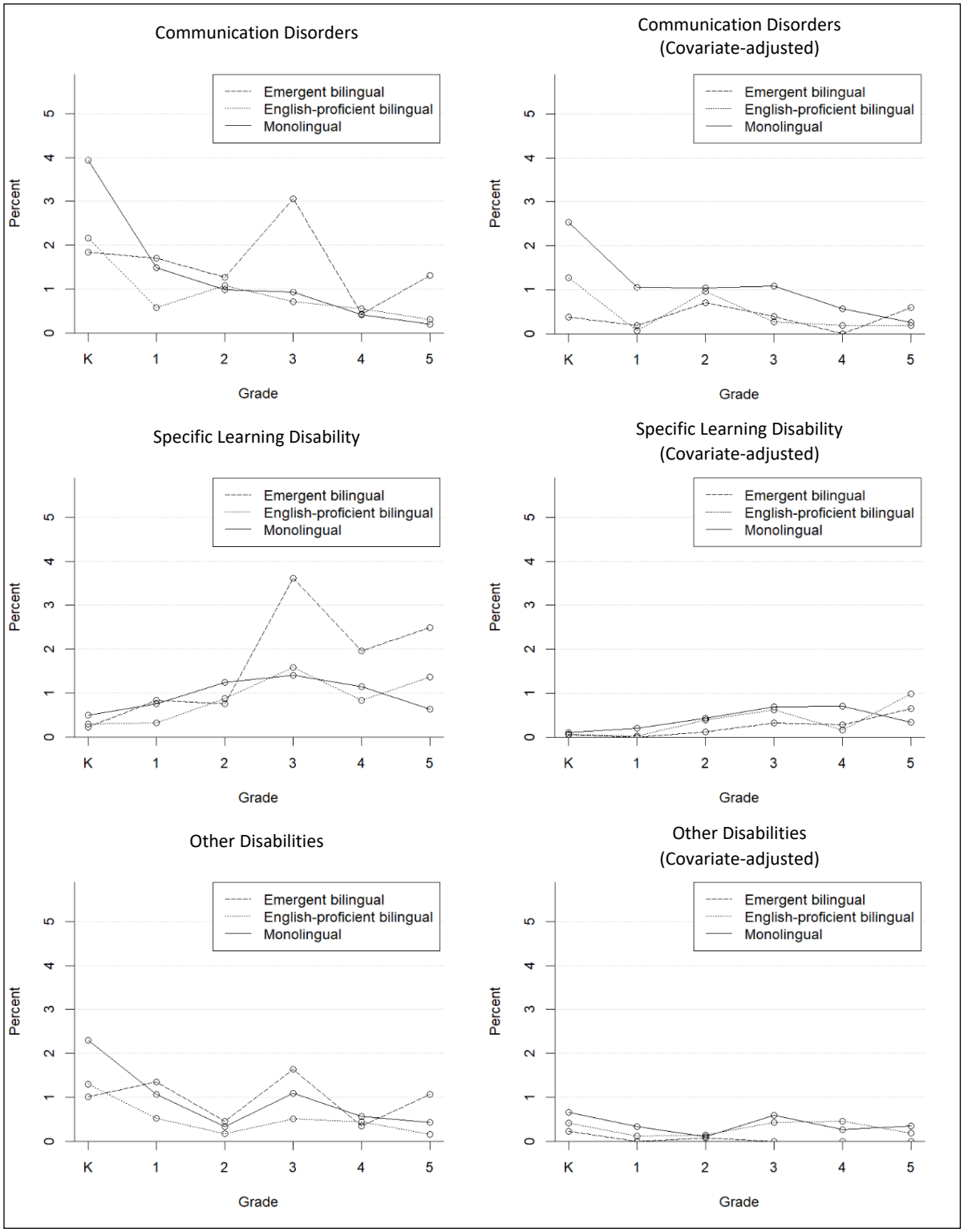


Figure 2

Incidence of Identified Disabilities from Reduced and Full (covariate-adjusted) Models



Chapter IV: Preventing the Disproportionate Identification of Language-Related Disorders in Bilingual Children

There are few greater challenges facing bilingual children in the US than the lack of equitable access to education services. Although they make up more than 10% of students in public schools, English language learners (ELLs), also referred to as emergent bilinguals (García et al., 2008), are at an increased risk for academic difficulties and exhibit poorer outcomes in the areas of reading, math, and science (National Center for Education Statistics, 2021). Emergent bilingual children are more likely to attend poorer performing schools with fewer resources (Peske & Haycock, 2006) and exhibit higher dropout rates than many of their peers (Rodriguez et al., 2020). Bilingual children are also often disproportionately represented in special education, as described in Chapter 3 and in previous studies (P. L. Morgan, Farkas, et al., 2017; Robinson & Norton, 2019; Skiba et al., 2011). The disproportionate identification of language-related disorders, including communication disorders (CD; otherwise referred to as speech or language impairments) and specific learning disability (SLD), is of particular importance given the potential for misidentification due to language differences (Bedore & Peña, 2008; Case & Taylor, 2005; Shenoy, 2014).

There is evidence of both overrepresentation and underrepresentation of bilingual children in language-related disability categories, as discussed in Chapter 3 and in previous studies (Artiles et al., 2005; P. L. Morgan, Farkas, et al., 2017; Sullivan, 2011; Umansky et al., 2017; Yamasaki & Luk, 2018). The problem of disproportionality is based on the presumption that rates of identification should be similar for bilinguals and monolinguals. Overrepresentation in special education reflects a situation in which more children from a certain group are identified with disabilities than would be expected given their relative population size, whereas underrepresentation is just the opposite. Neither of the scenarios is ideal given the long-term consequences associated with both. Overrepresentation suggests that

students may have been misidentified with disabilities and are receiving unnecessary special education services. Although additional services may sound like an innocuous and potentially beneficial occurrence, it is important to acknowledge the possible ramifications associated with being identified with a disability, such as lower graduation rates (Kemp, 2006) and an increased risk of future incarceration (Bell, 2016). Underrepresentation suggests that some bilingual students do not receive necessary services, because they were somehow missed in the disability identification process, resulting in a barrier to special education that other students may not experience. Although the existence of disproportionality does not provide definitive evidence of misidentification, it is telling that the likelihood of being identified with a disability systematically differs across groups, and thus raises questions about equitable access to services.

Despite the persistence of the disproportionate identification of language-related disorders in bilingual students, those individuals who are charged with making determinations about the need for special education services have few tools that they can use to combat the issue. For speech-language pathologists in particular, the role that school-based clinicians should take is unclear. Due to the systemic and multifaceted nature of disproportionality, there are substantial barriers to addressing the issue in a meaningful way. Although much has been written about best practices to reduce disproportionate disability identification for children from minority backgrounds, the solution to the problem remains elusive. To further complicate the issue, there is conflicting evidence regarding the nature of the problem, as over- or underidentification, making it difficult to know which practices to adopt. Although approaches such as nonbiased assessment have received substantial attention in the communication disorders literature, there is relatively little focus on prevention. School-based clinicians are well-positioned to effect positive change for at-risk students, but they must be provided with a set of tools that meets the unique needs of bilingual children.

Thus, the purpose of this paper is to provide school-based speech-language pathologists with an evidence-based approach to preventing the disproportionate identification of language-related disorders in bilingual children. It is important to acknowledge that practitioners from multiple disciplines play a role in making special education decisions. The goal of this paper is not to diminish the importance of practitioners in other fields, but rather to provide discipline-specific recommendations by focusing on a single profession. Although the focus of this paper is on bilingual children, much of the information below is applicable to students from minority backgrounds broadly. What follows is a review of relevant research on disproportionality in bilinguals, followed by a description of evidence-based methods that practitioners can adopt in order to strengthen their ability to provide preventative services.

A Review of Bilingual Disproportionality

Disproportionality has often been considered to be mainly a problem of overrepresentation (Artiles et al., 2002; Artiles & Trent, 1994; Dunn, 1968), a perspective which is reflected in federal policy (US Department of Education, 2016a, 2016b). Despite the prevailing narrative, there is evidence of both overrepresentation (De Valenzuela et al., 2006; Sullivan, 2011) as well as underrepresentation (Hibel & Jasper, 2012; P. L. Morgan et al., 2018; P. L. Morgan, Farkas, et al., 2017). In addition, several studies point to a more complex pattern of disproportionality, in which the risk for disability identification changes as students progress through school. Bilingual populations, in particular, may experience a shift from underrepresentation in early elementary school to overidentification in later grades as shown in Chapter 3 and in previous studies (Artiles et al., 2005; Samson & Lesaux, 2009; Yamasaki & Luk, 2018). Although over- and underrepresentation are often framed as mutually exclusive, both patterns can coexist depending on the group and grade being examined. There are also regional differences in identification rates that need to be considered, as discussed in Chapter 3. Different factors may

contribute to each of these patterns, and it is important to acknowledge how they may result in differences in reported outcomes.

Language Background

For bilingual children, language background is an important factor in predicting the likelihood of being identified with a disability. There is a great deal of diversity within the population of bilingual children in the US, which is not frequently considered in research and policy. Most studies of bilinguals in education have focused specifically on emergent bilinguals, or ELLs, with relatively little attention paid to English-proficient bilinguals. Evidence from a small number of studies has demonstrated how the likelihood of being identified with a language-related disorder differs across bilingual groups. Results from Chapter 3 provide evidence that emergent bilinguals experience growth in their likelihood of being identified from kindergarten to fifth grade, while English-proficient bilinguals experience underrepresentation in early elementary school, as shown in Figure 1. Yamasaki and Luk (2018) also found that emergent bilinguals faced an increasing likelihood of being identified with a language-related disorder as they progress through school, while English-proficient bilinguals may experience consistent underrepresentation in these disability categories across grades. Artiles et al. (2005) similarly found that emergent bilinguals were overrepresented in middle and high school, but that English-proficient bilinguals were consistently underrepresented. Although identification patterns vary somewhat across studies, when taken together, these outcomes suggest that emergent bilinguals face an increasing likelihood of being identified with a disability, particularly with language-related disorders. English-proficient bilinguals, on the other hand, are generally less likely than their monolingual peers to be identified, particularly in early elementary school.

It is also important to consider the relationship between ELL classification and the risk for disability identification. Umansky et al. (2017) examined the risk for two groups: children currently

classified as ELLs and children who had ever been classified as ELLs. Children in special education who were currently classified as ELLs were less likely to be reclassified as English-proficient than their ELL peers who were not receiving special education services. Because ELLs in special education were less likely to be reclassified, they made up a larger proportion of the special education population than would be expected by middle school, resulting in overrepresentation. When examining disability identification for children who had ever been classified as an ELL, on the other hand, the authors found consistent underrepresentation in this population. Results of studies such as these underscore the importance of considering how language background may affect the likelihood of disability identification.

Factors Contributing to Disproportionality

Disproportionality is a complex problem with no clear single cause. There are likely multiple factors that contribute to the issue, and it is a challenge to form a comprehensive picture of the underlying mechanism of both over- and underrepresentation. It is also important to acknowledge that the factors driving disproportionality for one group may be very different than those for another group. Although there is certainly some overlap across groups, disproportionality may manifest itself in different ways for children from different racial, ethnic, or linguistic backgrounds. The unique characteristics that distinguish bilingual children may result in patterns of disability identification that differ substantially from those observed for other groups. Family immigration status, for example, may contribute to underrepresentation in early elementary school, and is more likely to be a relevant factor for bilingual children than for their English-speaking peers (Hibel & Jasper, 2012)

There are several different mechanisms theorized to explain the existence of over- and underrepresentation. Overrepresentation is frequently attributed to systemic bias, which can take a variety of forms. Bias in language assessment may result in common aspects of bilingualism being

misidentified as symptoms of an underlying disorder. Assessment of areas related to language ability, such as expression, comprehension, reading, and writing, is a critical part of the identification of CD and SLD. For this reason, children who are in still in the process of acquiring a second language may be more likely to be misidentified with a language-related disorder, due to challenge of distinguishing between characteristics of bilingualism and symptoms of language-related disorders. Assessment of vocabulary without considering skills across languages, for example, will likely result in lower than expected lexical skills (Anaya et al., 2018; Bialystok et al., 2010). An analysis of phonology that does not consider phonological patterns across languages will lead to similarly inaccurate conclusions about speech sound development (Fabiano-Smith & Goldstein, 2010; Yavas & Goldstein, 1998). Overrepresentation may also be driven by an excessive number of special education referrals for bilingual students due to a lack of familiarity with typical bilingual development (A. A. Ortiz et al., 2011). Another possible mechanism underlying overrepresentation is poor quality of instruction. Because bilingual students often attend schools with fewer resources, they may be in instructional environments that are unable to meet their unique educational needs, thus increasing the risk for academic difficulties (Connor et al., 2007). When provided with access to higher quality instruction, at-risk bilingual students are often able to catch up to their peers (e.g., Gerber et al., 2004; McMaster et al., 2008; Vaughn, Mathes, et al., 2006).

The factors that contribute to underrepresentation likely differ from those related to overrepresentation. Underrepresentation may stem from cultural and linguistic barriers that families face, resulting in numerous challenges accessing education resources in an equitable manner. Immigrant families have limited access to healthcare, resulting in a lower likelihood of a referral for evaluation generated by a provider (Flores & The Committee on Pediatric Research, 2010). Because of fewer interactions with healthcare providers, families may also be generally unfamiliar with the process of disability identification (Calvo & Hawkins, 2015). Immigrant families may also be less engaged because

they feel ill-informed about the special education eligibility process, resulting in lower levels of participation (Wolfe & Durán, 2013). Previous negative experiences with special education evaluations may also result in lower levels of engagement (Hardin et al., 2009; Klingner & Harry, 2006; McHatton & Correa, 2005). Bilingual children may also be less likely to be referred for evaluation if they are in poor-performing schools (Hibel et al., 2010; P. L. Morgan et al., 2015). Additionally, teachers may be less likely to refer emergent bilinguals for a disability evaluation until they demonstrate improved English proficiency (Limbos & Geva, 2001).

Policy Issues

The Individuals with Disabilities Education Act (IDEA; Individuals with Disabilities Education Act, 2004) makes it clear that English proficiency should not be the basis for disability identification. This mandate is at odds with the evidence of disproportionate disability identification for bilinguals, which suggests that language background is associated with differences in identification rates. Despite the longstanding challenges faced by bilinguals, they are largely absent from federal policy on disproportionality, which has instead focused specifically on race and ethnicity. Over- and underrepresentation of children with limited English proficiency is considered to be outside of the scope of federal law (US Department of Education, 2016a). This narrow focus of federal policy is exemplified in the IDEA reporting requirements, in which states are required to report rates of disproportionate disability identification based on race and ethnicity. Disability identification based on English proficiency status, however, is not required and is frequently not publicly available. Although some states do provide this information, many do not, making the true severity of the problem very difficult to examine with any degree of certainty.

In addition to issues related to measurement and reporting, the scope of methods intended to address disproportionality in federal policy is limited to overrepresentation. The official position of the

Department of Education is that overidentification is the predominant problem, citing the "the well-documented and detrimental over-identification of certain students for special education services" (US Department of Education, 2016a, p. 1). Although overidentification is certainly problematic where it exists, the lack of acknowledgement of the possibility of underidentification may result in detrimental outcomes if not addressed in the correct manner. Ideally, interventions designed to address overrepresentation would also have an impact on underrepresentation, but this may not always be the case. By operating under the assumption that disproportionality is exclusively a problem of overidentification, policy may result in the development of practices that are counterproductive.

Shifting from Mitigation to Prevention

Before discussing the role of speech-language pathologists in reducing the disproportionate representation of bilinguals in special education, it is necessary to acknowledge the limitations of the traditional model of assessment and intervention. In the traditional model of service delivery the responsibility of clinicians is to identify existing language-related disorders and provide treatment when indicated (American Speech-Language-Hearing Association, 1988). With respect to disproportionality, this approach is somewhat limited as it does little to address the full range of factors that may impact over- and underrepresentation. One of the major weaknesses of this model is the lack of focus on the events that preceded a referral for evaluation. In effect, any attempt to address disproportionality in the traditional model would be a form of mitigation; clinicians are only able to make an impact after the point of referral, because their role is limited to assessment and treatment. The power of school-based clinicians to make a meaningful impact on disproportionate disability identification is fundamentally restricted if the only role that they play is after the point at which a referral has been generated.

One topic that has received substantial attention with respect to its role in disproportionality is nonbiased assessment, which is an attempt to reduce the degree to which diagnostic tools

systematically disadvantage certain cultural and linguistic groups (e.g., Campbell et al., 1997; Peña et al., 2001; Roseberry-McKibbin & O'Hanlon, 2005). The numerous alternative assessment methods proposed to reduce bias in language evaluation, such as dynamic assessment (Orellana et al., 2019) and nonword repetition (J. A. Ortiz, 2021a), have offered advancements in the ways in which many clinicians conduct assessment for children from culturally and linguistically diverse backgrounds. Despite the benefits that nonbiased assessment approaches offer, it is important to acknowledge their limitations in reducing disproportionality in special education. Assessment bias alone cannot fully explain the disproportionate representation of students from minority backgrounds in special education (Skiba, 2002). Because of its inability to address pre-referral events, an approach to reducing disproportionality which relies entirely on nonbiased assessment presumes that overrepresentation is the sole issue, and thus will be limited in its effectiveness. Reducing bias at the point of assessment, while helpful to distinguish between language difference and disorder, can do nothing to change the educational experiences that preceded an eventual referral for evaluation. An emergent bilingual child who is struggling in kindergarten, for example, may exhibit progressively more difficulty with each subsequent grade if instruction is not modified to meet their needs (e.g., through response to intervention/multi-tiered system of supports; RTI/MTSS). If teachers presume that difficulty is related to their English-proficiency, they may choose to delay referring, thus limiting early access to needed support. By the time they reach third grade, they may be so far behind expected grade-level performance that additional support may be necessary. This "wait-to-fail" approach is suboptimal and only results in a denial of access to equitable services.

Thus, we must recognize the importance of prevention in reducing disproportionality. Prevention is an essential component of service delivery in schools (American Speech-Language-Hearing Association, 1988) whose importance has been repeatedly emphasized in federal disability policy (Every Student Succeeds Act, 2015; Individuals with Disabilities Education Act, 2004; US Department of

Education, 2016a). Prevention is an opportunity for clinicians to ensure that children receive the appropriate level and type of support they need, by focusing on those areas that will lead to optimal outcomes before they have been referred for a disability evaluation. For this reason, in this paper, the term prevention will be used to refer to methods aimed at reducing disproportionate disability identification rates, and not to the prevention of the underlying disabilities themselves. Although prevention has been used to describe a variety of activities, including assessment (Sullivan, 2010), the following sections will detail a prevention model that includes only those activities occurring before the point of referral, including prereferral intervention, early identification, parent engagement, and collaboration. See Figure 2 for a graphical representation of this model.

Pre-referral Intervention

Pre-referral interventions provided through RTI/MTSS models, are underutilized in the speech-language pathologist toolkit (American Speech-Language-Hearing Association, 2020b; McKenna et al., 2021). The goal of interventions such as these is to provide students who are at-risk of academic difficulties with additional support, so that they can catch up to their peers. This is a particularly appealing prospect for bilingual children, who often have poorer academic outcomes than their monolingual peers. When provided with increased frequency and intensity of instruction, through tier 2 and tier 3 interventions in an RTI/MTSS model, bilingual children are indeed able to make substantial gains such that they are no longer at-risk (Vanderwood et al., 2014; Vaughn, Linan-Thompson, et al., 2006). Importantly, there is also evidence that RTI/MTSS may reduce disproportionate identification of SLD for bilinguals (Connor et al., 2014; O'Connor et al., 2013). Furthermore, there is evidence that RTI/MTSS may affect the prevalence of language-related disorders, though more research is warranted to fully explore the direction and degree of this effect (Hall-Mills, 2021).

As a method of prevention, RTI/MTSS offers some clear benefits. For children who would have not received early support, intervention can offer insight into a child's ability to respond to modifications to the instruction and offer some clarity regarding the need for additional services that would have otherwise been difficult to ascertain. Despite the acknowledgement of its importance by professional organizations and researchers, speech-language pathologist participation in RTI/MTSS is limited. School-based clinicians do not consistently conduct screenings, provide interventions, or have leadership roles in RTS/MTSS models (American Speech-Language-Hearing Association, 2020b; McKenna et al., 2021). When clinicians do not participate in RTI/MTSS, the first time that they may see a particular student is at the point of assessment. School-based speech-language pathologists have an opportunity to contribute to a meaningful method of prevention that will have a positive impact on disability identification rates, but currently play a limited role.

What do clinicians need to consider when participating in pre-referral interventions?

There are several key things that speech-language pathologists need to consider when participating in pre-referral interventions. One major consideration is that of the role of the clinicians on the intervention team. Many school-based clinicians are likely accustomed to operating primarily as direct service providers. Although typical clinical activities, such as screening, intervention design, and data collection, are all part of pre-referral interventions, the role of clinicians may be somewhat different than that of a more traditional pull-out intervention model due to the fact that they may not be directly providing the intervention. School-based clinicians who participate on intervention teams should consider what role they can play outside of providing direct intervention services. Regardless of who serves as the interventionist, speech-language pathologists have an opportunity to contribute to interventions that integrate the expertise of all stakeholders on the pre-referral intervention team.

Another consideration for school-based clinicians is that of intervention timing. The importance of when interventions are provided and how bilingual children may uniquely benefit from support provided in early elementary school is particularly evident when considering the spike in incidence of language-related disorders for emergent bilinguals in third grade. Because English proficiency at school entry has far-reaching effects, children who enter school with limited English proficiency may experience a disproportionate increase in identification rates, leading to overrepresentation in special education by fifth grade, as shown in Chapter 3. Given the challenge of distinguishing between language differences and disorders, pre-referral intervention teams will benefit from leveraging the expertise of speech-language pathologists as early as possible in the design and implementation of interventions. Clinicians should consider how they can support tier 2 and tier 3 interventions for bilingual children by using their knowledge of typical language development and common characteristics of bilingualism.

It is also important to acknowledge the unique set of challenges presented by providing services to bilingual children within an RTI/MTSS model to bilingual children. Both teachers and clinicians may find it difficult to implement teaching strategies they see as meaningful for children with limited English proficiency. For English-proficient bilinguals as well, intervention that does not consider abilities across languages may not yield optimal results. Despite these concerns, some intervention will benefit a child who is at-risk of academic difficulty far more than no intervention at all. This is not to say that schools should avoid providing services bilingually, if appropriate and possible, but rather that there is no advantage to withholding a tier 2 or tier 3 intervention for any student, regardless of their language background. Interventions provided in English, for example, can provide a substantial benefit to emergent bilingual children (Connor et al., 2014; Graves et al., 2011; Vaughn, Mathes, et al., 2006), as well as those provided in their native language (Cirino et al., 2009; Gerber et al., 2004; Vaughn, Linan-Thompson, et al., 2006).

Early Identification

The early identification of speech and language difficulties is a core principle of prevention (American Speech-Language-Hearing Association, 1988). Despite this, the lower-than-expected rates of language-related disorders in bilingual children in early elementary school suggest that early identification is lacking for this population. Figure 1 shows how rates of communication disorders in kindergarten are lower for both emergent and English-proficient bilinguals, compared to monolinguals. One possible explanation for this trend is that monolingual children may have higher rates of identified communication disorders before they enter kindergarten. Bilingual children, who are more likely to come from low socioeconomic status households, may have had limited opportunities to be identified with a communication disorder, due to limited access to the same health or education services as their monolingual peers. Because children with language impairment and speech sound disorders exhibit an increased risk of literacy difficulties the future (Boudreau & Hedberg, 1999; Peterson et al., 2009) the early detection of disorders such as these has the potential to make a substantial long-term impact on the outcomes of identified children.

Screening for speech and language difficulties is a common practice in preschool settings, a process with which many clinicians are familiar. Participation in universal screening in RTI/MTSS, on the other hand, may be less common. In school settings, the early identification of children at-risk of academic difficulties, through universal screening, is fundamental to RTI/MTSS. Screening for speech and language difficulties is invaluable in the identification of language-related disorders that may otherwise go undetected, and school-based clinicians play a very important role in this process. Despite its importance, speech-language pathologist participation in universal screening in RTI/MTSS is limited (American Speech-Language-Hearing Association, 2020b; McKenna et al., 2021). One of the potential barriers to participation in universal screening is the lack of clarity in terms of how clinicians can be most

effective. Clinicians may perceive their role as more focused on oral language than on literacy. RTI/MTSS, in contrast, is more often centered on literacy than oral language development. Most RTI/MTSS frameworks, for example, do not have a means of including oral language assessment (Cavazos & Ortiz, 2020). Furthermore, many clinicians may not feel adequately prepared to provide prevention, assessment, or intervention services in the domain of literacy (Loveall et al., 2022). Given the lack of emphasis of oral language ability in RTI/MTSS, it is not surprising that speech-language pathologists may perceive themselves as having a limited role in these types of tiered interventions. Despite this perception, the established link between oral and written language disorders provides clear support for the role of speech-language pathologists in the early identification of language and literacy difficulties.

What tools are available to assist in early identification?

For emergent bilinguals, limited English proficiency may present a barrier to screening language or literacy skills. This is a commonly cited concern among many school-based clinicians and teachers and may result in less support for emergent bilinguals. As previously mentioned, the misconception that limited English proficiency will inhibit any potential benefit from special education services may result in lower referral rates for emergent bilinguals (Limbos & Geva, 2001). Teachers may also assume that currently presenting areas of difficulty are simply related to limited English proficiency, and thus avoid a referral altogether. Teachers and clinicians may also simply be unsure about how to adequately conduct a language or literacy screening for a child with limited English proficiency.

Although language background presents several obstacles to early identification, clinicians can apply principles of nonbiased assessment to screening in order to obtain the highest degree of accuracy possible. To start, measures of language exposure and use should be collected through both and parent and teacher report, a common method of estimating proficiency across languages (Paradis et al., 2010;

Pratt et al., 2022). Information collected from parent/teacher report can be used to guide further screening decisions, such as the choice of direct measures. There are many widely available methods for direct screening of speech and language ability, including many commercial products, but clinicians should be cognizant of the potential for biased outcomes when using such tools on populations for which they were not designed. Direct screening methods ideally should be conducted using both of the child's languages, due to fact that monolingual screening may yield inadequate diagnostic accuracy for many bilingual children (Peña et al., 2011). Though, for children who have been in school for at least one year and who use English at least 30% of the time, language screening in English can be informative and reasonably accurate (Gillam et al., 2013).

Because direct screening in both languages may present a substantial barrier for children from certain language communities, parent report can also highlight any concerns that may be related to a delay or disorder. In addition to its utility in estimating language exposure, parent report can be a very useful tool in the identification of language impairment in bilingual children, particularly when used in conjunction with other measures (Guiberson et al., 2011; Guiberson & Rodríguez, 2010; Paradis et al., 2010; Pua et al., 2017). Parent report of language ability is equally useful for emergent bilinguals as well as English-proficient bilinguals. Although there may be less concern about the ability of English-proficient bilinguals to succeed in an English-only educational environment, thus reducing the perceived necessity of parent report, typical developmental aspects of bilingualism may be mistaken for language-based difficulties. In instances when there is a question of distinguishing between typical bilingualism and a language-related disorder, parent report can provide very valuable information.

Another tool which exhibits promise as a less-biased method of screening is nonword repetition, a measure of phonological working memory. In nonword repetition tasks, children are asked to repeat a series of word-like stimuli that align with the phonotactic rules of a specific language, and therefore

sound like plausible words. Children with CD and SLD frequently have difficulties with working memory (Montgomery et al., 2009; Peng & Fuchs, 2016). As a processing-dependent measure, nonword repetition is considered to be less-biased because it does not rely on prior linguistic knowledge, unlike many traditional forms of assessment. For bilingual children, nonword repetition is a particularly useful tool, because it is less dependent on English proficiency and is effective in the identification of language impairment (J. A. Ortiz, 2021a). Practically, the benefits of nonword repetition include its short administration time and ease of administration. For a guide to the clinical of application of nonword repetition see J. A. Ortiz (2021b).

Parent Engagement

Parent engagement is an essential component of service delivery for students with disabilities and is critical to academic success. Increased levels of parent engagement are associated with better academic outcomes (Boonk et al., 2018), reduction in behavioral problems (Sheridan et al., 2012), and decreased dropout rates (Barnard, 2004). For students with disabilities, schools have an obligation to ensure that parents are included in the development and provision of special education services (Every Student Succeeds Act, 2015; Individuals with Disabilities Education Act, 2004). Parents may also serve as useful partners in administering prevention, and their involvement in RTI/MTSS pre-referral intervention teams may also reduce the likelihood of needing special education services (Chen & Gregory, 2011).

Despite the importance of working with families, many parents of bilingual children may encounter numerous barriers to effective participation, thus, reducing their willingness to engage (Wolfe & Durán, 2013). Although it may not be surprising to learn that language differences impose burdens on participation, other types of barriers may exist as well, many of which are very concerning. Families from minority backgrounds report challenges such as inadequate information, disrespect, and negativity towards their children. These negative experiences inevitably create tension between families

and educators and make it difficult to maintain productive working relationships. Clinicians need to be cognizant of the cultural and linguistic barriers that families face, as well as the negative experiences that they may have had. Although there is nothing that one can do to erase these past experiences, it is critical to acknowledge that negative interactions such as these may be the source of lower levels of engagement.

How can school-based clinicians effectively work with families?

For school-based clinicians, parent engagement is an essential component of service delivery. Despite its importance, the inclusion of families in speech and language services provided in schools is far from universal (McKenna et al., 2021). Although clinicians report a high perceived importance of family inclusion in service delivery, approximately 43% of school-based clinicians report that a lack of family involvement was their greatest challenge (American Speech-Language-Hearing Association, 2020b). There is also a great deal of variability in the way that school-based clinicians initiate and maintain communication with families (Tambyraja, 2020).

Improving parent engagement is fundamental in the prevention of disproportionality, and there are many ways that clinicians can engage with parents and families as partners in prevention, assessment, and intervention. Although school-based speech-language pathologists are likely accustomed to communicating with families of the children on their caseload regularly, they may be far less likely to engage with families within the larger community. In the prevention model, parent engagement means stepping outside of one's own caseload and supporting families more broadly than in the traditional model of service delivery. For school-based clinicians the goal should be to use their specialized knowledge to support a specific area of need that they have identified at their site. Prevention-focused educational workshops are one means of engaging with families that can provide substantial benefits for school-based clinicians. Educational workshops are a well-established method of

engaging with families in schools. Despite their utility, workshops may not be part of the typical speech-language pathologist toolkit. One of the major benefits of family engagement through workshops is the benefit to families who may feel ill-informed about common school procedures, such as the disability identification process. Many families from minority backgrounds cite lack of information as a primary concern (Wolfe & Durán, 2013) and workshops can serve as an important tool to address this information gap. Workshops can be also an effective method of building support networks with the community, which many families may find beneficial (Kummerer, 2012). The content of workshops will depend on the education setting, the age range of children, and the expertise of the clinician and may include such topics as early language milestones, literacy development, and language enrichment activities.

Collaboration

A student-centered approach to service delivery necessitates that professionals work in a collaborative manner so that they are able share knowledge and make decisions that will result in optimal outcomes for a given student. In the prevention model, coordination between all stakeholders is required so that an intervention can be appropriately designed, and modified, to meet a particular child's needs. Interprofessional practice is a framework for collaboration that puts the individual receiving services at the center (for an overview, see American Speech-Language-Hearing Association, n.d.). The model for interprofessional practice was developed by the World Health Organization (2010) with the goal of improving health outcomes, and it holds promise for its potential to reduce inequities in schools (Pfeiffer et al., 2019; Rosa-Lugo et al., 2017). Interprofessional practice comprises four main competency areas: values and ethics, roles and responsibilities for collaborative practice, interprofessional communication, and interprofessional teams and teamwork (Interprofessional

Education Collaborative Expert Panel, 2011; Rosa-Lugo et al., 2017). One of the main distinctions of the interprofessional practice model is its emphasis on collaboration first, rather than as an afterthought.

Many speech-language pathologists working in schools cite limited time for collaboration as their greatest challenge and clinicians only spend about 5% of their time engaging in collaborative consultation (American Speech-Language-Hearing Association, 2022). Operating in siloed work environments has the potential to result in such negative consequences as diagnostic errors and redundancy in services across disciplines (American Speech-Language-Hearing Association, n.d.). For this reason, a central objective of interprofessional practice is to avoid fragmentation of services, so that professionals are not working in an isolated manner. A collaborative approach to early literacy support, for example, would utilize the expertise of multiple professionals including, but not limited to, the general education teacher and the speech-language pathologist. Unfortunately, many school-based clinicians exhibit very low levels of collaboration in assessment and intervention (Pfeiffer et al., 2019). It is not uncommon for school-based speech-language pathologists to work somewhat independently. In many school districts, for example, diagnostic clinicians may not be tied to a single school and may only briefly interact with the school-based support team before conducting the evaluation and submitting a report. In addition, the traditional model of pull-out intervention services also does not lend itself well to collaboration beyond a brief consultation with teachers between therapy sessions.

How can clinicians overcome common barriers to collaboration?

In a prevention model, collaboration is intentional, not incidental, and extends well beyond one-off conversations with colleagues. It may be difficult to envision what a collaborative model looks like in a school setting, given the multitude of forms that collaboration can take depending on the age-range, educational environment, and areas of difficulty being addressed. Effective collaboration begins with an acknowledgment of the limitations imposed by the traditional model of service delivery. Pull-out

services are common among school-based clinicians (American Speech-Language-Hearing Association, 2020a), but this model likely leads to less-than-ideal levels of collaboration. Because prevention-focused approaches, such as RTI/MTSS, are reliant on the contributions of multiple stakeholders, participation on a pre-referral intervention team may help facilitate improved communication across disciplines. In many cases, clinicians will need to seek out these opportunities as they may deviate from what is traditionally expected of school-based speech-language pathologists. For individuals serving on assessment teams, finding ways to collaborate may be a particularly challenging task. Assessment teams may perceive themselves as having a diminished role in prevention, given their responsibility to complete an evaluation after a referral has already been generated. Clinicians on assessment teams should consider what they can do outside of the traditional diagnostician role, to support goal of identifying children who truly need special education services. Clinicians need to be knowledgeable about the RTI/MTSS procedures at their sites and ensure that students who are referred for evaluation have been given an opportunity to demonstrate their learning capacity. This may mean having a voice earlier in the identification process and advocating for the provision of tier 2 and tier 3 services for children who need them. A referral generated without the use RTI/MTSS, for example, indicates the need for adjustments to the pre-referral process. In instances such as these, there is a clear opportunity for the school-based speech-language pathologist to collaborate with colleagues by advocating for a more prevention-oriented approach to special education eligibility procedures.

Clinicians should also take advantage of opportunities to actively participate in, and contribute to, professional development. The aim of prevention-focused professional development should be on finding ways to share knowledge and learn from colleagues, by leveraging the mutually shared objective of improving student outcomes. One particularly useful method of accomplishing this is with the community of practice model, a collaborative means of sharing information among individuals who have

a mutual interest (Wenger et al., 2002). Communities of practice serve as an effective method of professional learning among educators, by fostering improved collaboration and reducing feelings of isolation (Patton & Parker, 2017). There are numerous examples of how communities of practice can be been utilized across disciplines (e.g., Cruess et al., 2018; O'Donnell & Tobbell, 2007). Given the shared goal of reducing disproportionality across disciplines in schools, the community of practice model demonstrates promise for its potential to support collaboration in service of prevention.

Additional Considerations

The sections above describe the core elements of a prevention model whose goal is to reduce the disproportionate identification of language-related disorders in bilingual children. Although the focus here has been on individual action, it is also important to acknowledge the scope of the problem, which cannot be solved without the support of supervisors, principals, administrators, school boards, families, advocacy organizations, and politicians. The factors that drive disproportionality may appear intractable and clinicians may find it difficult to even implement the most basic of recommendations for best practice. Something so straightforward as reducing reliance on standardized assessment for children from culturally and linguistically diverse backgrounds, for example, may be a struggle in some school systems given the extensive use of standardized tests in language assessment (Arias & Friberg, 2017). Disproportionality represents a long-standing fundamental inequity in our education system, whose solution requires a change beyond what individual clinicians can do without the support of other stakeholders. This is not to say that clinicians should consider themselves to be inefficacious in reducing disproportionality, but rather that the solutions presented here should be adopted at the school-level, minimally. School-based speech-language pathologists have a unique skill set and are well positioned to invoke the type of change needed to improve educational equity for disadvantaged students.

For many clinicians, advocacy does not play a central role in the work that they do, and many do not feel equipped to participate in activities related to advocacy (Lugo et al., 2022). Advocacy can take a variety of different forms, and clinicians may be unsure of where to begin. As a starting point, school-based clinicians should ensure that they adhere to and promote best practice at their sites. On a larger scale, clinicians should consider participation in state-level professional associations, as members or in leadership positions. Many state-level associations have opportunities for members to participate in committees that actively engage in professional development and policy advocacy. One such example is the Multicultural Affairs Committee of the Maryland Speech-Language-Hearing Association, whose members comprise practicing speech-language pathologists and audiologists with a shared interest in resolving state-wide issues related to cultural and linguistic diversity. Organizations such as this provide opportunities for clinicians to engage in activities with a broad impact.

One additional concern for many clinicians is the additional responsibility associated with increased participation in the prevention methods described above (McKenna et al., 2021). This is a legitimate concern, and it is important to consider how increased use of prevention may be implemented in a way that is both effective and not overly burdensome. Rather than viewing prevention as an additional burden, we should envision an approach to service delivery that better serves the needs of the children we serve, by moving away from a model that relies solely on mitigation. For clinicians to be able to play a role in achieving the goal of reducing disproportionality in special education, a shift from the traditional assessment and treatment approach is needed, and resources should be allocated to prevention activities in a more meaningful way. It goes without saying reallocating resources may present a major challenge, considering the lack of service providers that many schools face. Decisions about the specific roles and responsibilities of school-based clinicians may also not be at the discretion of the clinicians themselves, but rather supervisors, principals, and administrators. For this reason, clinicians should work with those in leadership positions to help identify solutions to the specific barriers

to participation in prevention they may face. Every school is different, with a unique set of challenges, and thus solutions vary across sites. Solutions that do not consider prevention in a meaningful way, though, will do little to solve the lack of equity in access to special education services.

Lastly, one major challenge that must be acknowledged is the effect of poor quality of instruction on academic outcomes. Students who do not receive adequate instruction in early elementary are likely to have substantially more difficulty in later grades and may be at greater risk of academic failure, increasing the possibility of a special education referral. Bilingual students are at an elevated risk of academic difficulties, and their unique needs require instruction that considers their language background. The challenge that school-based speech-language pathologists face is that they often do not play a role in general education instruction. Despite this, it is important to consider how poor early academic preparation may lead to future areas of difficulty. RTI/MTSS may provide opportunities to provide early support to struggling students, but they are no substitute for high quality instruction. Clinicians should be aware of the degree to which instructional quality is related to disproportionate disability identification in their schools and utilize the tools that they have available to provide additional support wherever possible.

Conclusion

The disproportionate identification of language-related disorders in bilingual children in US schools is an ongoing issue. Despite being central to assessment and intervention services, school-based speech-language pathologists do not universally engage in disproportionality prevention activities. School-based clinicians can play a major role in reducing the disproportionate identification of disabilities in bilingual children. In order to meet the needs of the increasingly diverse populations of schools in the US, we need to shift our approach from disproportionality mitigation to prevention. Speech-language pathologists are well positioned to provide a meaningful contribution and should

adopt an approach that considers the importance of prevention in service delivery. Although the focus of this paper is on bilinguals, the concepts discussed here can be applied to children from culturally and linguistically diverse backgrounds broadly. The adoption and strengthening of practices such as early identification, pre-referral intervention, parent engagement, and collaboration will go a long way to improving outcomes for the clients we serve and to ensuring more equitable access to special education services for all children.

Figures

Figure 1

Prevalence of Language-Related Disorders in Bilingual and Monolingual Children in Schools

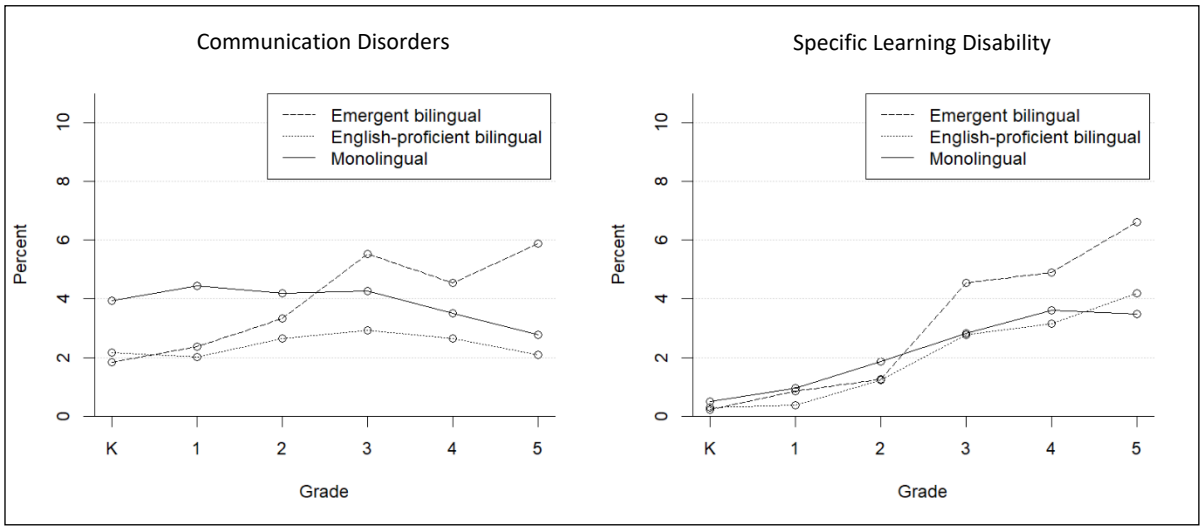
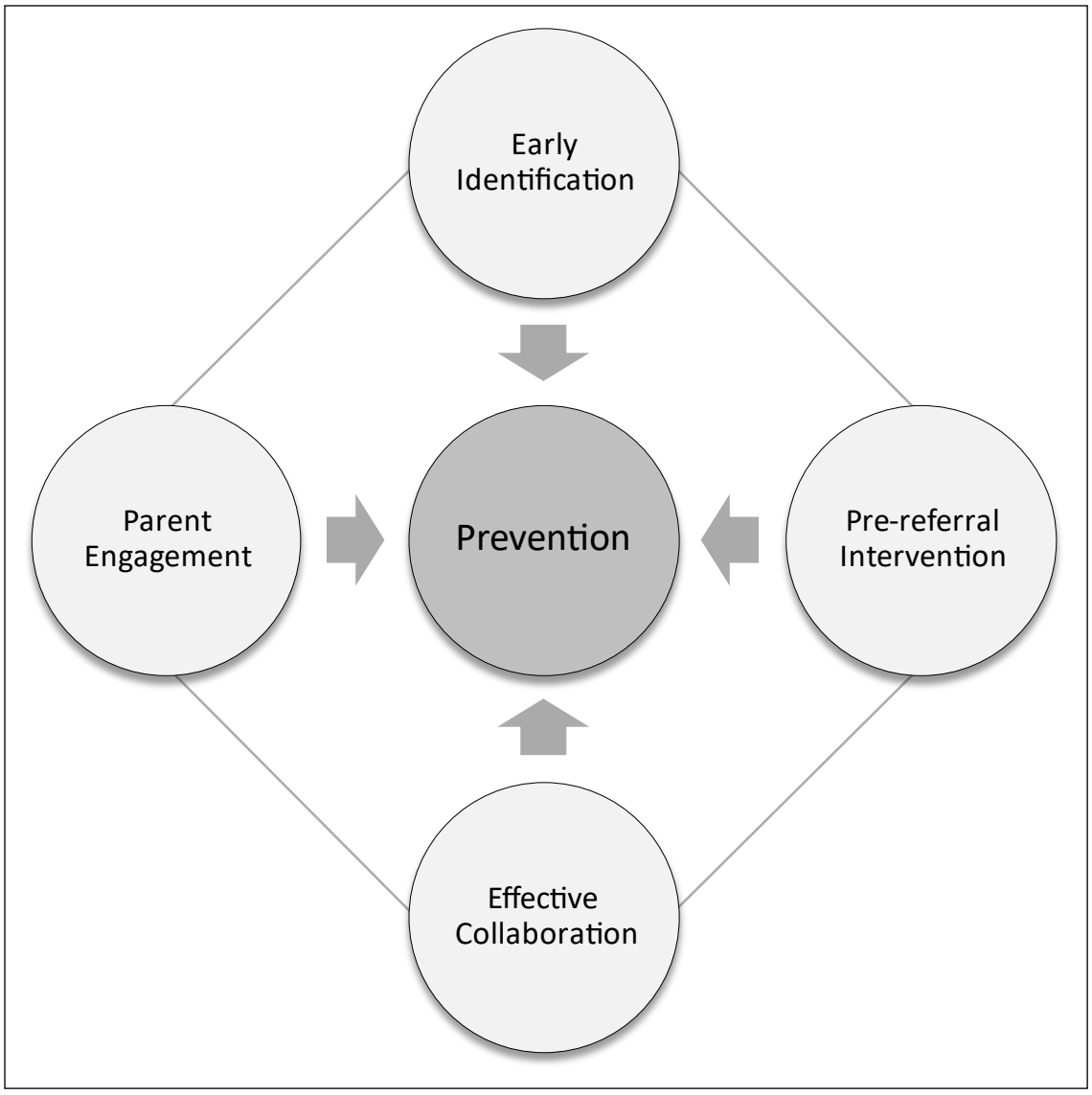


Figure 2

A Model for the Prevention of the Disproportionate Identification of Language-Related Disorders



Chapter V: Discussion and Implications

The focus of the papers presented in the preceding chapters was on the disproportionate identification of language-related disorders in bilingual children. Each of these papers was intended to address one dimension of disability identification in bilinguals. Chapter 2 comprised a meta-analysis of diagnostic accuracy for nonword repetition, a method of nonbiased language assessment. Chapter 3 comprised an empirical study of disproportionality in two disability categories: communication disorders (CD) and specific learning disability (SLD). Chapter 4 comprised a model of prevention for school-based speech-language pathologists, intended to reduce the disproportionate identification of language-related disorders in bilingual children. In the following sections, I will review the major findings from Chapters 2 and 3, the implications for practitioners discussed in Chapter 4, followed by a discussion of directions for future research.

Nonword Repetition in Bilingual Language Assessment

One of the major challenges in language assessment for bilingual children is finding diagnostic tools that have a sufficient level of precision to be useful in the identification of language impairment. Many commonly used assessment tools may result in inaccurate diagnostic conclusions when used with populations for which they were not intended. Tools that were developed for monolingual children, for example, are unlikely to exhibit adequate diagnostic accuracy if administered to bilingual children. The need for better assessment tools in bilingual language assessment is very important when considering the likelihood of misidentification of language impairment in bilingual children in public schools, and the potential effect on disproportionate disability identification in this population. Alternative assessment methods, such as dynamic assessment (Peña et al., 2001), conceptual vocabulary scoring (Anaya et al., 2018), and processing-dependent measures (Dollaghan & Campbell, 1998), may provide a more accurate representation of language ability compared to many traditional methods of assessment.

Nonword repetition, as a type of processing-dependent measure, is one method of alternative assessment for culturally and linguistically diverse children, for whom traditional forms of assessment may be inappropriate due to the potential for biased outcomes. The meta-analysis in Chapter 2 included 13 studies of the diagnostic accuracy of nonword repetition in bilinguals. Results suggest that nonword repetition is a useful tool in the identification of language impairment in bilingual children, but that it is best used alongside other tools. Across studies, the diagnostic accuracy of nonword repetition ranged from poor to good, reflecting the degree to which factors such as the stimuli, participant characteristics, and study design may affect study outcomes. Quasi-universal nonword tasks, which include words that may be phonologically plausible in a range of different languages, exhibited significantly better diagnostic accuracy than language-specific tasks. Although nonword repetition does not exhibit sufficient diagnostic accuracy to be used on its own, when used alongside other tools, nonword repetition can provide valuable diagnostic information.

This is the first study to systematically examine the evidence of diagnostic accuracy of nonword repetition for the purpose of identifying language impairment in bilinguals. Results from this study contribute to the growing research into less-biased assessment methods for bilingual children and have direct implications for any practitioner who conducts assessment on bilingual children, including speech-language pathologists and school psychologists.

Disproportionality of Language-Related Disorders in Bilinguals

The disproportionate representation of children from culturally and linguistically diverse backgrounds in special education is a problem that continues to persist, despite having first been identified over 50 years ago (Dunn, 1968). For bilingual children, there is evidence of both overrepresentation (Linn & Hemmer, 2011; Sullivan, 2011) as well as underrepresentation (P. L. Morgan et al., 2015; P. L. Morgan, Farkas, et al., 2017) in special education. The identification of language-

related disorders, including communication disorders and specific learning disability, is of particular concern for bilingual children given the difficulty in differentiating between language differences and disorders and the potential for misidentification. Bilingual disproportionately in language-related disorders is an area of research with a number of unanswered questions.

Chapter 3 comprises an empirical study of the disproportionate identification of language-related disorders in bilingual children. This study sought to answer questions related to the incidence and prevalence of communication disorders and specific learning disability in US schools, as well as predictors of their identification. Drawing from a nationally representative individual-level data set, this study examined the likelihood of being represented (i.e., prevalence) in language-related disorder categories, as well as rates of identification (i.e., incidence), for emergent bilinguals, English-proficient bilinguals, and monolinguals, from kindergarten to fifth grade. Emergent bilinguals experienced a spike in the incidence of language-related disorders in third grade, leading to overrepresentation in later grades. English-proficient bilinguals experienced under-representation in early elementary school grades, and more proportionate representation in later grades. Several predictors were significantly associated with language-related disorder identification, including gender, behavioral functioning, academic achievement, working memory, and geographic region.

Results from this study provide a novel contribution to the understanding of the identification of disabilities in bilingual children in elementary school. This study is among a small number of longitudinal studies of disproportionality in bilinguals. This is the first study to specifically examine both the prevalence and incidence of language-related disorders in emergent and English-proficient bilinguals.

Implications for Practitioners

Results from studies included in Chapters 2 and 3 have direct implications for education practitioners. Results from the meta-analysis of nonword repetition can be used to guide assessment

decisions, in order to attain the highest possible diagnostic accuracy, and served as the basis for a practitioner paper on the topic (J. A. Ortiz, 2021b). Chapter 3, similarly, served as the basis for the practitioner paper in Chapter 4.

The goal of the paper in Chapter 4 is to provide an evidence-based approach to preventing disproportionate disability identification in bilinguals for school-based speech-language pathologists. The role that prevention plays in disproportionality is central to the model presented in this chapter. This paper is novel in its central focus on the role that prevention plays in reducing disproportionality for school-based clinicians, whereas previous practitioner papers have largely focused on mitigation strategies, such as assessment bias. In addition, there are few discipline-specific papers that directly discuss the role of speech-language pathologists in addressing disproportionality. Lastly, this paper provides a meaningful contribution by clearly describing the nature of disability identification over time for different bilingual groups based on the most recently available research, and the implications for service delivery.

Future Research

Results from the studies presented in this dissertation provide a basis for further research into several topics related to the identification of language-related disorders in bilingual children. With respect to the use of nonword repetition in assessment, future studies should continue to examine the diagnostic accuracy of quasi-universal nonword repetition tasks, given the small number of studies that included tasks of this type. In addition, a greater variety of languages should be included in future studies to increase generalizability. Similarly, more children of different age ranges, particularly children under the age of five, should be included in future studies. This will strengthen the evidence supporting the use of nonword repetition as a tool in early childhood language screening. In addition, more research is needed into the diagnostic accuracy of nonword repetition when used in conjunction with

other assessment tools. Future meta-analytic studies of nonword repetition may also wish to include studies that report quantitative results other than diagnostic accuracy, such as mean differences. This will increase the sample size of future meta-analyses, providing additional insight into the clinical utility of the task.

Regarding research into the disproportionate identification of disabilities in bilingual children, there are several areas in which future studies can make novel contributions. Acknowledging and accounting for the heterogeneity of bilinguals is critical in the investigation of disability identification rates for bilingual children, as exemplified in the results from Chapter 3. Future research should continue to focus on emergent and English-proficient bilinguals as two distinct groups. Few studies draw a distinction between these groups, and most research focuses specifically on children classified as English language learners. The importance of accounting for variation in language background in the examination of rates of incidence, prevalence, as well as exit from special education is needed to better understand how patterns of representation across grades differ for children from different groups. Future longitudinal studies that include middle and high school will provide further insight into long-term outcomes. In addition, further examination of school-level variables, such as the teacher education, teacher demographics, school resources, and instructional practices, would provide a great deal of insight into additional factors that drive disproportionate disability identification rates.

With respect to methodology, disproportionality research has relied largely on inferential and descriptive statistics. Despite substantial advances in our understanding of the problem, the causal mechanisms underlying disproportionality require further examination. Future research, utilizing causal inference methodology, would be helpful to better understand the underlying mechanisms driving disproportionality. The growing number of studies utilizing techniques such as propensity score matching, to examine the effects of ELL classification (Umansky & Dumont, 2021) and special education

placement (P. L. Morgan, Frisco, et al., 2017), provide insight into how causal inference methodology may be applied to the study of disproportionality.

Lastly, one of the major limitations of current disproportionality research is the absence of information about misidentification. Future research into the role of misidentification is needed to better examine the accuracy of disability identification decisions. Few studies have focused on the accuracy of disability identification in schools, and more information regarding the degree to which misidentification affects children across different groups will provide insight into the mechanisms underlying disproportionality. There are substantial barriers to conducting this type of research, and many currently available nationally representative data sets, such as the ECLS-K:2011, do not provide the types of observations needed to draw conclusions about misidentification. Data that allow for the comparison of disability evaluation outcomes to a reference standard would be valuable in the examination of the role of misidentification on disproportionality.

References

Abedi, J. (2009). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology*, *10*(2). eric. http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ865585&site=ehost-live

American Speech-Language-Hearing Association. (n.d.). *Interprofessional Education/ Interprofessional Practice (IPE/IPP)*. American Speech-Language-Hearing Association. Retrieved June 17, 2022, from https://www.asha.org/practice/ipe-ipp/

American Speech-Language-Hearing Association. (1988). *Prevention of communication disorders*. American Speech-Language-Hearing Association; American Speech-Language-Hearing Association. https://www.asha.org/policy/ps1988-00228/

American Speech-Language-Hearing Association. (2020a). *2020 Schools survey: SLP caseload and workload characteristics*. 14.

American Speech-Language-Hearing Association. (2020b). *2020 Schools survey. Survey summary report: Numbers and types of responses, SLPs.* www.asha.org

American Speech-Language-Hearing Association. (2022). *2022 Schools survey. Survey summary report: Numbers and types of responses, SLPs.* www.asha.org

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of vocabulary testing in bilingual children. *Language, Speech & Hearing Services in Schools*, *49*(1), 85–97. https://doi.org/10.1044/2017_LSHSS-16-0081

Anderson, R. T. (2012). First language loss in Spanish-speaking children. *Bilingual Language Development and Disorders in Spanish-English Speakers*, *2*, 193–212.

Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools*, *48*(1), 1–15. https://doi.org/10.1044/2016_LSHSS-15-0090

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew: Accuracy of repetition tasks for identification of SLI in bilingual children. *International Journal of Language & Communication Disorders*, *51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Artiles, A. J., Harry, B., Reschly, D. J., & Chinn, P. C. (2002). Over-identification of students of color in special education: A critical overview. *Multicultural Perspectives*, *4*(1), 3–10.

Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., & Ortiz, A. (2010). Justifying and explaining disproportionality, 1968–2008: A critique of underlying views of culture. *Exceptional Children*, *76*(3), 279–299. https://doi.org/10.1177/001440291007600303

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, *71*(3), 283–300. https://doi.org/10.1177/001440290507100305

Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *The Journal of Special Education*, *27*(4), 410–437.

Bailey, P., Emad, A., Huo, H., Lee, M., Liao, Y., Lishinski, A., Nguyen, T., Xie, Q., Yu, J., & Zhang, T. (2021). *EdSurvey* (2.7.1).

Barnard, W. M. (2004). Parent involvement in elementary school and educational attainment. *Children and Youth Services Review*, *26*(1), 39–62. https://doi.org/10.1016/j.childyouth.2003.11.002

Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish. *Language Speech and Hearing Services in Schools*, *49*(2), 292. https://doi.org/10.1044/2017_LSHSS-17-0013

Basu, S. (2018). Age-of-arrival effects on the education of immigrant children: A sibling study. *Journal of Family and Economic Issues*, *39*(3), 474–493. https://doi.org/10.1007/s10834-018-9569-4

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–29. https://doi.org/10.2167/beb392.0

Bedore, L. M., Peña, E. D., Anaya, J. B., Nieto, R., Lugo-Neris, M. J., & Baron, A. (2018). Understanding disorder within variation: Production of English grammatical forms by English language learners. *Language, Speech, and Hearing Services in Schools*, *49*(2), 277–291.

Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools*, *36*(3), 188–200. https://doi.org/10.1044/0161-1461(2005/020)

Bell, C. (2016). Special needs under siege: From classrooms to incarceration. *Sociology Compass*, *10*(8), 698–705. https://doi.org/10.1111/soc4.12392

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *37*(4), 391–403.

Black, L. I., Vahratian, A., & Hoffman, H. J. (2015). *Communication disorders and use of intervention services among children aged 3-17 years: United States, 2012*. US Department of Health and Human Services, Centers for Disease Control and ….

Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, *66*, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1747–1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Boonk, L., Gijselaers, H. J. M., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, *24*, 10–30. https://doi.org/10.1016/j.edurev.2018.02.001

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *36*(4), 421–432.

Boudreau, D. M., & Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *American Journal of Speech-Language Pathology*, *8*(3), 249–260. https://doi.org/10.1044/1058-0360.0803.249

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, *22*(3), 441–469.

Buckley, J., Schneider, M., & Shang, Y. (2004). The effects of school facility quality on teacher retention in urban school districts. *National Clearinghouse for Educational Facilities*. https://eric.ed.gov/?id=ED539484

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, *332*(6033), 1049–1053. https://doi.org/10.1126/science.1201536

Calvo, R., & Hawkins, S. S. (2015). Disparities in quality of healthcare of children from immigrant families in the US. *Maternal and Child Health Journal*, *19*(10), 2223–2232. https://doi.org/10.1007/s10995-015-1740-z

Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, *9*(4), 312–322. https://doi.org/10.1080/14417040701624474

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment. *Journal of Speech, Language, and Hearing Research*, *40*(3), 519–525. https://doi.org/10.1044/jslhr.4003.519

Carreker, S., & Joshi, R. M. (2010). Response to intervention: Are the emperor's clothes really new? *Respuesta a La Intervención: ¿son Realmente Nuevos Los Vestidos Del Emperador?*, *22*(4), 943–948.

Case, R. E., & Taylor, S. S. (2005). Language difference or learning disability? Answers from a linguistic perspective. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *78*(3), 127–131. https://doi.org/10.3200/TCHS.78.3.127-131

Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology*, *29*(3), 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179

Cavazos, L. O., & Ortiz, A. A. (2020). Incorporating oral language assessment into MTSS/RTI frameworks: The potential of personal narrative assessment. *Bilingual Research Journal*, *43*(3), 323–344. https://doi.org/10.1080/15235882.2020.1826367

Chen, W.-B., & Gregory, A. (2011). Parental involvement in the prereferral process: Implications for schools. *Remedial and Special Education*, *32*(6), 447–457. https://doi.org/10.1177/0741932510362490

Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 125–150). Multilingual Matters. https://doi.org/10.21832/9781783093137-008

Chu, S.-Y., & Flores, S. (2011). Assessment of English language learners with learning disabilities. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *84*(6), 244–248. eric.

Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M., & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal*, *46*(3), 744–781. https://doi.org/10.3102/0002831208330214

Collins, K. M., Connor, D., Ferri, B., Gallagher, D., & Samson, J. F. (2016). Dangerous assumptions and unspoken limitations: A disability studies in education response to Morgan, Farkas, Hillemeier, Mattison, Maczuga, Li, and Cook (2015). *Multiple Voices for Ethnically Diverse Exceptional Learners*, *16*(1), 4–16.

Connor, C. M., Alberto, P. A., Compton, D. L., O'Connor, R. E., & National Center for Special Education Research (ED). (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.* (National Center for Special Education Research. 400 Maryland Avenue SW, Washington, DC 20202. Tel: 800-437-0833; Fax: 202-401-0689; Web site: http://ies.ed.gov/ncser/; NCESR 2014-3000). National Center for Special Education Research; eric. http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED544759&site=ehost-live

Connor, C. M., Morrison, F. J., & Underwood, P. S. (2007). A second chance in second grade: The independent and cumulative impact of first- and second-grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading*, *11*(3), 199–233. https://doi.org/10.1080/10888430701344314

Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies*, *9*(2), 135–156. https://doi.org/10.1023/A:1009462820157

Cruess, R. L., Cruess, S. R., & Steinert, Y. (2018). Medicine as a community of practice: Implications for medical education. *Academic Medicine*, *93*(2), 185–191. https://doi.org/10.1097/ACM.0000000000001826

Cruz, R. A., & Firestone, A. R. (2022). Understanding the empty backpack: The role of timing in disproportionate special education identification. *Sociology of Race and Ethnicity*, *8*(1), 95–113. https://doi.org/10.1177/23326492211034890

Cruz, R. A., & Rodl, J. E. (2018). An integrative synthesis of literature on disproportionality in special education. *The Journal of Special Education*, *52*(1), 50–63. https://doi.org/10.1177/0022466918758707

Cummings, K. D., & Smolkowski, K. (2015). Bridging the gap: Selecting students at risk of academic difficulties. *Assessment for Effective Intervention*, 7.

Cummings, K. P., & Hardin, B. J. (2017). Navigating disability and related services: Stories of immigrant families. *Early Child Development and Care*, *187*(1), 115–127. https://doi.org/10.1080/03004430.2016.1152962

de Almeida, L., Ferré, S., Morin, E., Prévost, P., Santos, C. dos, Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism*, *7*(3–4), 331–358. https://doi.org/10.1075/lab.15019.alm

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders*, *46*(6), 613–627. psyh. https://doi.org/10.1111/j.1460-6984.2011.00049.x

De Valenzuela, J. S., Copeland, S. R., Qi, C. H., & Park, M. (2006). Examining educational equity: Revisiting the disproportionate representation of minority students in special education. *Exceptional Children*, *72*(4), 425–441. https://doi.org/10.1177/001440290607200403

Deeks, J. J., Bossuyt, P. M., & Gatsonis, C. A. (Eds.). (2013). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Version 0.9). The Cochrane Collaboration. http://srdta.cochrane.org

Deeks, J. J., Macaskill, P., & Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, *58*(9), 882–893. https://doi.org/10.1016/j.jclinepi.2005.01.016

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Doebler, P., & Holling, H. (2015). *Meta-analysis of diagnostic accuracy with mada*. 21.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders.* Paul H Brookes Publishing.

Dollaghan, C. A., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1077–1088. https://doi.org/10.1044/1092-4388(2010/10-0093)

Donovan, M., & Cross, C. (2002). *Minority students in special and gifted education*. National Academies Press. https://doi.org/10.17226/10128

dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition: A Journal of Developmental Linguistics*, *25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Duncan, S. E., & De Avila, E. A. (1998). *PreLAS 2000 Cue Picture Book English Form C*. CTB/McGraw-Hill Companies, Inc.

Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children*, *35*(1), 5–22.

Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics*, *19*(2), 279–309.

Ehren, B. J., & Nelson, N. W. (2005). The responsiveness to intervention approach and language impairment. *Topics in Language Disorders*, *25*(2), 120–131. https://doi.org/10.1097/00011363-200504000-00005

Every Student Succeeds Act, Pub. L. No. Pub. L. No. 114-95 (2015).

Fabiano-Smith, L., & Goldstein, B. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research : JSLHR*, *53*, 160–178. https://doi.org/10.1044/1092-4388(2009/07-0064)

Flores, G., & The Committee on Pediatric Research. (2010). Racial and ethnic disparities in the health and health care of children. *Pediatrics*, *125*(4), e979–e1020. https://doi.org/10.1542/peds.2010-0188

García, O., Kleifgen, J. A., & Falchi, L. (2008). From English language learners to emergent bilinguals. Equity matters. Research review no. 1. *Campaign for Educational Equity, Teachers College, Columbia University*.

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

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259–272. https://doi.org/10.1007/BF03174081

Gerber, M., Jimenez, T., Leafstedt, J., Villaruz, J., Richards, C., & English, J. (2004). English reading effects of small-group intensive intervention in Spanish for K–1 English learners. *Learning Disabilities Research & Practice*, *19*(4), 239–251. https://doi.org/10.1111/j.1540-5826.2004.00109.x

Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of specific language impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1813–1823.

Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, *42*(1), 59–75. https://doi.org/10.1080/13682820600783210

Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish–English bilingual children with and without specific language impairment. *Journal of Communication Disorders*, *41*(2), 124–145. https://doi.org/10.1016/j.jcomdis.2007.07.001

Goldstein, B. A., & Fabiano, L. (2007). Assessment and intervention for bilingual children with phonological disorders. *The ASHA Leader*. https://leader.pubs.asha.org/doi/abs/10.1044/leader.FTR2.12022007.6

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis.

Journal of Speech, Language, and Hearing Research, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Graves, A. W., Brandon, R., Duesbery, L., McIntosh, A., & Pyle, N. B. (2011). The effects of tier 2 literacy instruction in sixth grade: Toward the development of a response-to-intervention model in middle school. *Learning Disability Quarterly*, *34*(1), 73–86. https://doi.org/10.1177/073194871103400105

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system (SSRS)*. American Guidance Service.

Guiberson, M., & Rodríguez, B. L. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *American Journal of Speech-Language Pathology*, *19*(3), 225–237. https://doi.org/10.1044/1058-0360(2010/09-0058)

Guiberson, M., & Rodríguez, B. L. (2013). Classification accuracy of nonword repetition when used with preschool-age Spanish-speaking children. *Language, Speech & Hearing Services in Schools*, *44*(2), 121–132. https://doi.org/10.1044/0161-1461(2012/12-0009)

Guiberson, M., & Rodríguez, B. L. (2016). Nonword repetition in Spanish-speaking toddlers with and without early language delays. *Folia Phoniatrica et Logopaedica*, *67*(5), 253–258. https://doi.org/10.1159/000442745

Guiberson, M., Rodríguez, B. L., & Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Language, Speech, and Hearing Services in Schools*, *42*(4), 536–549. https://doi.org/10.1044/0161-1461(2011/10-0076)

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language*, *30*(4), 234–245. https://doi.org/10.1055/s-0029-1241722

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English-speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice*, *25*(1), 48–58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Hall-Mills, S. (2019). A comparison of the prevalence rates of language impairment before and after response-to-intervention implementation. *Language, Speech, and Hearing Services in Schools*, *50*(4), 703–709. https://doi.org/10.1044/2019_LSHSS-18-0144

Hall-Mills, S. (2021). Shifting prevalence patterns for special educational needs in the era of response-to-intervention policy. *Frontiers in Education*, *6*. https://www.frontiersin.org/article/10.3389/feduc.2021.676646

Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, *2*. https://doi.org/10.3389/fcomm.2017.00016

Hardin, B. J., Mereoiu, M., Hung, H.-F., & Roach-Scott, M. (2009). Investigating parent and professional perspectives concerning special education services for preschool Latino children. *Early Childhood Education Journal*, *37*(2), 93–102. https://doi.org/10.1007/s10643-009-0336-x

Harrison, L. J., & McLeod, S. (2010). Risk and protective factors associated with speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language, and Hearing Research*, *53*(2), 508–529. https://doi.org/10.1044/1092-4388(2009/08-0086)

Hartung, J., & Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, *20*(24), 3875–3889. https://doi.org/10.1002/sim.1009

Heath, A., & Kilpi-Jakonen, E. (2012). *Immigrant children's age at arrival and assessment results*. *75*. https://doi.org/10.1787/5k993zsz6g7h-en

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who is placed into special education? *Sociology of Education*, *83*(4), 312–332. https://doi.org/10.1177/0038040710383518

Hibel, J., & Jasper, A. D. (2012). Delayed special education placement for learning disabilities among children of immigrants. *Social Forces*, *91*(2), 503–530. https://doi.org/10.1093/sf/sos092

Higgins, J. P. T., Barentsz, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. www.training.cochrane.org/handbook

Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, *23*(11), 1663–1682. https://doi.org/10.1002/sim.1752

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*, *327*(7414), 557–560.

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, *35*(04), 903. https://doi.org/10.1017/S0305000908008751

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education*, *84*(1), 175–196. https://doi.org/10.1080/00220973.2014.952397

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Mann, F. B., Barmer, A., Dilig, R., Nachazel, T., Barnett, M., & Purcell, S. (2020). *The condition of education 2020* (p. 348). National Center for Education Statistics.

Individuals with Disabilities Education Act, Pub. L. No. 108–446, 20 U.S.C. § 1400 (2004).

Interprofessional Education Collaborative Expert Panel. (2011). *Core competencies for interprofessional collaborative practice: Report of an expert panel*. Interprofessional Education Collaborative. https://doi.org/10.1044/persp3.SIG16.88

Justice, L. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech & Hearing Services in Schools*, *37*(4), 284–297. https://doi.org/10.1044/0161-1461(2006/033)

Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*, *51*(4), 337–347.

Kemp, S. E. (2006). Dropout policies and trends for students with and without disabilities. *Adolescence*, *41*(162), 235–250.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, *100*(4), 851–868. https://doi.org/10.1037/0022-0663.100.4.851

Klingner, J. K., & Harry, B. (2006). The special education referral and decision-making process for English language learners: Child study team meetings and placement conferences. *Teachers College Record*, *108*(11), 2247–2281. https://doi.org/10.1111/j.1467-9620.2006.00781.x

Knotek, S. (2003). Bias in problem solving and the social process of student study teams: A qualitative investigation. *The Journal of Special Education*, *37*(1), 2–14. https://doi.org/10.1177/00224669030370010101

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, *43*(6), 456–473. https://doi.org/10.1016/j.jcomdis.2010.02.002

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice*, *21*(1), 19–29. psyh. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Kummerer, S. E. (2012). Promising strategies for collaborating with Hispanic parents during family-centered speech-language intervention. *Communication Disorders Quarterly*, *33*(2), 84–95. https://doi.org/10.1177/1525740109358453

Laher, S., & Cockcroft, K. (2017). Moving from culturally biased to culturally responsive assessment practices in low-resource, multicultural settings. *Professional Psychology: Research and Practice*, *48*(2), 115–121. https://doi.org/10.1037/pro0000102

Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools*, *34*(1), 44–55. https://doi.org/10.1044/0161-1461(2003/005)

Larry P. v. Riles, C-71-2270 RFP (US District Court for the Northern District of California October 16, 1979). https://law.justia.com/cases/federal/district-courts/FSupp/495/926/2007878/.

Lazewnik, R., Creaghead, N. A., Smith, A. B., Prendeville, J.-A., Raisor-Becker, L., & Silbert, N. (2019). Identifiers of language impairment for Spanish-English dual language learners. *Language, Speech & Hearing Services in Schools*, *50*(1), 126–137. https://doi.org/10.1044/2018_LSHSS-17-0046

Lely, H. K. van der, & Howard, D. (1993). Children with specific language impairment: Linguistic impairment or short-term memory deficit? *Journal of Speech, Language, and Hearing Research*, *36*(6), 1193–1207.

Li'el, N., Williams, C., & Kane, R. (2018). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology*, *21*(6). https://doi.org/10.1080/17549507.2018.1513073

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities*, *34*(2), 136–151. https://doi.org/10.1177/002221940103400204

Linn, D., & Hemmer, L. (2011). English language learner disproportionality in special education: Implications for the scholar-practitioner. *Journal of Educational Research and Practice*, *1*(1), 14.

Loveall, S. J., Pitt, A. R., Rolfe, K. G., & Mann, J. (2022). Speech-language pathologist reading survey: Scope of practice, training, caseloads, and confidence. *Language, Speech, and Hearing Services in Schools*. https://doi.org/10.1044/2022_LSHSS-21-00135

Lugo, V. A., Wood, C., & Farquharson, K. (2022). Advocacy engagement and self-efficacy of school-based speech-language pathologists. *Language, Speech, and Hearing Services in Schools*, *53*(2), 479–493. https://doi.org/10.1044/2021_LSHSS-21-00137

Lumley, T. (2020). *Survey*. http://r-survey.r-forge.r-project.org/survey/

Macswan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teachers College Record*, *108*(11), 2304–2328. https://doi.org/10.1111/j.1467-9620.2006.00783.x

Martin Luther King Jr., ETC. v. Ann Arbor Sch. Dist., Civ. A. No. 7-71861. (US District Court for the Eastern District of Michigan August 24, 1979). https://law.justia.com/cases/federal/district-courts/FSupp/473/1371/2148458/

McGregor, K. K. (2020). How we fail children with developmental language disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

McHatton, P. A., & Correa, V. (2005). Stigma and discrimination: Perspectives from Mexican and Puerto Rican mothers of children with special needs. *Topics in Early Childhood Special Education*, *25*(3), 131–142. https://doi.org/10.1177/02711214050250030101

McInnes, M. D., Moher, D., Thombs, B. D., McGrath, T. A., Bossuyt, P. M., Clifford, T., Cohen, J. F., Deeks, J. J., Gatsonis, C., & Hooft, L. (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *Jama*, *319*(4), 388–396.

McKenna, M., Castillo, J., Dedrick, R. F., Ke Cheng, & Goldstein, H. (2021). Speech-language pathologist involvement in multi-tiered system of supports questionnaire: Advances in interprofessional practice. *Language, Speech & Hearing Services in Schools*, *52*, 597–611. https://doi.org/10.1044/2020_LSHSS-20-00084

McMaster, K. L., Kung, S.-H., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A "tier 1" approach to promoting English learners' response to intervention. *Exceptional Children*, *74*(2), 194–214. https://doi.org/10.1177/001440290807400204

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3.

Montgomery, J., Magimairaj, B., & Finney, M. (2009). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *19*, 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Morgan, G. P., Restrepo, M. A., & Auza, A. (2013). Comparison of Spanish morphology in monolingual and Spanish–English bilingual children with and without language impairment. *Bilingualism: Language and Cognition*, *16*(03), 578–596. https://doi.org/10.1017/S1366728912000697

Morgan, P. L., Farkas, G., Cook, M., Strassfeld, N. M., Hillemeier, M. M., Pun, W. H., Wang, Y., & Schussler, D. L. (2018). Are Hispanic, Asian, Native American, or language-minority children overrepresented in special education? *Exceptional Children*, *84*(3), 261–279. https://doi.org/10.1177/0014402917748303

Morgan, P. L., Farkas, G., Hillemeier, M. M., Li, H., Pun, W. H., & Cook, M. (2017). Cross-cohort evidence of disparities in service receipt for speech or language impairments. *Exceptional Children*, *84*(1), 27–41. https://doi.org/10.1177/0014402917718341

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, *44*(5), 278–292. https://doi.org/10.3102/0013189X15591157

Morgan, P. L., Frisco, M. L., Farkas2, G., & Hibel, J. (2017). Republication of "A propensity score matching analysis of the effects of special education services." *The Journal of Special Education*, *50*(4), 197–214. https://doi.org/10.1177/0022466916686105

Najarian, M., Nord, C., Wallner-Allen, K., Vaden-Kiernan, N., & Mulligan, G. M. (2020). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11: Third-Grade, Fourth-Grade, and Fifth-Grade Psychometric Report* (p. 304).

National Center for Education Statistics. (2021). *The condition of education 2021*. https://nces.ed.gov/programs/coe/

National Research Council. (2011). *Allocating federal funds for state programs for English language learners*. National Academies Press.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

O'Connor, R. E., Bocian, K. M., Beach, K. D., Sanchez, V., & Flynn, L. J. (2013). Special education in a 4-year Response to Intervention (RtI) environment: Characteristics of students with learning disability and grade of identification. *Learning Disabilities Research & Practice (Wiley-Blackwell)*, *28*(3), 98–112. https://doi.org/10.1111/ldrp.12013

O'Connor, R. E., Bocian, K. M., Sanchez, V., & Beach, K. D. (2014). Access to a responsiveness to intervention model: Does beginning intervention in kindergarten matter? *Journal of Learning Disabilities*, *47*(4), 307–328. https://doi.org/10.1177/0022219412459354

O'Donnell, V. L., & Tobbell, J. (2007). The transition of adult students to higher education: Legitimate peripheral participation in a community of practice? *Adult Education Quarterly*, *57*(4), 312–328.

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology*, *28*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Ortiz, A. A., Robertson, P. M., Wilkinson, C. Y., Liu, Y.-J., McGhee, B. D., & Kushner, M. I. (2011). The role of bilingual education teachers in preventing inappropriate referrals of ELLs to special education: Implications for response to intervention. *Bilingual Research Journal*, *34*(3), 316–333. eric.

Ortiz, J. A. (2021a). Using nonword repetition to identify language impairment in bilingual children: A meta-analysis of diagnostic accuracy. *American Journal of Speech-Language Pathology*, *30*(5), 2275–2295.

Ortiz, J. A. (2021b). Nonword repetition in bilingual assessment: A guide to clinical application. *Perspectives of the ASHA Special Interest Groups*, *6*(6), 1667–1677. https://doi.org/10.1044/2021_PERSP-21-00165

Paradis, J. (2017). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism: Language and Cognition*, *20*(1), 27–28. https://doi.org/10.1017/S136672891600033X

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*(6), 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning* (2nd ed.). Paul H. Brookes Publishing Co., Inc.

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*(3), 971–981. eric. https://doi.org/10.1044/1092-4388(2012/12-0050)

Patton, K., & Parker, M. (2017). Teacher education communities of practice: More than a culture of collaboration. *Teaching and Teacher Education*, *67*, 351–360. https://doi.org/10.1016/j.tate.2017.06.013

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, *51*(2), 192–202.

Peña, E. D., Gillam, R. B., Bedore, L. M., & Bohman, T. M. (2011). Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. *American Journal of Speech-Language Pathology*, *20*(4), 302–314.

Peña, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, *10*(2), 138–154. https://doi.org/10.1044/1058-0360(2001/014)

Peng, P., & Fuchs, D. (2016). A meta-analysis of working memory deficits in children with learning difficulties: Is there a difference between verbal domain and numerical domain? *Journal of Learning Disabilities*, *49*(1), 3–20. https://doi.org/10.1177/0022219414521667

Peske, H. G., & Haycock, K. (2006). Teaching inequality: How poor and minority students are shortchanged on teacher quality: A report and recommendations by the education trust. *Education Trust*.

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. *Journal of Speech, Language, and Hearing Research*, *60*(4), 983–998.

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research : JSLHR*, *52*(5), 1175–1188. https://doi.org/10.1044/1092-4388(2009/08-0024)

Pfeiffer, D. L., Pavelko, S. L., Hahs-Vaughn, D. L., & Dudding, C. C. (2019). A national survey of speech-language pathologists' engagement in interprofessional collaborative practice in schools: Identifying predictive factors and barriers to implementation. *Language, Speech, and Hearing Services in Schools*, *50*(4), 639–655. https://doi.org/10.1044/2019_LSHSS-18-0100

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Pratt, A. S., Adams, A. M., Peña, E. D., & Bedore, L. M. (2022). Exploring the use of parent and teacher questionnaires to screen for language disorders in bilingual children. *Topics in Early Childhood Special Education*, *42*(1), 77–90. https://doi.org/10.1177/0271121420942308

Pua, E. P. K., Lee, M. L. C., & Rickard Liow, S. J. (2017). Screening bilingual preschoolers for language difficulties: Utility of teacher and parent reports. *Journal of Speech, Language, and Hearing Research*, *60*(4), 950–968.

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J. P. M., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, *58*(10), 982–990. https://doi.org/10.1016/j.jclinepi.2005.02.022

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1398–1411. https://doi.org/10.1044/jslhr.4106.1398

Restrepo, M. A., & Silverman, S. W. (2001). Validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology*, *10*(4), 382–393. https://doi.org/10.1044/1058-0360(2001/032)

Robinson, G. C., & Norton, P. C. (2019). A decade of disproportionality: A state-level analysis of African American students enrolled in the primary disability category of speech or language impairment. *Language, Speech, and Hearing Services in Schools*, *50*(2), 267–282. https://doi.org/10.1044/2018_LSHSS-17-0149

Rodriguez, D., Carrasquillo, A., Garcia, E., & Howitt, D. (2020). Factors that challenge English learners and increase their dropout rates: Recommendations from the field. *International Journal of Bilingual Education and Bilingualism*, 1–17. https://doi.org/10.1080/13670050.2020.1722059

Rosa-Lugo, L., Mihai, F., & Nutta, J. (2017). Preparation of speech-language pathologists to work with English learners (ELs): Incorporating interprofessional education (IPE) and interprofessional collaborative practice (IPP) competencies. *Perspectives of the ASHA Special Interest Groups*, *2*, 103. https://doi.org/10.1044/persp2.SIG14.103

Roseberry-McKibbin, C., & O'Hanlon, L. (2005). Nonbiased assessment of English language learners: A tutorial. *Communication Disorders Quarterly*, *26*(3), 178–185. https://doi.org/10.1177/15257401050260030601

Ruiz, M. I. (2020). Beyond traditional response to intervention: Helping rural educators understand English learners' needs. *Rural Special Education Quarterly*, *39*(1), 35–53. https://doi.org/10.1177/8756870519894661

Rutjes, A. W. S., Reitsma, J. B., Vandenbroucke, J. P., Glas, A. S., & Bossuyt, P. M. M. (2005). Case–control and two-gate designs in diagnostic accuracy studies. *Clinical Chemistry*, *51*(8), 1335–1341. https://doi.org/10.1373/clinchem.2005.048595

Saenz, T. I., & Huer, M. B. (2003). Testing strategies involving least biased language assessment of bilingual children. *Communication Disorders Quarterly*, *24*(4), 184–193.

Samson, J. F., & Lesaux, N. K. (2009). Language-minority learners in special education: Rates and predictors of identification for services. *Journal of Learning Disabilities*, *42*(2), 148–162. eric.

Sanatullova-Allison, E., & Robison-Young, V. A. (2016). Overrepresentation: An overview of the issues surrounding the identification of English language learners with learning disabilities. *International Journal of Special Education*, *31*(2). http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1111073&site=ehost-live

Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandroni, R., & Faldella, G. (2010). Does preterm birth increase a child's risk for language impairment? *Early Human Development*, *86*(12), 765–772. https://doi.org/10.1016/j.earlhumdev.2010.08.014

Schuchardt, K., Maehler, C., & Hasselhorn, M. (2008). Working memory deficits in children with specific learning disorders. *Journal of Learning Disabilities*, *41*(6), 514–523. https://doi.org/10.1177/0022219408317856

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, *7*(3), 40–45.

Schwarzer, G., Antes, G., & Schumacher, M. (2007). A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine*, *26*(4), 721–733. https://doi.org/10.1002/sim.2588

Shakir, F. (2018). Proposed delay of Equity in IDEA regulations. *American Civil Liberties Union*, 7.

Shenoy, S. (2014). Assessment tools to differentiate between language differences and disorders in English language learners. *Berkeley Review of Education*, *5*(1), 33–52. eric.

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Andrew Garbacz, S., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of conjoint behavioral consultation and the mediating role of the parent–teacher relationship. *School Psychology Review*, *41*(1), 23–46. https://doi.org/10.1080/02796015.2012.12087374

Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of Learning Disabilities*, *44*(3), 246–257. https://doi.org/10.1177/0022219410374236

Sidik, K., & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine*, *21*(21), 3153–3159. https://doi.org/10.1002/sim.1262

Silva, M. R., Collier-Meek, M. A., Codding, R. S., Kleinert, W. L., & Feinberg, A. (2021). Data collection and analysis in response-to-intervention: A survey of school psychologists. *Contemporary School Psychology*, *25*(4), 554–571. https://doi.org/10.1007/s40688-020-00280-2

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Skiba, R. J. (2002). Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies*, 18.

Skiba, R. J., Artiles, A. J., Kozleski, E. B., Losen, D. J., & Harry, E. G. (2016). Risks and consequences of oversimplifying educational inequities: A response to Morgan et al.(2015). *Educational Researcher*, *45*(3), 221–225.

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85–107. https://doi.org/10.1080/02796015.2011.12087730

Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C.-G. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children*, *74*(3), 264–288. https://doi.org/10.1177/001440290807400301

Slama, R. B. (2014). Investigating whether and when English learners are reclassified into mainstream classrooms in the United States: A discrete-time survival analysis. *American Educational Research Journal*, *51*(2), 220–252. https://doi.org/10.3102/0002831214528277

Sullivan, A. L. (2010). Preventing disproportionality: A framework for culturally responsive assessment. *Communique*, *39*(3).

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, *77*(3), 317–334. https://doi.org/10.1177/001440291107700304

Sullivan, A. L., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children*, *79*(4), 475–494. https://doi.org/10.1177/001440291307900406

Sullivan, A. L., & Long, L. (2010). Examining the changing landscape of school psychology practice: A survey of school-based practitioners regarding Response to Intervention. *Psychology in the Schools*, *47*(10), 1059–1070. https://doi.org/10.1002/pits.20524

Tambyraja, S. R. (2020). Facilitating parental involvement in speech therapy for children with speech sound disorders: A survey of speech-language pathologists' practices, perspectives, and strategies. *American Journal of Speech-Language Pathology*, *29*(4), 1987–1996. https://doi.org/10.1044/2020_AJSLP-19-00071

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16. psyh. https://doi.org/10.1016/j.jcomdis.2012.08.002

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

Tourangeau, K., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., Najarian, M., & Mulligan, G. M. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011): User's Manual for the ECLS-K:2011 Kindergarten–Fifth Grade Data File and Electronic Codebook, Public Version*. 347.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, *53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

Umansky, I. M., & Dumont, H. (2021). English learner labeling: How English learner classification in kindergarten shapes teacher perceptions of student skills and the moderating role of bilingual instructional settings. *American Educational Research Journal*, *58*(5), 993–1031. https://doi.org/10.3102/0002831221997571

Umansky, I. M., Thompson, K. D., & Díaz, G. (2017). Using an ever–English learner framework to examine disproportionality in special education. *Exceptional Children*, *84*(1), 76–96. https://doi.org/10.1177/0014402917707470

US Department of Education. (2016a). Assistance to states for the education of children with disabilities; preschool grants for children with disabilities. *Federal Register*, *81*(41), 10968–10996.

US Department of Education. (2016b). *Racial and Ethnic Disparities in Special Education: A Multi-Year Disproportionality Analysis by State, Analysis Category, and Race/Ethnicity*. US Department of Education Washington, DC, USA.

Vanderwood, M. L., Tung, C., & Arellano, E. (2014). Application of a multitiered system of support with English language learners. *International Journal of School & Educational Psychology*, *2*(1), 45–53.

Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., Cardenas-Hagan, E., & Francis, D. J. (2006). Effectiveness of Spanish Intervention for First-Grade English Language Learners at Risk for Reading Difficulties. *Journal of Learning Disabilities*, *39*(1), 56–73. https://doi.org/10.1177/00222194060390010601

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal*, *107*(2), 153–180. https://doi.org/10.1086/510653

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, *36*(1), 1–48. https://doi.org/10.18637/jss.v036.i03

Wagner, M., Friend, M., Bursuck, W. D., Kutash, K., Duchnowski, A. J., Sumi, W. C., & Epstein, M. H. (2006). Educating students with emotional disturbances: A national perspective on school programs and services. *Journal of Emotional and Behavioral Disorders*, *14*(1), 12–30. https://doi.org/10.1177/10634266060140010201

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard business press.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, C. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, *19*(4), 298–310. psyh. https://doi.org/10.1044/1058-0360(2010/09-0064)

Wolfe, K., & Durán, L. K. 2. (2013). Culturally and linguistically diverse parents' perceptions of the IEP process. *Multiple Voices for Ethnically Diverse Exceptional Learners*, *13*(2), 4–18.

Woodcock, R. W., McGew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Riverside Publishing.

World Health Organization. (2010). *Framework for action on interprofessional education and collaborative practice*. World Health Organization.

Yamasaki, B. L., & Luk, G. (2018). Eligibility for special education in elementary school: The role of diverse language experiences. *Language, Speech, and Hearing Services in Schools*, *49*(4), 889–901. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0006

Yavas, M., & Goldstein, B. (1998). Phonological assessment and treatment of bilingual speakers. *American Journal of Speech-Language Pathology*, *7*(2), 49–60. https://doi.org/10.1044/1058-0360.0702.49

Ysseldyke, J. E., Vanderwood, M. L., & Shriner, J. (1997). Changes over the past decade in special education referral to placement probability: An incredibly reliable practice. *Diagnostique*, *23*(1), 193–201.