Technical Manual

Version 2.0

Lillian Durán, Ph.D. Mónica Zegers, Ph.D.
Julian M. Siebert, Ph.D. Cengiz Zopluoglu, Ph.D.
Heather Murphy, M.Ed. Phaedra Bell, Ph.D. Francesca Pei, Ph.D.
Maria Luisa Gorno-Tempini, M.D. Ph.D.

2026-04-14

Contents

Executive Summary

Multitudes is funded by the State of California with the primary goal of creating culturally and linguistically responsive measures that are both suitable for use and reflective of the diverse population of children across the state. The Multitudes digital platform includes a universal screener, additional follow-up assessments, professional development resources, and intervention suggestions— all available at no cost to California public schools. The goal of the screening tasks is to identify students in grades kindergarten through grade 2 who may develop reading problems, including dyslexia. The tasks were developed, calibrated, and validated over a three year period, from spring of 2022 to spring of 2025, with over 5,000 children across the state. The battery of assessments is available in Spanish and English and covers the domains of language, phonological awareness, alphabetic knowledge, reading and spelling, processing speed/automaticity, auditory short-term memory, and visuo-spatial processing.

The administration of the Multitudes tasks follows a two-step process. The first step is a short screening battery lasting ten to fifteen minutes, and consisting of three to four tasks found to be the most predictive of reading performance in each grade. Additional targeted assessment tasks are available to further explore children's strengths and needs across learning domains. These tasks provide actionable information for more effective classroom and individualized support strategies.

The screening is administered in a one-on-one format via two linked digital devices, with results immediately available on the administrator's dashboard.

Acknowledgements

The Multitudes project is a collaborative effort, made possible by the dedication and expertise of many individuals. We want to extend our sincere thanks to the large, multidisciplinary team for their invaluable contributions: Gabriella Parham-Cruzado, Javier Jasso, Amy Pratt, Melissa Brown, Minerva Meese for measure and intervention development. Mahalakshmi Ramamurthy and Jason Yeatman for visuo-spatial measure development. Hugh Catts, Yaacov Petscher, Jason Yeatman, John Gabrieli, Fumiko Hoeft, Nuria Gutiérrez, Julie Washington, Maryanne Wolf, Brandy Gatlin Nash, Eric Falke, Rebecca Silverman, Ashley Sanabria, Benjamin Domingue, and Aaron Scheffler for guidance and feedback. Phaedra Bell and Andrea Lynn Hartsough for leadership in partnership building, operations, and insight into diversity, inclusion, and representation; Meredith Kalmes, Jessica Clyne, Lucienne Vintaer, and Melina Flores for implementation and operational support. Elise Brenner, Joe Hess and the Vynyl team for platform development and technical support; Jeremy Lichtmacher, Chris Campbell, Anson Wong, Regina Juarez, Jesus Alfaro, Karla Santamaria, Nuvia Soto and Gabrielle Fall for task development technical support.

We are deeply grateful to our many partner schools, whose collaboration and participation were vital for the successful implementation and testing of the Multitudes platform. Teachers, children, families, school administrators, local education agencies and communities participated generously not only in the implementation, but also as active co-designers. We appreciate the efforts of all proctors from around the State of California who assisted with data collection, partnership building and operations. We also thank the dedicated members of our Advisory Council and Intervention Task Force.

We would also like to thank collaborators who contributed specifically to the Technical Manual: Lucy Yan for data analysis and presentation; Mahalakshmi Ramamurthy and Jason Yeatman for visuo-spatial measure descriptions; Benjamin W. Domingue for statistical support.

Copyrights

Part I

Introduction

1 Overview

The Multitudes literacy screening assessment and platform was developed by a multidisciplinary team within an academic non-profit context at the University of California, San Francisco (UCSF), in collaboration with multiple universities and research centers. All items on the 32 new parallel assessments available in English and Spanish were uniquely designed, tested, and incorporated into the platform by this team over three years, with funding from the California state legislature. The Spanish and English measures attend to the salient features of each language as well as cultural relevance, promoting fairness in testing and acknowledging the influence of cultural and language on assessment performance.

The development and norming process was conducted with more than 5,000 kindergarten, first, and second-grade students across California from 2022 to 2024. Given our unique affiliation within an academic research institution, the Multitudes screener is not a static tool–as more data is collected, Multitudes continues to be updated, refined, and improved. This technical manual describes the development process and technical adequacy of all the measures included in the current Multitudes suite of assessments.

2 Mission and Vision

2.1 Mission

Our mission is to bring the latest insights from neuroscience into the classroom by creating literacy assessments that are accurate, fair, and inclusive. We design tools that reflect the cultural and linguistic diversity of the communities they serve, helping educators identify learning needs early and support every child in becoming a confident reader.

2.2 Vision

We envision a future in which every child, no matter their language or background, has the chance to build strong reading skills from the start. By bridging science and education, we aim to reduce barriers to learning, give teachers the resources they need, and ensure all students have the opportunity to succeed in school and beyond.

3 Fairness in Testing

The United States has a diverse population of children in grades K-2, and assessment developers should attend to and measure how equitably and accurately children's skills are estimated across populations. The Multitudes suite of measures was designed with fairness as a core design and development principle. In 2014, the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014) added a chapter dedicated to fairness in testing. The chapter presented four views of fairness: (a) lack of bias, (b) equitable treatment in the testing process, (c) equality in outcomes of testing, and (d) opportunity to learn. The Multitudes development team was committed to integrating these concepts of fairness into all aspects of the assessment development process. The following section describes how we attended to each.

3.1 Lack of Bias

The large corpora of words and images selected for each measure were reviewed by an internal diversity and equity committee as well as teachers and administrators from districts across the state. We removed or revised items that lacked broad cultural resonance or that were found to be linguistically inappropriate. We intentionally included broad racial and ethnic representation in the images selected for the measures. Importantly, we also analyzed classification accuracy across populations to check for bias in identification of risk (see the section titled "Universal Screening").

3.2 Equitable Treatment in the Testing Process

The Multitudes team developed an extensive administration manual that outlines administration and scoring directions for each measure to support standardization. Student-facing directions for each measure are recorded and embedded in the assessment platform to support standardized administration of the tests. Feedback is provided via the platform to encourage engagement. Receptive measures are scored automatically through the platform, thus reducing testing bias and administrator interpretation (e.g., Elision Receptive, Listening Comprehension, Spelling, Semantic Mapping).

We provide significant guidance for how to score verbal responses provided in African American/Black English (AAE) to reduce the likelihood that Black children will be unfairly evaluated. AAE follows regular phonological, semantic, and morpho-syntactical patterns specified and included in our scoring schemes (J. A. Washington and Seidenberg 2021). For more, please see the UCSF Multitudes Administrator Guides for Language Variation.

The administration manual also provides comprehensive guidance about how to score multilingual children's expressive responses and, if administering measures in English, describes how English language production might be influenced by other language(s) to which a child is exposed. Any child with emerging English proficiency may speak English in ways influenced by their heritage language; some sounds, words, and grammatical elements might transfer, and some might result in cross-linguistic interference. Specific recommendations are provided for how to score Arabic, Mandarin, Tagalog, and Vietnamese-influenced English. Likewise, a child's Spanish language production will be reflective of regional dialects. Regional variation is natural and expected, and responses should be scored accordingly. For example, in the Spanish expressive vocabulary measure many items have multiple correct responses as different words may be used by individuals from various countries or even areas of the same country who speak Spanish. For more, please see the UCSF Multitudes Administrator Guides for Language Variation.

All children deserve to be comfortable and supported during the assessment process. There is considerable attention given in the administration manual to creating a testing environment and administrator-child interactions that respect the child's developmental level and their need for security. The manual provides specific recommendations for warm-up activities that can be incorporated into the testing process to build rapport, verbal prompts that can promote a child's engagement, and guidance about what to do if a child expresses distress or disinterest.

3.3 Equality in Outcomes of Testing

The testing of historically marginalized children and emergent bilinguals with assessments that were not designed to accurately reflect their abilities is a concern (Randall et al. 2023); (Solano-Flores 2023). The accuracy of universal screening identification was compared across key groups to ensure that these populations would not be over- or under-identified (see the Section titled "Universal Screening"). UCSF Multitudes was designed to elevate the skills and abilities that these children bring to the classroom through an asset-based lens and a development process that centered on the lived experiences of these communities and the linguistic assets they bring in their native language(s).

3.4 Opportunity to Learn

Although we cannot control for the varied educational experiences of all children who take the Multitudes screener, we can provide evidence-based resources to promote the delivery of effective instruction. There is a critical need for professional development, coaching, and ongoing technical support to ensure that the performance of each child is accurately and fairly interpreted and that the next steps taken are appropriate and effective. As such, the Multitudes platform includes instructional recommendations and professional development opportunities related to administration and instruction. Instructional recommendations come with a menu of student learning activities in English and Spanish aligned with reading development goals identified through screening.

In summary, fairness in assessment is now considered to be integral to arguments about validity (Sireci and Randall 2021); (Solano-Flores 2023). The rich diversity in the United States requires a

measurement approach that is culturally and linguistically responsive and aims to reduce bias when estimating children's abilities. Importantly, absolute fairness in assessment is impossible to achieve, because no measurement instrument demonstrates perfect reliability and any validity judgment is a matter of context and degree. Given this reality, Multitudes adopts a growth mindset, with plans for ongoing research to continually improve fairness and accuracy in reading assessment, ultimately with the mission and vision of contributing to enhanced equity in educational outcomes across all populations.

4 Multitudes Development Participants

Each stage of development including piloting, calibration, validation, and prediction modeling drew on a sample of participants from schools across the state. Specific sample populations specific to each measure and analysis are detailed accordingly in subsequent chapters. Table 4.1 below shows a snapshot of all children who took part in the development of UCSF Multitudes. The sample is representative of the diveristy in the student population in California.

Table 4.1: Participants in Multitudes Research (N = 5544)

Characteristic	n	Percentage
Gender		
Male	2,292	41.3
Female	2,711	48.9
Unknown	541	9.8
Race		
Hispanic or Latino	3,593	64.8
Non-Hispanic Black	261	4.7
Non-Hispanic Asian	235	4.2
Non-Hispanic Filipino	98	1.8
Non-Hispanic White	515	9.3
Non-Hispanic American Indian or Alaska Native	30	0.5
Non-Hispanic Pacific Islander	36	0.6
Multiple	94	1.7
Intentionally Left Blank	42	0.8
Unknown	640	11.5
English Proficiency		
English Only	2,062	37.2
English Learner	2,336	42.1
IFEP or RFEP	412	7.4
Unknown	734	13.2
IEP/504[a]		
Yes	338	6.1
No	2,560	46.2
Unknown	2,646	47.7

[a]Participants that have ever had an Individualized Education Program or Section 504 plan.

5 Multitudes Screening Sequence

Multitudes is designed to identify students at risk of reading difficulties and to provide valuable information for instructional and intervention planning. For this purpose, we describe a two-step process for screening. The first step involves a carefully selected set of measures that have been found to most accurately identify students who require additional instructional support. The second includes a battery of follow up assessments that can further identify and pinpoint a child's instructional needs.

5.1 Step 1: Universal Screening

The set of measures included in universal screening was selected following empirical analyses exploring the sensitivity and specificity of identification (see the Section titled "Universal Screening"). The battery in each grade takes about 10 minutes and is designed for one-on-one administration between a teacher and child using linked devices.

- **Kindergarten**: Letter Naming Fluency (LNF), Expressive Vocabulary (EVO), Rapid Automatized Naming-Objects (RANO), and Elision-Receptive (ELIR)
- **Grade 1**: Letter Sound Fluency (LSF), Expressive Vocabulary (EVO), Rapid Automatized Naming-Objects (RANO), and Word Reading (WRE)
- **Grade 2**: Spelling (SPE), Expressive Vocabulary (EVO), Rapid Automatized Naming-Letters (RANL), and Word Reading (WRE)

5.2 Step 2: Follow-up Measures

The follow-up measures were designed to be instructionally relevant, supporting teachers and other educational professionals in developing student profiles that identify specific strengths and areas of need to guide decision-making. This is a unique feature of Multitudes that moves beyond identification of risk, to guide intervention.

For children whose universal screening scores fall within the "support needed" range, the Multitudes platform will present the follow-up step as 'recommended' (highlighted in red on Table 1.2). It is important to note that any child may take these additional measures, not just those identified as needing support. In fact, we encourage teachers and other educational professionals to use the available Multitudes measures at their discretion to gain a deeper understanding of their students' ability levels Table 5.1.

Table 5.1: Measures by Grade Level

	Universal screener			Follow-up and Additional		
Measure	K	G1	G2	K	G1	G2
Expressive Vocabulary	•	•	•			
Word Reading		•	•			
Rapid Automatized Naming – Objects	•	•				(∘)
Rapid Automatized Naming – Letters			•	∘	∘	
Elision – Receptive	•				(∘)	
Letter Naming Fluency	•					
Letter Sound Fluency		•				
Spelling			•			
Sentence Repetition				∘	∘	∘
Listening Comprehension				∘	(∘)	(∘)
Narrative Story Production				(∘)	(∘)	
Nonword Reading					(∘)	(∘)
Nonword Repetition				(∘)	(∘)	(∘)
Oral Reading Fluency						∘
Rapid Visual Processing Letters				(∘)	(∘)	(∘)
Rapid Visual Processing Symbols				(∘)	(∘)	(∘)
Receptive Vocabulary				(∘)	(∘)	(∘)
Semantic Mapping				(∘)	(∘)	(∘)
Digit Span				(∘)	(∘)	(∘)
Elision – Expressive				(∘)	∘	∘

Note:
• indicates measures included in the universal screener.
∘ indicates measures administered as follow-up assessments.
(∘) indicates additional measures administered beyond the follow-up step.

6 Domains and Measures

Multitudes includes measures of domains that underlie reading development. They were chosen for their importance in predicting risk of reading difficulty, including dyslexia in English and in Spanish (Riva et al. 2021; Taran et al. 2022; Julie A. Washington et al. 2020; Wren and Herrera 2021). We realize that no single measure fully isolates a single domain, nor the neural networks subserving it, and thus have chosen a taxonomy that integrates concepts from both cognitive and educational sciences and considers the practical relevance for instruction.

The sections that follow present a description of each domain, the Multitudes tasks that measure them, and the empirical relationships to reading and risk prediction that ground and inform domain selection and task development.

6.1 Language

Language includes many subdomains, including syntax, semantics, morphology, phonology. and pragmatics. It refers to the components of spoken words that support communication, both receptive (understanding others) and expressive (sharing our thoughts). Language has been found to be strongly related to reading comprehension among monolingual speakers of English (Deacon and Kieffer 2018; M. J. Snowling and Hulme 2020; Storch and Whitehurst 2002a), particularly as children move into the upper elementary grades (Catts, Hogan, and Adlof 2005; B. R. Foorman et al. 2015; García and Cain 2014; Vellutino et al. 2007). Among Spanish-English bilingual children, English oral language skills have been found to be predictive of both English decoding and reading comprehension skills (Miller et al. 2006a; Nakamoto, Lindsey, and Manis 2006; Swanson, Xinhua Zheng, and Jerman 2009). Spanish oral language skills, like vocabulary and syntax, have also been shown to predict small, but important amounts of variance in English reading outcomes (Miller et al. 2006a; Proctor et al. 2006; Proctor, Harring, and Silverman 2017; Sun-Alperin and Wang 2009). Spanish oral language has also been found to contribute to Spanish reading comprehension in bilingual children (Nakamoto, Lindsey, and Manis 2008).

Multitudes measures language with the following five tasks:

6.1.1 Expressive Vocabulary (EVO)

Vocabulary has been defined as the body of words used in a particular language (Nagy and Scott 2000). In educational research, the measurement of vocabulary is often used to provide an estimate of an individual's language development and is one of the most important predictors of reading comprehension in both English and Spanish (Biemiller and Slonim 2001; Kieffer and Lesaux 2007; Proctor et al. 2006; Stahl and Nagy 2007). Vocabulary development is important for all children

and especially for children who are in the process of acquiring English (Mancilla-Martinez et al. 2020; Proctor et al. 2006). Therefore, measuring expressive vocabulary is foundational to identifying children who might struggle with reading comprehension in both English and Spanish. Multitudes uniquely provides a conceptually scored vocabulary measure for bilingual children,that recognizes responses in either Spanish or English.

6.1.2 Listening Comprehension (LCO)

Listening comprehension is a critical skill for reading and for academic success, as well as an excellent measure of a child's ability to understand connected text when it is presented orally. Listening comprehension tasks tap multiple domains of oral language skills, including phonology, morphology, semantics, and syntactic skills, all of which are related to reading development. Listening comprehension has been shown to be highly related to reading comprehension among monolingual English speakers (Adlof, Catts, and Lee 2010; Catts et al. 2001; Hogan, Adlof, and Alonzo 2014; Nation et al. 2004, 2010; Storch and Whitehurst 2002b). Listening comprehension in Spanish has also been found to be related to reading comprehension (Goodwin, August, and Calderon 2015; Nakamoto, Lindsey, and Manis 2008). Similar to many other components of language, among bilingual children, within-language relationships between listening and reading comprehension are strong, while cross-language relationships are weak or not found (Jeon and Yamashita 2014; Oh, Mancilla-Martinez, and Hwang 2023; Proctor et al. 2006).

6.1.3 Narrative Story Production (NSP)

Narrative story production is a naturalistic measure of storytelling and functional language use (Fiestas and Penã 2004; Heilmann et al. 2010; Uccelli and Paéz 2007). Both macrostructure (i.e., character, setting, problem, solution) and microstructure elements (i.e., morphosyntax, vocabulary) can be scored to measure a child's language. Narrative story production tasks have long been used to determine language delays and disorders. More recently, they have been used in reading screening (Petersen and Spencer 2012). Children at risk of language disorders consistently show lower performance across a range of narrative measures (Winters et al. 2022). Numerous experts have recommended using oral narratives when assessing the language abilities of bilingual children, given the natural communication involved in the task (Fiestas and Penã 2004; Uccelli and Paéz 2007). Narrating a story may also feel more familiar and comfortable for children from diverse backgrounds compared to formal language tests (Mandler 1980), as it is an opportunity to emulate the narratives produced by their families and their culture (Gutierrez-Clellen, Peña, and Quinn 1995; Melzi 2000). Narrative story production has been found to be related to reading achievement in both Spanish and English (Miller et al. 2006b; Reese et al. 2009). Bilingual children can also use both Spanish and English in their narrative response and will be scored based on their overall content and not what they produced in isolation in each language.

6.1.4 Semantic Mapping (SMT)

Semantic mapping provides a measure of semantic depth versus the breadth measured by the Expressive Vocabulary (EVO) task. Measures of semantic depth show slower growth in children

with language disorders across school age (Karla K. McGregor et al. 2013) and can be used to distinguish bilingual children with and without language difficulties (Javier Jasso et al. 2020). In such tasks, individuals identify underlying relationships between objects; this measures not only measures semantic knowledge but also underlying concept development. Concept development undergirds reading comprehension and is crucial across languages (Y.-S. G. Kim 2023).

6.1.5 Sentence Repetition (SRT)

Sentence repetition involves the ability to recall and repeat sentences of varying length and complexity. Such tasks measure syntactic skills, as well as lexical knowledge and memory (Marinis and Armon-Lotem 2015; Polišenská, Chiat, and Roy 2014). Repeating sentences requires processing, analyzing, and reconstructing abstract linguistic information (Marinis and Armon-Lotem 2015). Sentence repetition also measures verbal short-term memory (Pratt, Peña, and Bedore 2020), and the ability to retain information momentarily is essential for reading comprehension. For bilingual populations, existing studies show the importance of testing individuals in both their languages (Simon-Cereijido and Gutiérrez-Clellen 2017). While most existing studies of sentence repetition examine its utility in the context of oral language disorders, sentence repetition has been found to be related to reading in both English and Spanish in several studies (MOLL et al. 2013).

6.2 Phonological Awareness

Phonological awareness (PA) refers to the ability to perceive and manipulate sounds in language (Lonigan 2006). It includes the manipulation of sounds at different units that increase in difficulty from word to syllable to onset-rime to phoneme awareness (Anthony et al. 2011). There is much research that links PA to word reading in English (Vellutino et al. 2004; Maryanne Wolf et al. 2002; Ziegler and Goswami 2005) and in Spanish (Martínez and Goikoetxea 2019; Míguez-Álvarez, Cuevas-Alonso, and Saavedra 2021). In neuroscience literature, children with dyslexia have shown differences in auditory processing compared to typically developing readers (Qi et al. 2023).

Significant correlations have also been found between PA development in English and Spanish (Melby-Lervåg and Lervåg 2011). Phonological awareness is generally assessed in both English and Spanish with measures of alliteration (Car, tree, cat: “Which two start with the same sound?”), blending (“What is m (pause) oon?”), segmenting (“How many syllables are in the word ma-ni-pu-late?”), and elision (“What is butter without /b/?”) (Kilpatrick 2012). Elision was selected for the Multitudes suite of measures as it has demonstrated the strongest relationship with reading ability in both English and Spanish(Kilpatrick 2012).

Multitudes measures phonological awareness with the following two tasks:

6.2.1 Elision Expressive (ELIE) & Elision Receptive (ELIR)

Elision, sometimes called deletion, refers to manipulating the sounds in a word by removing them (Semel et al. 2006a). Difficulty can depend on factors such as what is removed, i.e., a syllable or a phoneme (Anthony et al. 2011, 2009), as well as the location of the removed sounds, i.e., the

beginning, middle, or end of the word (McBride-Chang 1995). For example, a relatively easier item might be, "What is baseball without ball?" whereas more difficult items might be, "What is money without /mun/?" or "What is toy without /t/?". Elision tasks in assessment batteries account for unique variance in reading abilities, explaining more variance in reading ability than blending and segmenting in English with first and second graders (Kilpatrick 2012). Elision has been found at kindergarten entrance to be predictive of later reading outcomes and/or risk for reading difficulties; this predictiveness, however, does not persist after second grade (Hogan, Catts, and Little 2005). Elision also predicts reading (identification, word attack) and spelling of nonwords (Swank and Catts 1994). Targeting elision skills through reading instruction appears to improve English reading skills, and the inverse has also been shown: elision skills improve as a result of more reading (Clayton et al. 2019).

Elision has also been used with Spanish-speaking children to measure their phonological awareness in the early elementary grades (Anthony et al. 2011). In studies of cross-linguistic transfer of literacy skills between Spanish and English, elision measures have been found to be related to both Spanish and English reading outcomes(Dickinson et al. 2004; Kremin et al. 2016; Pasquarella et al. 2015). Elision is a relatively difficult phonological awareness task in Spanish, making it a good candidate for students in early elementary up until second grade (Anthony et al. 2011). Elision has also been included in other tests of phonological awareness in Spanish (e.g., Test of Phonological Awareness in Spanish, Riccio et al. (2004a); Get Ready to Read!, Grover J. Whitehurst and Lonigan (2001)).

6.3 Alphabetic Knowledge

The acquisition of alphabetic knowledge, or knowledge of letter names and sounds, is foundational in children's early literacy development (Graver J. Whitehurst and Lonigan 1998) and recognized as the strongest predictor of later reading ability in English and Spanish (National Research Council 1998; Scarborough 1998; Schatschneider et al. 2004). Knowing letter names is related to reading performance, and children who demonstrate higher letter name knowledge tend to demonstrate higher reading acquisition (National Research Council 1998; O'Connor and Jenkins 1999; M. J. Snowling, Gallagher, and Frith 2003; Torppa et al. 2006). Knowing letter sounds is a critical and predictive skill, and children who can accurately connect graphemes to their corresponding sounds are also better decoders and readers (Piasta, Purpura, and Wagner 2009). Among bilingual children, kindergarten English letter naming fluency significantly predicted English oral reading fluency through the end of first grade (Roberts 2005; Yesil-Dagli 2011). The knowledge of letter names and letter sounds has also been found to be strongly related to reading development in Spanish (Signorini 1997).

Multitudes measures alphabetic knowledge with the following two tasks:

6.3.1 Letter Naming Fluency (LNF) & Letter Sound Fluency (LSF)

Identifying letter sounds and names remain two of the most predictive measures of Spanish and English oral reading fluency (Anthony et al. 2006; Kremin et al. 2016; Lindsey, Manis, and Bailey 2003; Pasquarella et al. 2015; Solari et al. 2013). Letter names and letter sounds have remained stable and valuable contributors over time to screening for reading difficulties in both English and

Spanish (Genesee et al. 2006; Ozernov-Palchik et al. 2017). Letter naming and letter sounds are both important predictors of reading in Spanish, and it is important to measure both (Lindsey, Manis, and Bailey 2003); there is no reason to choose one over the other. In kindergarten, children are more likely to know letter names than letter sounds based on exposure to letter names in preschool and in popular children's media. Letter names are emphasized in the United States, given their relationship to reading in English, which makes this different than the context of monolingual Spanish-speaking countries. At the same time, letter naming has also been shown to be a strong predictor of reading with pre-literate monolingual Spanish-speaking students (De la Calle 2018). (Lindsey, Manis, and Bailey 2003) also found a cross-linguistic relationship such that the knowledge of Spanish letter names in kindergarten predicted English reading fluency in first grade. The knowledge of letter sounds becomes more stable in first grade based on having a year of literacy instruction, making letter sounds a more appropriate measure for first grade rather than kindergarten.

6.4 Reading and Spelling

The Reading and Spelling domain is comprised of subdomains including decoding, reading fluency, and spelling. Decoding refers to the process of translating printed words to speech (B. Foorman 2023). This process requires that a child has accurate and fluent knowledge of letter-sound correspondences. Decoding is foundational to fluent reading and is highly correlated with reading comprehension, especially early in the process of learning to read (Ehri 2020). Decoding is central to both English and Spanish reading (Goodwin, August, and Calderon 2015). Word reading is fundamentally important to reading comprehension (Garcia et al. 2006) as word reading encompasses the ability to recognize words, whether by decoding or by having a consolidated orthographic representation, allowing for automatic recognition of that word (i.e., sight word reading). The ability to accurately and efficiently read words underpins the broader skill of understanding text. Major theories of reading comprehension include word reading as a critical skill set for both monolingual and Spanish-English bilingual children (Hoover and Gough 1990; Y.-S. G. Kim 2020; Scarborough, Neuman, and Dickinson 2001).

Oral reading fluency is the ability to read a connected text aloud with speed, accuracy, and expression. This involves not just reading the words correctly but doing so in a way that is effcient and sounds natural and conveys the meaning of the text.

Spelling, often referred to as encoding, is a skill that has been shown to be highly related to reading in monolingual (Ellis and Cataldo 1990; Treiman and Kessler 2021) and bilingual (Vettori et al. 2023) children. Caravolas and Samara (2015) identify three core skills that underpin this relationship: knowledge of the alphabet, phoneme awareness, and rapid automatized naming. The developmental sequence of writing skills acquisition has been shown to be remarkably similar in monolingual and bilingual children (Ferreiro and Teberosky 1982; Gentry 1982, 2000; Rubin and Carlan 2005).

Multitudes measures Reading & Spelling with the following four tasks:

6.4.1 Nonword Reading (NRE)

Nonword reading is a strong predictor of reading ability in both Spanish and English (Diana L. Baker, Park, and Baker 2010; Durgunoğlu, Nagy, and Hancin-Bhatt 1993; Genesee et al. 2006; Y.-S. Kim 2012; Leafstedt and Gerber 2005; Proctor et al. 2006). Nonword reading also has the benefit of removing the variable of familiarity with words to isolate the skill of decoding. Nowords are carefully constructed to capture a range of decoding ability and to adhere to the syntactical rules of the target language. Nonword reading has been found to capture the initial abilities and growth of both Spanish-and English-speaking children (Doris Luft Baker, Park, and Baker 2010; Y.-S. Kim and Pallante 2010).

6.4.2 Oral Reading Fluency (ORF)

Among monolingual English-speaking children, oral reading fluency has been found to be moderately to highly correlated with reading comprehension outcomes (Adams 1990; Cutting and Scarborough 2006; Fuchs et al. 2001; National Reading Panel (US) 2000; Reschly et al. 2009; Yeo 2009). Among bilingual children, as with monolinguals, research has found that oral reading fluency in the language of instruction is moderately to highly correlated with reading comprehension outcomes (Ives Wiley and Deno 2005; Riedel 2007), but may be moderated by oral language proficiency in the target language (Crosson and Lesaux 2009). Oral reading fluency has been found to correlate across languages in bilingual children (De Ramírez and Shapiro 2007); however, within-language relationships between foundational skills and oral reading fluency are stronger than cross-language relationships (Solari et al. 2013). Importantly, since Spanish has a more transparent orthography than English, children who are instructed in Spanish often reach a higher level of word reading accuracy at an early age (Seymour, Aro, and Erskine 2003); therefore oral reading fluency which accounts for speed and accuracy among Spanish monolinguals has been a more sensitive measure of dyslexia or other reading problems (Y.-S. Kim and Pallante 2010; Serrano and Defior 2008; Verhoeven and Keuning 2017).

6.4.3 Spelling (SPE)

Spelling has been shown to be a good diagnostic marker for detecting reading difficulty (Chua, Rickard Liow, and Yeong 2014). Spelling errors in English and Spanish have been found to be indicative of children with reading and writing problems (Serrano and Defior 2010). Spelling development follows a known sequence that can be tapped in the construction of a spelling assessment (Defior and Serrano 2005). First, children learn to recognize words based on their visual features and do not relate the sounds to the letters in the words. After this pre-reading phase, children begin to learn phoneme-to-grapheme correspondence. The second stage is further divided into three substages. The first of these is the "semiphonetic" stage, where children initially demonstrate certain understandings of the relations between spelling and the sounds of words. In the second substage, "phonetics," children develop the ability to segment words and represent all the sounds they hear in a word. In the final substage, "transitional," children follow certain spelling conventions but are still acquiring irregular or exceptional words. In the third stage, once the code is mastered, spelling is orthographically correct, and children have mastered orthographically irregular words (Ellis 1994).

Understanding these stages and issues, such as the spelling of diphthongs and consonant clusters (Serrano and Defior 2010) facilitated the careful curation of a corpus of words and tasks in the Multitudes spelling test that map onto these stages.

6.4.4 Word Reading (WRE)

Word reading in Spanish and English has been found to be positively correlated in bilingual children (Durgunoğlu, Nagy, and Hancin-Bhatt 1993; Gottardo 2002), but studies show that the magnitude of correlation between L1 and L2 word reading skills depends on several factors, including the degree of similarity between scripts (Geva and Siegel 2000; Melby-Lervåg and Lervåg 2011) and the instructional context in which children learn these skills (Gottardo, Chen, and Huo 2021). Word reading skills have been shown to be predictive of reading comprehension outcomes within languages for both English monolinguals and Spanish-English bilinguals (Gottardo and Mueller 2009; Silverman et al. 2015).

6.5 Processing Speed | Automaticity

Rapid Automatized Naming (RAN) measures processing speed and how quickly an individual can label a repeating set of familiar stimuli (Denckla and Rudel 1974; Maryanne Wolf and Bowers 1999). RAN mimics the skills needed for automatic reading; both processes, reading and rapid naming, require serial processing of visual information with visual or orthographic representations, access to phonological representations or labels, and oral articulation of the stimuli that are presented visually (Georgiou and Parrila 2020; Georgiou et al. 2013). Furthermore, both processes require similar fluency and integration skills (Kirby et al. 2008). RAN taps into a language-universal cognitive mechanism involved in reading alphabetic orthographies that is independent of complexity and is a good predictor of reading across alphabetic languages (Landerl et al. 2018). Learning to read can enhance the automaticity in retrieval and labeling, with particular benefit for more difficult lexical items (Araújo and Faísca 2019); therefore, more proficient readers become more efficient labelers. Extensive research suggests that low naming speed is a characteristic of poor readers or individuals with dyslexia (Denckla and Rudel 1976; Heikkilä et al. 2009; Willburger et al. 2008; Maryanne Wolf et al. 2002) and that RAN tasks can be a particularly useful task to identify risk of dyslexia in languages other than English (Kirby et al. 2010).

Multitudes measures processing speed | automaticity with the following two tasks:

6.5.1 Rapid Automatized Naming – Letters (RANL)

Rapid Automatized Naming – Letters (RANL) is a strong predictor of reading and spelling (Chen et al. 2021). Research shows that rapid naming of letter tasks successfully predict reading ability beyond kindergarten and are more strongly related to future reading performance in English-speaking children compared to non-alphanumeric RAN tasks (McWeeny et al. 2022). Importantly, assessing rapid naming of letters has been found to make a significant, unique contribution to reading prediction beyond phonological awareness (Katzir et al. 2006; McWeeny et al. 2022).

Neuroimaging studies have demonstrated that rapid naming of letters activates key components of the reading network, including the angular gyrus, superior parietal lobule, and medial extrastriate areas, making it a more informative measure for assessing and predicting reading skills, especially as children advance in their literacy development (Misra et al. 2004). Rapid letter naming shows unique regions of activation over rapid naming of objects, particularly in semantic and articulatory regions (Cummine et al. 2014).

6.5.2 Rapid Automatized Naming – Objects (RANO)

Rapid Automatized Naming – Objects (RANO) can be a useful predictor of reading ability in the early stages of literacy development. Neurologically, rapid naming of objects activates similar brain areas as reading, including motor planning (e.g., cerebellum), semantic access (middle temporal gyrus), articulation (supplementary motor association, motor/pre-motor, anterior cingulate cortex), and grapheme–phoneme mapping (ventral supramarginal gyrus) (Cummine et al. 2014). This task is often used in kindergarten assessments to avoid biasing results against children with limited alphabetic knowledge (Norton and Wolf 2012). However, the predictive power of rapid object naming tends to decrease after kindergarten, once children become more familiar with letter names (Misra et al. 2004).

6.6 Auditory Short-Term Memory

Auditory short-term memory (ASTM) is defined as, "[the] capacity for temporarily maintaining verbal information when the external stimulus is no longer available to sensory systems" (Yue and Martin 2021, 72). This skill is often assessed with digit span and nonword repetition.

It is unclear what effect ASTM has on reading acquisition. While some studies suggest that ASTM does not predict reading when other skills (e.g., phonological awareness and naming speed) are considered (McDougall et al. 1994; Parrila, Kirby, and McQuarrie 2004), others propose that it is a direct predictor of early word-level reading from ages 4 to 6 (Cunningham et al. 2020). There is more agreement in the literature that ASTM is typically reduced in children and adults with dyslexia (Brady 1986; Brady, Shankweiler, and Mann 1983; Majerus and Cowan 2016; M. Snowling et al. 1997; Swanson and Siegel 2011), especially serial order Short Term Memory (STM) impairment.

Multitudes measures auditory short-term memory with the following two tasks:

6.6.1 Digit Span (DGS)

Digit span is a measure of working memory and auditory short-term memory. Digit span is a brief store of acoustic information as children are read a series of numbers in increasing length and are asked to repeat the string of numbers accurately. Performance on digit span tasks has been found to distinguish children with dyslexia from children with general learning disabilities, with lower span scores in children with dyslexia (J. Torgesen and Goldman 1977; J. K. Torgesen and Houck 1980). Among several models of working memory, the classical Multicomponent Model of Working Memory (A. Baddeley 2003; A. D. Baddeley and Hitch 1974) has proven useful in understanding

the role of working memory in reading and writing. The model consists of the "central executive'' and its two component systems: the "phonological loop'' and the "visuospatial sketchpad.'' The speech-based "phonological loop'' is presented as a system that comprises a brief acoustic store and an articulatory rehearsal process (A. Baddeley 2003). The relationship of digit span tasks to reading in Spanish has not been as well studied, but digit span has been related to reading rate in Spanish (Naveh-Benjamin and Ayres 1986).

6.6.2 Nonword Repetition (NWR)

Nonword repetition is a task that requires repeating a nonsense or nonword (Gathercole et al. 1994). Nonword items can be mono- or multisyllabic but must follow the phonotactic structure of the language, since this will influence accuracy. While usually accepted as a measure of phonological working memory, nonword repetition requires multiple language processes, including language perception, phonological encoding, phonological memory, and articulation (Coady and Evans 2008).

Nonword repetition has been shown to have close developmental links with vocabulary, reading, and comprehensive language skills in children (Gathercole et al. 1994). It is considered one of the most effective predictors of language learning ability in childhood (Gathercole 2006). Nonword repetition and existing vocabulary knowledge both contribute to children's word learning, but their relative influence depends on how word learning is measured. Nonword repetition is a stronger predictor of phonological recall, phonological recognition, and semantic recognition, while vocabulary knowledge is a stronger predictor of verbal semantic recall (Adlof and Patten 2017). Moreover, nonword repetition and vocabulary development have reciprocal relationships in preschoolers, although the predictive relationship from vocabulary to nonword repetition is stronger than vice versa (Verhagen et al. 2019). Nevertheless, these tasks appear to be less influenced by children's language proficiency.

While the skills measured through nonword reading tasks are crucial for early language acquisition, they remain important for word learning across the lifespan (Gathercole 2006). However, the long-term relationship between nonword repetition and word reading is less clear: while learning to read predicts growth in nonword repetition between ages 6 and 7, nonword repetition is not a longitudinal predictor of reading growth (Nation and Hulme 2010).

6.7 Visual-spatial Processing

Visual processing deficit theories in dyslexia date back to the 1850s, when dyslexia was first conceived as a visual problem (M. Snowling et al. 1997). Ever since, many theories have been proposed in the field and have a contentious history (Hulme 1988; Pennington 2011; Stein and Walsh 1997; Vellutino et al. 2004; M. Wolf et al. 2024). Despite years of research and many theories, a fundamental methodological impediment to understanding the underlying integration of visual factors associated with dyslexia has been sample size. Hundreds of studies have tested models of various deficits that contribute to reading difficulties, but samples are not sufficiently large and diverse to discern the contribution of different risk factors, with only a couple of studies exceeding N>100(O'Brien and Yeatman 2021; Talcott et al. 2002; Valdois et al. 2021).

Multitudes measures visual-spatial processing with the following two tasks:

6.7.1 Rapid Visual Processing - Letters (RVPL) & Rapid Visual Processing – Symbols (RVPS)

Rapid visual processing is the ability to rapidly encode and recall multiple visual elements simultaneously in a brief glimpse (Bosse, Tainturier, and Valdois 2007; Valdois, Bosse, and Tainturier 2004). It involves processing visual information quickly without taxing working memory (Pelli et al. 2006; Sperling 1960). Extensive research links rapid visual processing to reading ability across various languages, including French, English, Dutch, and Chinese (Bosse, Tainturier, and Valdois 2007; Van Den Boer, Van Bergen, and Jong 2015). This skill has been shown to i) correlate with reading ability (Ramamurthy, White, and Yeatman 2024); ii) differ in children with dyslexia (Ramamurthy, White, and Yeatman 2024); iii) be independent of phonological awareness deficits; and iv) identify a subgroup of poor readers with intact phonological skills (Bosse and Valdois 2009; Lobier, Zoubrinetzky, and Valdois 2012). Rapid visual processing is assessed using two tasks: Rapid Visual Processing with Letters (RVPL) and Rapid Visual Processing with Symbols (RVPS). RVPL measures the ability to rapidly identify letters in 2- and 4-letter strings, while RVPS assesses the ability to rapidly locate and identify non-namable visual symbols, making it language-agnostic. These tasks are considered promising tools for early identification of struggling readers not captured by conventional phonological awareness measures.

Part II

Measure Development

7 Psychometric Approach

The Multitudes measures can be grouped into two categories, based on the way in which they are administered, analyzed, and scored. The measures include computerized adaptive tests (CAT) and fixed-form measures. Below, we explain the process of developing and gathering reliability and validity evidence for each type of measure.

7.1 Computerized Adaptive Testing Measures

Ten measures were developed for computerized adaptive testing (CAT; Gershon 2005; Wainer et al. 2000; Weiss 1985) using Rasch modeling to calibrate item parameters: Elision Expressive, Elision Receptive, Expressive Vocabulary, Listening Comprehension, Nonword Reading, Nonword Repetition, Semantic Mapping, Sentence Repetition, Spelling, and Word Reading. These parameters were then used to create algorithms that select the items the child encounters on the test. CAT measures increase efficiency and precision by presenting children with items at their ability level and engaging them in testing only until their performance is established, computed as a theta value. The applicable analyses for each measure are reported in the section titled "Measure Development."

7.1.1 Rasch Scaling

Rasch scaling is a widely used mathematical method to model the measurement of an unobserved (latent) trait or ability (Andrich 1988). The Rasch model for dichotomous data is often regarded as an item response theory (IRT) model with one item parameter. However, rather than being a particular IRT model, it possesses a property that distinguishes it from other IRT models. Specifically, the defining property of Rasch scaling is the principle of invariant measurement. With the invariant measurement property, different subsets of items can be administered to different children, and comparable objective scale scores can be obtained. When the outcome of Rasch scaling is combined with computerized adaptive testing, items are most informative of the child's proficiency level, and scale scores are reported in real time.

Rasch scaling has four basic assumptions that must be met (Andrich 1988):

1. **Unidimensionality**. The assumption that only one latent dimension determines how a child responds to a given item, namely the child's ability or skill level.

2. **Monotonicity**. Higher proficiency levels translate to a higher probability of responding correctly to an item.

3. **Local Independence**. Children's response to one item is conditionally independent of their responses to all other items after accounting for the latent dimension that these items measure.

4. **Invariance**. Item parameters can be obtained from any group of children, regardless of where they fall on the ability scale.

Rasch scaling allows constructing unitary scales across grades, allowing for more accurate and consistent comparisons across children and grades on the same construct. In choosing Rasch scaling as our statistical model, we focused on developing items that are evidence-based representations of each unitary construct (Wilson 2023).

7.1.2 Data Collection Design

Different grade-level forms were constructed after creating the initial pool for each task. For each grade level, we constructed six forms. Each form included a set of unique items and anchor items. The ratio of anchor items to the total number of items in any form ranged from 20% to 30%. Anchor items were purposefully spread across different forms to create enough overlap across forms within a grade as well as across grade levels. Each form was administered to at least 100 children. The form construction and data collection design allowed us to calibrate item parameters for a large number of items across grades.

7.1.3 Classical Item Analysis and Item Parameter Calibration

A classical item analysis was conducted before Rasch scaling to identify problematic items. Basic item statistics were computed for each item, such as the proportion of correct responses and point-biserial correlations. Items with point-biserial correlations lower than 0.20 were flagged and removed from the data before Rasch scaling. The Rasch model was fitted using the TAM package for R, using marginal maximum likelihood estimation (MLE) to estimate item parameters. After estimating item parameters, we estimated person parameters using the maximum likelihood estimate. In addition, we generated a Wright Map to examine the alignment between the difficulty of items and the distribution of children's proficiency. We also calculated the residuals for each person on each item based on the estimated model parameters.

7.1.4 Assessing Unidimensionality

One way to evaluate the assumption of unidimensionality after fitting a Rasch model is to calculate the proportion of variance in responses attributed to the primary latent variable (G. Jr. Engelhard and Wang 2020; Linacre 2006). In the IRT literature, it is typical to use the variance accounted for by the latent factor as evidence of unidimensionality. For instance, Reckase (1979) argued that the first factor should account for at least 20% of the total variance to produce reasonable person parameter estimates. We adopted the approach described by Linacre (2006). We first found the variance associated with the observed responses (VO) and the variance associated with residuals (VR) after fitting the model. Then, the percent of variance explained by the Rasch model can be found by (VO – VR)/ VO.

7.1.5 Item Fit Statistics

Item fit can be categorized according to the framework suggested by G. Engelhard, Wang, and Wind (2018). According to this framework, items can be put into four different categories based on their Infit and Outfit Mean-Square values, as shown in Table 7.1.

Infit means inlier-sensitive or information-weighted fit. This is more sensitive to the pattern of responses to items targeted at the child's ability level, and vice versa. Outfit, or outlier-sensitive fit, is more sensitive to responses to items with difficulty far from a child's ability, and vice versa. Mean-square (MSE) fit statistics show the size of the randomness, i.e., the amount of distortion of the measurement system, with an expected value of 1. Values less than 1.0 indicate observations are too predictable (redundancy, data overfit the model). Values greater than 1.0 indicate unpredictability (unmodeled noise, data underfit the model). In general, values near 1.0 indicate little distortion of the measurement system, regardless of the standardized value. We evaluated mean-squares high above 1.0 before mean-squares much below 1.0, because the average mean-square is usually forced to be near 1.0. Outfit problems are less of a threat to measurement than Infit issues and are easier to manage. To evaluate the fit statistics of each item, we used the thresholds outlined in Table 7.1. Items that fell into category D were removed from the final item pool.

Table 7.1: Infit-Outfit Mean-Square Values

Infit.Outfit.MSE.Value	Interpretation	Fit.Category
$0.5 < \text{MSE} < 1.5$	Productive for measurement	A
$\text{MSE} < 0.5$	Less productive, but not distorting of measures	B
$1.5 < \text{MSE} < 2$	Unproductive, but not distorting of measures	C
$\text{MSE} > 2$	Unproductive and distorting of measures	D

7.1.6 Ongoing Qualitative Item Review

In addition to evaluating item-level statistics, our team conducted a thorough review of items for cultural and linguistic relevance, representation, and appropriateness. This review was carried out by a multidisciplinary group of project members with varied cultural, linguistic, disciplinary, and lived experiences, who examined test items for potential concerns that could arise in the field or for specific student populations. We also held several focus groups with teachers and administrators to gather feedback on the test items. In addition, selected experts in the field reviewed images, wording, administration procedures, and scoring. During calibration studies, proctors recorded feedback in a shared spreadsheet while administering items, noting any concerns raised by children across the state. Proctors were also invited to internal executive meetings to share their experiences.

Our team carefully considered data from all of these sources. Items were removed not only based on poor statistical performance but also when they were found to be potentially offensive in certain communities, unfamiliar to some groups in ways that could introduce bias, ambiguous in meaning, or

reinforcing of harmful stereotypes. These extensive procedures gave the team greater confidence that the final pool of items was both psychometrically robust and culturally and linguistically appropriate for use throughout California.

7.1.7 Computerized Adaptive Testing (CAT)

Computerized Adaptive Testing (CAT) aims to construct an optimal test for each child by estimating ability after each item administration and selecting subsequent items from an item pool based on the child's estimated proficiency. (Meijer and Nering 1999; Wainer et al. 2000). CAT has several advantages over paper-and-pencil tests, including greater efficiency and precision, with scores immediately available.

After calibrating item parameters for each task, we designed a CAT algorithm to administer these tasks. There are five important decisions to make during a CAT algorithm:

How to:

- start a test session?
- estimate proficiency during the test after every item administration?
- select the next item?
- stop the test?
- estimate the proficiency at the end of the session once the test is terminated?

The number of options for each decision poses a challenge for an optimal design. Practical considerations (e.g., the number of items a kindergartener responds to without frustration) guided certain decisions. We also ran computer simulations to inform our approach. Below, we present a summary of parameters currently used in the CAT algorithm in the Multitudes platform. We continuously consider these design choices to improve our measurement.

- **Starting a CAT session.** The algorithm randomly selects one of the 30 items that give the most information about average proficiency, meaning items whose difficulty levels are relatively close to the average ability level. The session starts by administering this randomly selected item.
- **Proficiency estimation during the test.** The algorithm uses the maximum a posteriori (MAP) estimation after the first item and uses MAP estimates after every item during the test.
- **Next item selection.** The algorithm uses Maximum Fisher Information (MFI) to select the next item based on the MAP proficiency estimate from the previous items.
- **Final proficiency estimate.** The algorithm uses Maximum Likelihood Estimation (MLE) to obtain the final proficiency estimate after the test is terminated.
- **Terminating the test.** The algorithm uses a fixed test length rule to stop the test. We ran extensive computer simulations to find an optimal test length for each test. The simulations were carried out using the catR (Magis and Raîche 2012) package, which was revised to fix a seed usage issue (Cui 2020). The person parameters were generated for 1,000 simulees using the empirical PDF estimated from the distribution observed for each task. We simulated two independent CAT sessions for each simulee by manipulating the test length from six to fifteen items, with one based on the true person parameter and item parameters. From each

simulated CAT session, we saved the final person parameter estimate. Then, we calculated three outcomes:

a. The correlation between the final person parameter estimates from the first CAT session and the true person parameter estimates.

b. The correlation between the final person parameter estimates from the second CAT session and the true person parameter estimates.

c. The correlation between the final person parameter estimates from the first CAT session and the final person parameter estimates from the second CAT session

The correlation in (a) and (b) can be considered as an estimate of the validity coefficient, and the correlation in (c) can be considered as an estimate of the test-retest reliability coefficient. For each task, we find the optimal test length such that the test-retest reliability coefficient reaches 0.8 and the validity coefficient reaches 0.9 from these simulations.

7.2 Fixed Form Measures

Seven measures are administered in standard forms that do not vary and were developed using classical test theory: Digit Span, Letter Naming Fluency, Letter Sound Fluency, Narrative Story Production, Oral Reading Fluency, Rapid Automatized Naming Letters, and Rapid Automatized Naming Objects. Some were analyzed and scored using the raw number of correct responses (e.g., Digit Span). Some were timed, with the final score being the number of correct items over the total time taken to complete the task (e.g., RANL, RANO). Others followed a different scoring scheme. Basic descriptive statistics are provided for these measures, including the distribution of scores, the mean level of performance, and the standard deviation.

8 Digit Span Forward

8.1 Task Description

Children listen to a string of single-digit numbers and are asked to repeat it verbatim.

8.2 Construct

The Digit Span task measures the construct of verbal short-term memory by assessing children's ability to remember and repeat a sequence of numbers of increasing length.

8.3 Item Development

Ten items were developed. Item length ranges from 2 to 6 digits per item, and for each level of length, a total of two items were developed. The same string of items was used for the English and Spanish measures.

8.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones. This assessment has a stop rule embedded: if the child is unable to repeat correctly the two items for a span length (e.g., the two items with four numbers), then the assessment ends.

8.5 Samples

Table 8.1: Demographic Characteristics of Samples for the English and Spanish Digit Span Forward Tasks

	English			Spanish		
Characteristic	**K** N = 249	**G1** N = 329	**G2** N = 383	**K** N = 45	**G1** N = 65	**G2** N = 72
Timepoint						
Fall 2023	248 (100%)	326 (99%)	346 (91%)	45 (100%)	65 (100%)	71 (99%)
Fall 2024	1 (0.4%)	2 (0.6%)	36 (9.4%)	0 (0%)	0 (0%)	1 (1.4%)
Unknown	0	1	1			
Administration Format						
Not applicable	249 (100%)	329 (100%)	383 (100%)	45 (100%)	65 (100%)	72 (100%)
Race						
American/Alaskan Native	5 (2.0%)	5 (1.5%)	1 (0.3%)			
Asian	24 (9.8%)	35 (11%)	28 (7.8%)	0 (0%)	0 (0%)	1 (1.4%)
Black/African American	35 (14%)	37 (11%)	40 (11%)			
Not reported	34 (14%)	64 (20%)	70 (20%)	28 (62%)	58 (89%)	60 (83%)
Other	74 (30%)	46 (14%)	44 (12%)	16 (36%)	1 (1.5%)	3 (4.2%)
White	72 (30%)	140 (43%)	175 (49%)	1 (2.2%)	6 (9.2%)	8 (11%)
Unknown	5	2	25			
Ethnicity						
Hispanic/Latin(o/a)	90 (36%)	151 (46%)	176 (50%)	43 (98%)	63 (97%)	66 (92%)
Intentional nonreport	8 (3.2%)	3 (0.9%)	2 (0.6%)			
Not Hispanic/Latin(o/a)	151 (61%)	173 (53%)	172 (49%)	1 (2.3%)	2 (3.1%)	6 (8.3%)
Unknown	0	2	33	1	0	0
Gender						
Female	127 (51%)	161 (49%)	169 (47%)	24 (53%)	40 (62%)	35 (49%)
Male	122 (49%)	166 (51%)	188 (53%)	21 (47%)	25 (38%)	37 (51%)
Unknown	0	2	26			
Home Language						
English	191 (80%)	246 (76%)	276 (78%)	0 (0%)	1 (1.5%)	5 (6.9%)
Spanish	32 (13%)	55 (17%)	53 (15%)	45 (100%)	64 (98%)	67 (93%)
Other	17 (7.1%)	24 (7.4%)	24 (6.8%)	0 (0%)	0 (0%)	0 (0%)
Unknown	9	4	30			
English Proficiency Label						
(Re-)Classified Proficient	9 (4.6%)	28 (8.8%)	20 (5.6%)	4 (15%)	6 (9.5%)	6 (8.5%)
English Learner	37 (19%)	52 (16%)	54 (15%)	22 (85%)	56 (89%)	60 (85%)
English-only	148 (76%)	237 (75%)	280 (79%)	0 (0%)	1 (1.6%)	5 (7.0%)
Unknown	55	12	29	19	2	1
Ever IEP/504	11 (5.9%)	28 (11%)	29 (13%)	2 (4.8%)	5 (9.1%)	2 (6.9%)
Unknown	64	67	159	3	10	43

8.6 Score distribution

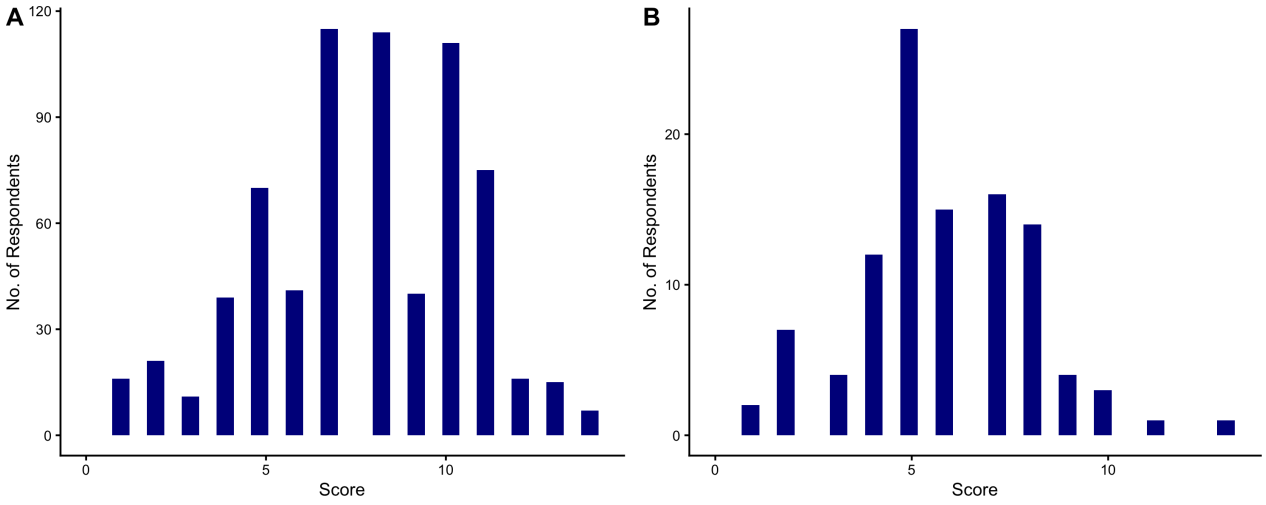

Figure 8.1: Score Distribution of the English and Spanish Digit Span Forward Tasks

8.7 Criterion Validity Evidence

Forthcoming.

9 Elision-Expressive

9.1 Task Description

Children hear a word and are prompted to take away part of it. They are then asked to say aloud what remains after the specified unit has been removed.

9.2 Construct

The Expressive Elision task measures phonological awareness and auditory manipulation skills. Students hear a word and are asked to delete a specific linguistic unit (compound word, syllable, or phoneme), then verbally produce the portion of the word or non-word that remains. This task requires accurate auditory processing, lexical retrieval, and articulation, providing insight into students' phonological processing and expressive language development.

9.3 Item Development

An original list of words was derived from curricula used in both English and dual-language programs in California. The list was later extended by researchers by selecting words that fit the selection criteria presented below for Elision-Receptive and Elision-Expressive to increase the item pool.

For Elision-Receptive researchers selected words for which, when deleting a unit of language (at the word, syllable, and phoneme levels), a new word was generated. This word could then be imaged and the child could select what word was left (i.e. What is cowboy without cow?)The deleted unit of language could be located at the beginning of the word, at the end of the word, or in middle positions. To ensure phonetic representation of different sounds of the language, researchers excluded some items in which the same syllables and phonemes were repeated multiple times (e.g., syllable blends with [ ] and /l/ phonemes).

In the Elision-Expressive, children listen to a word, remove the designated part, and verbally produce the remaining segment of the word. The item involved a mix of real and nonwords, the location of the deletion varied as did the unit of deletion from compund word, to syllable, to phoneme.

9.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

9.5 Calibration Samples

Table 9.1: Demographic Characteristics of Calibration Samples for the English and Spanish Expressive Elision Tasks

	English			Spanish		
Characteristic	**K** N = 603	**G1** N = 1,769	**G2** N = 1,369	**K** N = 594	**G1** N = 636	**G2** N = 548
Timepoint						
Spring 2023	0 (0%)	0 (0%)	0 (0%)	594 (100%)	636 (100%)	0 (0%)
Fall 2023	592 (98%)	658 (37%)	689 (50%)	0 (0%)	0 (0%)	332 (61%)
Winter 2024	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	216 (39%)
Fall 2024	11 (1.8%)	1,111 (63%)	680 (50%)	0 (0%)	0 (0%)	0 (0%)
Administration Format						
CAT	11 (1.8%)	1,111 (63%)	680 (50%)			
Forms	592 (98%)	658 (37%)	689 (50%)	594 (100%)	636 (100%)	548 (100%)
Race						
American/Alaskan Native	15 (2.5%)	63 (3.6%)	26 (2.0%)	9 (1.6%)	7 (1.2%)	3 (0.5%)
Asian	73 (12%)	166 (9.6%)	132 (10%)	7 (1.2%)	6 (1.0%)	5 (0.9%)
Black/African American	73 (12%)	190 (11%)	163 (12%)	5 (0.9%)	4 (0.7%)	3 (0.5%)
Not reported	131 (22%)	303 (17%)	230 (18%)	357 (63%)	381 (63%)	353 (65%)
Other	111 (19%)	257 (15%)	138 (11%)	46 (8.1%)	48 (7.9%)	24 (4.4%)
White	190 (32%)	755 (44%)	616 (47%)	141 (25%)	161 (27%)	159 (29%)
Unknown	10	35	64	29	29	1
Ethnicity						
Hispanic/Latin(o/a)	325 (54%)	1,180 (69%)	839 (65%)	516 (97%)	549 (97%)	503 (94%)
Intentional nonreport	10 (1.7%)	6 (0.4%)	3 (0.2%)	0 (0%)	1 (0.2%)	2 (0.4%)
Not Hispanic/Latin(o/a)	262 (44%)	521 (31%)	447 (35%)	15 (2.8%)	18 (3.2%)	30 (5.6%)
Unknown	6	62	80	63	68	13
Gender						
Female	306 (51%)	880 (52%)	636 (49%)	314 (59%)	339 (60%)	297 (54%)
Male	292 (49%)	804 (48%)	649 (51%)	222 (41%)	229 (40%)	250 (46%)
Non-binary	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.2%)
Unknown	5	85	84	58	68	0
Home Language						
English	413 (70%)	925 (55%)	791 (63%)	76 (13%)	74 (12%)	50 (9.2%)
Spanish	110 (19%)	660 (40%)	373 (30%)	473 (84%)	525 (87%)	484 (89%)
Other	63 (11%)	83 (5.0%)	91 (7.3%)	16 (2.8%)	6 (1.0%)	7 (1.3%)
Unknown	17	101	114	29	31	7
English Proficiency Label						
(Re-)Classified Proficient	38 (7.2%)	114 (7.2%)	108 (8.7%)	47 (9.5%)	48 (8.7%)	81 (16%)
English Learner	145 (28%)	599 (38%)	354 (29%)	393 (79%)	440 (80%)	394 (76%)
English-only	342 (65%)	878 (55%)	776 (63%)	57 (11%)	61 (11%)	44 (8.5%)
Unknown	78	178	131	97	87	29
Ever IEP/504	38 (8.0%)	144 (10%)	120 (12%)	30 (11%)	23 (12%)	35 (11%)
Unknown	128	334	339	314	448	237

9.6 Psychometric Analysis

9.6.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 0 items from the English task and 0 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 9.2 summarizes the basic item characteristics, Figure 9.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 9.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Expressive Elision Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 139**	**N = 139**	**N = 117**	**N = 117**
No. of Responses	280 (184)	280 (184)	227 (176)	227 (176)
Proportion Correct	0.36 (0.18)	0.36 (0.18)	0.31 (0.13)	0.31 (0.13)
Point-biserial Correlation	0.54 (0.12)	0.54 (0.12)	0.61 (0.11)	0.61 (0.11)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

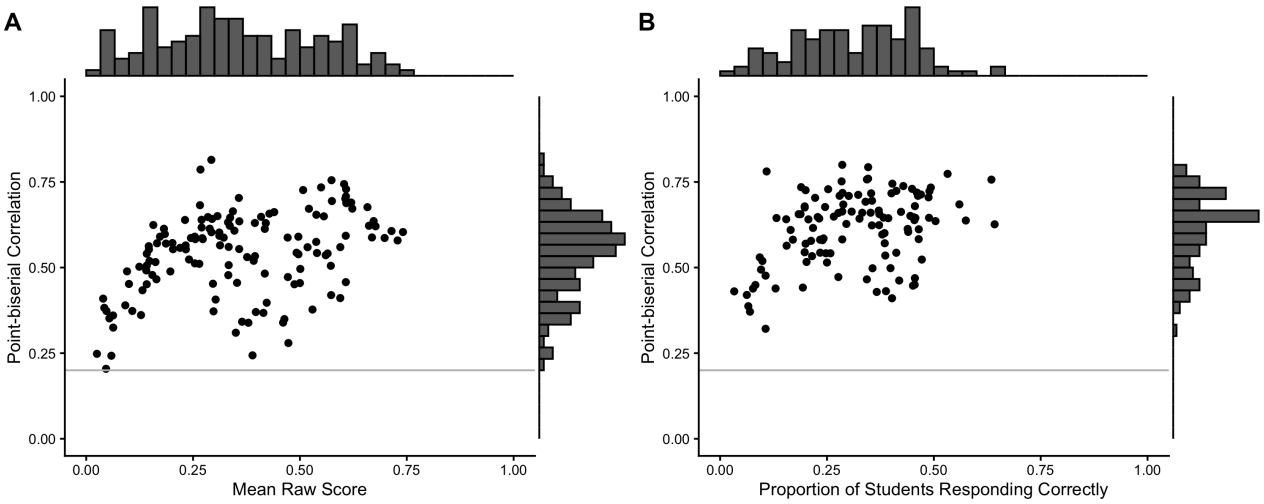

Figure 9.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Expressive Elision Tasks

9.6.2 Rasch Analysis

9.6.2.1 Item Location Estimates

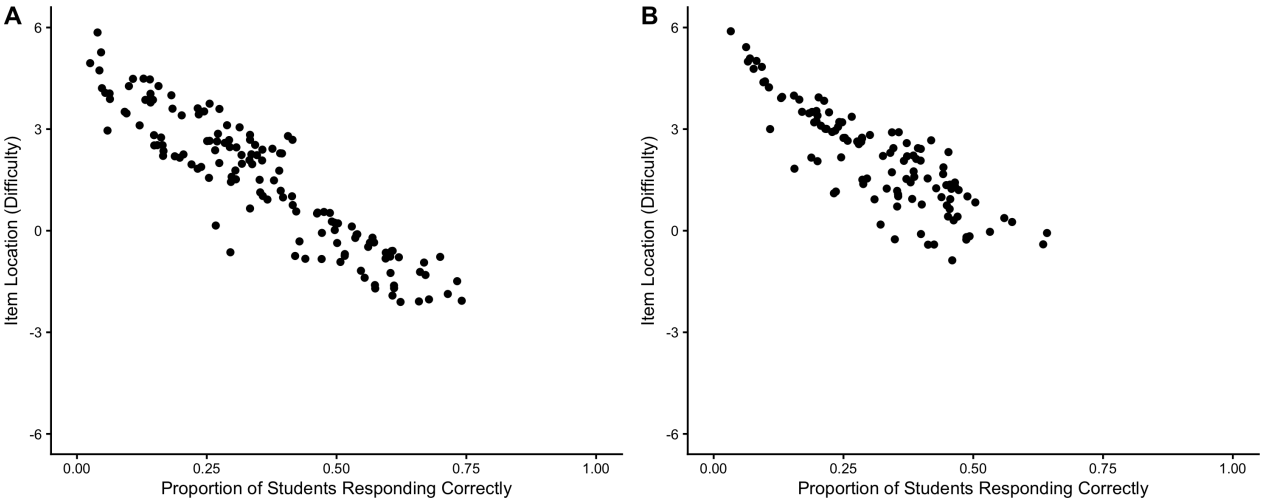

Figure 9.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Expressive Elision Tasks

9.6.2.2 Item Fit Statistics

Table 9.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Expressive Elision Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	120	0	0	0	120	103	0	0	0	103
B	1	0	0	0	1	8	0	0	0	8
C	8	0	0	0	8	4	0	0	0	4
D	9	0	1	0	10	2	0	0	0	2
Total	138	0	1	0	139	117	0	0	0	117

9.6.2.3 Person Location Estimates

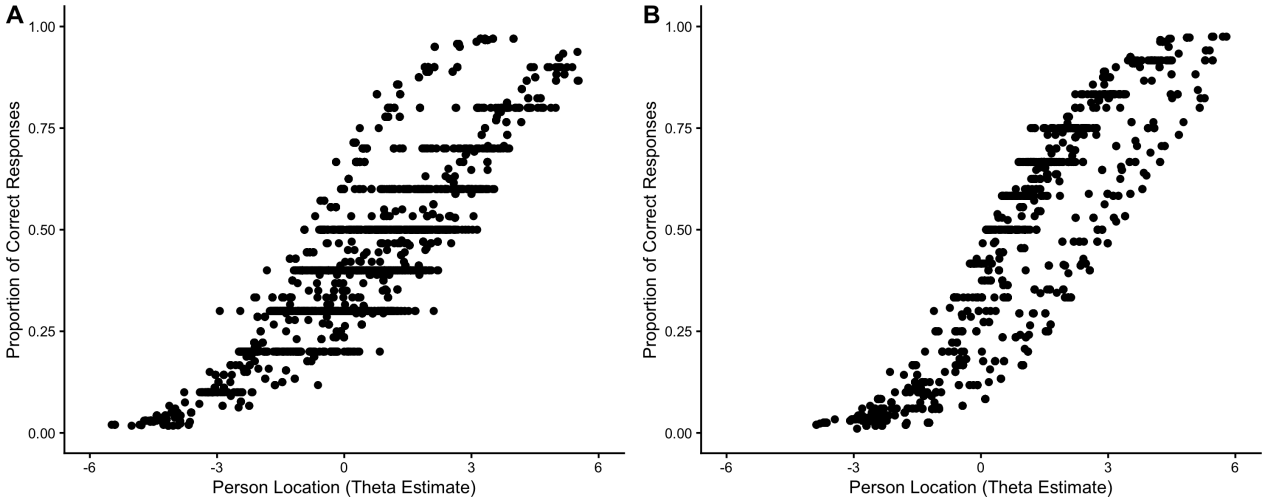

Figure 9.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Expressive Elision Tasks

9.6.2.4 Person Fit Statistics

Table 9.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Expressive Elision Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	1,713	0	15	0	1,728	1,245	0	5	0	1,250
B	746	782	0	0	1,528	208	814	0	0	1,022
C	75	0	38	3	116	79	0	25	0	104
D	99	0	53	36	188	5	0	19	7	31
Total	2,633	782	106	39	3,560	1,537	814	49	7	2,407

9.6.2.5 Distribution of Theta Estimates

9.6.2.6 Wright Maps

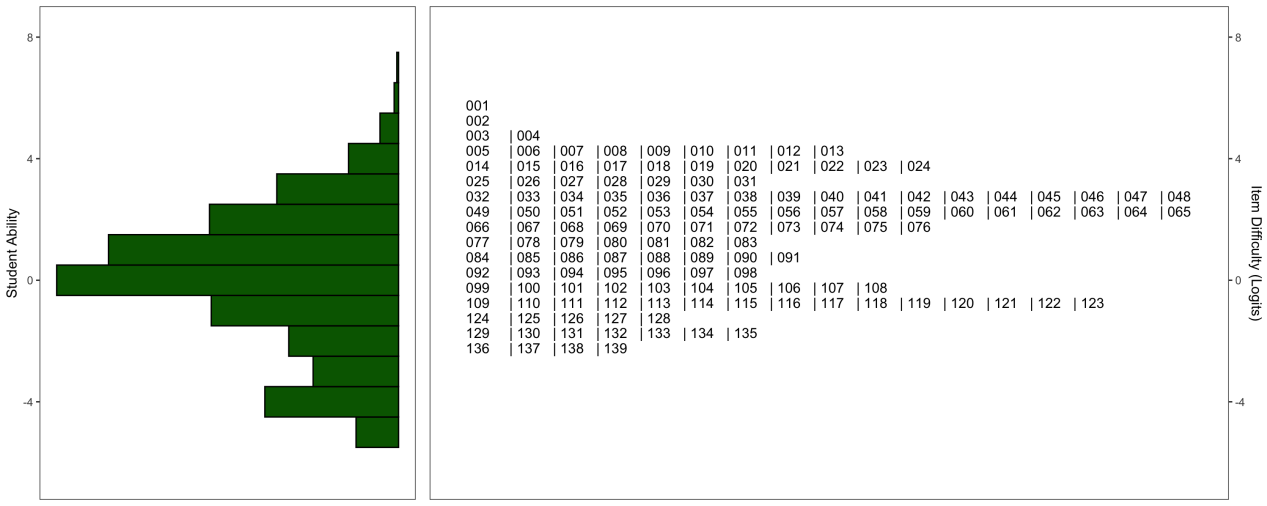

Figure 9.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Expressive Elision Task

9.6.2.7 Model Summary

Table 9.5: Summary of Rasch Model Statistics for the English and Spanish Expressive Elision Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 139	N = 3,560	N = 117	N = 2,407
Logit Scale Location	1.46 (1.95)	0.27 (-1.37, 1.45)	2.16 (1.46)	0.33 (-1.77, 1.77)
Outfit	1.16 (0.68)	0.55 (0.36, 0.85)	0.95 (0.37)	0.60 (0.08, 0.91)
Infit	1.00 (0.13)	0.74 (0.53, 0.98)	0.99 (0.12)	0.72 (0.10, 0.94)
Reliability of Separation	0.8529	0.7923	0.8268	0.6924

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 139 and 117 for the English and Spanish task, respectively.

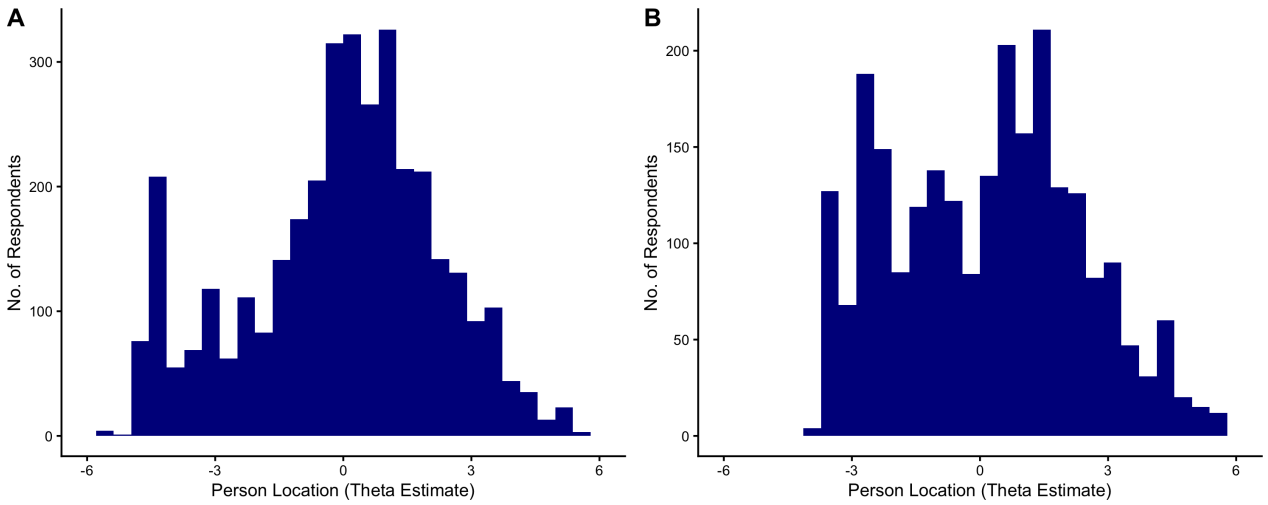

Figure 9.4: Distribution of Theta Estimates for the English and Spanish Expressive Elision Tasks

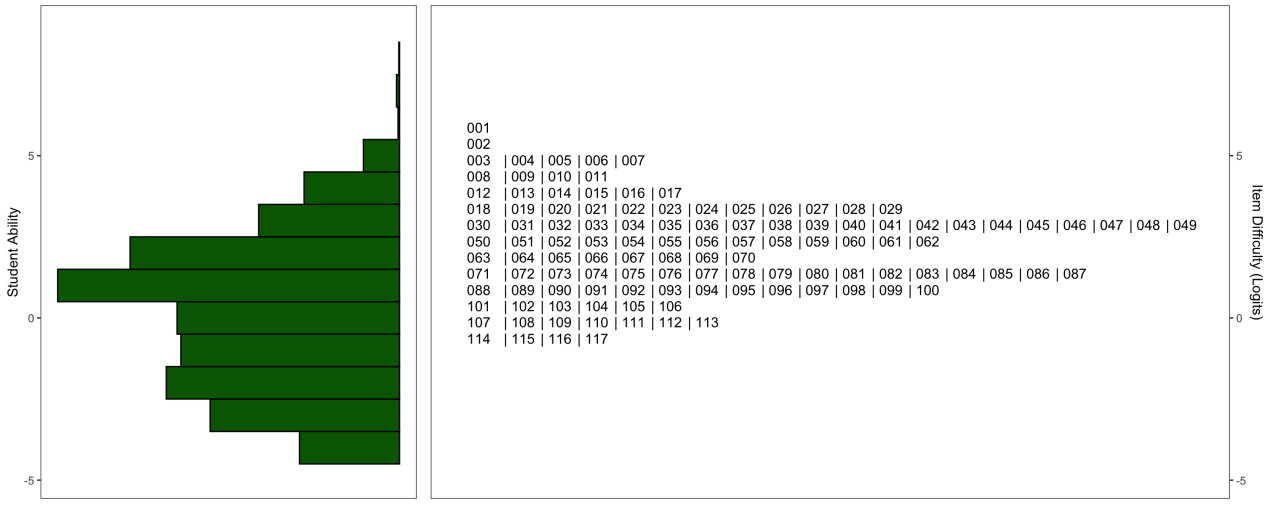

Figure 9.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Expressive Elision Task

9.7 Criterion Validity Evidence

9.7.1 Sample

Table 9.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Expressive Elision Tasks

	English			Spanish		
Characteristic	**K** N = 248	**G1** N = 222	**G2** N = 202	**K** N = 51	**G1** N = 167	**G2** N = 190
Timepoint						
Winter 2024	248 (100%)	222 (100%)	202 (100%)	51 (100%)	167 (100%)	190 (100%)
Race						
American/Alaskan Native	5 (2.0%)	3 (1.4%)	1 (0.5%)	0 (0%)	4 (2.4%)	1 (0.5%)
Asian	33 (13%)	36 (16%)	8 (4.3%)	2 (3.9%)	2 (1.2%)	0 (0%)
Black/African American	26 (11%)	27 (12%)	34 (18%)			
Not reported	28 (11%)	29 (13%)	13 (7.1%)	28 (55%)	116 (70%)	120 (63%)
Other	72 (29%)	44 (20%)	3 (1.6%)	11 (22%)	5 (3.0%)	10 (5.3%)
White	81 (33%)	83 (37%)	125 (68%)	10 (20%)	38 (23%)	58 (31%)
Unknown	3	0	18	0	2	1
Ethnicity						
Hispanic/Latin(o/a)	99 (40%)	93 (42%)	120 (60%)	46 (90%)	156 (93%)	178 (99%)
Intentional nonreport	7 (2.8%)	2 (0.9%)	0 (0%)	0 (0%)	0 (0%)	2 (1.1%)
Not Hispanic/Latin(o/a)	142 (57%)	127 (57%)	81 (40%)	5 (9.8%)	11 (6.6%)	0 (0%)
Unknown	0	0	1	0	0	10
Gender						
Female	125 (50%)	102 (46%)	97 (48%)	32 (63%)	83 (50%)	100 (53%)
Male	123 (50%)	120 (54%)	105 (52%)	19 (37%)	84 (50%)	90 (47%)
Home Language						
English	183 (75%)	165 (75%)	126 (82%)	6 (12%)	17 (10%)	21 (11%)
Spanish	30 (12%)	24 (11%)	23 (15%)	44 (86%)	147 (89%)	168 (89%)
Other	31 (13%)	32 (14%)	5 (3.2%)	1 (2.0%)	1 (0.6%)	0 (0%)
Unknown	4	1	48	0	2	1
English Proficiency Label						
(Re-)Classified Proficient	10 (4.9%)	17 (7.8%)	11 (7.1%)	6 (13%)	19 (12%)	35 (20%)
English Learner	46 (23%)	40 (18%)	17 (11%)	40 (83%)	130 (79%)	119 (68%)
English-only	148 (73%)	160 (74%)	126 (82%)	2 (4.2%)	16 (9.7%)	21 (12%)
Unknown	44	5	48	3	2	15
Ever IEP/504	17 (8.9%)	21 (12%)	17 (11%)	1 (2.3%)	11 (7.1%)	6 (6.5%)
Unknown	58	47	48	8	12	97

English Expressive Elision was correlated with the Elision subtest from the Comprehensive Test of Phonological Processing, 2nd Edition (CTOPP-2) test (Wagner et al. 2013). Spanish Expressive Elision was correlated with the Deletion subtest from the Test of Phonological Awareness in Spanish (TPAS)(Riccio et al. 2004b).

Table 9.7: Concurrent Criterion Validity Correlations for the English and Spanish Expressive Elision Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	248	0.78 [0.72, 0.82]	46	0.73 [0.56, 0.84]	51	0.58 [0.36, 0.74]
G1	222	0.71 [0.64, 0.77]	40	0.66 [0.44, 0.80]	167	0.47 [0.35, 0.58]
G2	202	0.78 [0.73, 0.83]	NA	NA	190	0.51 [0.39, 0.61]

10 Elision-Receptive

10.1 Task Description

Children are shown three pictures, each of which is named for them. They then hear a word and are prompted to take away part of it. Children are asked to select the picture that represents the resulting word.

10.2 Construct

The Elision-Receptive task measures phonological awareness and auditory manipulation skills. Students hear a word and are asked to delete a specific linguistic unit (compound word, syllable, or phoneme), then select a picture representing the portion of the word that remains. This task emphasizes auditory discrimination, phonological processing, and receptive language abilities, providing insight into students' ability to manipulate language units without requiring verbal production.

10.3 Item Development

An original list of words was developed from English and Dual Language Program curricula commonly used in California. Researchers later extended the list to increase the item pool for the Elision-Receptive task.

Words were chosen such that, when a specified linguistic unit (word, syllable, or phoneme) was removed, the remainder formed a new, valid word that was also imageable. The removable element could appear at the beginning, the end, or in the middle of the word.

To ensure a diverse range of sounds, items were screened to minimize repetition of syllables and phonemes (e.g., blends with / / or /l/).

The Elision Receptive target words were selected for easy imaginability and concreteness, with pictures designed to be universally recognizable to reduce cultural bias.

10.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

10.5 Calibration Samples

Table 10.1: Demographic Characteristics of Calibration Samples for the English and Spanish Elision-Receptive Tasks

	English		Spanish	
Characteristic	**K** N = 2,306	**G1** N = 668	**K** N = 1,140	**G1** N = 650
Timepoint				
Spring 2023	0 (0%)	0 (0%)	606 (53%)	644 (99%)
Fall 2023	605 (26%)	660 (99%)	0 (0%)	0 (0%)
Fall 2024	1,701 (74%)	8 (1.2%)	534 (47%)	6 (0.9%)
Administration Format				
CAT	1,701 (74%)	8 (1.2%)	534 (47%)	6 (0.9%)
Forms	605 (26%)	660 (99%)	606 (53%)	644 (99%)
Race				
American/Alaskan Native	88 (4.1%)	12 (1.8%)	38 (3.4%)	8 (1.3%)
Asian	154 (7.1%)	86 (13%)	21 (1.9%)	7 (1.1%)
Black/African American	203 (9.4%)	78 (12%)	11 (1.0%)	4 (0.6%)
Not reported	259 (12%)	127 (19%)	462 (42%)	385 (62%)
Other	521 (24%)	76 (11%)	226 (20%)	52 (8.4%)
White	944 (44%)	288 (43%)	347 (31%)	164 (26%)
Unknown	137	1	35	30
Ethnicity				
Hispanic/Latin(o/a)	1,388 (70%)	347 (52%)	954 (97%)	559 (96%)
Intentional nonreport	19 (1.0%)	3 (0.4%)	3 (0.3%)	1 (0.2%)
Not Hispanic/Latin(o/a)	563 (29%)	317 (48%)	23 (2.3%)	20 (3.4%)
Unknown	336	1	160	70
Gender				
Female	997 (50%)	315 (47%)	520 (53%)	346 (60%)
Male	997 (50%)	352 (53%)	459 (47%)	234 (40%)
Unknown	312	1	161	70
Home Language				
English	1,277 (63%)	490 (74%)	113 (10%)	72 (12%)
Spanish	671 (33%)	105 (16%)	975 (88%)	539 (87%)
Other	82 (4.0%)	70 (11%)	16 (1.4%)	7 (1.1%)
Unknown	276	3	36	32
English Proficiency Label				
(Re-)Classified Proficient	76 (4.1%)	62 (9.5%)	72 (7.6%)	49 (8.8%)
English Learner	606 (33%)	123 (19%)	789 (84%)	451 (81%)
English-only	1,161 (63%)	471 (72%)	83 (8.8%)	59 (11%)
Unknown	463	12	196	91
Ever IEP/504	123 (7.8%)	64 (11%)	71 (10.0%)	25 (13%)
Unknown	725	106	427	452

10.6 Psychometric Analysis

10.6.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 0 items from the English task and 0 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 10.2 summarizes the basic item characteristics, Figure 10.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 10.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Elision-Receptive Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 59**	**N = 59**	**N = 62**	**N = 62**
No. of Responses	636 (768)	636 (768)	274 (158)	274 (158)
Proportion Correct	0.75 (0.12)	0.75 (0.12)	0.57 (0.09)	0.57 (0.09)
Point-biserial Correlation	0.50 (0.10)	0.50 (0.10)	0.54 (0.11)	0.54 (0.11)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

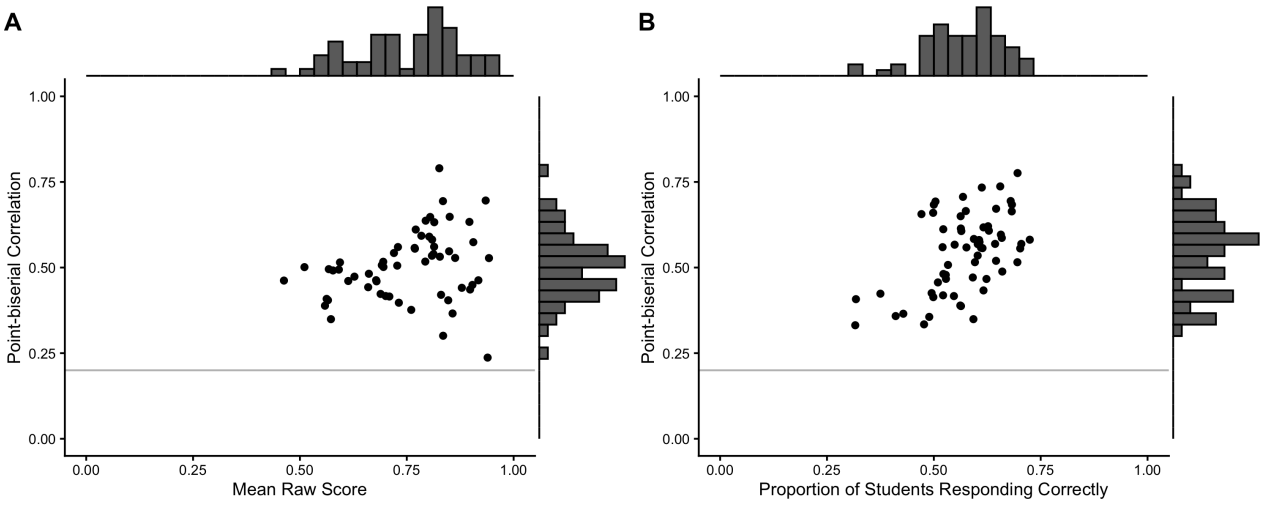

Figure 10.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Elision-Receptive Tasks

10.6.2 Rasch Analysis

10.6.2.1 Item Location Estimates

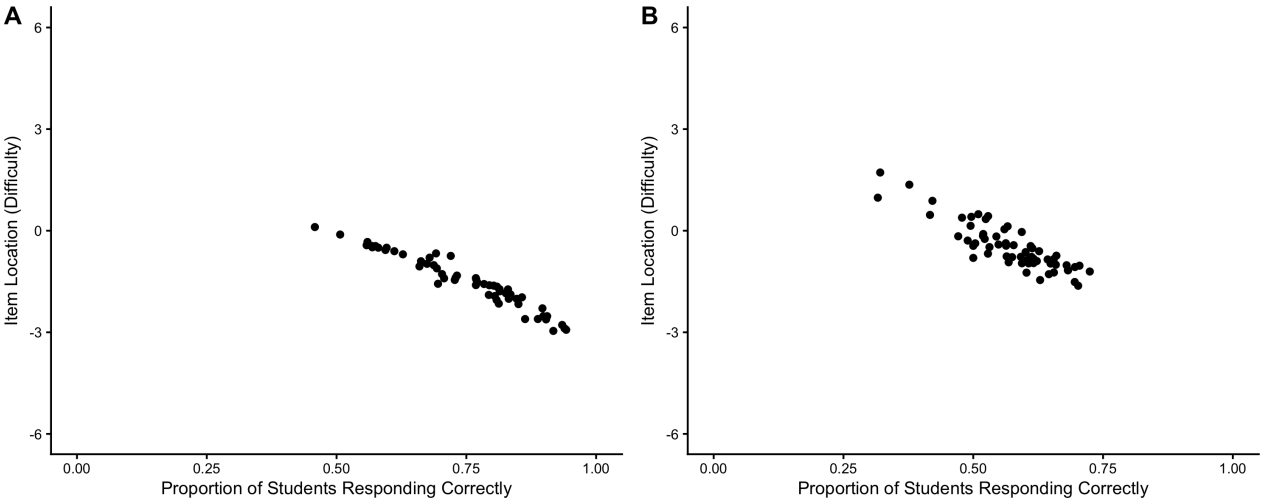

Figure 10.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Elision-Receptive Tasks

10.6.2.2 Item Fit Statistics

Table 10.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Elision-Receptive Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	57	0	0	0	57	62	0	0	0	62
B	1	0	0	0	1	0	0	0	0	0
C	1	0	0	0	1	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0
Total	59	0	0	0	59	62	0	0	0	62

10.6.2.3 Person Location Estimates

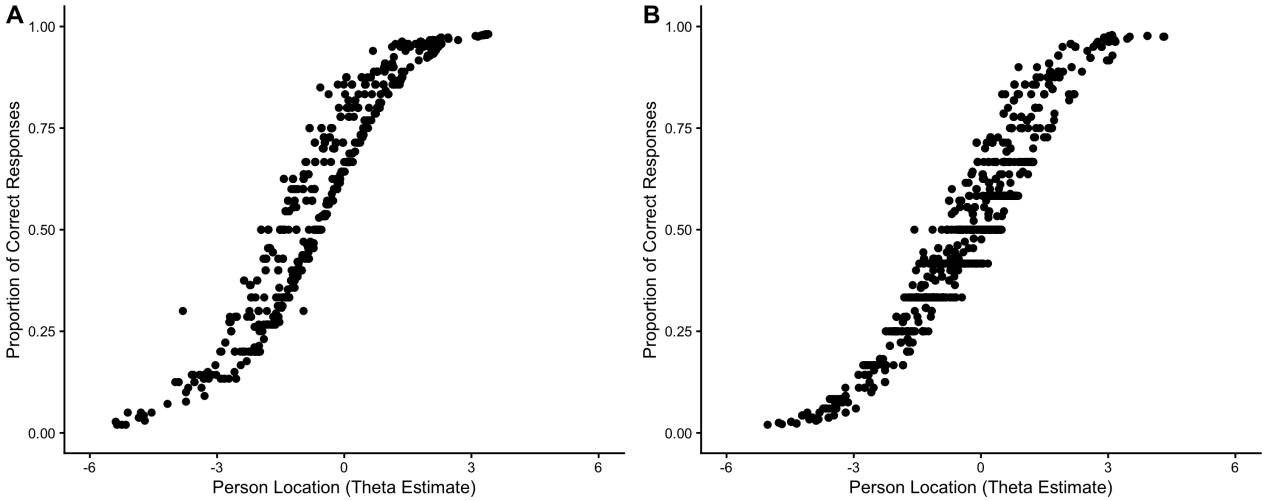

Figure 10.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Elision-Receptive Tasks

10.6.2.4 Person Fit Statistics

Table 10.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Elision-Receptive Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	2,220	0	0	0	2,220	1,387	0	0	0	1,387
B	88	566	0	0	654	45	231	0	0	276
C	39	0	5	0	44	48	0	7	0	55
D	5	0	6	1	12	4	0	4	0	8
Total	2,352	566	11	1	2,930	1,484	231	11	0	1,726

10.6.2.5 Distribution of Theta Estimates

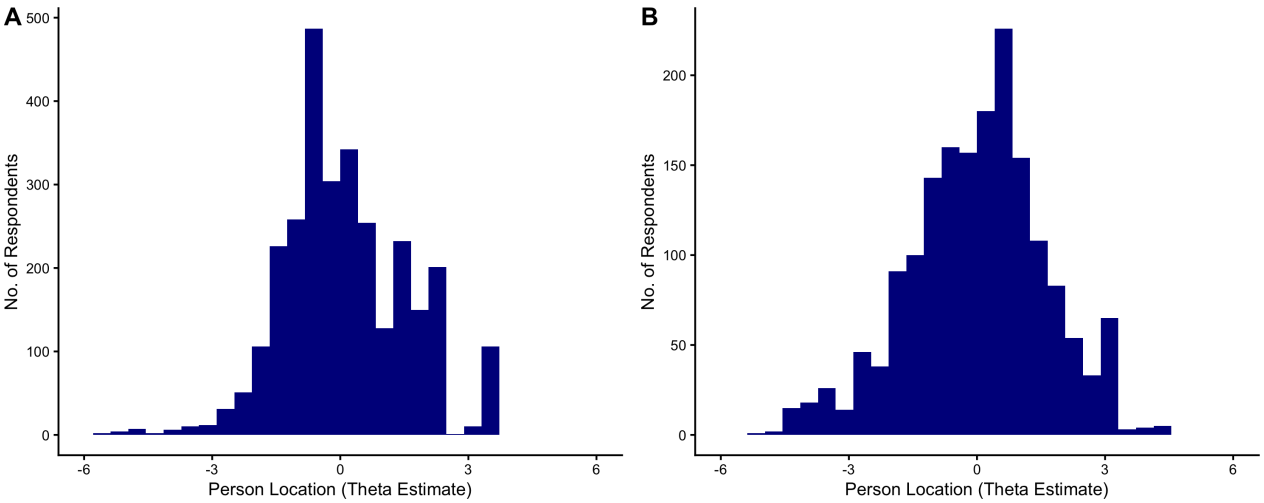

Figure 10.4: Distribution of Theta Estimates for the English and Spanish Elision-Receptive Tasks

10.6.2.6 Wright Maps

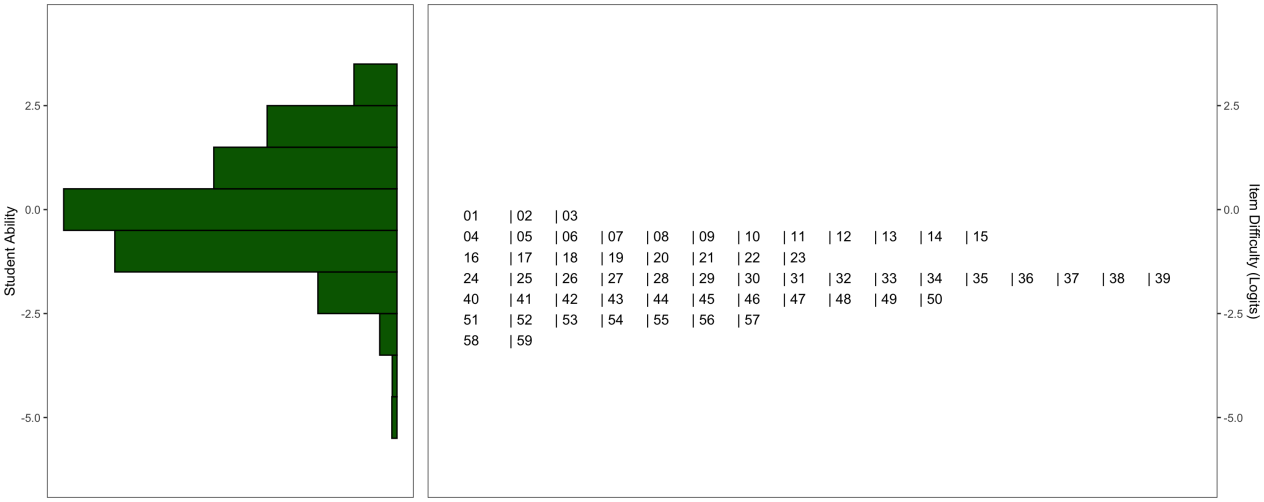

Figure 10.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Elision-Receptive Task

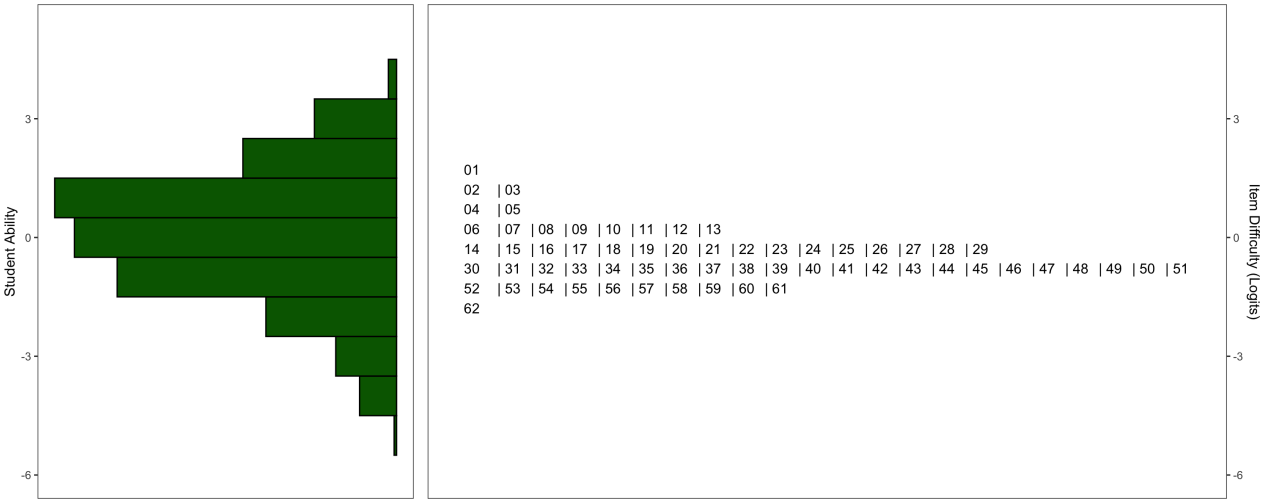

Figure 10.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Elision-Receptive Task

10.6.2.7 Model Summary

Table 10.5: Summary of Rasch Model Statistics for the English and Spanish Elision-Receptive Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 59	N = 2,930	N = 62	N = 1,726
Logit Scale Location	-1.51 (0.78)	-0.09 (-0.80, 0.98)	-0.48 (0.69)	0.18 (-1.05, 0.97)
Outfit	0.94 (0.17)	0.88 (0.56, 1.03)	0.99 (0.17)	0.86 (0.66, 1.01)
Infit	0.98 (0.08)	0.90 (0.73, 1.01)	0.98 (0.11)	0.88 (0.74, 0.99)
Reliability of Separation	0.6754	0.4677	0.7243	0.6226

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 59 and 62 for the English and Spanish task, respectively.

10.7 Criterion Validity Evidence

10.7.1 Sample

Table 10.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Elision-Receptive Tasks

	English		Spanish	
Characteristic	**K** N = 260	**G1** N = 228	**K** N = 52	**G1** N = 166
Timepoint				
Winter 2024	260 (100%)	228 (100%)	52 (100%)	166 (100%)
Race				
American/Alaskan Native	5 (1.9%)	3 (1.3%)	0 (0%)	4 (2.4%)
Asian	35 (14%)	37 (16%)	2 (3.8%)	2 (1.2%)
Black/African American	29 (11%)	28 (12%)		
Not reported	29 (11%)	32 (14%)	28 (54%)	116 (71%)
Other	73 (28%)	44 (19%)	11 (21%)	5 (3.0%)
White	86 (33%)	84 (37%)	11 (21%)	37 (23%)
Unknown	3	0	0	2
Ethnicity				
Hispanic/Latin(o/a)	106 (41%)	96 (42%)	47 (90%)	155 (93%)
Intentional nonreport	7 (2.7%)	2 (0.9%)		
Not Hispanic/Latin(o/a)	147 (57%)	130 (57%)	5 (9.6%)	11 (6.6%)
Gender				
Female	130 (50%)	105 (46%)	33 (63%)	82 (49%)
Male	130 (50%)	123 (54%)	19 (37%)	84 (51%)
Home Language				
English	191 (75%)	170 (75%)	6 (12%)	17 (10%)
Spanish	33 (13%)	25 (11%)	45 (87%)	146 (89%)
Other	31 (12%)	32 (14%)	1 (1.9%)	1 (0.6%)
Unknown	5	1	0	2
English Proficiency Label				
(Re-)Classified Proficient	11 (5.1%)	18 (8.1%)	6 (12%)	19 (12%)
English Learner	48 (22%)	40 (18%)	41 (84%)	129 (79%)
English-only	156 (73%)	165 (74%)	2 (4.1%)	16 (9.8%)
Unknown	45	5	3	2
Ever IEP/504	20 (10.0%)	22 (12%)	1 (2.3%)	11 (7.1%)
Unknown	59	48	9	11

English Elision-Receptive was correlated with the Elision subtest from the Comprehensive Test of Phonological Processing, 2nd Edition (Wagner et al. 2013). Spanish Elision-Receptive was correlated with the Deletion subtest from the Test of Phonological Awareness in Spanish (TPAS) (Riccio et al. 2004b).

Table 10.7: Concurrent Criterion Validity Correlations for the English and Spanish Elision-Receptive Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	260	0.50 [0.40, 0.58]	48	0.54 [0.31, 0.72]	52	0.57 [0.35, 0.73]
G1	228	0.49 [0.38, 0.58]	40	0.31 [-0.01, 0.56]	166	0.40 [0.27, 0.52]

11 Expressive Vocabulary

11.1 Task Description

Children are shown pictures and are asked to name them.

11.2 Construct

The Expressive Vocabulary task measures the construct of semantic knowledge. Students are prompted to name images, thereby measuring the ability to accurately label nouns and verbs.

11.3 Item Development

The Multitudes development process involved developing parallel measures in English and Spanish when possible. To be able to develop a linguistically and culturally appropriate measure scored conceptually, we needed to select a list of words that were relevant for the population being studied, with comparable levels of difficulty, and easily represented with an image.

For the selection of the words being targeted with our measure, we first reviewed reading curricula commonly used in California, specifically, and in the United States more broadly, in addition to other literacy materials sourced from Chile, Mexico, and Panama. To extend the list of selected words, we used Age of Acquisition (AoA) databases for English (Brysbaert and Biemiller 2017) and Spanish (Alonso, Fernández, and Díez 2015; Alonso, Díez, and Fernández 2016) to compile a comprehensive list with somewhat comparable AoA in both languages. The AoA was used as a proxy of difficulty, hypothesizing that words acquired earlier in development were more likely to yield easier items. In addition to AoA information, we used Clearpond (Marian et al. 2012) to retrieve information on the word's frequency, orthographic and phonological length, and neighborhood frequency.

We used iStock to choose the real pictures of the selected words. The chosen pictures underwent a rigorous selection process to meet specific criteria:

- **Easily Recognizable.** Emphasis was placed on selecting images that could be easily identified.
- **Minimal Construct Irrelevant Features.** Preference was given to pictures with a clean and unobtrusive background, and a white background was opted for whenever possible.
- **Removal of Irrelevant Information.** Unnecessary elements, such as a leaf in a pear or surrounding mountains, were eliminated through cropping, focusing solely on the essential components of the image.

- **Culturally Representative.** Images that were representative of a culturally diverse population. The final pool of images was reviewed by the Justice, Equity, Diversity, and Inclusion committee, and any images that were identified as not representative of the target word or that had potential concerns were deliberately excluded from the final pool.

11.3.1 Dialectal considerations.

- **English.** The researchers listed all acceptable responses, and an additional online search was conducted to find other regional terms for specific items in English.
- **Spanish.** A group of six Spanish-English bilingual research assistants was shown the images and requested to label them. To avoid penalizing the use of regional terms, we asked research assistants from different countries or regions (including Chile, Colombia, three different states in Mexico, and Venezuela) to list all the different terms known to them to label said pictures. An additional search was conducted to find other regional terms for both English and Spanish.

Additionally, a blank box was available for examiners to annotate other possible acceptable responses given by the participating children during piloting. We then reviewed all of those responses and decided which ones should be added to the correct response list for each item.

11.4 Scoring

11.4.1 English

A list of accepted terms for English was provided. The assessment uses a dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

11.4.2 Spanish-English Bilingual

A list of accepted words for Spanish and English is provided. The assessment was calibrated using conceptual scoring meaning a child can respond in either English or Spanish. The assessment uses a dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

11.5 Calibration Samples

Table 11.1: Demographic Characteristics of Calibration Samples for the English and Spanish Expressive Vocabulary Tasks

	English			Spanish		
Characteristic	**K** N = 2,127	**G1** N = 2,805	**G2** N = 2,887	**K** N = 1,164	**G1** N = 1,270	**G2** N = 1,013
Timepoint						
Spring 2023	0 (0%)	0 (0%)	0 (0%)	621 (53%)	630 (50%)	0 (0%)
Fall 2023	611 (29%)	668 (24%)	703 (24%)	0 (0%)	0 (0%)	345 (34%)
Winter 2024	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	264 (26%)
Fall 2024	1,516 (71%)	2,137 (76%)	2,184 (76%)	543 (47%)	640 (50%)	404 (40%)
Administration Format						
CAT	1,516 (71%)	2,137 (76%)	2,184 (76%)			
Forms	611 (29%)	668 (24%)	703 (24%)	621 (100%)	630 (100%)	609 (100%)
Race						
American/Alaskan Native	71 (3.6%)	97 (3.7%)	64 (2.4%)	40 (3.5%)	45 (3.6%)	13 (1.3%)
Asian	155 (7.8%)	223 (8.6%)	197 (7.4%)	20 (1.8%)	32 (2.6%)	24 (2.4%)
Black/African American	208 (10%)	283 (11%)	305 (11%)	11 (1.0%)	19 (1.5%)	11 (1.1%)
Not reported	196 (9.9%)	263 (10%)	246 (9.2%)	484 (42%)	552 (45%)	465 (47%)
Other	442 (22%)	343 (13%)	350 (13%)	230 (20%)	151 (12%)	67 (6.7%)
White	917 (46%)	1,395 (54%)	1,503 (56%)	355 (31%)	438 (35%)	419 (42%)
Unknown	138	201	222	24	33	14
Ethnicity						
Hispanic/Latin(o/a)	1,245 (68%)	1,806 (69%)	1,851 (70%)	979 (96%)	1,150 (97%)	926 (94%)
Intentional nonreport	16 (0.9%)	8 (0.3%)	5 (0.2%)	3 (0.3%)	3 (0.3%)	2 (0.2%)
Not Hispanic/Latin(o/a)	573 (31%)	787 (30%)	783 (30%)	34 (3.3%)	37 (3.1%)	55 (5.6%)
Unknown	293	204	248	148	80	30
Gender						
Female	946 (51%)	1,290 (50%)	1,303 (50%)	537 (53%)	662 (57%)	508 (51%)
Male	920 (49%)	1,307 (50%)	1,322 (50%)	477 (47%)	501 (43%)	481 (49%)
Non-binary	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.1%)
Unknown	261	208	262	150	107	23
Home Language						
English	1,295 (70%)	1,633 (68%)	1,633 (66%)	114 (10%)	126 (10%)	101 (10%)
Spanish	471 (25%)	672 (28%)	740 (30%)	1,009 (89%)	1,100 (89%)	858 (89%)
Other	82 (4.4%)	87 (3.6%)	110 (4.4%)	16 (1.4%)	5 (0.4%)	9 (0.9%)
Unknown	279	413	404	25	39	45
English Proficiency Label						
(Re-)Classified Proficient	60 (3.5%)	111 (4.8%)	258 (11%)	74 (7.6%)	96 (8.6%)	121 (13%)
English Learner	462 (27%)	651 (28%)	593 (24%)	813 (84%)	912 (82%)	718 (77%)
English-only	1,179 (69%)	1,574 (67%)	1,606 (65%)	84 (8.7%)	107 (9.6%)	93 (10.0%)
Unknown	426	469	430	193	155	81
Ever IEP/504	118 (8.1%)	201 (9.8%)	212 (10%)	74 (10%)	70 (9.3%)	76 (11%)
Unknown	665	752	818	452	515	330
Unknown				543	640	404

11.6 Psychometric Analysis

11.6.1 Basic Item Statistics

We excluded 0 items from the English task and 2 items from the Spanish task based on low response counts ($n < 90$). 4 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 10 items from the English task and 5 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 11.2 summarizes the basic item characteristics, Figure 11.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 11.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Expressive Vocabulary Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 151**	**N = 141**	**N = 156**	**N = 147**
No. of Responses	704 (480)	739 (476)	280 (253)	292 (255)
Proportion Correct	0.54 (0.23)	0.55 (0.21)	0.49 (0.25)	0.50 (0.23)
Point-biserial Correlation	0.44 (0.15)	0.46 (0.13)	0.45 (0.13)	0.46 (0.11)
Excluded (n < 90)	0 (0%)	0 (0%)	2 (1.3%)	0 (0%)
Excluded (pbis < .2)	10 (6.6%)	0 (0%)	5 (3.3%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	4 (2.6%)	0 (0%)

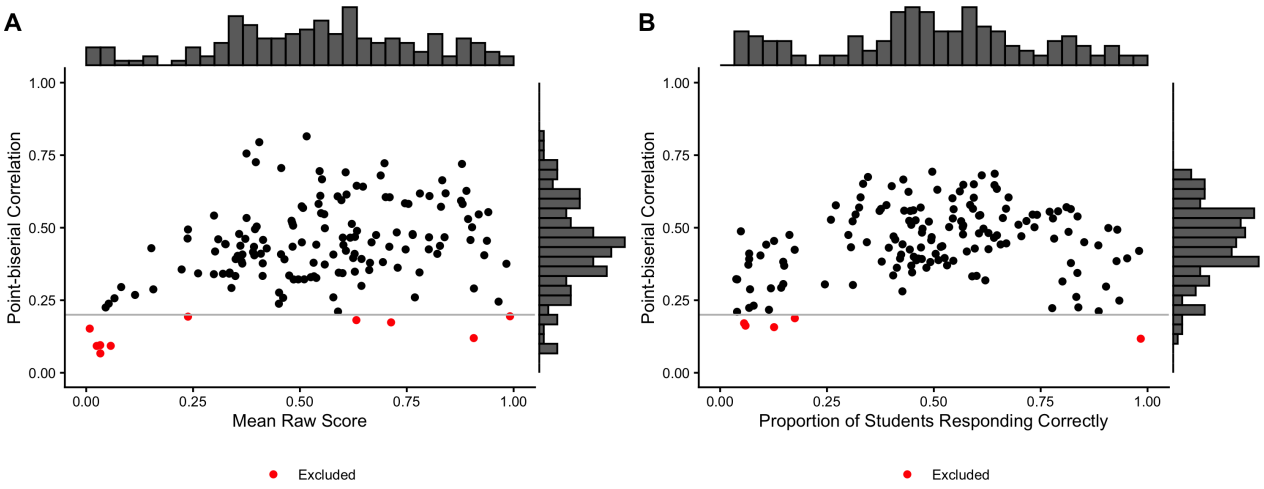

Figure 11.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Expressive Vocabulary Tasks

11.6.2 Rasch Analysis

11.6.2.1 Item Location Estimates

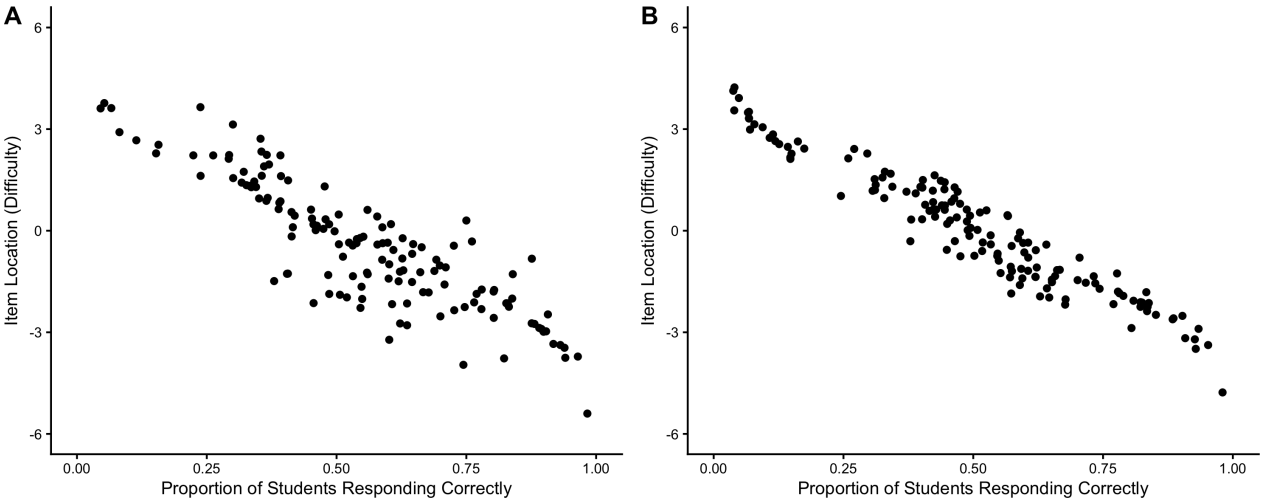

Figure 11.3: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Expressive Vocabulary Tasks

11.6.2.2 Item Fit Statistics

Table 11.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Expressive Vocabulary Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	131	0	0	0	131	139	0	0	0	139
B	3	0	0	0	3	4	0	0	0	4
C	5	0	0	0	5	4	0	0	0	4
D	2	0	0	0	2	0	0	0	0	0
Total	141	0	0	0	141	147	0	0	0	147

11.6.2.3 Person Location Estimates

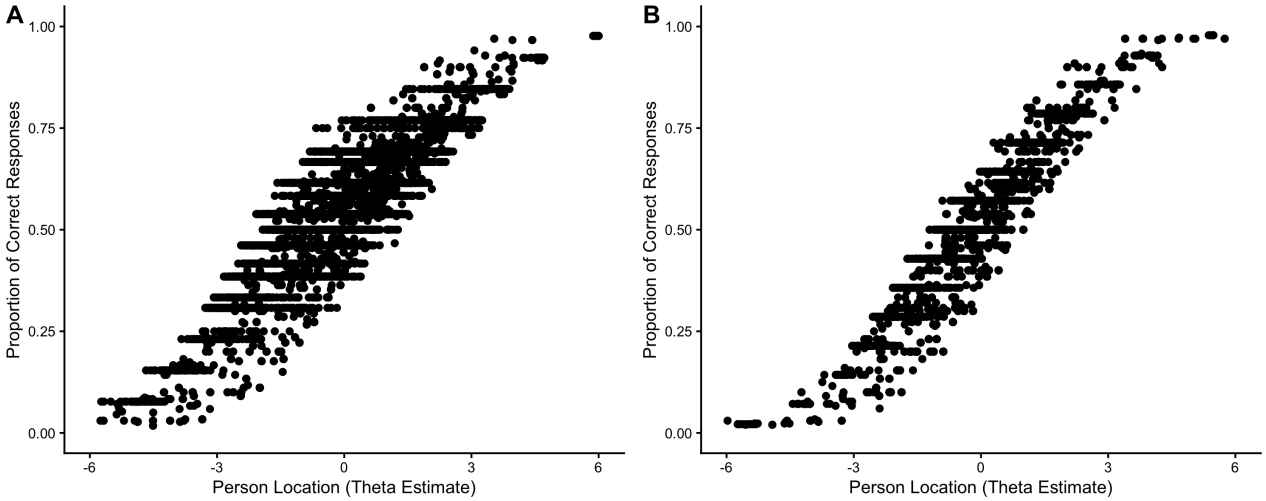

Figure 11.4: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Expressive Vocabulary Tasks

11.6.2.4 Person Fit Statistics

Table 11.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Expressive Vocabulary Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	5,932	0	41	0	5,973	2,434	0	13	1	2,448
B	921	319	0	0	1,240	375	233	0	0	608
C	205	0	86	3	294	107	0	40	3	150
D	141	0	83	28	252	58	0	48	6	112
Total	7,199	319	210	31	7,759	2,974	233	101	10	3,318

11.6.2.5 Distribution of Theta Estimates

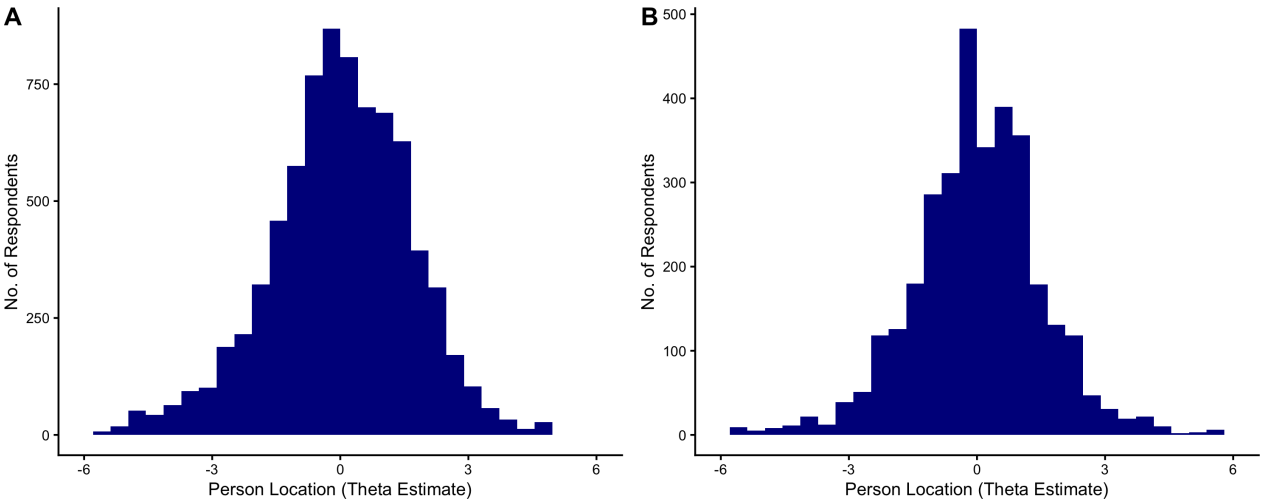

Figure 11.5: Distribution of Theta Estimates for the English and Spanish Expressive Vocabulary Tasks

11.6.2.6 Wright Maps

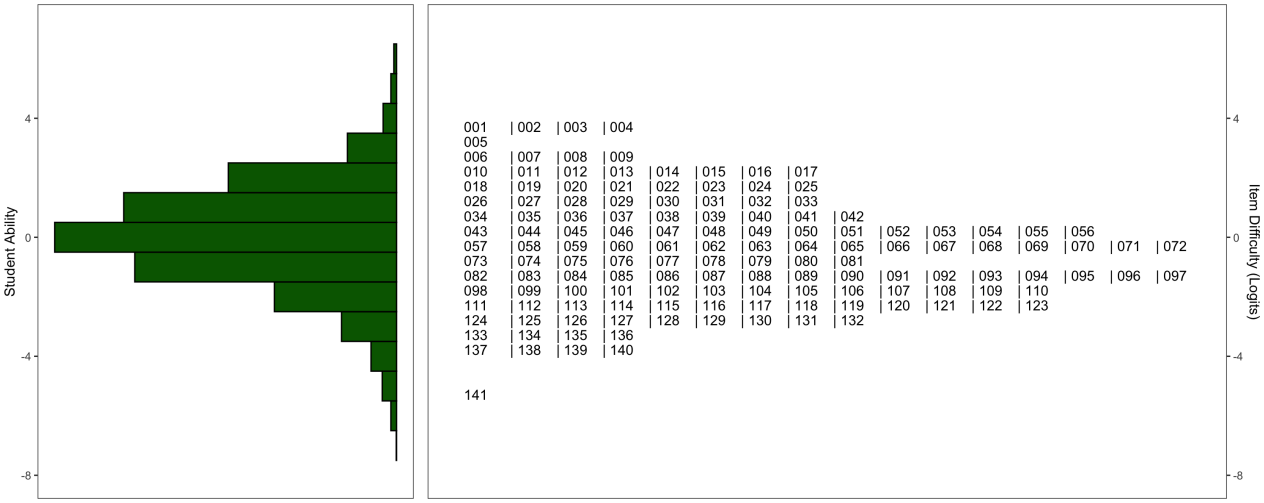

Figure 11.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Expressive Vocabulary Task

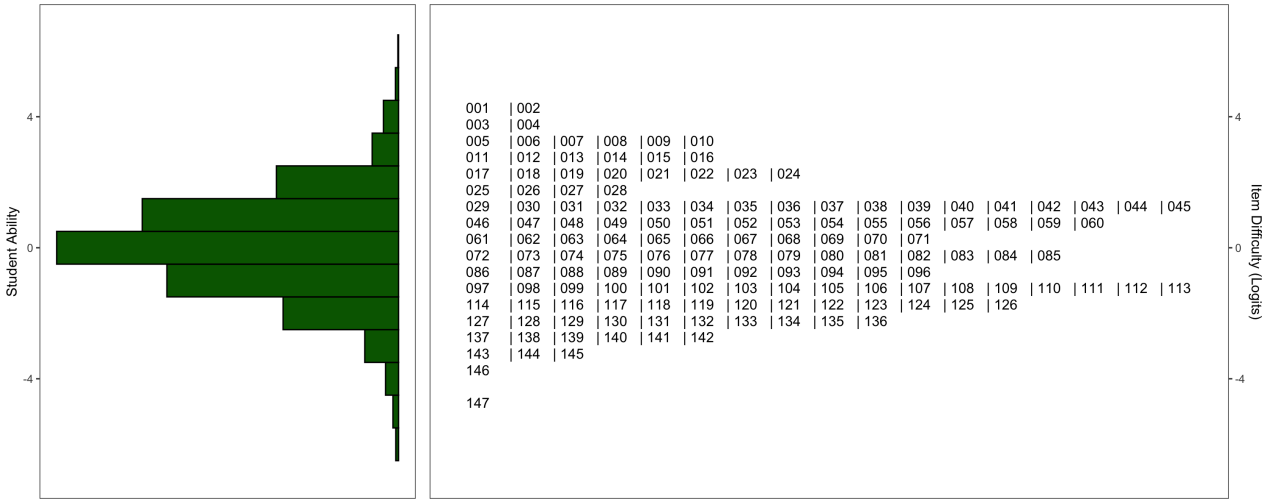

Figure 11.7: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Expressive Vocabulary Task

11.6.2.7 Model Summary

Table 11.5: Summary of Rasch Model Statistics for the English and Spanish Expressive Vocabulary Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 141	N = 7,759	N = 147	N = 3,318
Logit Scale Location	-0.47 (1.88)	0.03 (-0.98, 1.10)	0.01 (1.84)	-0.01 (-0.89, 0.99)
Outfit	1.02 (0.30)	0.71 (0.56, 0.93)	1.01 (0.22)	0.77 (0.59, 0.96)
Infit	0.98 (0.11)	0.85 (0.71, 1.02)	0.99 (0.08)	0.87 (0.74, 1.02)
Reliability of Separation	0.8292	0.8206	0.7708	0.7491

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 141 and 147 for the English and Spanish task, respectively.

11.7 Criterion Validity Evidence

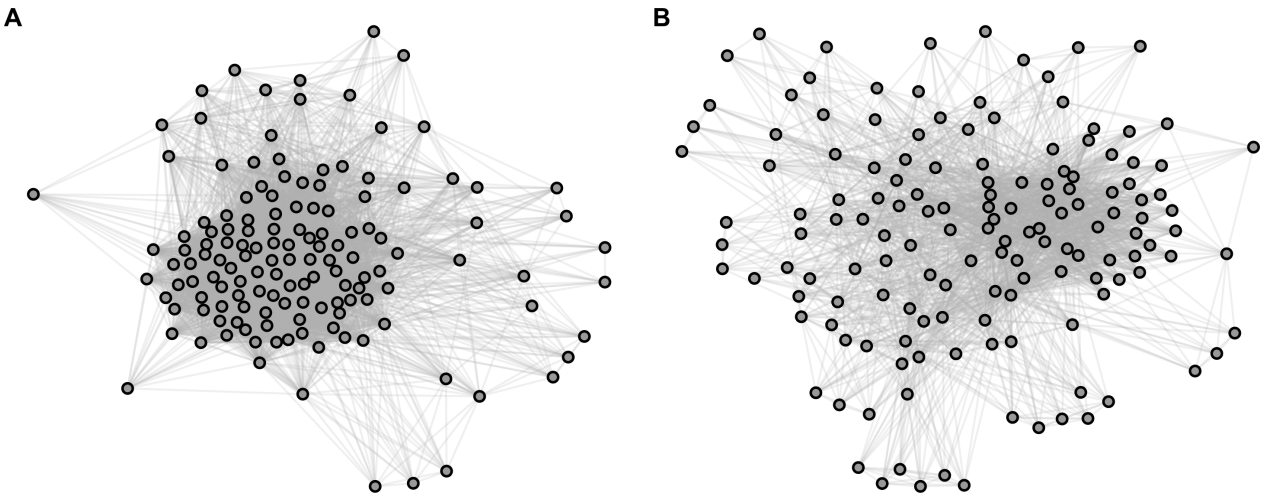

Figure 11.2: Network Graphs Showing Connections Between Items on the English (Panel A) and Spanish (Panel B) Expressive Vocabulary Tasks

11.7.1 Sample

Table 11.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Expressive Vocabulary Tasks

	English			Spanish		
Characteristic	**K** N = 262	**G1** N = 231	**G2** N = 204	**K** N = 240	**G1** N = 225	**G2** N = 265
Timepoint						
Winter 2024	262 (100%)	231 (100%)	204 (100%)	240 (100%)	225 (100%)	265 (100%)
Race						
American/Alaskan Native	5 (1.9%)	3 (1.3%)	1 (0.5%)	2 (0.8%)	4 (1.8%)	4 (1.5%)
Asian	36 (14%)	37 (16%)	8 (4.3%)	8 (3.4%)	2 (0.9%)	0 (0%)
Black/African American	28 (11%)	30 (13%)	34 (18%)	1 (0.4%)	0 (0%)	0 (0%)
Not reported	30 (12%)	32 (14%)	13 (7.0%)	132 (55%)	153 (69%)	172 (65%)
Other	74 (29%)	44 (19%)	3 (1.6%)	41 (17%)	8 (3.6%)	18 (6.8%)
White	86 (33%)	85 (37%)	127 (68%)	54 (23%)	55 (25%)	69 (26%)
Unknown	3	0	18	2	3	2
Ethnicity						
Hispanic/Latin(o/a)	106 (40%)	96 (42%)	121 (60%)	217 (92%)	209 (93%)	248 (98%)
Intentional nonreport	8 (3.1%)	2 (0.9%)	0 (0%)	1 (0.4%)	0 (0%)	2 (0.8%)
Not Hispanic/Latin(o/a)	148 (56%)	133 (58%)	82 (40%)	19 (8.0%)	16 (7.1%)	2 (0.8%)
Unknown	0	0	1	3	0	13
Gender						
Female	132 (50%)	106 (46%)	98 (48%)	126 (53%)	110 (49%)	137 (52%)
Male	130 (50%)	125 (54%)	106 (52%)	114 (48%)	115 (51%)	128 (48%)
Home Language						
English	190 (74%)	172 (75%)	128 (82%)	29 (12%)	23 (10%)	23 (8.7%)
Spanish	33 (13%)	25 (11%)	23 (15%)	206 (87%)	198 (89%)	239 (91%)
Other	34 (13%)	33 (14%)	5 (3.2%)	2 (0.8%)	1 (0.5%)	1 (0.4%)
Unknown	5	1	48	3	3	2
English Proficiency Label						
(Re-)Classified Proficient	12 (5.5%)	18 (8.0%)	11 (7.1%)	31 (14%)	24 (11%)	42 (17%)
English Learner	50 (23%)	41 (18%)	17 (11%)	184 (81%)	176 (80%)	178 (73%)
English-only	155 (71%)	167 (74%)	128 (82%)	11 (4.9%)	19 (8.7%)	23 (9.5%)
Unknown	45	5	48	14	6	22
Ever IEP/504	19 (9.4%)	22 (12%)	18 (12%)	20 (9.5%)	23 (11%)	16 (12%)
Unknown	59	50	48	30	14	127

English Expressive Vocabulary was correlated with the Expressive One-Word Picture Vocabulary Test, 4th Edition (Martin and Brownell 2011). Spanish Expressive Vocabulary was correlated with the Expressive One-Word Picture Vocabulary Test, 4th Edition, Bilingual Edition (Martin 2013).

Table 11.7: Concurrent Criterion Validity Correlations for the English and Spanish Expressive Vocabulary Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	262	0.75 [0.70, 0.80]	50	0.70 [0.52, 0.82]	240	0.59 [0.50, 0.67]
G1	231	0.76 [0.70, 0.81]	41	0.80 [0.66, 0.89]	225	0.51 [0.41, 0.60]
G2	204	0.72 [0.65, 0.78]	NA	NA	265	0.48 [0.39, 0.57]

12 Listening Comprehension

12.1 Task Description

Children listen to a sentence and are shown three or four pictures. They are asked to choose the picture that best represents the meaning of the sentence they heard.

12.2 Construct

The Listening Comprehension task measures the construct of the grammatical comprehension of sentences. Sentences cover three broad areas of grammar: phrasal syntax, temporal relationships, and modified noun phrases.

12.3 Item Development

While many available tests use production tasks to tap into children's grammatical knowledge in their language(s), considerably fewer tests use receptive tasks to measure these skills. Researchers conducted a review of existing receptive grammar tests, as well as a review of the literature regarding easy and difficult grammatical constructions, and used this to inform item design. Grammatical constructions were considered for inclusion if (1) they had the potential to differentiate low vs. high comprehension and (2) they could be tested using a receptive format (e.g., with appropriate foils). This resulted in a blueprint of selected grammatical constructions, presented below:

12.3.1 English

- **Phrasal syntax:** Correct parsing of the sentence required correctly linking the subject ('the doer') and object ('the one being acted upon') in the sentence. This included alternations in ditransitive sentences (direct object indirect object vs indirect object direct object); passive constructions; subject/object relative clauses with reversible and nonreversible noun phrases; and interrogatives (direct vs. embedded questions).
- **Temporal comprehension:** Comprehension required linking the order of events and ranged from easy (with events linearly matching the real-world chronological order) to difficult (mismatch). This was done using causal clauses (because/so), temporal clauses, future tense (will), and conditional clauses (if).

- **Complex noun phrases:** Comprehension required identifying modified noun phrases. This was done using prepositional phrases (with the red hat), adjectives (long striped), and quantifiers (none, all).

12.3.2 Spanish

- **Phrasal syntax:** Correct parsing of the sentence required correctly linking the subject ('the doer') and object ('the one being acted upon') in the sentence. This included alternations in ditransitive sentences (direct object indirect object vs indirect object direct object); passive constructions; subject/object relative clauses with reversible and nonreversible noun phrases; and interrogatives (direct vs. embedded questions).
- **Temporal comprehension:** Comprehension required linking the order of events and ranged from easy (with events linearly matching the real-world chronological order) to difficult (mismatch). This was done using causal clauses (porque/para que); temporal clauses (hasta que); future tense (va a), and conditional clauses (si, mientras).
- **Complex noun phrases:** Comprehension required identifying modified noun phrases. This was done using prepositional phrases (entre el árbol y la casa); adjectives (en el carro rojo); and quantifiers (todo, ninguno).

The sentences were represented by illustrations created specifically for the assessment. The following guidelines were provided for the development of the illustrations:

- **Easily Recognizable.** Emphasis was placed on developing illustrations that could be easily identified.
- **Removal of Irrelevant Information.** Unnecessary elements were not included in the image, focusing solely on the essential components of the image related to the content of the sentence.
- **Diversity Representation.** The illustrations were designed to target various aspects of diversity, including different racial backgrounds (through variations in skin tones and hair textures), and diverse abilities such as featuring characters who use wheelchairs, prosthetics, or hearing devices. Cultural representations were carefully considered, encompassing a range of clothing styles, skin tones and physical features to reflect various backgrounds. The Justice, Equity, Diversity, and Inclusion (JEDI) team reviewed all the developed illustrations to enhance diversity in representation.

Dialectal considerations.

12.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

12.5 Calibration Samples

Table 12.1: Demographic Characteristics of Calibration Samples for the English and Spanish Listening Comprehension Tasks

	English			Spanish		
Characteristic	**K** N = 294	**G1** N = 301	**G2** N = 290	**K** N = 239	**G1** N = 278	**G2** N = 260
Timepoint						
Fall 2023	294 (100%)	301 (100%)	290 (100%)	239 (100%)	278 (100%)	260 (100%)
Administration Format						
Forms	294 (100%)	301 (100%)	290 (100%)	239 (100%)	278 (100%)	260 (100%)
Race						
American/Alaskan Native	7 (2.4%)	3 (1.0%)	3 (1.1%)	2 (0.8%)	3 (1.1%)	4 (1.6%)
Asian	39 (13%)	49 (16%)	13 (4.8%)	6 (2.5%)	3 (1.1%)	0 (0%)
Black/African American	29 (10.0%)	31 (10%)	55 (20%)	1 (0.4%)	0 (0%)	0 (0%)
Not reported	44 (15%)	49 (16%)	20 (7.4%)	134 (57%)	183 (67%)	171 (66%)
Other	77 (26%)	50 (17%)	13 (4.8%)	43 (18%)	11 (4.0%)	20 (7.8%)
White	95 (33%)	119 (40%)	165 (61%)	51 (22%)	75 (27%)	63 (24%)
Unknown	3	0	21	2	3	2
Ethnicity						
Hispanic/Latin(o/a)	133 (45%)	127 (42%)	170 (59%)	215 (91%)	255 (92%)	243 (98%)
Intentional nonreport	6 (2.0%)	2 (0.7%)	0 (0%)	1 (0.4%)	0 (0%)	2 (0.8%)
Not Hispanic/Latin(o/a)	155 (53%)	172 (57%)	116 (41%)	20 (8.5%)	22 (7.9%)	2 (0.8%)
Unknown	0	0	4	3	1	13
Gender						
Female	145 (49%)	139 (46%)	144 (50%)	128 (54%)	138 (50%)	139 (53%)
Male	149 (51%)	162 (54%)	143 (50%)	111 (46%)	139 (50%)	121 (47%)
Unknown	0	0	3	0	1	0
Home Language						
English	213 (74%)	225 (75%)	189 (79%)	27 (11%)	33 (12%)	16 (6.2%)
Spanish	39 (14%)	36 (12%)	40 (17%)	207 (88%)	241 (88%)	241 (93%)
Other	36 (13%)	38 (13%)	10 (4.2%)	2 (0.8%)	1 (0.4%)	1 (0.4%)
Unknown	6	2	51	3	3	2
English Proficiency Label						
(Re-)Classified Proficient	25 (10%)	23 (7.8%)	16 (6.7%)	29 (13%)	43 (16%)	33 (14%)
English Learner	58 (23%)	56 (19%)	34 (14%)	184 (83%)	204 (74%)	189 (79%)
English-only	165 (67%)	215 (73%)	188 (79%)	9 (4.1%)	27 (9.9%)	16 (6.7%)
Unknown	46	7	52	17	4	22
Ever IEP/504	20 (8.4%)	27 (10%)	27 (11%)	20 (9.8%)	21 (9.8%)	15 (13%)
Unknown	56	43	51	35	63	143

12.6 Psychometric Analysis

12.6.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 12 items from the English task and 14 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 12.2 summarizes the basic item characteristics, Figure 12.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 12.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Listening Comprehension Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 117**	**N = 105**	**N = 120**	**N = 106**
No. of Responses	171 (101)	172 (103)	151 (91)	152 (90)
Proportion Correct	0.81 (0.15)	0.80 (0.14)	0.72 (0.18)	0.72 (0.16)
Point-biserial Correlation	0.38 (0.13)	0.41 (0.10)	0.36 (0.12)	0.39 (0.09)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	12 (10%)	0 (0%)	14 (12%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

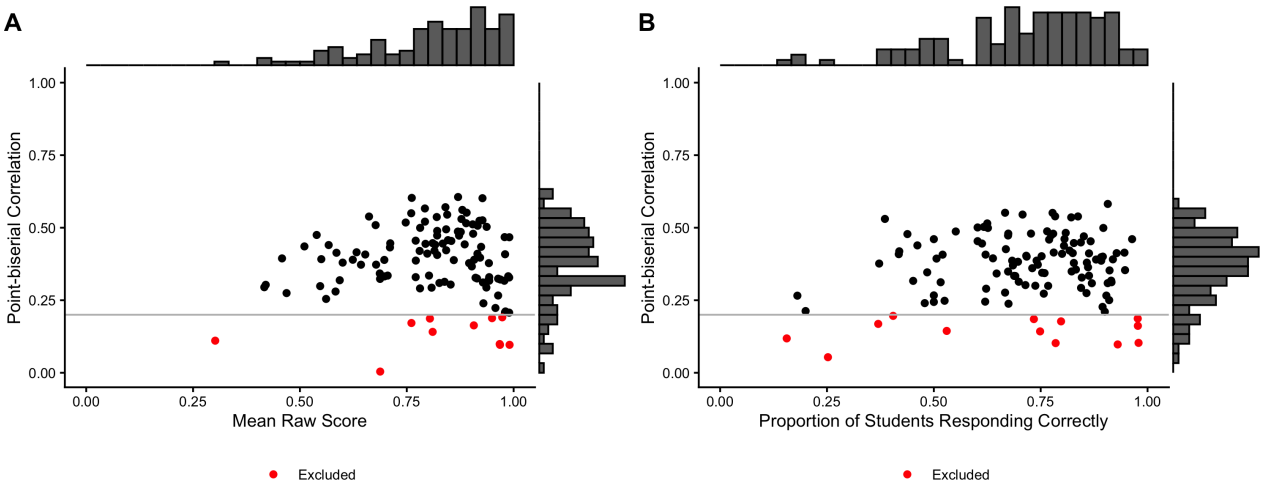

Figure 12.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Listening Comprehension Tasks

12.6.2 Rasch Analysis

12.6.2.1 Item Location Estimates

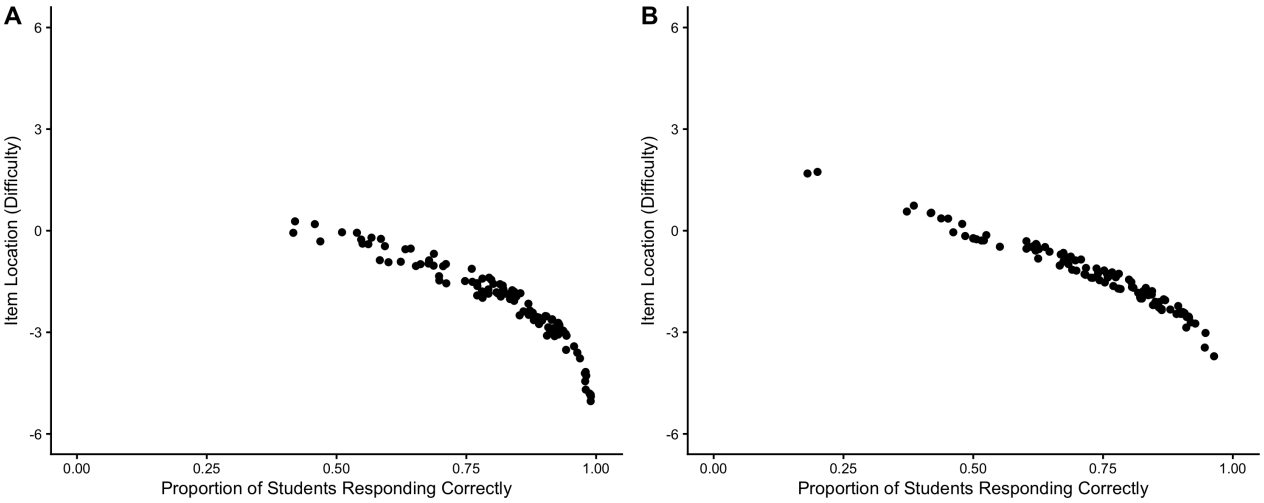

Figure 12.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Listening Comprehension Tasks

12.6.2.2 Item Fit Statistics

Table 12.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Listening Comprehension Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	98	0	0	0	98	105	0	0	0	105
B	6	0	0	0	6	1	0	0	0	1
C	1	0	0	0	1	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0
Total	105	0	0	0	105	106	0	0	0	106

12.6.2.3 Person Location Estimates

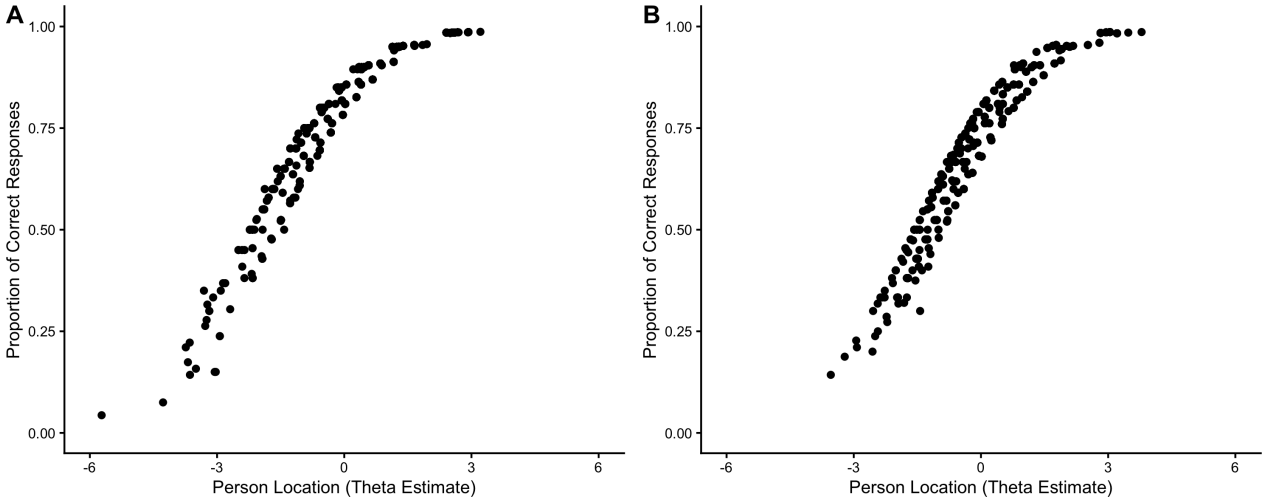

Figure 12.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Listening Comprehension Tasks

12.6.2.4 Person Fit Statistics

Table 12.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Listening Comprehension Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	557	0	0	0	557	641	0	0	0	641
B	187	97	0	0	284	60	39	0	0	99
C	29	0	4	0	33	25	0	1	0	26
D	3	0	2	1	6	1	0	1	0	2
Total	776	97	6	1	880	727	39	2	0	768

12.6.2.5 Distribution of Theta Estimates

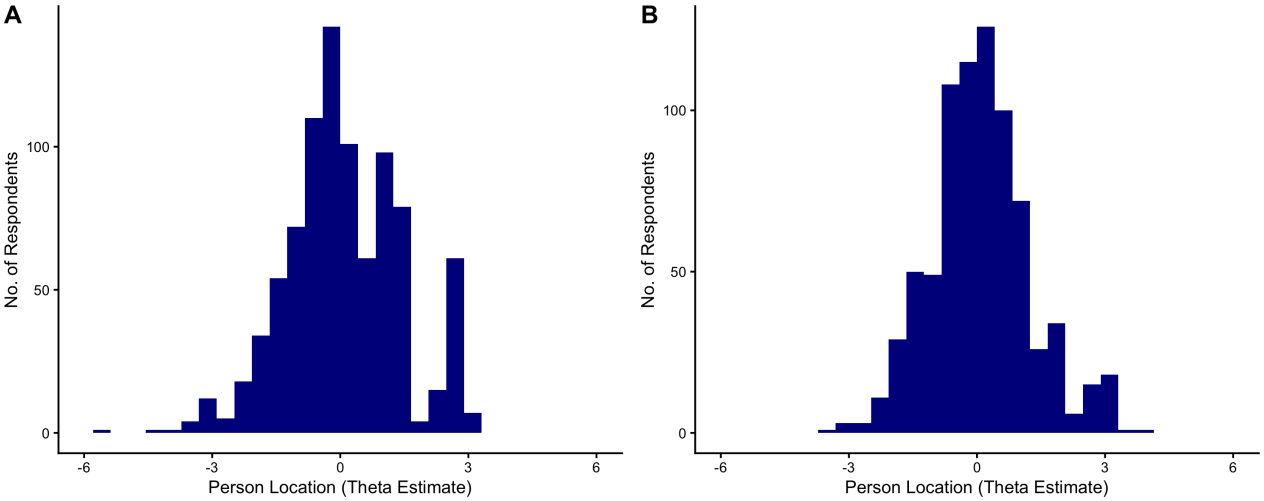

Figure 12.4: Distribution of Theta Estimates for the English and Spanish Listening Comprehension Tasks

12.6.2.6 Wright Maps

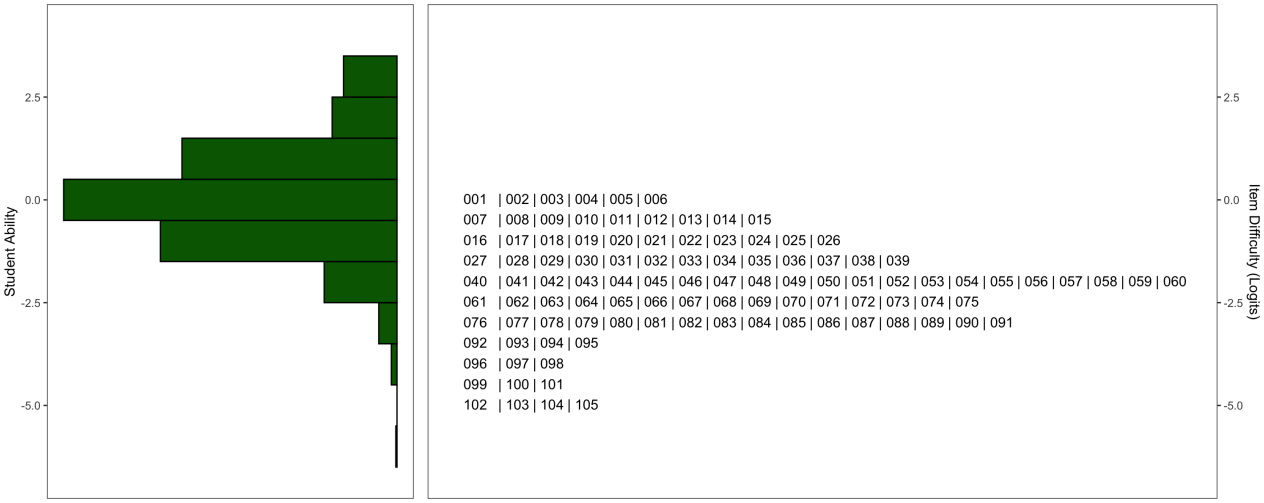

Figure 12.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Listening Comprehension Task

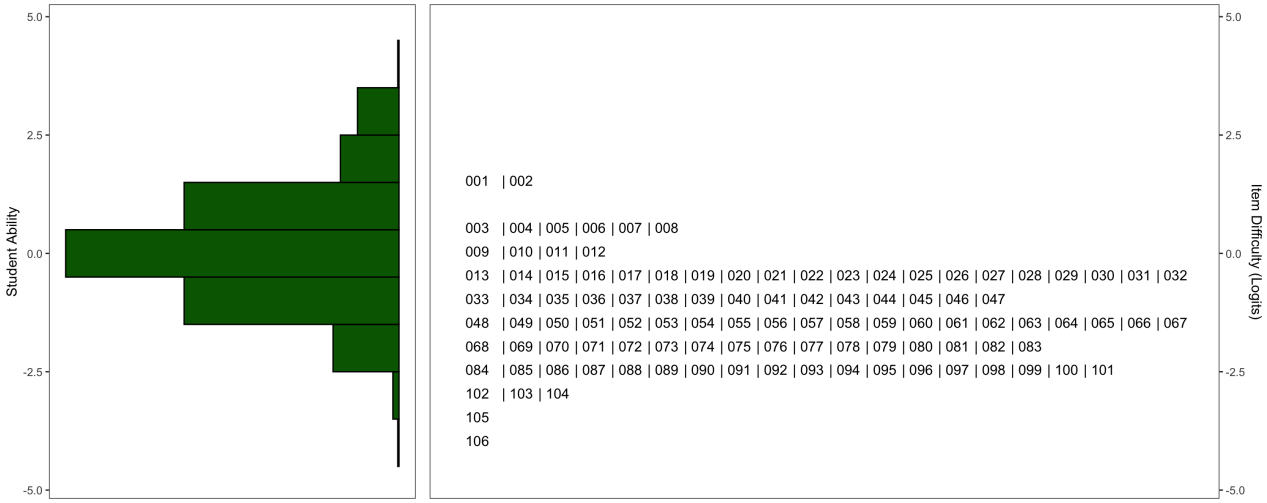

Figure 12.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Listening Comprehension Task

12.6.2.7 Model Summary

Table 12.5: Summary of Rasch Model Statistics for the English and Spanish Listening Comprehension Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 105	N = 880	N = 106	N = 768
Logit Scale Location	-2.07 (1.22)	-0.06 (-0.81, 1.17)	-1.28 (1.02)	0.06 (-0.66, 0.77)
Outfit	0.93 (0.26)	0.69 (0.39, 0.98)	0.98 (0.15)	0.86 (0.66, 1.06)
Infit	0.99 (0.09)	0.86 (0.71, 1.02)	1.00 (0.07)	0.92 (0.79, 1.05)
Reliability of Separation	0.6992	0.5419	0.7060	0.6316

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 105 and 106 for the English and Spanish tasks, respectively.

12.7 Criterion Validity Evidence

12.7.1 Sample

Table 12.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Listening Comprehension Tasks

	English			Spanish		
Characteristic	**K** N = 261	**G1** N = 231	**G2** N = 201	**K** N = 242	**G1** N = 226	**G2** N = 261
Timepoint						
Winter 2024	261 (100%)	231 (100%)	201 (100%)	242 (100%)	226 (100%)	261 (100%)
Race						
American/Alaskan Native	5 (1.9%)	3 (1.3%)	1 (0.5%)	2 (0.8%)	4 (1.8%)	4 (1.5%)
Asian	35 (14%)	36 (16%)	8 (4.4%)	8 (3.3%)	2 (0.9%)	0 (0%)
Black/African American	27 (10%)	30 (13%)	32 (17%)	1 (0.4%)	0 (0%)	0 (0%)
Not reported	30 (12%)	32 (14%)	13 (7.1%)	133 (55%)	154 (69%)	168 (65%)
Other	73 (28%)	45 (19%)	3 (1.6%)	40 (17%)	8 (3.6%)	18 (6.9%)
White	88 (34%)	85 (37%)	126 (69%)	56 (23%)	55 (25%)	69 (27%)
Unknown	3	0	18	2	3	2
Ethnicity						
Hispanic/Latin(o/a)	109 (42%)	98 (42%)	121 (61%)	218 (91%)	210 (93%)	244 (98%)
Intentional nonreport	7 (2.7%)	2 (0.9%)	0 (0%)	1 (0.4%)	0 (0%)	2 (0.8%)
Not Hispanic/Latin(o/a)	145 (56%)	131 (57%)	79 (40%)	20 (8.4%)	16 (7.1%)	2 (0.8%)
Unknown	0	0	1	3	0	13
Gender						
Female	129 (49%)	110 (48%)	97 (48%)	128 (53%)	110 (49%)	134 (51%)
Male	132 (51%)	121 (52%)	104 (52%)	114 (47%)	116 (51%)	127 (49%)
Home Language						
English	190 (74%)	174 (76%)	126 (82%)	29 (12%)	23 (10%)	23 (8.9%)
Spanish	34 (13%)	24 (10%)	22 (14%)	208 (87%)	199 (89%)	235 (91%)
Other	32 (13%)	32 (14%)	5 (3.3%)	2 (0.8%)	1 (0.4%)	1 (0.4%)
Unknown	5	1	48	3	3	2
English Proficiency Label						
(Re-)Classified Proficient	12 (5.6%)	17 (7.5%)	11 (7.2%)	31 (14%)	24 (11%)	39 (16%)
English Learner	49 (23%)	40 (18%)	16 (10%)	185 (81%)	177 (80%)	177 (74%)
English-only	155 (72%)	169 (75%)	126 (82%)	11 (4.8%)	19 (8.6%)	23 (9.6%)
Unknown	45	5	48	15	6	22
Ever IEP/504	20 (10%)	23 (13%)	18 (12%)	20 (9.4%)	23 (11%)	16 (12%)
Unknown	63	47	48	30	15	123

English Listening Comprehension was correlated with the Sentence Comprehension subtest of the Clinical Evaluation of Language

Fundamentals, 5th Edition (CELF 5) test (Wiig, Semel, and Secord 2013). Spanish Listening Comprehension was correlated with the Sentence Comprehension subtest of the Clinical Evaluation of Language Fundamentals, 4th Edition, Spanish (CELF 4 Spanish) test (Semel et al. 2006b).

Table 12.7: Concurrent Criterion Validity Correlations for the English and Spanish Listening Comprehension Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	261	0.51 [0.41, 0.59]	49	0.58 [0.35, 0.74]	242	0.45 [0.35, 0.55]
G1	231	0.37 [0.26, 0.48]	40	0.44 [0.15, 0.66]	226	0.42 [0.31, 0.52]
G2	201	0.42 [0.29, 0.52]	NA	NA	261	0.40 [0.29, 0.50]

13 Letter Naming Fluency

13.1 Task Description

Children are shown a page of assorted lowercase and uppercase letters, arranged in five rows of ten. They are asked to name as many letters as they can within 45 seconds.

13.2 Construct

The Letter Naming Fluency task measures children's fluency in accessing their alphabetic knowledge.

13.3 Item Development

13.3.1 English

All the letters in the English alphabet were included. Letters were written in both upper and lower-case format and were randomly organized in a 5x10 grid.

13.3.2 Spanish

All the letters of the Spanish alphabet were included except for the four letters rarely used in the Spanish language: "q" (qu), "w" (doble v, uve doble), "x" (equis), and "y" (y griega). Letters were written in both upper and lower-case format and were randomly organized in a 5x10 grid.

13.4 Scoring

Participating children were presented with a 5x10 grid containing 50 letters and were asked to name as many letters as they could within 45 seconds. The final score was calculated as the number of correctly named letters.

13.5 Samples

Table 13.1: Demographic Characteristics of Samples for the English and Spanish Letter Naming Fluency Tasks

	English		Spanish	
Characteristic	**K** N = 2,932	**G1** N = 258	**K** N = 1,362	**G1** N = 202
Timepoint				
Fall 2023	460 (18%)	0 (0%)	433 (38%)	0 (0%)
Fall 2024	2,157 (82%)	258 (100%)	699 (62%)	202 (100%)
Unknown	315	0	230	0
Administration Format				
Not applicable	2,932 (100%)	258 (100%)	1,362 (100%)	202 (100%)
Race				
American/Alaskan Native	101 (3.8%)	12 (4.7%)	45 (3.5%)	8 (4.0%)
Asian	187 (7.0%)	34 (13%)	48 (3.7%)	9 (4.5%)
Black/African American	241 (9.1%)	40 (16%)	19 (1.5%)	1 (0.5%)
Not reported	329 (12%)	47 (18%)	460 (35%)	61 (30%)
Other	622 (23%)	30 (12%)	302 (23%)	38 (19%)
White	1,182 (44%)	93 (36%)	424 (33%)	85 (42%)
Unknown	270	2	64	0
Ethnicity				
Hispanic/Latin(o/a)	1,727 (71%)	146 (57%)	1,137 (95%)	197 (98%)
Intentional nonreport	31 (1.3%)	2 (0.8%)	5 (0.4%)	1 (0.5%)
Not Hispanic/Latin(o/a)	686 (28%)	107 (42%)	58 (4.8%)	3 (1.5%)
Unknown	488	3	162	1
Gender				
Female	1,233 (50%)	141 (56%)	625 (52%)	111 (56%)
Male	1,257 (50%)	113 (44%)	573 (48%)	88 (44%)
Unknown	442	4	164	3
Home Language				
English	1,570 (63%)	231 (91%)	181 (14%)	0 (0%)
Spanish	809 (33%)	8 (3.1%)	1,101 (85%)	202 (100%)
Other	96 (3.9%)	15 (5.9%)	6 (0.5%)	0 (0%)
Unknown	457	4	74	0
English Proficiency Label				
(Re-)Classified Proficient	85 (3.8%)	7 (2.8%)	100 (8.7%)	25 (14%)
English Learner	743 (33%)	25 (9.9%)	919 (80%)	154 (85%)
English-only	1,410 (63%)	220 (87%)	129 (11%)	2 (1.1%)
Unknown	694	6	214	21
Ever IEP/504	165 (8.4%)	19 (8.9%)	104 (9.6%)	16 (9.3%)
Unknown	965	45	284	30

13.6 Score distribution

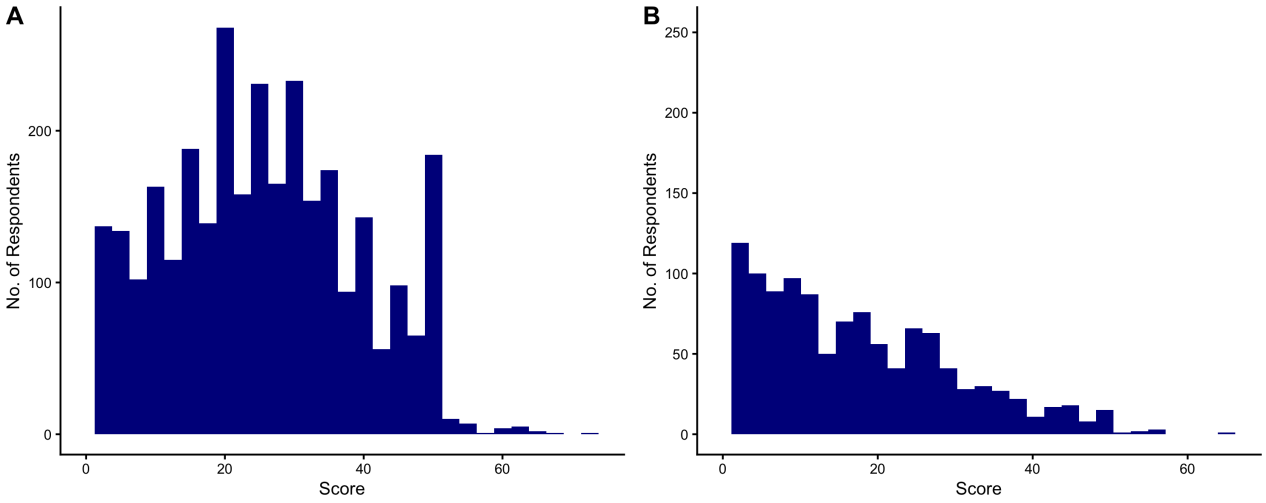

Figure 13.1: Score Distribution of the English and Spanish Letter Naming Fluency Tasks

13.7 Criterion Validity Evidence

13.7.1 Sample

Table 13.2: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Letter Naming Fluency Tasks

	English	**Spanish**
Characteristic	**K** N = 60	**K** N = 139
Timepoint		
Spring 2024	60 (100%)	139 (100%)
Race		
American/Alaskan Native	1 (1.7%)	6 (4.3%)
Asian	8 (13%)	11 (7.9%)
Black/African American	8 (13%)	6 (4.3%)
Not reported	12 (20%)	37 (27%)
Other	7 (12%)	22 (16%)
White	24 (40%)	57 (41%)
Ethnicity		
Hispanic/Latin(o/a)	32 (53%)	122 (88%)
Intentional nonreport	1 (1.7%)	1 (0.7%)
Not Hispanic/Latin(o/a)	27 (45%)	15 (11%)
Gender		
Female	34 (57%)	77 (55%)
Male	26 (43%)	62 (45%)
Home Language		
English	43 (72%)	29 (21%)
Spanish	14 (23%)	110 (79%)
Other	3 (5.0%)	
English Proficiency Label		
(Re-)Classified Proficient	4 (7.7%)	12 (9.4%)
English Learner	8 (15%)	87 (68%)
English-only	40 (77%)	29 (23%)
Unknown	8	11
Ever IEP/504	1 (5.3%)	12 (9.2%)
Unknown	41	9
Unknown		1

Letter Naming Fluency English was correlated with the Receptive One-Word Picture Vocabulary Test, 4th Edition (Martin and Brownell 2011). Letter Naming Fluency Spanish was correlated with the Receptive One-Word Picture Vocabulary Test, 4th Edition, Bilingual Edition (Martin 2013) test.

Table 13.3: Concurrent Criterion Validity Correlations for the English and Spanish Letter Naming Fluency Tasks

English	Spanish

	All		All	
Grade	n	r [CI]	n	r [CI]
K	60	0.60 [0.41, 0.74]	139	0.88 [0.83, 0.91]

14 Letter Sound Fluency

14.1 Task Description

Children are shown a page of assorted lowercase and uppercase letters, arranged in five rows of ten. They are asked to say the sounds of as many letters as they can within 45 seconds.

14.2 Construct

The Letter Sound Fluency task measures children's fluency in saying the sounds of the alphabet.

14.3 Item Development

14.3.1 English

All the letters of the English alphabet were used. For letters representing more than one phoneme, all corresponding phonemes were accepted as correct, see Table 14.1. It was common for children to add "uh" to the end of consonant sounds, and that was scored as correct.

Table 14.1: Acceptable Responses for English Letters

Scenario	Child says
Letter "a"	/a/ as in "hat" or "hate"
Letter "e"	/e/ as in "get" or "he"
Letter "i"	/i/ as in "bit" or "bite"
Letter "o"	/o/ as in "got" or "go"
Letter "u"	/u/ as in "hug", "rule" or "mule"
Letter "c"	/k/ as in "cat" or /s/ as in "nice"
Letter "g"	/g/ as in "goat" or /j/ as in "page"
Letter "y"	/e/ as in "yellow" or "happy" or /ai/ as in "cry" or /i/ as in "gym"

14.3.2 Spanish

All the phonemes of the Spanish alphabet were used, including the phonemes that are not present in the English language (e.g., /ñ/) and phonemes that are represented with more than one letter (e.g., /ch/and /ll/). The only exceptions were the letter “h”, as it does not have a sound in Spanish, and letters “q” and “x”, as they cannot be represented as single phonemes. For letters representing more than one phoneme, all corresponding phonemes were accepted as correct, see Table 14.2.

Table 14.2: Acceptable Responses for Spanish Letters

Scoring Events	Child says
Letter ”c”	/k/ as in ”casa” or /s/ as in ”cebolla”
Letter ”g”	soft /g/ as in ”gato” or hard /x/ as in ”gente”
Letter ”r”	trilled sound as in ”ratón” or soft sound as in ”cara”
Letter ”y”	/i/ as in ”rey” or /ll/ as in ”yema”

14.4 Scoring

Participating children were presented with a 5x10 grid containing 50 letters and were asked to produce the sounds of the letters as fast as they could within 45 seconds. The final score was calculated as the number of correct responses.

14.5 Samples

Table 14.3: Demographic Characteristics of Samples for the English and Spanish Letter Sound Fluency Tasks

	English			Spanish		
Characteristic	**K** N = 212	**G1** N = 3,215	**G2** N = 564	**K** N = 9	**G1** N = 1,283	**G2** N = 240
Timepoint						
Fall 2023	0 (0%)	396 (14%)	346 (62%)	0 (NA%)	381 (37%)	71 (30%)
Fall 2024	70 (100%)	2,426 (86%)	216 (38%)	0 (NA%)	640 (63%)	168 (70%)
Unknown	142	393	2	9	262	1
Administration Format						
Not applicable	212 (100%)	3,215 (100%)	564 (100%)	9 (100%)	1,283 (100%)	240 (100%)
Race						
American/Alaskan Native	11 (5.7%)	111 (3.8%)	3 (0.6%)	0 (0%)	47 (3.7%)	4 (1.7%)
Asian	12 (6.3%)	232 (8.0%)	52 (9.8%)	1 (11%)	39 (3.1%)	4 (1.7%)
Black/African American	9 (4.7%)	285 (9.8%)	65 (12%)	1 (11%)	20 (1.6%)	1 (0.4%)
Not reported	34 (18%)	362 (12%)	98 (18%)	0 (0%)	513 (41%)	116 (49%)
Other	71 (37%)	460 (16%)	64 (12%)	0 (0%)	132 (10%)	12 (5.0%)
White	55 (29%)	1,468 (50%)	249 (47%)	7 (78%)	510 (40%)	102 (43%)
Unknown	20	297	33	0	22	1
Ethnicity						
Hispanic/Latin(o/a)	117 (82%)	2,017 (69%)	261 (50%)	8 (89%)	1,182 (95%)	226 (95%)
Intentional nonreport	0 (0%)	14 (0.5%)	2 (0.4%)	0 (0%)	2 (0.2%)	0 (0%)
Not Hispanic/Latin(o/a)	26 (18%)	897 (31%)	258 (50%)	1 (11%)	63 (5.1%)	12 (5.0%)
Unknown	69	287	43	0	36	2
Gender						
Female	100 (47%)	1,459 (50%)	236 (45%)	5 (56%)	648 (53%)	122 (51%)
Male	112 (53%)	1,450 (50%)	294 (55%)	4 (44%)	572 (47%)	116 (49%)
Unknown	0	306	34	0	63	2
Home Language						
English	87 (68%)	1,686 (64%)	431 (82%)	0 (0%)	150 (12%)	8 (3.3%)
Spanish	37 (29%)	886 (33%)	55 (10%)	9 (100%)	1,103 (88%)	231 (97%)
Other	4 (3.1%)	79 (3.0%)	38 (7.3%)	0 (0%)	4 (0.3%)	0 (0%)
Unknown	84	564	40	0	26	1
English Proficiency Label						
(Re-)Classified Proficient	0 (0%)	141 (5.6%)	27 (5.1%)	1 (11%)	147 (12%)	48 (21%)
English Learner	35 (59%)	799 (32%)	65 (12%)	8 (89%)	933 (77%)	171 (74%)
English-only	24 (41%)	1,582 (63%)	434 (83%)	0 (0%)	129 (11%)	12 (5.2%)
Unknown	153	693	38	0	74	9
Ever IEP/504	15 (26%)	205 (9.2%)	48 (12%)	1 (11%)	106 (9.8%)	11 (6.0%)
Unknown	154	994	172	0	201	58

14.6 Score distribution

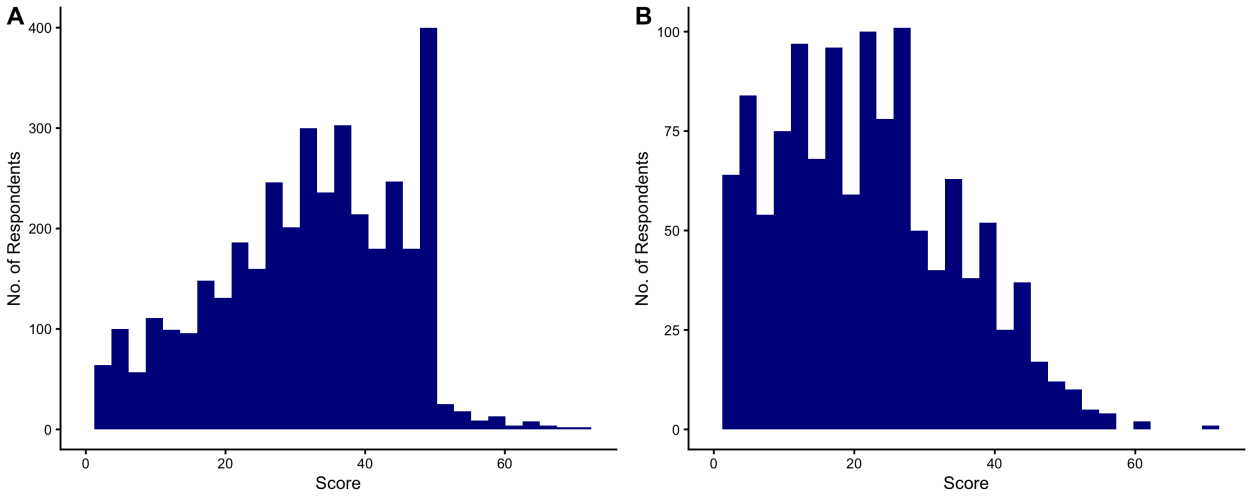

Figure 14.1: Score Distribution of the English and Spanish Letter Sound Fluency Tasks

14.7 Criterion Validity Evidence

14.7.1 Sample

Table 14.4: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Letter Sound Fluency Tasks

	English			Spanish
Characteristic	**K** N = 70	**G1** N = 129	**G2** N = 1	**G1** N = 140
Timepoint				
Spring 2024	0 (0%)	59 (46%)	1 (100%)	140 (100%)
Spring 2025	70 (100%)	70 (54%)	0 (0%)	
Race				
American/Alaskan Native	4 (6.3%)	3 (2.4%)	0 (0%)	4 (2.9%)
Asian	4 (6.3%)	13 (10%)	0 (0%)	6 (4.3%)
Black/African American	3 (4.7%)	13 (10%)	0 (0%)	4 (2.9%)
Not reported	13 (20%)	18 (15%)	0 (0%)	36 (26%)
Other	22 (34%)	28 (23%)	0 (0%)	7 (5.1%)
White	18 (28%)	49 (40%)	1 (100%)	81 (59%)
Unknown	6	5	0	2
Ethnicity				
Hispanic/Latin(o/a)	41 (84%)	49 (39%)	0 (0%)	124 (89%)
Intentional nonreport	0 (0%)	1 (0.8%)	0 (0%)	
Not Hispanic/Latin(o/a)	8 (16%)	75 (60%)	1 (100%)	16 (11%)
Unknown	21	4	0	
Gender				
Female	31 (44%)	59 (46%)	0 (0%)	69 (49%)
Male	39 (56%)	70 (54%)	1 (100%)	71 (51%)
Home Language				
English	27 (64%)	80 (73%)	1 (100%)	32 (23%)
Spanish	14 (33%)	14 (13%)	0 (0%)	104 (75%)
Other	1 (2.4%)	15 (14%)	0 (0%)	2 (1.4%)
Unknown	28	20	0	2
English Proficiency Label				
(Re-)Classified Proficient	0 (0%)	8 (9.9%)	0 (0%)	17 (12%)
English Learner	13 (62%)	20 (25%)	0 (0%)	90 (65%)
English-only	8 (38%)	53 (65%)	1 (100%)	31 (22%)
Unknown	49	48	0	2
Ever IEP/504	6 (29%)	4 (5.6%)	0 (NA%)	13 (11%)
Unknown	49	58	1	23

English Letter Sound Fluency was correlated with the Letter-Word Identification subtest from the Woodcock-Johnson IV (WJ IV ACH) test (Schrank, McGrew, and Mather 2014). Letter Sound Fluency Spanish was correlated with the Identificación de Letras y Palabras subtest from the Batería IV Woodcock-Muñoz (Batería IV APROV) test (Woodcock et al. 2019). The Letter-Word Identification and Identificación de Letras y Palabras subtests require the reading of words, moving beyond letter-sound fluency. Given that the tasks do not completely align, the lower correlations are not surprising.

Table 14.5: Concurrent Criterion Validity Correlations for the English and Spanish Letter Sound Fluency Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	70	0.62 [0.45, 0.74]	NA	NA	140	0.52 [0.39, 0.63]
G1	129	0.20 [0.03, 0.36]	20	0.09 [-0.37, 0.51]	140	0.52 [0.39, 0.63]

15 Nonword Reading

15.1 Task Description

Children are asked to read nonsense words that follow regular sound-spelling rules.

15.2 Construct

The Nonword Reading task measures the construct of decoding accuracy, the ability to translate print into speech by correctly pairing graphemes (letters) with their corresponding phonemes (sounds) using pronounceable nonsense words.

15.3 Item Development

15.3.1 English

The constructed items followed the English phonotactic structure, featuring diverse syllable constructions that reflect the language's phonological patterns. Specific orthographic patterns that matched curriculum-based learning goals were targeted in the development of the items. Below, we provide the list of targets and the considerations for nonword development for each one of those targets.

- **Predictable consonants:** Nonwords with predictable consonants (i.e., m, s, t, l; p, f, c (/k/), n; b, r, j, k; v, g (/g/), w, d; h, y, z, x), and predictable short vowels.
- **Consonant digraphs:** Nonwords with consonant digraphs (i.e., ch, wh, th, ng) and predictable short vowels.
- **Two-consonant blends:** Nonwords with two-syllable blends (e.g., qu, st, sm, sn, -st, -ft, - lp; sr, sl, cr, cl, tr, dr) and predictable short vowels.
- **Single consonants:** Nonwords with single consonants (e.g., /s/ for c, s; /z/ for s, z; /k/ for k, c, -ck after a short vowel; /g/ for j, g)
- **Hard and soft c and g:** Nonwords with hard and soft c and g, with predictable short vowels, and VCe long vowel pattern in single-syllable words.
- **Final consonant blends with nasals:** Nonwords with final nasal consonant blends (i.e., nt, nd, mp, nk) with predictable short vowels.
- **VCe long vowel pattern in single-syllable words:** Nonwords with VCe long vowel pattern in single-syllable words, including also predictable consonants, consonant diagraphs, and/or two consonant blends.

- **Vowel teams for long vowel sounds:** Nonwords with vowel teams for long vowel sounds (e.g., ee, ea; ai, ay; oa, ow, oe; igh), including also predictable consonants, consonant digraphs, and/or two consonant blends.
- **Vowel-r combinations with single syllables:** Nonwords with vowel-r combinations and single syllables (i.e., er, ar, or, ir, ur) with predictable short vowels.
- **Other vowel-r combinations:** Nonwords with other vowel-r combinations (e.g., are, air, our, ore, ear, eer, ure), including also predictable consonants.
- **Diphthongs and vowels:** Nonwords with diphthongs and vowels (e.g., /aw/ and /oo/: oi, oy; ou, ow; au, aw; oo, u), including also predictable consonants, consonant digraphs, and/or two consonant blends.
- **Use of y:** Nonwords with y as consonant /y/, as /ɪ/ on ends of one-syllable words like fly, as /e/ on ends of multisyllabic words like lobby, or as /ɪ/ in a few words like myth.
- **Use of -ild, -ost, -old, -olt, -ind pattern:** Nonwords with -ild, -ost, -old, -olt, -ind, including also predictable consonants and/or consonant digraphs.
- **Digraphs:** Nonwords with figraphs (e.g., ph (/f/), gh (/f/), ch (/k/ and /sh/))
- **Trigraphs:** Nonwords with trigraphs (e.g., -tch (/ch/), -dge (/j/)).
- **Three-consonant blends and blends with digraphs:** Nonwords with three-consonant blends and blends with digraphs (e.g., squ, str, scr, thr, shr) and predictable short vowels.

15.3.2 Spanish

The constructed items followed the Spanish phonotactic structure, featuring diverse syllable constructions that reflect the language's phonological patterns. The length of the items ranged from succinct two-phoneme combinations to more intricate six-phoneme combinations. While two-phoneme words are relatively uncommon in the Spanish lexicon, we deliberately incorporated two-phoneme nonwords into our dataset to include easier items. Below, we provide the list of targets and the considerations for nonword development for each one of those targets.

- **Predictable consonants:** Nonwords with predictable consonants (i.e., m, s, t, l; p, f, n; b, r, j, v, d, y, z).
- **Consonant digraphs:** Nonwords with consonant digraphs (i.e., ch, ll, rr).
- **Two-consonant blends:** Nonwords with two-syllable blends (e.g., br, bl, cr, cl, tr, dr).
- **Single consonants:** Nonwords with single consonants (e.g., /s/ for c, z; /g/ for j, g)
- **Hard and soft c and g:** Nonwords with hard and soft c and g.
- **Diphthongs and vowels:** Nonwords with diphthongs and vowels (e.g., /ai/, /ei/), including also predictable consonants, consonant digraphs, and/or two consonant blends.
- **Hiatus:** sequence of vowels belonging to different syllables (e.g., /ia/, /ua/).
- **Stress types:** words that with antepenultimate stress, penultimate stress, and ultimate stress.

15.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

15.5 Calibration Samples

Table 15.1: Demographic Characteristics of Calibration Samples for the English and Spanish Nonword Reading Tasks

	English		Spanish	
Characteristic	**G1** N = 1,019	**G2** N = 1,062	**G1** N = 779	**G2** N = 670
Timepoint				
Winter 2024	607 (88%)	648 (89%)	0 (NA%)	436 (100%)
Fall 2024	81 (12%)	78 (11%)	0 (NA%)	0 (0%)
Unknown	331	336	779	234
Administration Format				
CAT	412 (40%)	414 (39%)	202 (26%)	234 (35%)
Forms	607 (60%)	648 (61%)	577 (74%)	436 (65%)
Race				
American/Alaskan Native	20 (2.0%)	11 (1.1%)	4 (2.0%)	8 (1.2%)
Asian	124 (12%)	114 (11%)	6 (3.0%)	7 (1.1%)
Black/African American	110 (11%)	143 (14%)	4 (2.0%)	7 (1.1%)
Not reported	195 (19%)	223 (21%)	79 (40%)	381 (57%)
Other	134 (13%)	95 (9.2%)	10 (5.0%)	31 (4.7%)
White	429 (42%)	452 (44%)	97 (49%)	232 (35%)
Unknown	7	24	579	4
Ethnicity				
Hispanic/Latin(o/a)	527 (52%)	575 (55%)	183 (91%)	597 (91%)
Intentional nonreport	7 (0.7%)	4 (0.4%)	0 (0%)	4 (0.6%)
Not Hispanic/Latin(o/a)	480 (47%)	469 (45%)	19 (9.4%)	56 (8.5%)
Unknown	5	14	577	13
Gender				
Female	487 (48%)	522 (50%)	109 (54%)	374 (56%)
Male	531 (52%)	527 (50%)	93 (46%)	295 (44%)
Non-binary	0 (0%)	0 (0%)	0 (0%)	1 (0.1%)
Unknown	1	13	577	0
Home Language				
English	738 (75%)	715 (72%)	44 (22%)	103 (16%)
Spanish	154 (16%)	164 (17%)	154 (77%)	547 (83%)
Other	97 (9.8%)	112 (11%)	2 (1.0%)	7 (1.1%)
Unknown	30	71	579	13
English Proficiency Label				
(Re-)Classified Proficient	80 (8.4%)	82 (8.3%)	33 (17%)	99 (16%)
English Learner	183 (19%)	188 (19%)	131 (66%)	441 (70%)
English-only	689 (72%)	719 (73%)	36 (18%)	94 (15%)
Unknown	67	73	579	36
Ever IEP/504	81 (9.7%)	83 (12%)	18 (11%)	49 (10%)
Unknown	182	353	613	194

15.6 Psychometric Analysis

15.6.1 Basic Item Statistics

We excluded 16 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 1 items in the English task. Additionally, we excluded 0 items from the English task and 0 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 15.2 summarizes the basic item characteristics, Figure 15.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 15.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Nonword Reading Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 143**	**N = 127**	**N = 170**	**N = 170**
No. of Responses	211 (141)	234 (132)	231 (148)	231 (148)
Proportion Correct	0.44 (0.18)	0.44 (0.16)	0.51 (0.14)	0.51 (0.14)
Point-biserial Correlation	0.63 (0.08)	0.63 (0.08)	0.64 (0.08)	0.64 (0.08)
Excluded (n < 90)	16 (11%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (no variation)	1 (0.7%)	0 (0%)	0 (0%)	0 (0%)

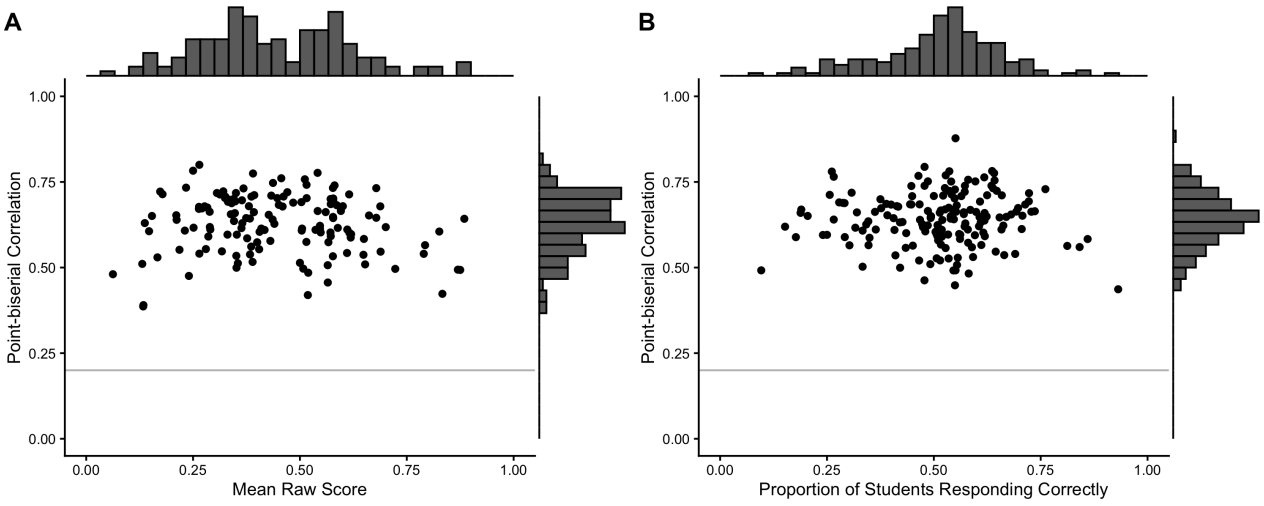

Figure 15.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Nonword Reading Tasks

15.6.2 Rasch Analysis

15.6.2.1 Item Location Estimates

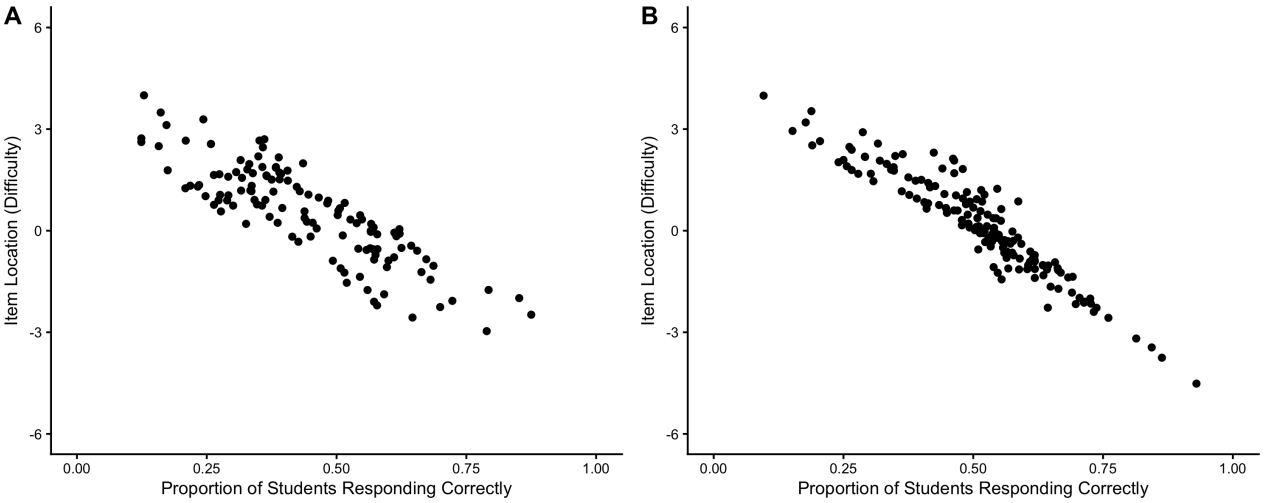

Figure 15.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Nonword Reading Tasks

15.6.2.2 Item Fit Statistics

Table 15.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Nonword Reading Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	111	0	0	0	111	153	0	0	0	153
B	5	0	0	0	5	4	0	0	0	4
C	8	0	0	0	8	9	0	0	0	9
D	3	0	0	0	3	4	0	0	0	4
Total	127	0	0	0	127	170	0	0	0	170

15.6.2.3 Person Location Estimates

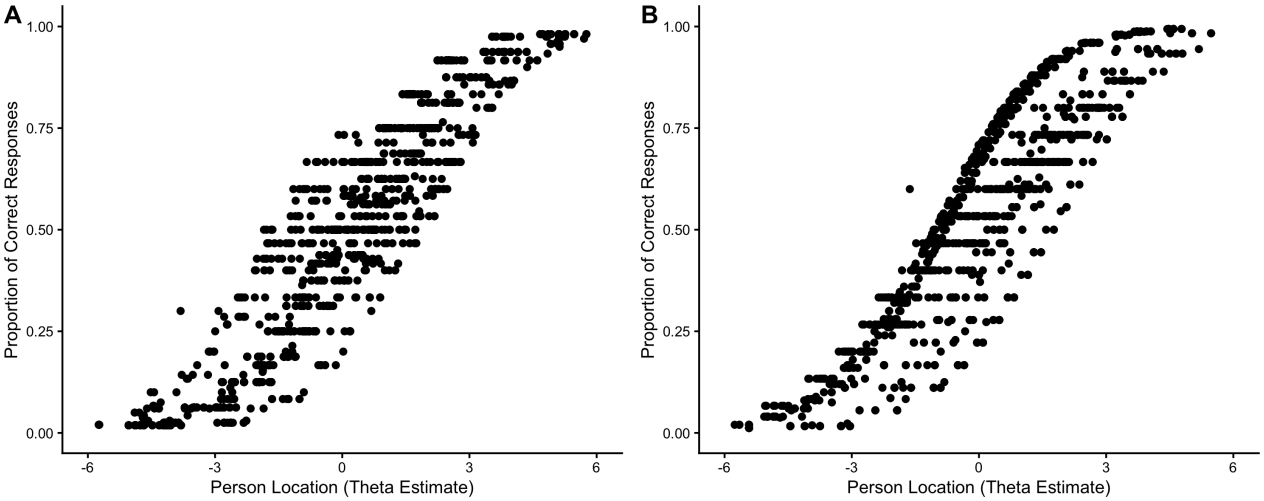

Figure 15.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Nonword Reading Tasks

15.6.2.4 Person Fit Statistics

Table 15.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Nonword Reading Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	1,204	0	4	0	1,208	1,028	0	1	0	1,029
B	279	421	0	0	700	143	172	0	0	315
C	77	0	11	0	88	43	0	5	0	48
D	43	0	18	2	63	33	0	9	0	42
Total	1,603	421	33	2	2,059	1,247	172	15	0	1,434

15.6.2.5 Distribution of Theta Estimates

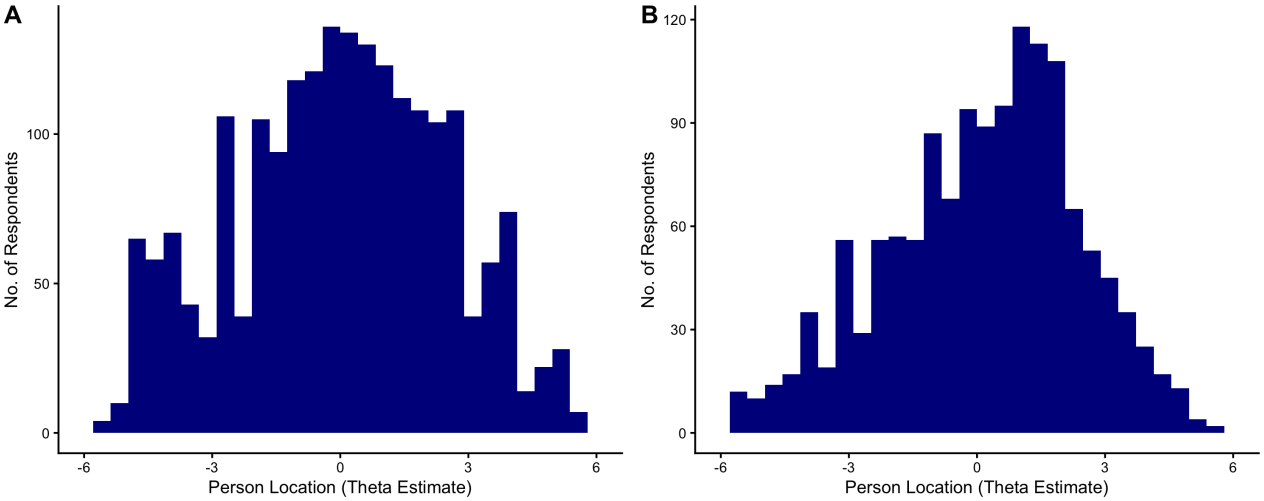

Figure 15.4: Distribution of Theta Estimates for the English and Spanish Nonword Reading Tasks

15.6.2.6 Wright Maps

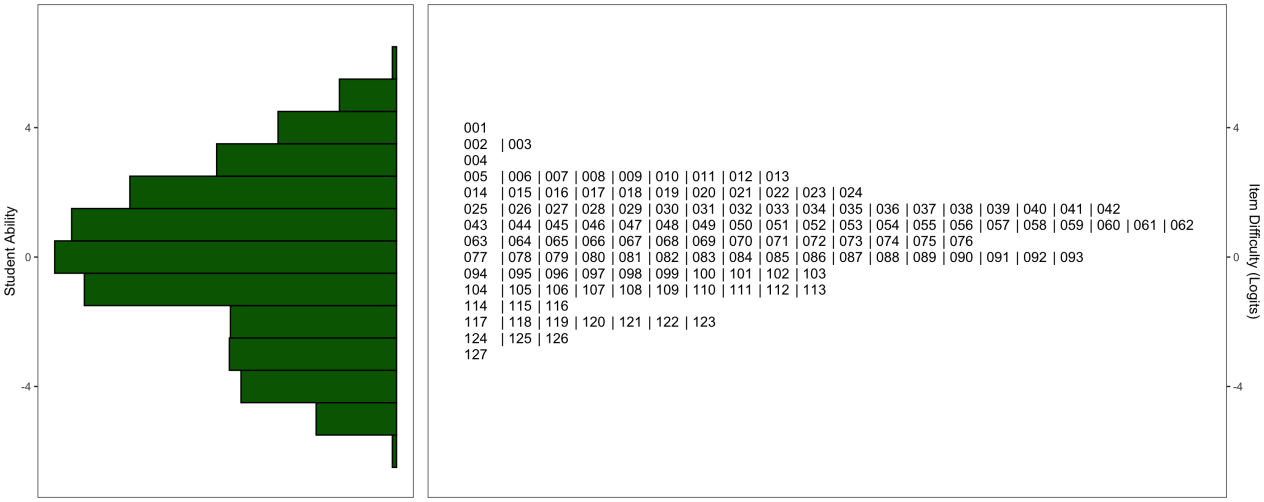

Figure 15.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Nonword Reading Task

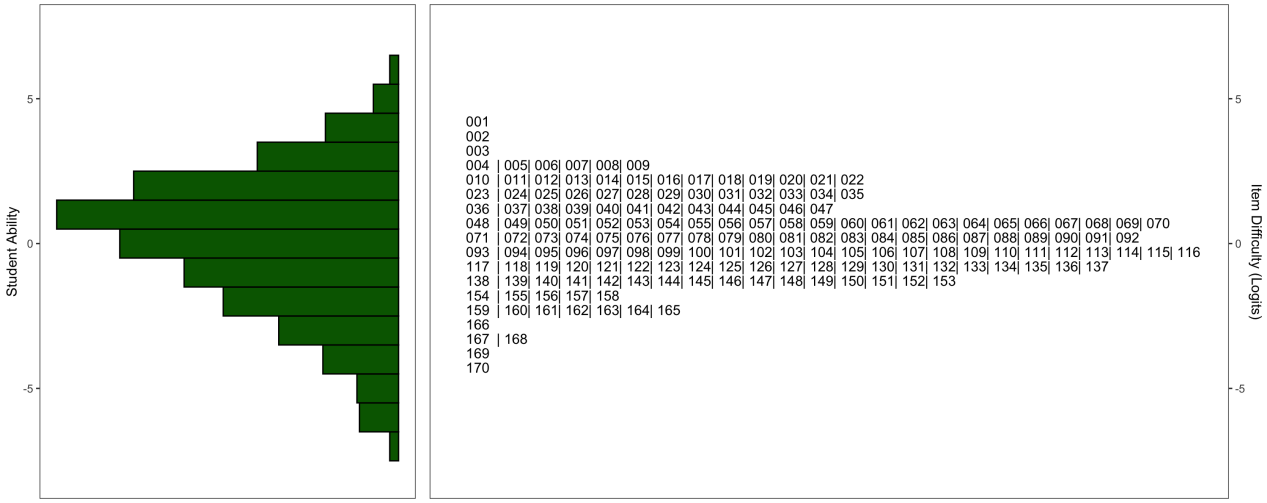

Figure 15.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Nonword Reading Task

15.6.2.7 Model Summary

Table 15.5: Summary of Rasch Model Statistics for the English and Spanish Nonword Reading Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 127	N = 2,059	N = 170	N = 1,434
Logit Scale Location	0.55 (1.42)	0.05 (-1.69, 1.83)	0.16 (1.46)	0.34 (-1.53, 1.76)
Outfit	0.99 (0.40)	0.66 (0.35, 0.93)	0.96 (0.35)	0.80 (0.55, 0.96)
Infit	0.99 (0.12)	0.82 (0.58, 1.00)	0.98 (0.12)	0.88 (0.75, 0.99)
Reliability of Separation	0.8891	0.8279	0.9269	0.8892

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 127 and 170 for the English and Spanish task, respectively.

15.7 Criterion Validity Evidence

15.7.1 Sample

Table 15.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Nonword Reading Tasks

	English			Spanish	
Characteristic	**K** N = 79	**G1** N = 223	**G2** N = 258	**G1** N = 194	**G2** N = 229
Timepoint					
Spring 2024	0 (0%)	223 (100%)	258 (100%)	194 (100%)	229 (100%)
Spring 2025	79 (100%)	0 (0%)	0 (0%)		
Race					
American/Alaskan Native	4 (5.5%)	5 (2.2%)	1 (0.4%)	4 (2.1%)	4 (1.8%)
Asian	5 (6.8%)	25 (11%)	33 (13%)	6 (3.1%)	3 (1.3%)
Black/African American	4 (5.5%)	27 (12%)	32 (12%)	4 (2.1%)	4 (1.8%)
Not reported	13 (18%)	55 (25%)	68 (26%)	75 (39%)	97 (43%)
Other	22 (30%)	35 (16%)	26 (10%)	10 (5.2%)	15 (6.6%)
White	25 (34%)	76 (34%)	98 (38%)	93 (48%)	104 (46%)
Unknown	6	0	0	2	2
Ethnicity					
Hispanic/Latin(o/a)	49 (84%)	102 (46%)	140 (54%)	175 (90%)	201 (88%)
Intentional nonreport	0 (0%)	2 (0.9%)	0 (0%)	0 (0%)	2 (0.9%)
Not Hispanic/Latin(o/a)	9 (16%)	119 (53%)	118 (46%)	19 (9.8%)	26 (11%)
Unknown	21	0	0		
Gender					
Female	36 (46%)	98 (44%)	126 (49%)	103 (53%)	129 (56%)
Male	43 (54%)	125 (56%)	132 (51%)	91 (47%)	100 (44%)
Home Language					
English	27 (53%)	161 (73%)	177 (69%)	44 (23%)	57 (25%)
Spanish	23 (45%)	37 (17%)	41 (16%)	146 (76%)	166 (74%)
Other	1 (2.0%)	23 (10%)	39 (15%)	2 (1.0%)	1 (0.4%)
Unknown	28	2	1	2	5
English Proficiency Label					
(Re-)Classified Proficient	1 (3.3%)	22 (10%)	23 (9.0%)	33 (17%)	37 (17%)
English Learner	21 (70%)	47 (22%)	59 (23%)	123 (64%)	131 (59%)
English-only	8 (27%)	148 (68%)	173 (68%)	36 (19%)	54 (24%)
Unknown	49	6	3	2	7
Ever IEP/504	7 (23%)	22 (11%)	29 (14%)	16 (10%)	17 (9.2%)
Unknown	49	24	47	36	45

English Nonword Reading was correlated with the Word Attack subtest from the Woodcock-Johnson IV (WJ IV ACH) test (Schrank, McGrew, and Mather 2014). Spanish Nonword Reading was correlated with the Análisis de palabras subtest from the Batería IV Woodcock-Muñoz (Batería IV APROV) test (Woodcock et al. 2019).

Table 15.7: Concurrent Criterion Validity Correlations for the English and Spanish Nonword Reading Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
G1	222	0.75 [0.69, 0.80]	47	0.77 [0.62, 0.86]	194	0.82 [0.77, 0.87]
G2	258	0.71 [0.65, 0.77]	59	0.77 [0.64, 0.86]	229	0.75 [0.69, 0.81]

16 Narrative Story Production

16.1 Task Description

Children watch a short video. The video tells a story without any words. After, children are asked to tell everything that happened in the story. Then, they answer a conceptually related question.

16.2 Construct

The Narrative Story Production task evaluates the child's understanding of a story structure and their ability to develop and verbally communicate a narrative.

16.3 Item Development

Five initial videos were proposed to assess children's narrative production. The script blueprint for the animation of these videos were developed with the following considerations in mind: (1) each video had 5 or 6 events, structured around a description of a problem, a resolution attempt, and the actual resolution of the problem, (2) the story of the video had to include a clear emotional response, and (3) the characters and scenarios had to be familiar for the majority of the children in California. Four videos are currently available for selection: one involves building a sandcastle, another depicts packing up toys in preparation for a beach outing, a third shows a bear obtaining an ice cream cone, and the final video features a beach umbrella blowing away. The task is available in two formats: a bilingual version in which prompts are delivered in both Spanish and English, and an English-only version designed for monolingual English speakers.

16.4 Scoring

A complex scoring schema was developed focused on the macrostructure of the narrative. The scoring schema had three main sections: problem, solution, and a semantic question. The problem section consisted of naming the main problem and four possible types of supporting details, yielding a final score between 0 and 5. The solution section consisted of naming the main solution and four possible types of supporting details, yielding a final score between 0 and 5. Finally, the semantic question which involved a prompt such as, "name as many flavors of ice cream as you can." A partial scoring model was used to capture variablity in the number of items children could name. This item was scored as incorrect if children provided an incorrect or off topic response (0), if the child

provided 1-2 examples they achieved a score of (1), and if they named 2 or more items the response was scored a (2).

16.5 Samples

Table 16.1: Demographic Characteristics of Samples for the English and Spanish Narrative Story Production Tasks

	English		Spanish	
Characteristic	**K** N = 134	**G1** N = 120	**K** N = 109	**G1** N = 94
Timepoint				
Fall 2024	2 (100%)	2 (100%)	1 (100%)	0 (NA%)
Unknown	132	118	108	94
Administration Format				
Not applicable	134 (100%)	120 (100%)	109 (100%)	94 (100%)
Race				
American/Alaskan Native	3 (2.2%)	0 (0%)	0 (0%)	2 (2.1%)
Asian	16 (12%)	16 (13%)	4 (3.7%)	1 (1.1%)
Black/African American	20 (15%)	14 (12%)		
Not reported	16 (12%)	33 (28%)	61 (56%)	63 (67%)
Other	28 (21%)	20 (17%)	17 (16%)	3 (3.2%)
White	51 (38%)	36 (30%)	27 (25%)	25 (27%)
Unknown	0	1		
Ethnicity				
Hispanic/Latin(o/a)	59 (44%)	44 (37%)	101 (94%)	89 (95%)
Intentional nonreport	3 (2.2%)	2 (1.7%)		
Not Hispanic/Latin(o/a)	72 (54%)	73 (61%)	7 (6.5%)	5 (5.3%)
Unknown	0	1	1	0
Gender				
Female	78 (58%)	52 (44%)	62 (57%)	47 (50%)
Male	56 (42%)	67 (56%)	47 (43%)	47 (50%)
Unknown	0	1		
Home Language				
English	100 (75%)	86 (74%)	21 (19%)	15 (16%)
Spanish	24 (18%)	18 (15%)	88 (81%)	79 (84%)
Other	9 (6.8%)	13 (11%)	0 (0%)	0 (0%)
Unknown	1	3		
English Proficiency Label				
(Re-)Classified Proficient	12 (10%)	10 (8.6%)	19 (19%)	23 (25%)
English Learner	23 (20%)	30 (26%)	75 (77%)	60 (65%)
English-only	82 (70%)	76 (66%)	4 (4.1%)	9 (9.8%)
Unknown	17	4	11	2
Ever IEP/504	7 (6.5%)	10 (9.7%)	8 (9.4%)	8 (9.9%)
Unknown	26	17	24	13

16.6 Score distribution

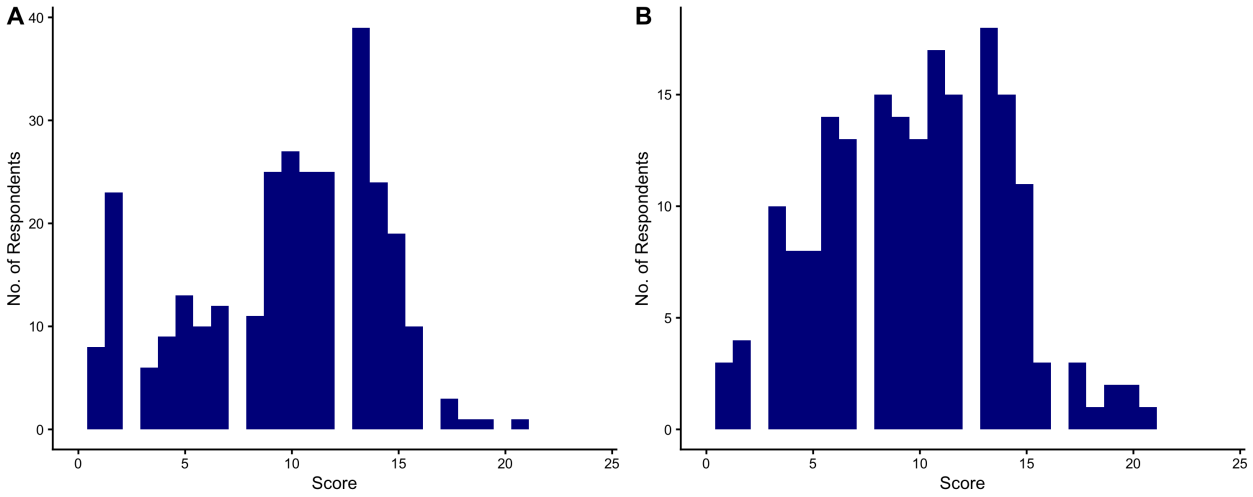

Figure 16.1: Score Distribution of the English and Spanish Narrative Story Production Tasks

16.7 Criterion Validity Evidence

16.7.1 Sample

Forthcoming.

17 Nonword Repetition

17.1 Task Description

Children listen to nonsense words and are asked to repeat them verbatim.

17.2 Construct

The Nonword Repetition task measures the construct of auditory short-term memory. Students repeat a series of pseudowords of differing syllable length and complexity of sound combinations, thereby assessing linguistic abilities that have not been taught or learned previously and that are less culturally and linguistically biased.

17.3 Item Development

17.3.1 English

The constructed items followed the English phonotactic structure, featuring diverse syllable constructions that reflect the language's phonological patterns. The length of the items ranged from one to five syllables. For each level of number of syllables, a mix of light and heavy items was developed based on phoneme density. We conceptualized nonwords as light when less than half of the syllables contained complex onsets or codas and had no diphthongs. Light items used the following syllable patterns: CVC, CV, and/or VC. We conceptualized nonwords as heavy when more than half of the syllables contain complex onsets or codas, or a diphthong. Heavy items used the following syllable patterns: CCVC, CCVCC, CVVC, and/or CVCC.

Dialectal considerations. To avoid possible mis-scoring due to a lack of consideration of dialectal differences, we avoided using word-final "th" in item development, as it is can often fronted ("breve" for "breathe") or stopped ("wit" for "with") in African American English and in English influenced by other languages. We also avoided clusters of nasals and stops (/nt/). Additionally, proctors were instructed to be cognizant that voiced word-final consonants can be devoiced and/or spirantized in some dialects (e.g., "bed" as "beth" or "bet") and should be scored as correct.

17.3.2 Spanish

The constructed items followed the Spanish phonotactic structure, featuring diverse syllable constructions that reflect the language's phonological patterns. The length of the items ranged from two-syllable nonwords with five or six phonemes to five-syllable nonwords of 11 to 13 phonemes in length. Given the scarcity of one-syllable words in Spanish, the task started at two-syllable words, in line with existing nonword repetition tasks. For the blueprint used for item development, see Table 17.1.

Table 17.1: Blueprint for item design for Nonword Repetition

Number of syllables	CV structure	Nonword example
Two syllable nonwords	CV CVC	g a u β e ɾ
	CVC CV	m e ɾ ɣ u i
	CVC CV	t i ŋ k u a
	CVC CVC	m e ɾ f a s
Three syllable nonwords	CV CV CV	ɾ u ʧ e t u a
	CV CV CV	ʧ e ɾ u ɣ u a
	CVC CV CVC	t i n ʧ a u β e l
	CVC CV CVC	d u ɾ b i e p o s
Four syllable nonwords	CV CV CV CV	k i ʧ e ɾ u p i a
	CV CV CV CV	f i ɾ u ʧ e p i a
	CV CV CV CVC	m a ʊ t e i p o i t i n
	CV CVC CV CVC	m e ɾ a n ʧ u t i n
Five syllable nonwords	CVC CV CV CV CV	b i ŋ k u a m i ɛ f e ɣ u i
	CVC CVC CV CV CV	h u s t i n r u n a i ʧ e
	CV CVC CV CV CVC	n u e β u r d a i ɣ u i p o s
	CVC CV CV CVC CVC	p e n ʧ u f a i t i n β e l

17.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

17.5 Calibration Samples

Table 17.2: Demographic Characteristics of Calibration Samples for the English and Spanish Nonword Repetition Tasks

	English		Spanish	
Characteristic	**K** N = 601	**G1** N = 665	**K** N = 527	**G1** N = 569
Timepoint				
Spring 2023	0 (0%)	0 (0%)	527 (100%)	569 (100%)
Fall 2023	598 (100%)	663 (100%)	0 (0%)	0 (0%)
Fall 2024	3 (0.5%)	2 (0.3%)	0 (0%)	0 (0%)
Administration Format				
CAT	3 (0.5%)	2 (0.3%)		
Forms	598 (100%)	663 (100%)	527 (100%)	569 (100%)
Race				
American/Alaskan Native	16 (2.7%)	12 (1.8%)	9 (1.8%)	7 (1.3%)
Asian	73 (12%)	87 (13%)	7 (1.4%)	5 (0.9%)
Black/African American	74 (13%)	79 (12%)	4 (0.8%)	4 (0.7%)
Not reported	132 (22%)	127 (19%)	353 (69%)	366 (67%)
Other	105 (18%)	75 (11%)	14 (2.7%)	14 (2.6%)
White	192 (32%)	284 (43%)	125 (24%)	150 (27%)
Unknown	9	1	15	23
Ethnicity				
Hispanic/Latin(o/a)	323 (54%)	345 (52%)	496 (97%)	522 (96%)
Intentional nonreport	10 (1.7%)	3 (0.5%)	0 (0%)	1 (0.2%)
Not Hispanic/Latin(o/a)	264 (44%)	316 (48%)	14 (2.7%)	19 (3.5%)
Unknown	4	1	17	27
Gender				
Female	305 (51%)	317 (48%)	307 (60%)	325 (60%)
Male	293 (49%)	347 (52%)	208 (40%)	217 (40%)
Unknown	3	1	12	27
Home Language				
English	416 (71%)	493 (74%)	75 (15%)	70 (13%)
Spanish	106 (18%)	100 (15%)	422 (82%)	468 (86%)
Other	63 (11%)	69 (10%)	15 (2.9%)	6 (1.1%)
Unknown	16	3	15	25
English Proficiency Label				
(Re-)Classified Proficient	38 (7.3%)	61 (9.3%)	44 (10.0%)	42 (8.7%)
English Learner	138 (26%)	120 (18%)	341 (77%)	385 (80%)
English-only	346 (66%)	472 (72%)	56 (13%)	57 (12%)
Unknown	79	12	86	85
Ever IEP/504	36 (7.6%)	62 (11%)	27 (9.9%)	23 (12%)
Unknown	127	108	253	378

17.6 Psychometric Analysis

17.6.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 0 items from the English task and 0 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 17.3 summarizes the basic item characteristics, Figure 17.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 17.3: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Nonword Repetition Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 118**	**N = 118**	**N = 145**	**N = 145**
No. of Responses	178 (139)	178 (139)	142 (103)	142 (103)
Proportion Correct	0.57 (0.19)	0.57 (0.19)	0.65 (0.18)	0.65 (0.18)
Point-biserial Correlation	0.54 (0.08)	0.54 (0.08)	0.50 (0.09)	0.50 (0.09)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

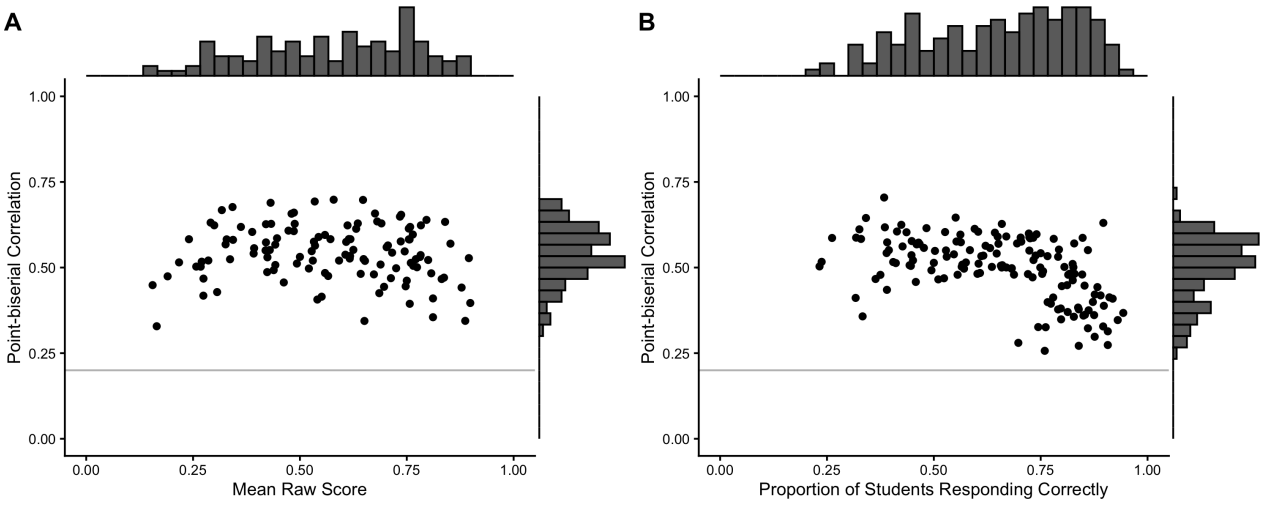

Figure 17.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Nonword Repetition Tasks

17.6.2 Rasch Analysis

17.6.2.1 Item Location Estimates

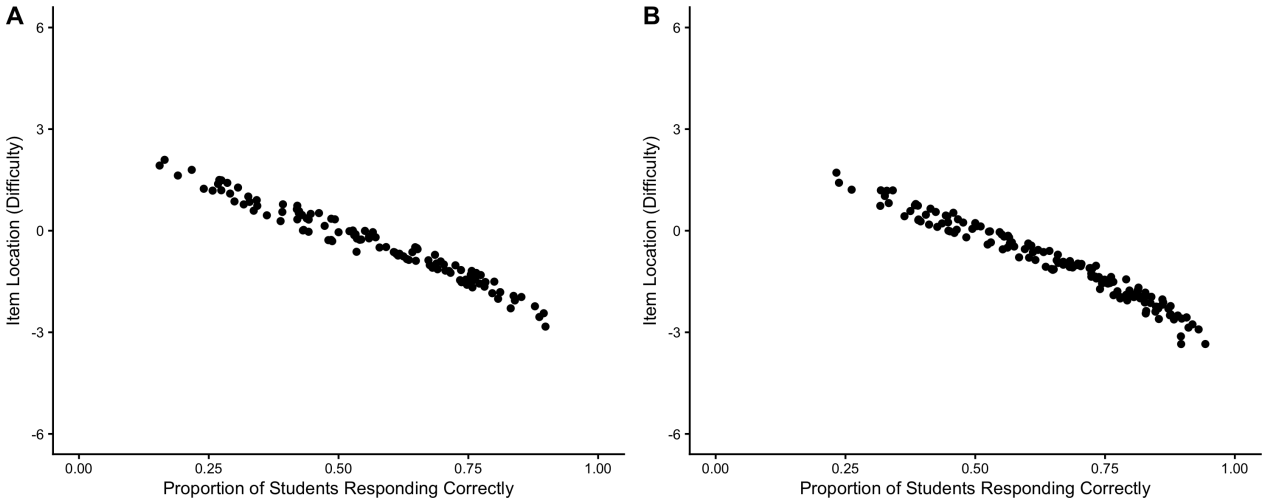

Figure 17.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Nonword Repetition Tasks

17.6.2.2 Item Fit Statistics

Table 17.4: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Nonword Repetition Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	117	0	0	0	117	138	0	0	0	138
B	0	0	0	0	0	0	0	0	0	0
C	1	0	0	0	1	7	0	0	0	7
D	0	0	0	0	0	0	0	0	0	0
Total	118	0	0	0	118	145	0	0	0	145

17.6.2.3 Person Location Estimates

17.6.2.4 Person Fit Statistics

Table 17.5: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Nonword Repetition Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	994	0	1	0	995	780	0	0	0	780
B	90	119	0	0	209	112	101	0	0	213
C	40	0	8	0	48	39	0	13	0	52
D	4	0	2	0	6	11	0	5	0	16
Total	1,128	119	11	0	1,258	942	101	18	0	1,061

17.6.2.5 Distribution of Theta Estimates

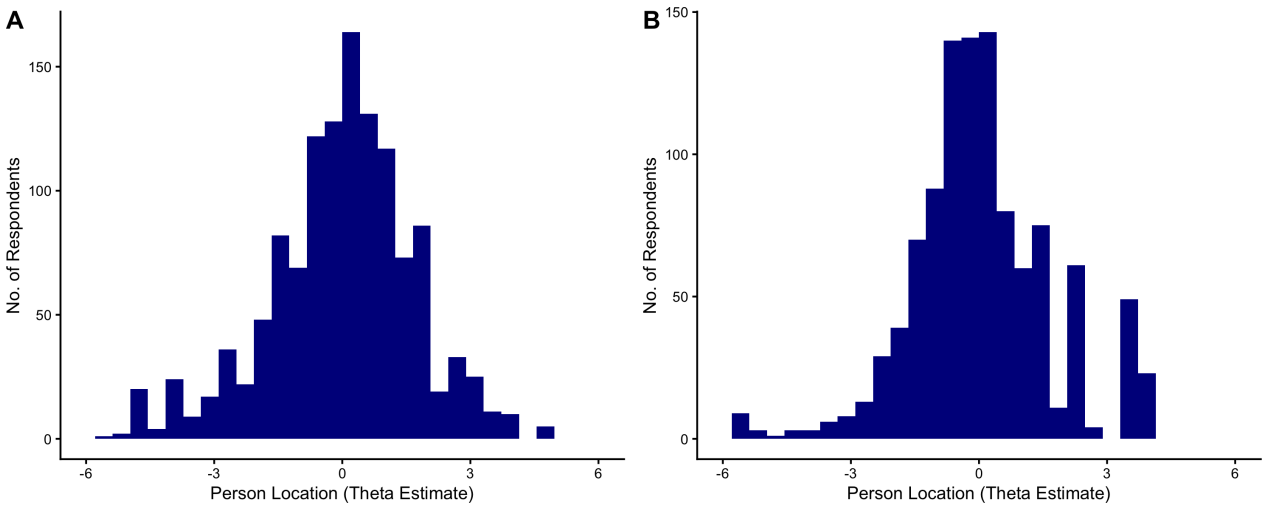

Figure 17.4: Distribution of Theta Estimates for the English and Spanish Nonword Repetition Tasks

17.6.2.6 Wright Maps

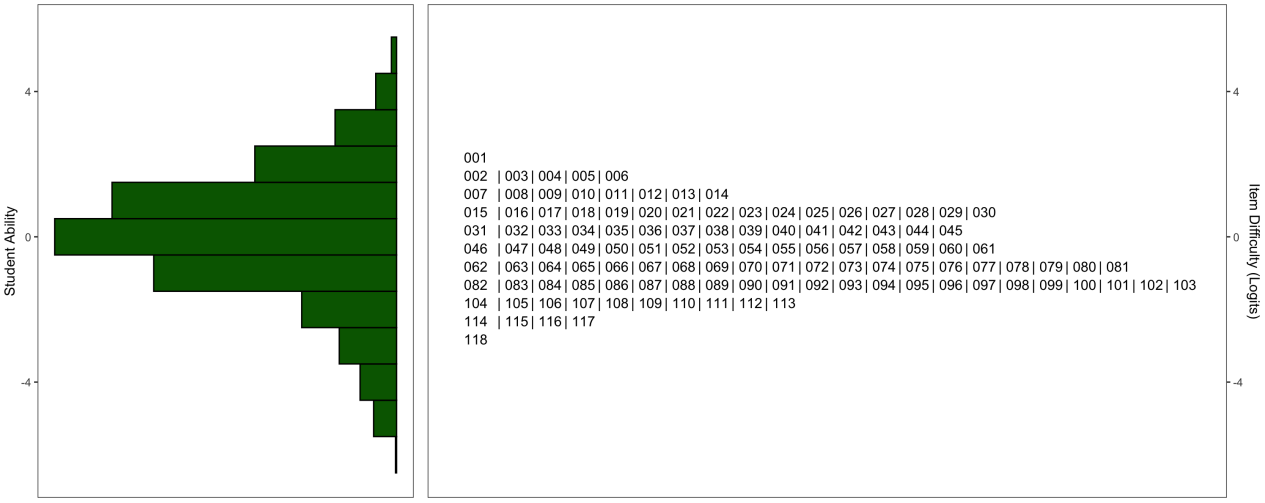

Figure 17.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Nonword Repetition Task

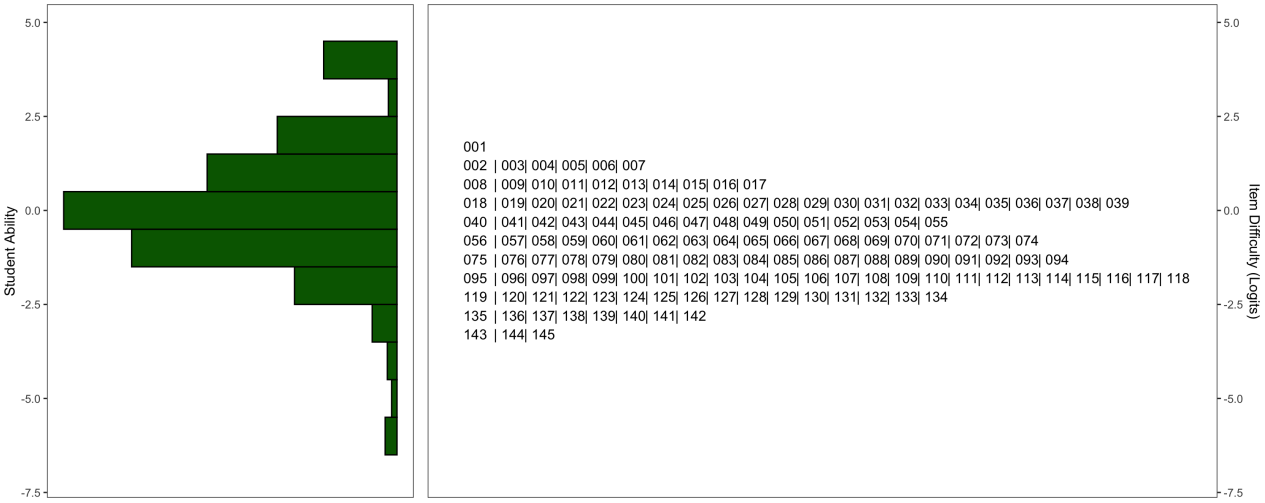

Figure 17.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Nonword Repetition Task

17.6.2.7 Model Summary

Table 17.6: Summary of Rasch Model Statistics for the English and Spanish Nonword Repetition Tasks

	English	Spanish

	Item	Person	Item	Person
Characteristic	N = 118	N = 1,258	N = 145	N = 1,061
Logit Scale Location	-0.41 (1.09)	0.05 (-0.89, 1.02)	-0.95 (1.14)	-0.07 (-0.90, 0.96)
Outfit	0.98 (0.18)	0.84 (0.63, 1.04)	1.01 (0.23)	0.81 (0.57, 1.11)
Infit	0.99 (0.09)	0.90 (0.77, 1.04)	1.00 (0.10)	0.89 (0.74, 1.06)
Reliability of Separation	0.8244	0.7596	0.8272	0.7669

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 118 and 145 for the English and Spanish task, respectively.

17.7 Criterion Validity Evidence

17.7.1 Sample

Table 17.7: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Nonword Repetition Tasks

	English		Spanish	
Characteristic	**K** N = 251	**G1** N = 215	**K** N = 236	**G1** N = 221
Timepoint				
Winter 2024	251 (100%)	215 (100%)	236 (100%)	221 (100%)
Race				
American/Alaskan Native	5 (2.0%)	3 (1.4%)	2 (0.9%)	4 (1.8%)
Asian	36 (15%)	35 (16%)	8 (3.4%)	2 (0.9%)
Black/African American	25 (10%)	26 (12%)	1 (0.4%)	0 (0%)
Not reported	28 (11%)	29 (13%)	130 (56%)	149 (68%)
Other	71 (29%)	44 (20%)	38 (16%)	8 (3.7%)
White	83 (33%)	78 (36%)	55 (24%)	55 (25%)
Unknown	3	0	2	3
Ethnicity				
Hispanic/Latin(o/a)	99 (39%)	91 (42%)	214 (92%)	205 (93%)
Intentional nonreport	7 (2.8%)	2 (0.9%)	1 (0.4%)	0 (0%)
Not Hispanic/Latin(o/a)	145 (58%)	122 (57%)	18 (7.7%)	16 (7.2%)
Gender				
Female	124 (49%)	97 (45%)	124 (53%)	108 (49%)
Male	127 (51%)	118 (55%)	112 (47%)	113 (51%)
Home Language				
English	183 (74%)	159 (74%)	28 (12%)	21 (9.6%)
Spanish	31 (13%)	24 (11%)	203 (87%)	196 (90%)
Other	32 (13%)	31 (14%)	2 (0.9%)	1 (0.5%)
Unknown	5	1	3	3
English Proficiency Label				
(Re-)Classified Proficient	11 (5.4%)	17 (8.1%)	31 (14%)	24 (11%)
English Learner	47 (23%)	39 (19%)	180 (81%)	174 (81%)
English-only	147 (72%)	154 (73%)	10 (4.5%)	17 (7.9%)
Unknown	46	5	15	6
Ever IEP/504	18 (9.4%)	21 (13%)	20 (9.7%)	23 (11%)
Unknown	59	47	29	12
Unknown			3	0

English Nonword Repetition was correlated with the Nonword Repetition subtest from the Woodcock-Johnson IV (WJ IV COG) test (Schrank, McGrew, and Mather 2014). Spanish Nonword Repetition was correlated with the Repetición de Palabras Sin Sentido subtest from the Batería IV Woodcock-Muñoz (Batería IV COG) test (Woodcock et al. 2019).

Table 17.8: Concurrent Criterion Validity Correlations for the English and Spanish Nonword Repetition Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	251	0.45 [0.34, 0.54]	47	0.25 [-0.04, 0.50]	236	0.49 [0.38, 0.58]
G1	215	0.53 [0.42, 0.62]	39	0.60 [0.35, 0.77]	221	0.48 [0.37, 0.57]

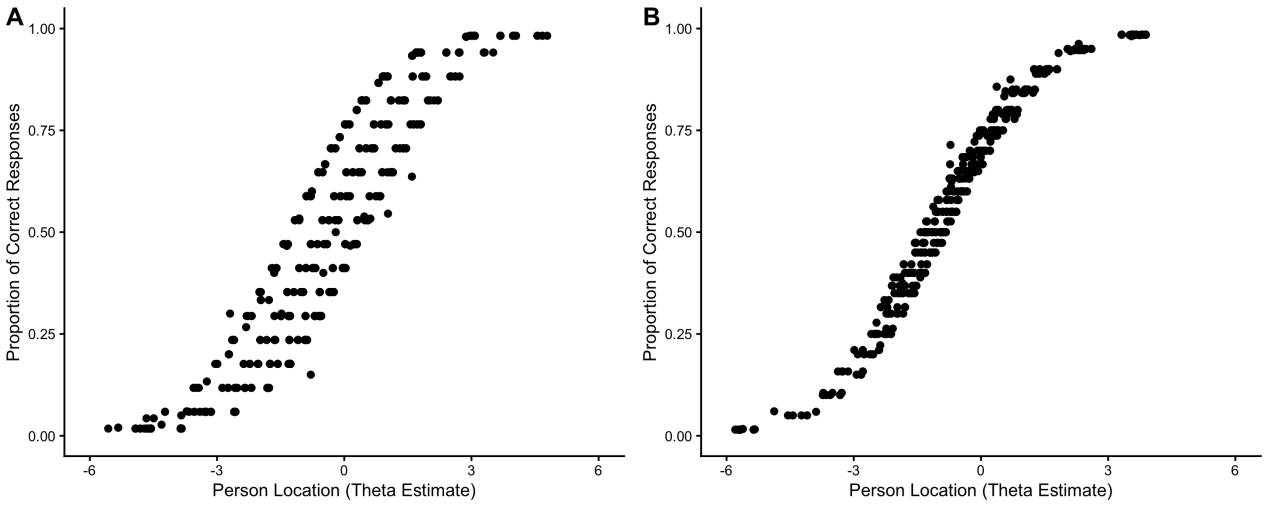

Figure 17.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Nonword Repetition Tasks

18 Oral Reading Fluency

18.1 Task Description

Children are shown a passage and are asked to read aloud as fast and as accurately as they can within a two-minute time limit.

18.2 Construct

The Oral Reading Fluency task measures children's ability to read smoothly, accurately, and quickly. This involves not just reading the words correctly but doing so in a way that sounds natural and conveys the meaning of the text.

18.3 Item Development

Two classrooms of second-grade students were asked about topics of interest for the ORF passages. Multiple topics emerged, but animals were a consistent topic across classrooms. Other proposed topics (e.g., famous actresses, musicians, and sportspeople) were excluded to avoid proper names. A total of nine passages on California animals were written in each language.

18.3.1 English

All passages were written to have a Lexile level of 410L-600L. This Lexile range was targeted based on the national standards for this grade, see **?@tbl-item-design-orf_en**. To ensure further comparability across texts, additional indicators were obtained using Lexile Text Analyzer (2024). These results are presented below.

- **Decoding:** Easier texts have words with fewer syllables and simpler sounds (e.g., "net" and "shop"). Harder texts have words with more syllables and more complex sounds (e.g., "balloon" and "ceremony"). Ranges from 1-5. All passages were scored as Neutral (3/5)
- **Vocabulary:** Easier texts have more common, familiar, and concrete words. Harder texts have more rare, unfamiliar, and abstract words. Ranges from 1-5. Five passages were scored as Neutral (3/5), and four were scored as Difficult (4/5)
- **Sentences:** Easier texts have shorter sentences and more words that overlap between sentences. Harder texts have longer sentences and fewer words that overlap between sentences. Ranges from 1-5. All passages were scored as Difficult (4/5).

- **Patterns:** Easier texts have more repeated words and phrases. Harder texts have fewer repeated words and phrases. Ranges from 1-5. Two passages were scored as Difficult (4/5) and seven were scored as Very Difficult (5/5).

18.3.2 Spanish

Like in English, all passages were written to have a Lexile level of 410L-600L. To ensure further comparability across texts, additional indicators were obtained using Legible (Legible 2024). For the text statistics and average, see Table 18.1.

Table 18.1: Spanish Passages Statistics

	Blue Belly Lizard	Sea Lion	Kit Fox	Raccoon	Grizzly Bear	Quail	Condor	White Shark	Banana Slug	Average
Text legibility - Value										
Fernández Huerta	91.92	84.62	86.56	85.69	90.2	88.43	88.54	84.32	85.83	87.35
Gutiérrez	52.56	48.16	49.24	49.42	50.81	50.25	51.51	47.81	49.11	49.87
Szigriszt-Pazos	87.89	80.43	82.18	81.65	86.14	84.34	84.67	79.94	81.93	83.24
INFLESZ	87.89	80.43	82.18	81.65	86.14	84.34	84.67	79.94	81.93	83.24
legibility	87.85	91.6	75.68	89.32	83.1	88.97	89.87	96.37	74.95	86.41
Text legibility - Difficulty										
Fernández Huerta	very easy	easy	easy	easy	very easy	easy	easy	easy	easy	
Gutiérrez	normal	normal	normal	normal	normal	normal	normal	normal	normal	
Szigriszt-Pazos	very easy	easy	easy	easy	very easy	easy	easy	easy	easy	
INFLESZ	very easy	easy	very easy	easy	easy	easy	very easy	easy	very easy	
legibility	easy	very easy	a little easy	easy	easy	easy	easy	very easy	a little easy	
Estimated grade (Crawford):	2.9	3.5	3	3.3	3.2	3.3	3.2	3.7	3.4	3.28
Estimated time of reading:	0.9	0.9	0.8	0.9	1	1	1.1	0.9	0.9	0.93
Text statistics										
characters	904	997	871	970	1042	1107	1148	1036	1067	1015.78
letters	699	785	689	761	816	866	895	822	843	797.33
syllables	303	327	290	323	344	366	389	341	350	337
words	171	173	154	172	193	202	214	181	190	183.33
phrases	19	19	20	20	19	20	23	18	20	19.78
paragraphs	3	3	2	4	4	5	5	5	4	3.89
letters per word	4.09	4.54	4.47	4.42	4.23	4.29	4.18	4.54	4.44	4.36
syllables per word	1.77	1.89	1.88	1.88	1.78	1.81	1.82	1.88	1.84	1.84
words per sentence	8.55	8.65	7.33	8.19		9.62	8.92	9.53	9.05	8.73

18.4 Scoring

The final score consisted of a rate between the number of accurately read words (i.e., the total number of words read minus the incorrectly read words) divided by the total allocated time (per second) of two minutes or the total time it took the child to read the whole passage, if under two minutes.

18.5 Samples

Table 18.2: Demographic Characteristics of Samples for the English and Spanish Oral Reading Fluency Tasks

	English		Spanish	
Characteristic	**G1** N = 314	**G2** N = 774	**G1** N = 280	**G2** N = 337
Timepoint				
Fall 2024	30 (100%)	469 (100%)	0 (NA%)	114 (100%)
Unknown	284	305	280	223
Administration Format				
Not applicable	314 (100%)	774 (100%)	280 (100%)	337 (100%)
Race				
American/Alaskan Native	7 (2.3%)	19 (2.6%)	4 (1.4%)	6 (1.8%)
Asian	42 (14%)	77 (11%)	23 (8.2%)	8 (2.4%)
Black/African American	32 (11%)	81 (11%)	10 (3.6%)	7 (2.1%)
Not reported	45 (15%)	129 (18%)	58 (21%)	116 (35%)
Other	61 (20%)	94 (13%)	108 (39%)	24 (7.2%)
White	114 (38%)	330 (45%)	77 (28%)	174 (52%)
Unknown	13	44	0	2
Ethnicity				
Hispanic/Latin(o/a)	184 (60%)	534 (74%)	226 (81%)	307 (91%)
Intentional nonreport	3 (1.0%)	5 (0.7%)	3 (1.1%)	1 (0.3%)
Not Hispanic/Latin(o/a)	119 (39%)	187 (26%)	49 (18%)	28 (8.3%)
Unknown	8	48	2	1
Gender				
Female	151 (53%)	346 (48%)	150 (54%)	173 (51%)
Male	135 (47%)	368 (52%)	128 (46%)	163 (49%)
Unknown	28	60	2	1
Home Language				
English	163 (63%)	390 (59%)	101 (36%)	78 (23%)
Spanish	77 (30%)	237 (36%)	170 (61%)	251 (75%)
Other	18 (7.0%)	34 (5.1%)	9 (3.2%)	4 (1.2%)
Unknown	56	113	0	4
English Proficiency Label				
(Re-)Classified Proficient	7 (4.1%)	50 (7.8%)	31 (12%)	33 (10%)
English Learner	64 (38%)	217 (34%)	134 (51%)	223 (68%)
English-only	98 (58%)	372 (58%)	96 (37%)	73 (22%)
Unknown	145	135	19	8
Ever IEP/504	13 (6.3%)	84 (16%)	21 (8.8%)	29 (9.9%)
Unknown	109	242	42	45

18.6 Score distribution

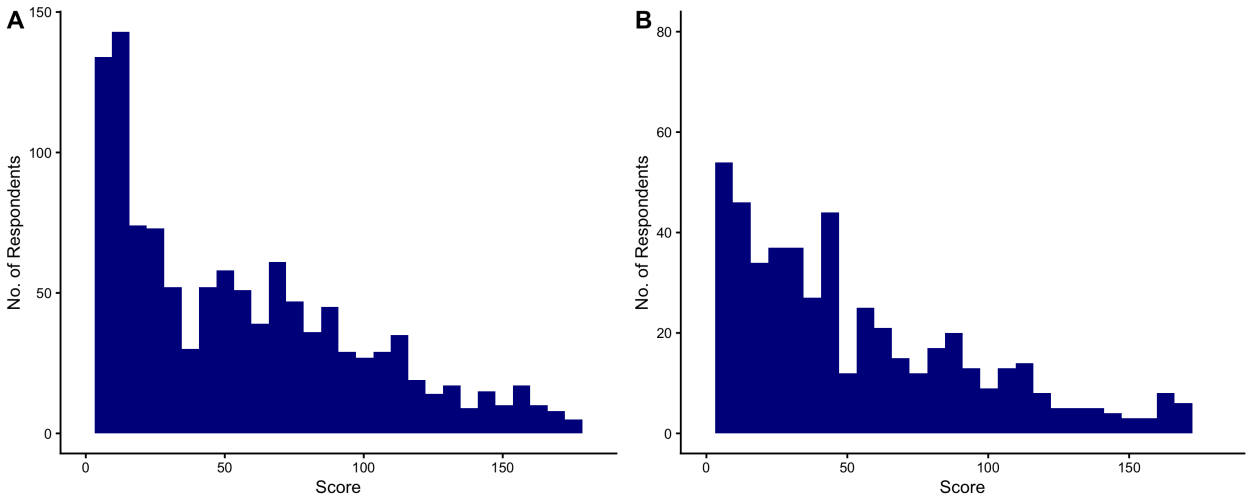

Figure 18.1: Score Distribution of the English and Spanish Oral Reading Fluency Tasks

18.7 Criterion Validity Evidence

18.7.1 Sample

Table 18.3: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Oral Reading Fluency Tasks

	English	Spanish
Characteristic	**G2** N = 297	**G2** N = 241
Timepoint		
Spring 2024	297 (100%)	241 (100%)
Race		
American/Alaskan Native	4 (1.3%)	4 (1.7%)
Asian	28 (9.4%)	3 (1.3%)
Black/African American	34 (11%)	4 (1.7%)
Not reported	84 (28%)	112 (47%)
Other	21 (7.1%)	13 (5.4%)
White	126 (42%)	103 (43%)
Ethnicity		
Hispanic/Latin(o/a)	192 (65%)	216 (90%)
Intentional nonreport	1 (0.3%)	2 (0.8%)
Not Hispanic/Latin(o/a)	103 (35%)	23 (9.5%)
Unknown	1	
Gender		
Female	151 (51%)	139 (58%)
Male	146 (49%)	102 (42%)
Home Language		
English	175 (59%)	55 (23%)
Spanish	92 (31%)	178 (75%)
Other	28 (9.5%)	4 (1.7%)
Unknown	2	4
English Proficiency Label		
(Re-)Classified Proficient	37 (13%)	36 (15%)
English Learner	82 (28%)	149 (63%)
English-only	174 (59%)	52 (22%)
Unknown	4	4
Ever IEP/504	19 (10%)	18 (9.0%)
Unknown	110	42
Unknown		2

English Oral Reading Fluency was correlated with the Oral Reading subtest from the Woodcock-Johnson IV (WJ IV ACH) test (Schrank, McGrew, and Mather 2014). Spanish Oral Reading Fluency was correlated with the Lectura Oral subtest from the Batería IV Woodcock-Muñoz (Batería IV APROV) test (Woodcock et al. 2019).

Table 18.4: Concurrent Criterion Validity Correlations for the English and Spanish Oral Reading Fluency Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
G2	297	0.11 [-0.01, 0.22]	82	0.25 [0.03, 0.44]	241	0.13 [0.00, 0.25]

19 Rapid Automatized Naming of Letters

19.1 Task Description

Children are shown a page with 5 unique letters repeated randomly, arranged in five rows of ten. They are asked to name as many letters as quickly as possible.

19.2 Construct

The Rapid Automatized Naming - Letters task measures the construct of automatic processing and retrieval of letter names. It assesses how quickly and accurately children can name familiar letters, skills that are closely linked to automatic reading.

19.3 Item Development

19.3.1 English

For letter selection, we selected a mix of vowels and consonants, prioritizing letters that were typically acquired earlier in literacy development. Additionally, the selected letters fulfilled the following selection criteria:

- **Letters could not be reversible:** letter reversals are a common developmental mistake of early readers. To avoid eliciting a developmental error, we excluded the following letters: "b," "d," "p," and "q."
- **Letters should be clearly visually distinguishable:** depending on the font selected, some letters can be easily confused. Consequently, we excluded the letter "v" for its potential to confuse with the letter "u," and we excluded the letter "a" for its potential to confuse with the letter "o."

The final set of letters selected for the English measure was: **c, e, o, s, u**.

19.3.2 Spanish

For letter selection, we wanted to select a mix of vowels and consonants, prioritizing letters that were typically acquired earlier in literacy development. Additionally, the selected letters needed to fulfill the following selection criteria:

- **Letters had to be monosyllabic:** to be able to compare the performance in English and Spanish, letters needed to have similar phonological length. While most of the letters in English are monosyllabic, that is not the case in Spanish. Most of the letters acquired earlier in literacy development, because they are simpler and consistent, tend to be disyllabic (e.g., m: e-me, n: e-ne, l: e-le, s: e-se, f: e-fe). These letters were excluded.
- **Letters could not be reversible:** letter reversals are a common, developmentally expected mistake that early readers commit. To avoid over-penalizing children for making mistakes due to letter reversal, we excluded the following letters: "b," "d," "p," and "q."
- **Letters also need to be clearly visually identifiable:** depending on the font selected, some letters can be easily confused. Consequently, we excluded the letter "v" for its potential to confuse with the letter "u," and we excluded the letter "a" for its potential to confuse with the letter "o."

The final set of letters selected for the Spanish measure was: **c, e, o, t, u**.

19.4 Scoring

Participating children were presented with a 5x10 grid containing 50 letters and were asked to name as fast as they could, and the time taken by the participant to name all letters was recorded. The final score consisted of a rate between the number of accurately named letters (i.e., the total number of letters minus the incorrectly named letters) divided by the total time it took the child to complete the grid.

19.5 Samples

Table 19.1: Demographic Characteristics of Samples for the English and Spanish Rapid Automatized Naming of Letters Tasks

	English			Spanish		
Characteristic	**K** N = 1,225	**G1** N = 1,613	**G2** N = 3,019	**K** N = 786	**G1** N = 922	**G2** N = 748
Timepoint						
Fall 2023	456 (48%)	396 (30%)	417 (15%)	429 (69%)	378 (54%)	365 (49%)
Fall 2024	500 (52%)	942 (70%)	2,277 (85%)	193 (31%)	328 (46%)	380 (51%)
Unknown	269	275	325	164	216	3
Administration Format						
Not applicable	1,225 (100%)	1,613 (100%)	3,019 (100%)	786 (100%)	922 (100%)	748 (100%)
Race						
American/Alaskan Native	31 (2.6%)	53 (3.4%)	62 (2.2%)	26 (3.3%)	32 (3.5%)	9 (1.2%)
Asian	114 (9.5%)	146 (9.3%)	204 (7.4%)	29 (3.7%)	27 (3.0%)	23 (3.1%)
Black/African American	129 (11%)	177 (11%)	283 (10%)	9 (1.2%)	8 (0.9%)	11 (1.5%)
Not reported	193 (16%)	267 (17%)	310 (11%)	354 (45%)	409 (45%)	271 (37%)
Other	303 (25%)	239 (15%)	370 (13%)	155 (20%)	76 (8.3%)	48 (6.5%)
White	430 (36%)	692 (44%)	1,546 (56%)	209 (27%)	362 (40%)	373 (51%)
Unknown	25	39	244	4	8	13
Ethnicity						
Hispanic/Latin(o/a)	679 (61%)	993 (65%)	1,929 (70%)	699 (94%)	858 (94%)	680 (93%)
Intentional nonreport	23 (2.1%)	7 (0.5%)	6 (0.2%)	2 (0.3%)	1 (0.1%)	0 (0%)
Not Hispanic/Latin(o/a)	403 (36%)	536 (35%)	815 (30%)	40 (5.4%)	49 (5.4%)	50 (6.8%)
Unknown	120	77	269	45	14	18
Gender						
Female	568 (50%)	790 (52%)	1,344 (49%)	393 (53%)	480 (54%)	362 (50%)
Male	562 (50%)	732 (48%)	1,392 (51%)	344 (47%)	409 (46%)	363 (50%)
Unknown	95	91	283	49	33	23
Home Language						
English	731 (62%)	923 (61%)	1,655 (64%)	97 (13%)	104 (11%)	94 (13%)
Spanish	365 (31%)	524 (35%)	813 (32%)	671 (87%)	804 (88%)	604 (86%)
Other	80 (6.8%)	67 (4.4%)	109 (4.2%)	6 (0.8%)	4 (0.4%)	6 (0.9%)
Unknown	49	99	442	12	10	44
English Proficiency Label						
(Re-)Classified Proficient	63 (6.1%)	109 (7.5%)	270 (11%)	82 (11%)	124 (14%)	83 (12%)
English Learner	352 (34%)	470 (33%)	649 (25%)	567 (79%)	679 (76%)	520 (75%)
English-only	619 (60%)	867 (60%)	1,628 (64%)	65 (9.1%)	85 (9.6%)	86 (12%)
Unknown	191	167	472	72	34	59
Ever IEP/504	64 (6.8%)	129 (9.9%)	224 (11%)	48 (7.6%)	69 (9.0%)	61 (11%)
Unknown	280	306	887	158	154	169

19.6 Score distribution

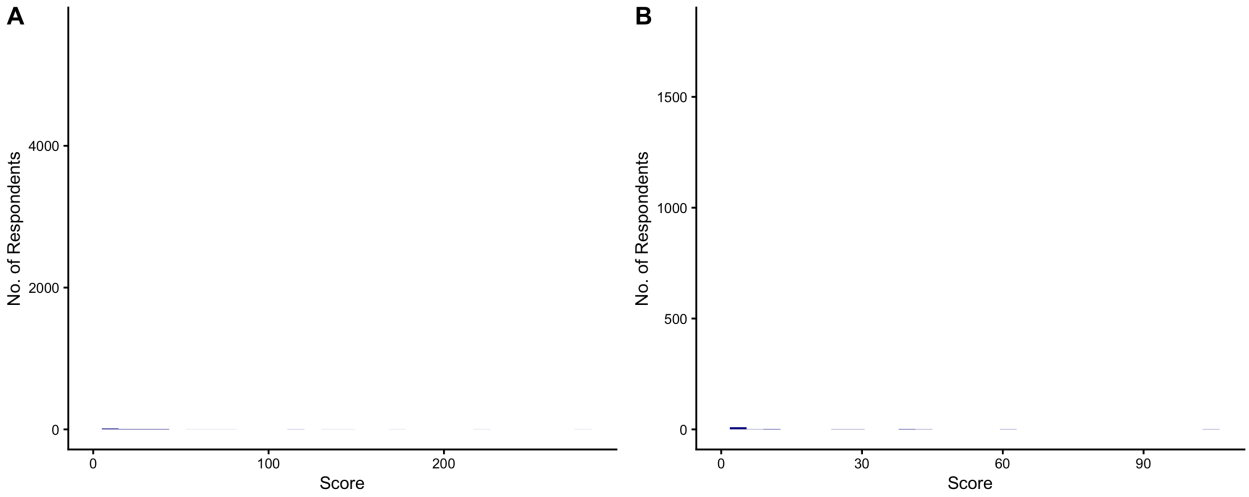

Figure 19.1: Score Distribution of the English and Spanish Rapid Automatized Naming of Letters Tasks

19.7 Criterion Validity Evidence

19.7.1 Sample

Table 19.2: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Rapid Automatized Naming of Letters Tasks

	English		Spanish	
Characteristic	**K** N = 190	**G1** N = 209	**K** N = 129	**G1** N = 175
Timepoint				
Spring 2024	190 (100%)	209 (100%)	129 (100%)	175 (100%)
Race				
American/Alaskan Native	2 (1.1%)	4 (1.9%)	4 (3.1%)	4 (2.3%)
Asian	29 (15%)	26 (12%)	9 (7.0%)	5 (2.9%)
Black/African American	23 (12%)	25 (12%)	2 (1.6%)	1 (0.6%)
Not reported	29 (15%)	52 (25%)	48 (37%)	78 (45%)
Other	43 (23%)	27 (13%)	22 (17%)	7 (4.0%)
White	64 (34%)	75 (36%)	44 (34%)	79 (45%)
Ethnicity				
Hispanic/Latin(o/a)	84 (44%)	89 (43%)	115 (89%)	163 (93%)
Intentional nonreport	10 (5.3%)	2 (1.0%)	1 (0.8%)	0 (0%)
Not Hispanic/Latin(o/a)	96 (51%)	118 (56%)	13 (10%)	12 (6.9%)
Gender				
Female	101 (53%)	105 (50%)	73 (57%)	99 (57%)
Male	89 (47%)	104 (50%)	56 (43%)	76 (43%)
Home Language				
English	137 (72%)	154 (74%)	28 (22%)	28 (16%)
Spanish	34 (18%)	28 (14%)	101 (78%)	144 (83%)
Other	18 (9.5%)	25 (12%)	0 (0%)	2 (1.1%)
Unknown	1	2	0	1
English Proficiency Label				
(Re-)Classified Proficient	14 (8.4%)	25 (12%)	25 (20%)	32 (18%)
English Learner	31 (19%)	38 (18%)	81 (65%)	119 (69%)
English-only	121 (73%)	143 (69%)	18 (15%)	22 (13%)
Unknown	24	3	5	2
Ever IEP/504	8 (5.9%)	20 (11%)	6 (5.4%)	12 (8.3%)
Unknown	55	19	18	30
Unknown			0	1

English Rapid Automatized Naming of Letters was correlated with two measures: the Letters subtest from the Acadience Learning RAN (Powell-Smith et al. 2020a) in kindergarten and first grade, and the RAN Letters subtest from the Rapid Automatized Naming and Rapid Alternating Stimulus Tests (M. Wolf and Denckla 2003a) in second grade.

Spanish Rapid Automatized Naming of Letters was correlated with the Letters subtest from the Acadience Learning RAN (Powell-Smith et al. 2020a) in kindergarten and first grade. No rapid naming of letters task could be found in Spanish that was normed on second-grade-aged children in the United States.

These analyses are problematic because of differences in how the measures are scored. The external assessments are scored based on total time, with RAN Objects and RAN Letters being completed in tandem. The Multitudes RANL measure is scored as the number correct divided by the total time and is distinct from RANO. Given these scoring differences, the correlations are difficult to interpret.

Table 19.3: Concurrent Criterion Validity Correlations for the English and Spanish Rapid Automatized Naming of Letters Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	190	-0.66 [-0.73, -0.57]	31	-0.74 [-0.87, -0.53]	129	-0.69 [-0.77, -0.59]
G1	207	-0.71 [-0.77, -0.63]	38	-0.80 [-0.89, -0.65]	175	-0.61 [-0.69, -0.50]

20 Rapid Automatized Naming of Objects

20.1 Task Description

Children are shown a page with 5 unique objects repeated randomly, arranged in five rows of ten. They are asked to name as many objects as quickly as possible.

20.2 Construct

The Rapid Automatized Naming - Objects task measures the construct of automaticity in processing and retrieval of object names. It assesses how quickly and accurately children can name familiar objects, skills that are closely linked to automatic reading.

20.3 Item Development

Target concepts were selected to be appropriate for both English and Spanish speakers in their respective languages. The selection process involved identifying words acquired early in development, characterized by brevity (one or two syllables) in both languages. These words were chosen based on their likely familiarity to students, irrespective of their backgrounds. Additional consideration was given to selecting words from diverse semantic categories, while avoiding phonemes or phonemic clusters known for difficulty in pronunciation in both languages. Additional information about the selected words and their semantic categories, length, frequency, and age of acquisition was also taken into consideration. The five items selected for the English ad Spanish subtest were hand, chair, cheese, moon, and cat.

It is important to note that student performance for completing this task in English and Spanish is not directly comparable. This is due to the rarity of highly imaginable monosyllabic nouns in Spanish. Consequently, while the majority of items in English are monosyllabic –thus, faster to pronounce–, their Spanish counterparts are often disyllabic and phonetically longer, making their labeling slower.

20.4 Scoring

Participating children were presented with a 5x10 grid containing 50 illustrations of the selected objects and were asked to name as fast as they could, and the time taken by the participant to name all objects was recorded. The final score consisted of a rate between the number of accurately named

objects (i.e., the total number of objects minus the incorrectly named objects) divided by the total time it took the child to complete namig all items in the grid.

20.5 Samples

Table 20.1: Demographic Characteristics of Samples for the English and Spanish Rapid Automatized Naming of Objects Tasks

	English			Spanish		
Characteristic	**K** N = 2,751	**G1** N = 3,104	**G2** N = 798	**K** N = 1,202	**G1** N = 1,192	**G2** N = 376
Timepoint						
Fall 2023	461 (18%)	396 (14%)	418 (79%)	438 (44%)	388 (41%)	370 (99%)
Fall 2024	2,054 (82%)	2,381 (86%)	113 (21%)	568 (56%)	560 (59%)	3 (0.8%)
Unknown	236	327	267	196	244	3
Administration Format						
Not applicable	2,751 (100%)	3,104 (100%)	798 (100%)	1,202 (100%)	1,192 (100%)	376 (100%)
Race						
American/Alaskan Native	90 (3.6%)	108 (3.8%)	3 (0.4%)	39 (3.4%)	41 (3.5%)	2 (0.5%)
Asian	181 (7.3%)	222 (7.9%)	83 (11%)	47 (4.1%)	38 (3.2%)	6 (1.6%)
Black/African American	236 (9.5%)	280 (9.9%)	84 (11%)	12 (1.1%)	17 (1.5%)	3 (0.8%)
Not reported	295 (12%)	347 (12%)	179 (23%)	422 (37%)	475 (41%)	218 (58%)
Other	549 (22%)	418 (15%)	92 (12%)	258 (23%)	114 (9.7%)	12 (3.2%)
White	1,139 (46%)	1,441 (51%)	323 (42%)	361 (32%)	485 (41%)	134 (36%)
Unknown	261	288	34	63	22	1
Ethnicity						
Hispanic/Latin(o/a)	1,605 (70%)	1,948 (69%)	377 (50%)	1,001 (95%)	1,100 (95%)	345 (92%)
Intentional nonreport	32 (1.4%)	12 (0.4%)	3 (0.4%)	5 (0.5%)	1 (<0.1%)	0 (0%)
Not Hispanic/Latin(o/a)	669 (29%)	862 (31%)	374 (50%)	49 (4.6%)	59 (5.1%)	31 (8.2%)
Unknown	445	282	44	147	32	0
Gender						
Female	1,158 (50%)	1,408 (50%)	360 (47%)	557 (53%)	608 (54%)	195 (52%)
Male	1,170 (50%)	1,395 (50%)	403 (53%)	496 (47%)	527 (46%)	181 (48%)
Unknown	423	301	35	149	57	0
Home Language						
English	1,531 (66%)	1,655 (65%)	560 (74%)	167 (15%)	144 (12%)	48 (13%)
Spanish	702 (30%)	834 (33%)	115 (15%)	956 (85%)	1,018 (87%)	315 (85%)
Other	93 (4.0%)	75 (2.9%)	81 (11%)	6 (0.5%)	4 (0.3%)	6 (1.6%)
Unknown	425	540	42	73	26	7
English Proficiency Label						
(Re-)Classified Proficient	87 (4.1%)	139 (5.6%)	61 (8.1%)	94 (9.3%)	138 (12%)	54 (15%)
English Learner	647 (30%)	758 (31%)	136 (18%)	803 (79%)	861 (77%)	270 (74%)
English-only	1,394 (66%)	1,569 (64%)	559 (74%)	115 (11%)	122 (11%)	42 (11%)
Unknown	623	638	42	190	71	10
Ever IEP/504	154 (8.2%)	203 (9.4%)	77 (13%)	79 (8.5%)	93 (9.3%)	33 (12%)
Unknown	882	938	227	270	187	109

20.6 Score distribution

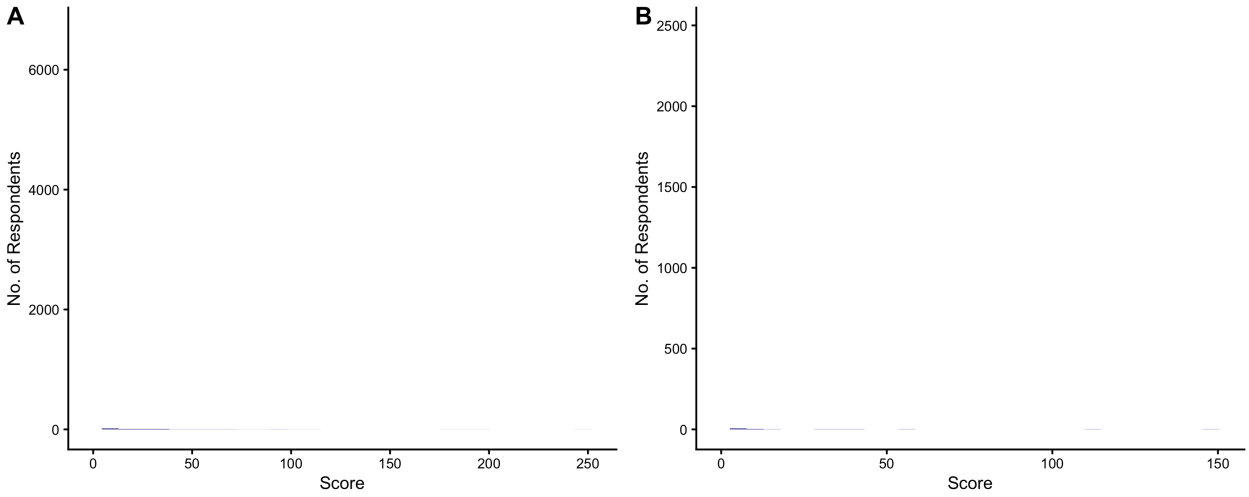

Figure 20.1: Score Distribution of the English and Spanish Rapid Automatized Naming of Objects Tasks

20.7 Criterion Validity Evidence

20.7.1 Sample

Table 20.2: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Rapid Automatized Naming of Objects Tasks

	English		Spanish	
Characteristic	**K** N = 209	**G1** N = 227	**K** N = 163	**G1** N = 214
Timepoint				
Spring 2024	209 (100%)	227 (100%)	163 (100%)	214 (100%)
Race				
American/Alaskan Native	2 (1.0%)	4 (1.8%)	5 (3.1%)	3 (1.4%)
Asian	29 (14%)	30 (13%)	9 (5.5%)	5 (2.4%)
Black/African American	23 (11%)	26 (11%)	1 (0.6%)	2 (0.9%)
Not reported	38 (18%)	56 (25%)	64 (39%)	100 (47%)
Other	50 (24%)	30 (13%)	29 (18%)	8 (3.8%)
White	67 (32%)	81 (36%)	55 (34%)	94 (44%)
Ethnicity				
Hispanic/Latin(o/a)	99 (47%)	97 (43%)	151 (93%)	199 (93%)
Intentional nonreport	10 (4.8%)	2 (0.9%)	1 (0.6%)	0 (0%)
Not Hispanic/Latin(o/a)	100 (48%)	128 (56%)	11 (6.7%)	15 (7.0%)
Gender				
Female	111 (53%)	108 (48%)	95 (58%)	117 (55%)
Male	98 (47%)	119 (52%)	68 (42%)	97 (45%)
Home Language				
English	151 (73%)	165 (73%)	28 (17%)	30 (14%)
Spanish	39 (19%)	32 (14%)	135 (83%)	180 (85%)
Other	18 (8.7%)	28 (12%)	0 (0%)	2 (0.9%)
Unknown	1	2	0	2
English Proficiency Label				
(Re-)Classified Proficient	17 (9.2%)	26 (12%)	23 (15%)	37 (18%)
English Learner	38 (21%)	45 (20%)	114 (75%)	149 (71%)
English-only	129 (70%)	152 (68%)	15 (9.9%)	24 (11%)
Unknown	25	4	11	4
Ever IEP/504	9 (5.9%)	22 (11%)	8 (5.7%)	17 (9.3%)
Unknown	57	22	23	32
Unknown			0	2

English Rapid Automatized Naming of Objects was correlated with two measures: the Objects subtest from the Acadience Learning RAN (Powell-Smith et al. 2020b) in kindergarten and first grade, and the RAN Objects subtest from the Rapid Automatized Naming and Rapid Alternating Stimulus Tests (M. Wolf and Denckla 2003b) in second grade.

Spanish Rapid Automatized Naming of Objects was correlated with the Objects subtest from the Acadience Learning RAN (Powell-Smith et al. 2020b) in kindergarten and first grade. No rapid naming of objects task could be found in Spanish that was normed on second-grade-aged children in the United States.

These analyses are problematic because of differences in how the measures are scored. The external assessments are scored based on total time, with RAN Objects and RAN Letters being completed in tandem. The Multitudes RANO measure is scored as the number correct divided by the total time and is distinct from RANL. Given these scoring differences, the correlations are difficult to interpret.

Table 20.3: Concurrent Criterion Validity Correlations for the English and Spanish Rapid Automatized Naming of Objects Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	209	-0.50 [-0.60, -0.39]	38	-0.37 [-0.62, -0.06]	163	-0.58 [-0.67, -0.47]
G1	225	-0.57 [-0.65, -0.47]	45	-0.67 [-0.81, -0.47]	214	-0.61 [-0.69, -0.52]

21 Semantic Mapping

21.1 Task Description

Children are shown a group of pictures and are asked to choose the picture that doesn't belong with the others. The constructs represented by the items become more complex to increase item difficulty.

21.2 Construct

The Semantic Mapping task requires knowledge of categories and features that create a relationship between items. It requires doing a fast mapping of the objects depicted to identify within group belonging/exclusion. This task can be administered in any language given that it is a receptive task and children simply point to the picture that does not belong. The test has been calibrated in English and Spanish, but the administration instructions are available in Arabic, Madarin Tagalog, and Vietnamese.

21.3 Theoretical and Empirical Foundations

Semantic mapping provides a measure of semantic depth versus the breadth measured by the Expressive Vocabulary (EVO) task. Measures of semantic depth show slower growth in children with language disorders across school-age (K. K. McGregor et al. 2013) and can be used to distinguish bilingual children with and without language difficulties (J. Jasso et al. 2020). Individuals must identify underlying relationships between objects which measures not only semantic knowledge but also underlying concept development. Concept development undergirds reading comprehension and is crucial for providing the resources for reading development across languages (Y.-S. Kim 2023).

21.4 Item Development

A list of semantic categories was developed by the research team, using the words targeted by the curricula used in dual language programs as a reference. These categories were both ordinate (e.g., animals) and subordinate (e.g., farm animals), and had different levels of concreteness, ranging from highly concrete (e.g., fruits) to abstract (e.g., things that produce artificial light).

For each category, one foil was selected. These foils could be radically different for easier items (e.g., a rocket for a clothing category including a t-shirt and a dress) to moderately different for harder items (e.g., a spoon for a category of gardening tools including a rake and a shovel).

The number of pictures per item also varied as a proxy of difficulty, ranging from 3 pictures (2 targets and 1 foil) for easier items to 5 pictures (4 targets and 1 foil) for harder items. Researchers used iStock to choose the real pictures to represent both the target and foils. The chosen pictures underwent a rigorous selection process to meet specific criteria:

- **Easily Recognizable.** Emphasis was placed on selecting images that could be easily identified.
- **Consistent Background.** Preference was given to pictures with a clean and unobtrusive background, and a white background was opted for whenever possible. For those cases in which at least one of the pictures required a background, pictures with backgrounds were selected for all the items, to ensure visual consistency and avoid children's use of the background as influencing their selection.
- **Diversity Representation.** The Justice, Equity, Diversity, and Inclusion (JEDI) team reviewed all the selected pictures to ensure diversity in representation, based on race/ethnicity, gender identity, age, cultural artifacts, etc.
- **Cultural responsiveness.** Items that were potentially culturally unfamiliar or inappropriate were deliberately excluded from the final selection.

21.5 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

21.6 Calibration Samples

Table 21.1: Demographic Characteristics of Calibration Samples for the English and Spanish Semantic Mapping Tasks

	English			Spanish		
Characteristic	**K** N = 435	**G1** N = 486	**G2** N = 36	**K** N = 1,000	**G1** N = 1,050	**G2** N = 0
Timepoint						
Spring 2023	0 (0%)	0 (0%)	0 (0%)	608 (61%)	646 (62%)	0 (NA%)
Winter 2024	303 (70%)	338 (70%)	0 (0%)	0 (0%)	0 (0%)	0 (NA%)
Fall 2024	132 (30%)	148 (30%)	36 (100%)	392 (39%)	404 (38%)	0 (NA%)
Administration Format						
CAT	132 (30%)	148 (30%)	36 (100%)	392 (39%)	404 (38%)	0 (NA%)
Forms	303 (70%)	338 (70%)	0 (0%)	608 (61%)	646 (62%)	0 (NA%)
Race						
American/Alaskan Native	14 (3.3%)	14 (2.9%)	0 (0%)	35 (3.6%)	33 (3.2%)	0 (NA%)
Asian	49 (11%)	60 (12%)	0 (0%)	18 (1.9%)	18 (1.8%)	0 (NA%)
Black/African American	66 (15%)	81 (17%)	0 (0%)	9 (0.9%)	7 (0.7%)	0 (NA%)
Not reported	60 (14%)	66 (14%)	0 (0%)	458 (47%)	517 (51%)	0 (NA%)
Other	105 (24%)	73 (15%)	8 (100%)	170 (18%)	117 (12%)	0 (NA%)
White	135 (31%)	189 (39%)	0 (0%)	277 (29%)	325 (32%)	0 (NA%)
Unknown	6	3	28	33	33	0
Ethnicity						
Hispanic/Latin(o/a)	194 (48%)	233 (49%)	0 (NA%)	847 (97%)	939 (97%)	0 (NA%)
Intentional nonreport	6 (1.5%)	1 (0.2%)	0 (NA%)	2 (0.2%)	2 (0.2%)	0 (NA%)
Not Hispanic/Latin(o/a)	204 (50%)	246 (51%)	0 (NA%)	20 (2.3%)	28 (2.9%)	0 (NA%)
Unknown	31	6	36	131	81	0
Gender						
Female	209 (50%)	234 (48%)	2 (25%)	462 (53%)	553 (59%)	0 (NA%)
Male	212 (50%)	249 (52%)	6 (75%)	407 (47%)	388 (41%)	0 (NA%)
Unknown	14	3	28	131	109	0
Home Language						
English	318 (76%)	378 (79%)	8 (100%)	75 (7.8%)	73 (7.2%)	0 (NA%)
Spanish	66 (16%)	73 (15%)	0 (0%)	876 (91%)	935 (92%)	0 (NA%)
Other	37 (8.8%)	28 (5.8%)	0 (0%)	16 (1.7%)	7 (0.7%)	0 (NA%)
Unknown	14	7	28	33	35	0
English Proficiency Label						
(Re-)Classified Proficient	14 (4.0%)	32 (7.1%)	0 (0%)	72 (8.7%)	94 (10%)	0 (NA%)
English Learner	76 (22%)	70 (16%)	0 (0%)	696 (84%)	760 (83%)	0 (NA%)
English-only	259 (74%)	348 (77%)	8 (100%)	57 (6.9%)	61 (6.7%)	0 (NA%)
Unknown	86	36	28	175	135	0
Ever IEP/504	24 (7.8%)	49 (13%)	0 (NA%)	62 (10%)	54 (9.9%)	0 (NA%)
Unknown	127	96	36	394	507	0

21.7 Psychometric Analysis

21.7.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 11 items from the English task and 6 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 21.2 summarizes the basic item characteristics, Figure 21.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 21.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Semantic Mapping Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 124**	**N = 113**	**N = 124**	**N = 118**
No. of Responses	124 (91)	130 (93)	249 (162)	256 (163)
Proportion Correct	0.67 (0.22)	0.71 (0.18)	0.64 (0.20)	0.66 (0.19)
Point-biserial Correlation	0.44 (0.18)	0.48 (0.14)	0.48 (0.17)	0.50 (0.15)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	11 (8.9%)	0 (0%)	6 (4.8%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

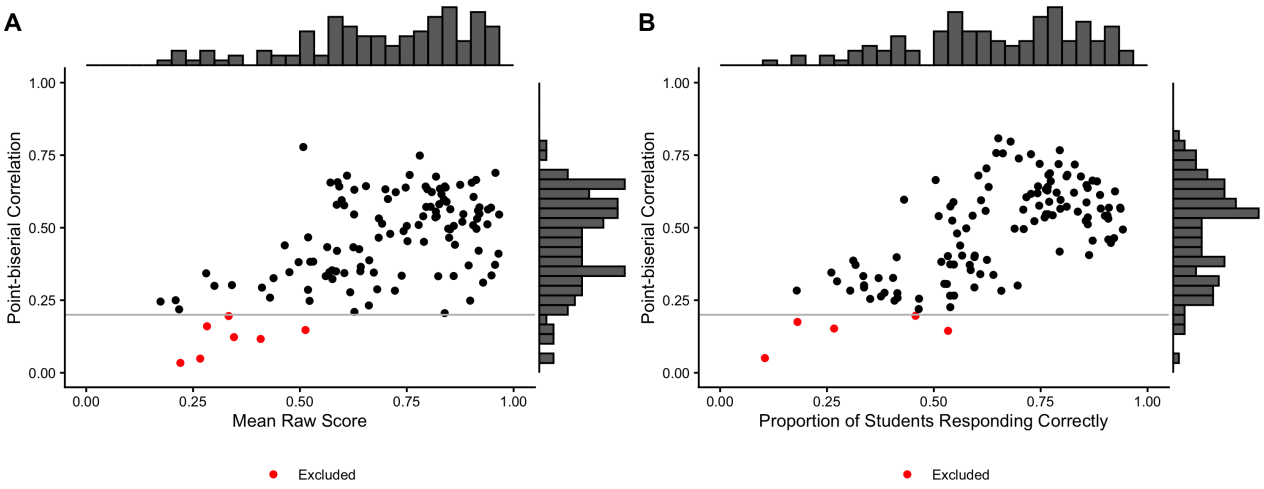

Figure 21.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Semantic Mapping Tasks

21.7.2 Rasch Analysis

21.7.2.1 Item Location Estimates

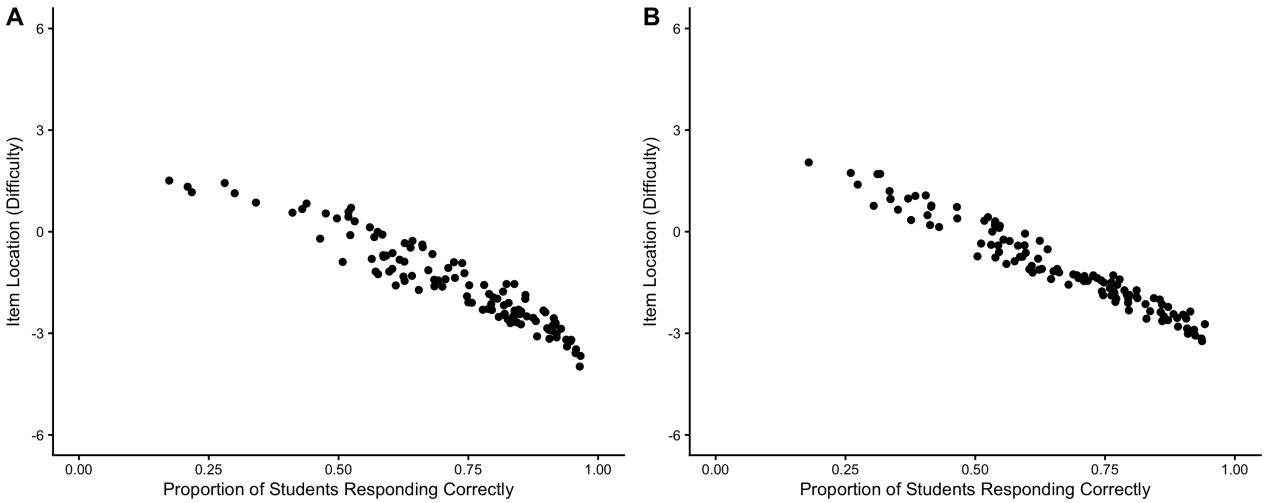

Figure 21.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Semantic Mapping Tasks

21.7.2.2 Item Fit Statistics

Table 21.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Semantic Mapping Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	89	0	0	0	89	103	0	0	0	103
B	14	0	0	0	14	7	0	0	0	7
C	10	0	0	0	10	8	0	0	0	8
D	0	0	0	0	0	0	0	0	0	0
Total	113	0	0	0	113	118	0	0	0	118

21.7.2.3 Person Location Estimates

21.7.2.4 Person Fit Statistics

Table 21.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English (Panel A) and Spanish (Panel B) Semantic Mapping Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	640	0	0	0	640	1,396	0	4	0	1,400
B	156	89	0	0	245	269	155	0	0	424
C	23	0	5	0	28	86	0	13	0	99
D	8	0	8	4	20	20	0	10	2	32
Total	827	89	13	4	933	1,771	155	27	2	1,955

21.7.2.5 Distribution of Theta Estimates

21.7.2.6 Wright Maps

21.7.2.7 Model Summary

Table 21.5: Summary of Rasch Model Statistics for the English and Spanish Semantic Mapping Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 113	N = 933	N = 118	N = 1,955
Logit Scale Location	-1.49 (1.30)	0.16 (-0.74, 0.98)	-1.06 (1.28)	0.27 (-0.84, 1.05)
Outfit	0.93 (0.36)	0.79 (0.49, 1.00)	0.95 (0.32)	0.83 (0.54, 1.04)
Infit	0.97 (0.14)	0.88 (0.73, 1.04)	0.97 (0.15)	0.89 (0.75, 1.05)
Reliability of Separation	0.7647	0.6828	0.7908	0.7474

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 113 and 118 for the English and Spanish task, respectively.

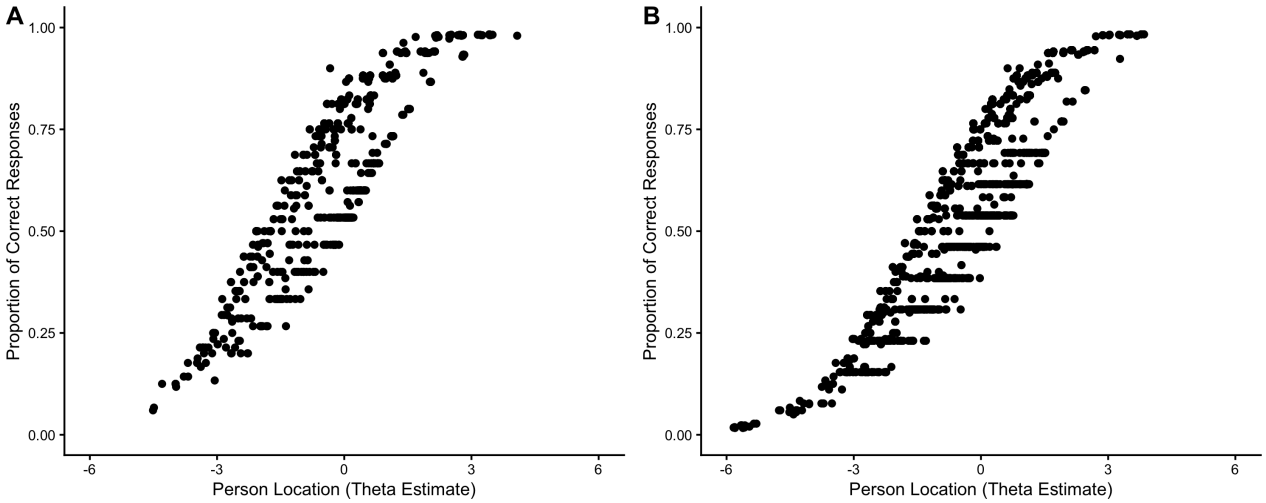

Figure 21.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responess for English and Spanish Semantic Mapping Tasks

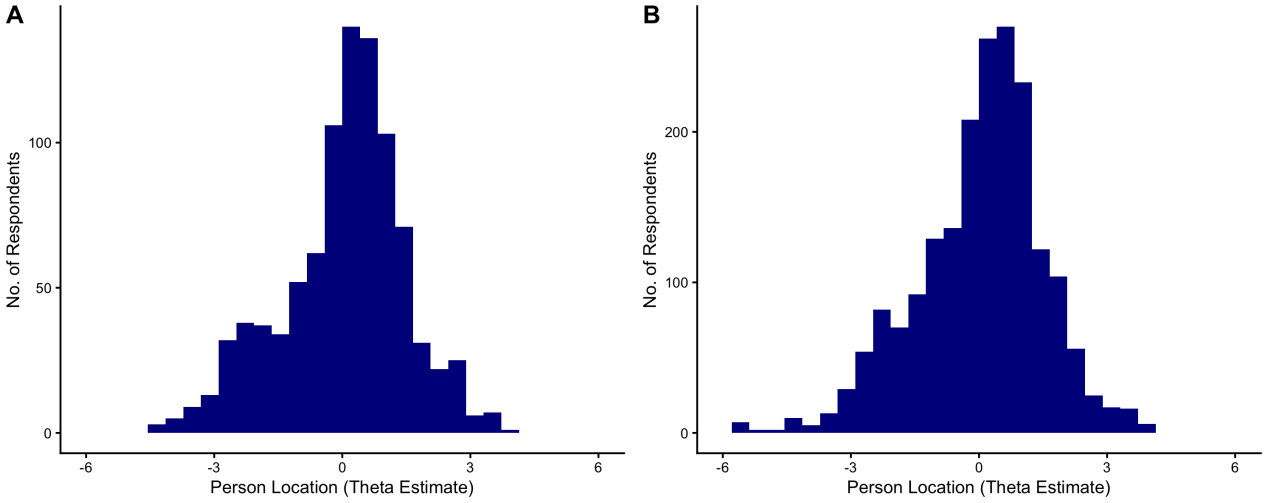

Figure 21.4: Distribution of Theta Estimates for the English and Spanish Semantic Mapping Tasks

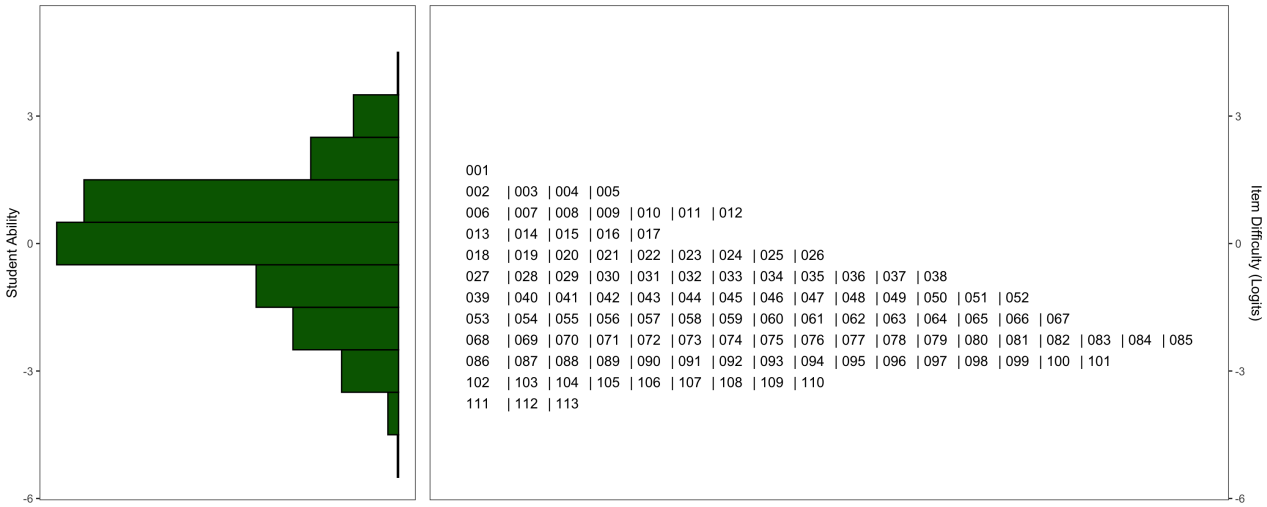

Figure 21.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Semantic Mapping Task

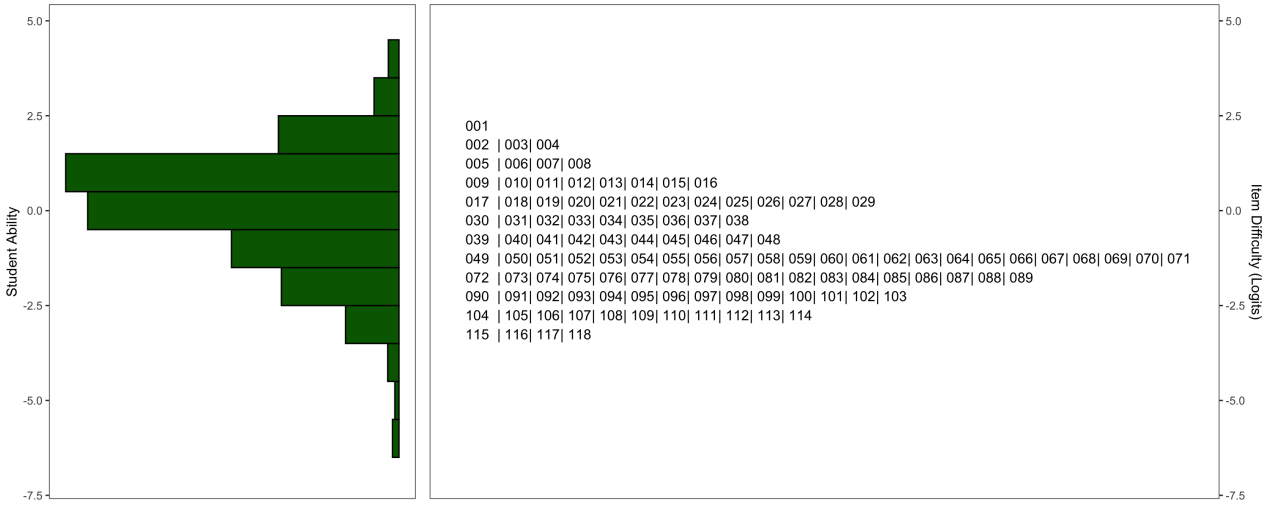

Figure 21.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Semantic Mapping Task

21.8 Criterion Validity Evidence

21.8.1 Sample

Forthcoming.

#| label: tbl-criterion-validity-smt #| tbl-cap: “Concurrent Criterion Validity Correlations for the English and Spanish Semantic Mapping Tasks” tbl.critval <- fun.criterion_validity_table(“SMT”) tbl.critval “

22 Spelling

22.1 Task Description

The child hears a word and spells it by selecting the correct letters among the foils available on the screen and drags them to the bottom of the screen in the correct order.

22.2 Construct

The Spelling task measures the children's ability to put the alphabetic principle into action and encode speech sounds into print.

22.3 Item Development

22.3.1 English

For the development of the item pool, the research team reviewed multiple curricula to build up a list of frequent, decodable words, including curricula used in the United States, like McGraw-Hill's "Wonders", Benchmark's "Benchmark Advance", and HMH's "Journey".

From this pool of items, Clearpond (Marian et al. 2012) was used to retrieve information on the word's frequency, orthographic and phonological length, and neighborhood frequency. This information was used to select a sample of hihg frequency words whose semantic meaning was overall easily accessed by the target population, with varying orthographic and phonological length. The final list of words also targeted the following characteristics: short, long, and variant vowels; r-controlled vowels; use of soft c and g; silent letters (e.g., /bm/, /sc/); diphthongs; consonant digraphs; two and three-letter blends; closed, open, and CVC syllables.

Using the letters of each target word as reference, the research team selected between 3 and 5 foil letters to be included among the correct letters for spelling the word. Foil letters were selected based on different criteria:

- **Phonological foils:** letters with similarly-sounding phonemes as the target letter (e.g., z for s; c for k; v for b)
- **Visual foils:** letters visually similar to the target letter (e.g., d for b; m for n)
- **Vocalic foils:** alternative vowels to the targeted ones (e.g., o for a; e for i)
- **Morphological foils:** alternative spelling of a conventional morpheme (e.g., z for s, t for ed for past tense verbs)

- **Unrelated foils:** additional foils were included in the pool that were not easily confused with the letters needed to spell the target word.

22.3.2 Spanish

For the development of the item pool, the research team reviewed multiple curricula to build up a list of frequent, decodable words, including curricula used in dual language programs in California, including McGraw-Hill Maravillas, Estrellita, Houghton Mifflin Lectura. Curricular materials from Mexico, Panama, and Chile were also reviewed.

From this pool of items, Clearpond (Marian et al. 2012) was used to retrieve information on the word's frequency, orthographic and phonological length, and neighborhood frequency. This information was used to select a sample of words that were high frequency, whose semantic meaning was overall easily accessed by the target population, and that had varying orthographic and phonological lengths.

Using the letters of each target word as reference, the research team selected between 3 and 5 foil letters to be included among the correct letters for spelling the word. Foils letters were based on different criteria:

- **Phonological foils:** letters that sound similar to the target letter (e.g., z for s; c for k; v for b)
- **Visual foils:** letters that are visually similar to the target letter (e.g., d for b; m for n)
- **Vocalic foils:** alternative vowels to the targeted ones (e.g., o for a; e for i)
- **Stress foils:** Spanish language uses accents for stressed letters and the child had to discern if the word contained the accented or nonaccented letter (e.g., é for e; a for á)
- **Crosslinguistic English phonology foils:** phonemes that are represented with a different letter in English than they are in Spanish (e.g., th for d)
- **Unrelated foils:** additional foils were included in the pool that were not easily confused with the letters needed to speall the target word.

22.4 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

22.5 Calibration Samples

Table 22.1: Demographic Characteristics of Calibration Samples for the English and Spanish Spelling Tasks

Characteristic	English **G2** N = 2,805	Spanish **G2** N = 299
Timepoint		
Fall 2024	2,191 (100%)	0 (NA%)
Unknown	614	299
Administration Format		
CAT	2,191 (78%)	
Forms	614 (22%)	299 (100%)
Race		
American/Alaskan Native	54 (2.1%)	4 (1.3%)
Asian	179 (6.9%)	3 (1.0%)
Black/African American	259 (10%)	4 (1.3%)
Not reported	310 (12%)	156 (53%)
Other	365 (14%)	14 (4.7%)
White	1,420 (55%)	116 (39%)
Unknown	218	2
Ethnicity		
Hispanic/Latin(o/a)	1,886 (74%)	264 (88%)
Intentional nonreport	5 (0.2%)	1 (0.3%)
Not Hispanic/Latin(o/a)	668 (26%)	34 (11%)
Unknown	246	
Gender		
Female	1,255 (49%)	173 (58%)
Male	1,291 (51%)	126 (42%)
Unknown	259	
Home Language		
English	1,426 (59%)	71 (24%)
Spanish	890 (37%)	216 (74%)
Other	91 (3.8%)	5 (1.7%)
Unknown	398	7
English Proficiency Label		
(Re-)Classified Proficient	268 (11%)	40 (14%)
English Learner	701 (30%)	186 (64%)
English-only	1,405 (59%)	66 (23%)
Unknown	431	7
Ever IEP/504	210 (10%)	24 (9.5%)
Unknown	784	46

22.6 Psychometric Analysis

22.6.1 Basic Item Statistics

We excluded 0 items from the English task and 9 items from the Spanish task based on low response counts ($n < 90$). 2 items were excluded because they had no variance in the Spanish task, and 2 items in the English task. Additionally, we excluded 1 items from the English task and 3 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 22.2 summarizes the basic item characteristics, Figure 22.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 22.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Spelling Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 90**	**N = 87**	**N = 102**	**N = 88**
No. of Responses	440 (371)	452 (372)	120 (84)	132 (83)
Proportion Correct	0.42 (0.22)	0.42 (0.21)	0.42 (0.21)	0.43 (0.19)
Point-biserial Correlation	0.57 (0.12)	0.58 (0.11)	0.53 (0.16)	0.54 (0.13)
Excluded (n < 90)	0 (0%)	0 (0%)	9 (8.8%)	0 (0%)
Excluded (pbis < .2)	1 (1.1%)	0 (0%)	3 (3.0%)	0 (0%)
Excluded (no variation)	2 (2.2%)	0 (0%)	2 (2.0%)	0 (0%)

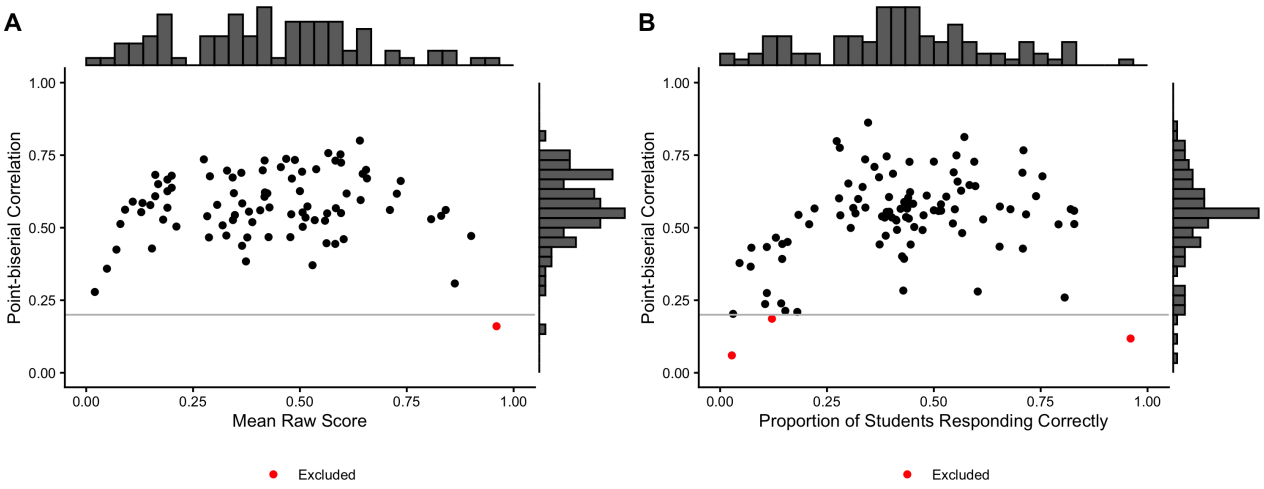

Figure 22.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Spelling Tasks

22.6.2 Rasch Analysis

22.6.2.1 Item Location Estimates

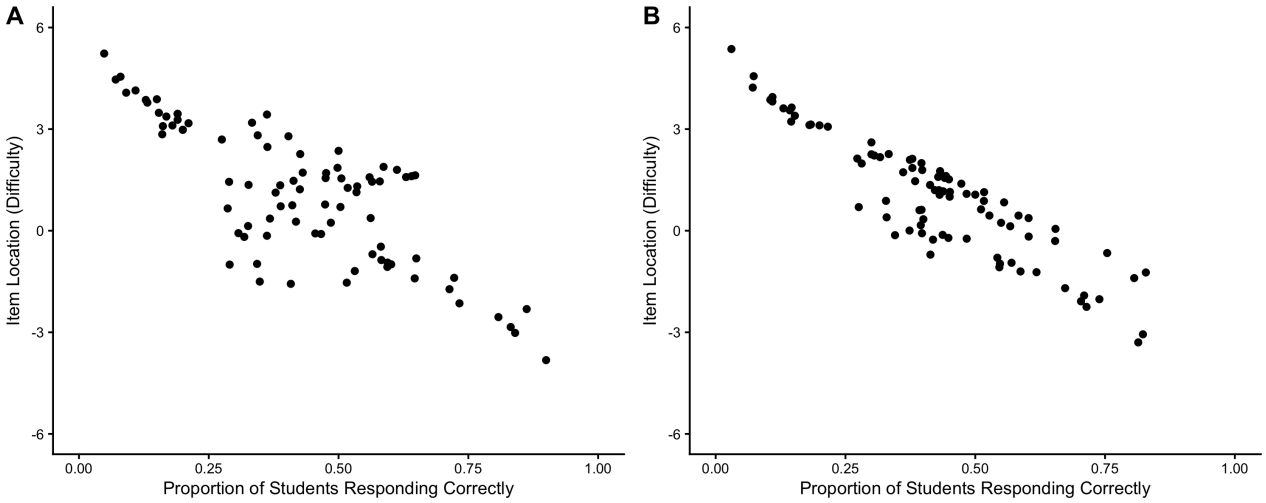

Figure 22.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Spelling Tasks

22.6.2.2 Item Fit Statistics

Table 22.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Spelling Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	67	0	0	0	67	70	0	0	0	70
B	10	0	0	0	10	4	0	0	0	4
C	5	0	0	0	5	9	0	0	0	9
D	5	0	0	0	5	3	0	2	0	5
Total	87	0	0	0	87	86	0	2	0	88

22.6.2.3 Person Location Estimates

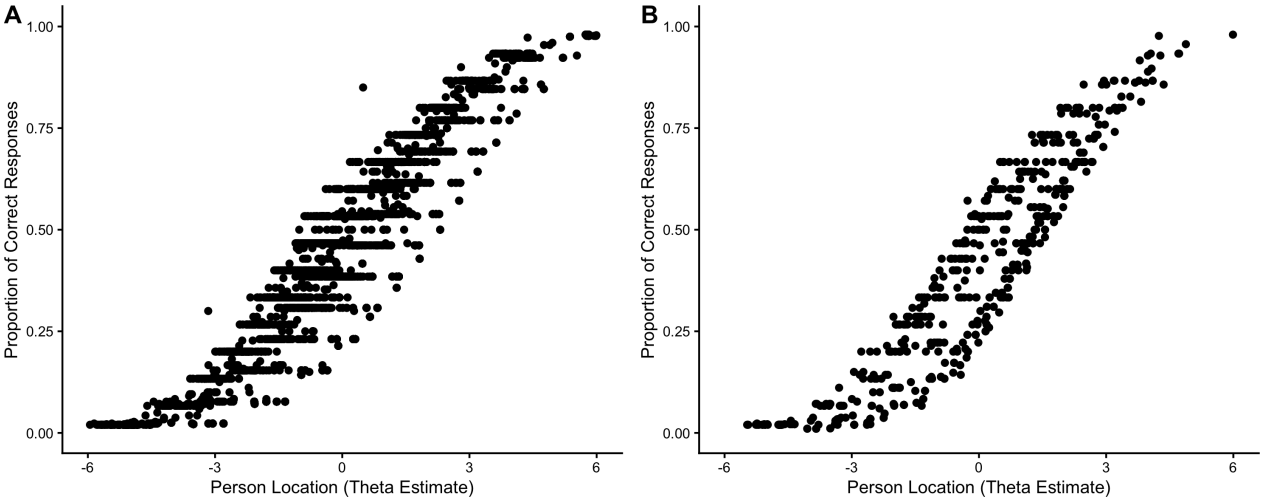

Figure 22.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Spelling Tasks

22.6.2.4 Person Fit Statistics

Table 22.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Spelling Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	1,846	0	9	2	1,857	438	0	1	0	439
B	319	364	0	0	683	67	102	0	0	169
C	70	0	14	1	85	24	0	6	0	30
D	78	0	28	7	113	15	0	8	0	23
Total	2,313	364	51	10	2,738	544	102	15	0	661

22.6.2.5 Distribution of Theta Estimates

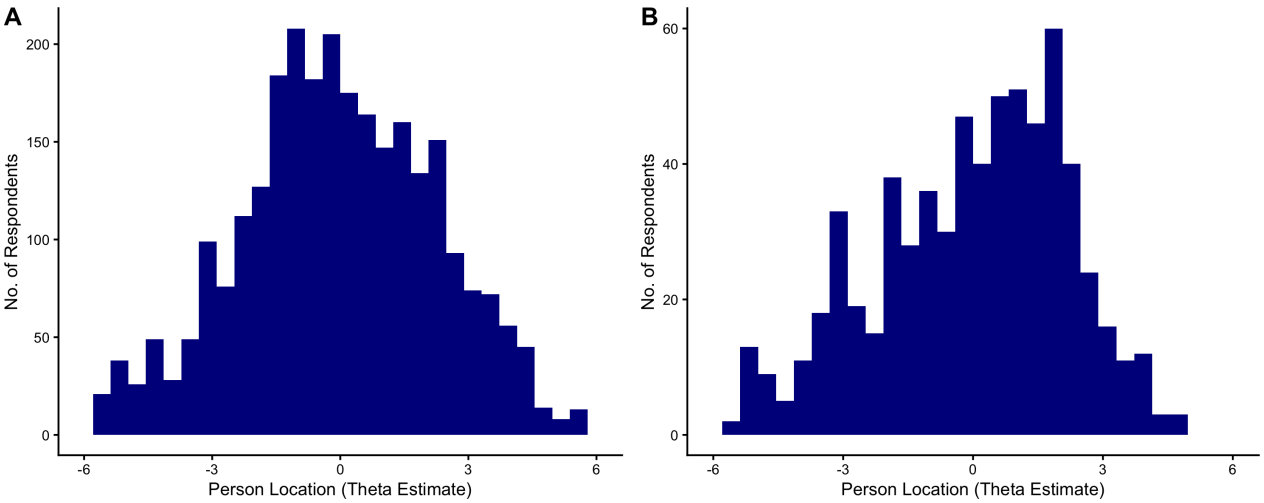

Figure 22.4: Distribution of Theta Estimates for the English and Spanish Spelling Tasks

22.6.2.6 Wright Maps

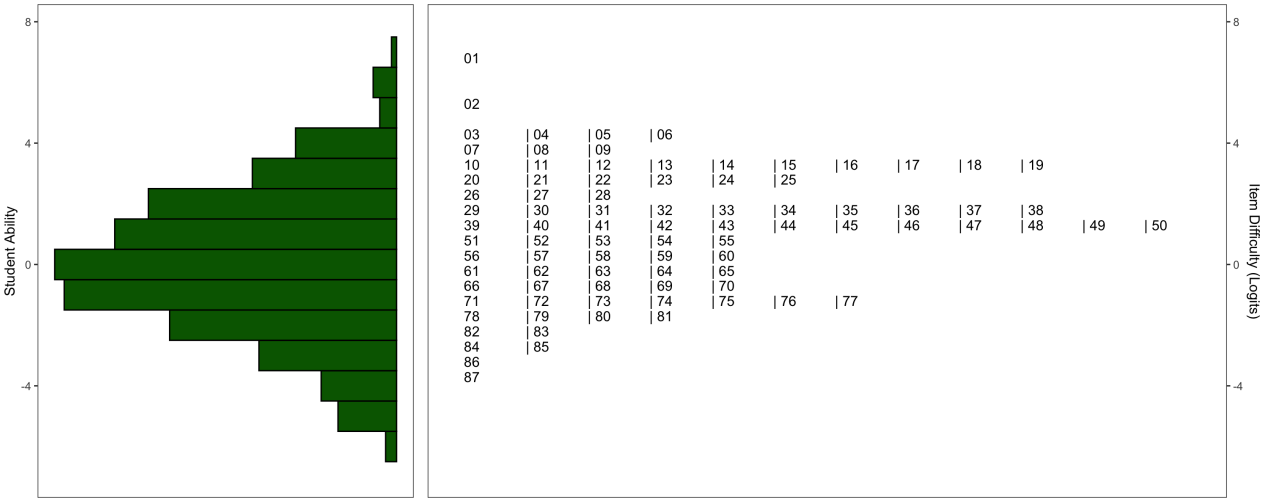

Figure 22.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Spelling Task

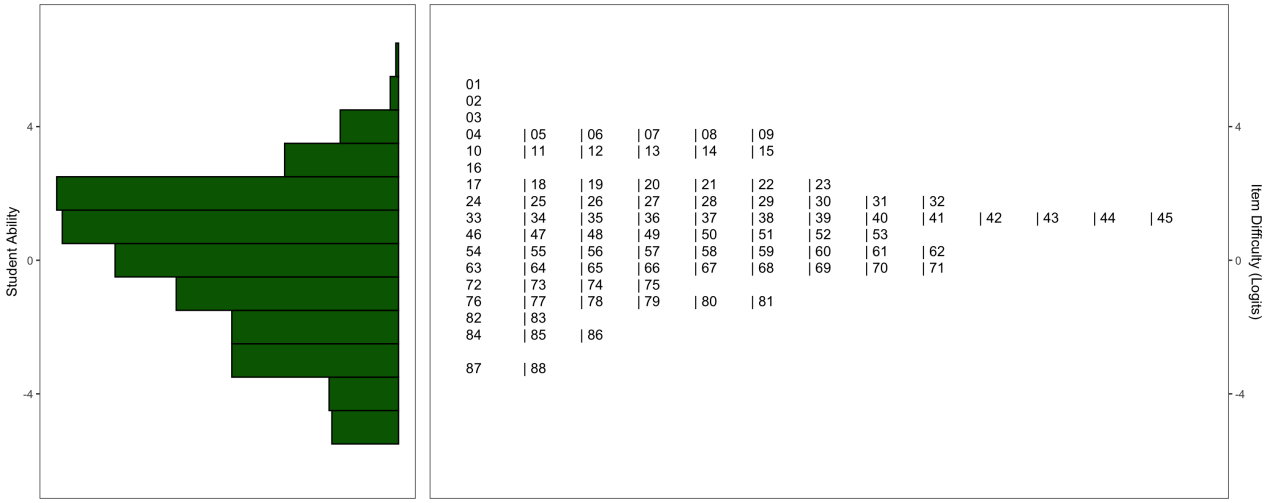

Figure 22.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Spelling Task

22.6.2.7 Model Summary

Table 22.5: Summary of Rasch Model Statistics for the English and Spanish Spelling Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 87	N = 2,738	N = 88	N = 661
Logit Scale Location	1.15 (2.08)	-0.09 (-1.54, 1.66)	0.98 (1.77)	0.25 (-1.57, 1.76)
Outfit	1.07 (0.85)	0.69 (0.50, 0.89)	1.10 (0.60)	0.70 (0.49, 0.97)
Infit	0.98 (0.13)	0.84 (0.68, 0.99)	1.02 (0.20)	0.85 (0.67, 1.01)
Reliability of Separation	0.8937	0.8560	0.8871	0.8307

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 87 and 88 for the English and Spanish task, respectively.

22.7 Criterion Validity Evidence

22.7.1 Sample

Table 22.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Spelling Tasks

	English	Spanish
Characteristic	**G2** N = 212	**G2** N = 221
Timepoint		
Spring 2024	212 (100%)	
Race		
American/Alaskan Native	2 (0.9%)	4 (1.8%)
Asian	15 (7.1%)	9 (4.1%)
Black/African American	27 (13%)	1 (0.5%)
Not reported	34 (16%)	50 (23%)
Other	45 (21%)	73 (33%)
White	89 (42%)	84 (38%)
Ethnicity		
Hispanic/Latin(o/a)	102 (48%)	202 (92%)
Intentional nonreport	3 (1.4%)	
Not Hispanic/Latin(o/a)	106 (50%)	18 (8.2%)
Unknown	1	1
Gender		
Female	93 (44%)	110 (50%)
Male	119 (56%)	110 (50%)
Home Language		
English	148 (70%)	54 (24%)
Spanish	38 (18%)	159 (72%)
Other	24 (11%)	8 (3.6%)
Unknown	2	
English Proficiency Label		
(Re-)Classified Proficient	17 (8.1%)	41 (19%)
English Learner	42 (20%)	136 (62%)
English-only	152 (72%)	42 (19%)
Unknown	1	2
Ever IEP/504	20 (13%)	16 (8.9%)
Unknown	54	41
Spring 2025		221 (100%)
Unknown		1

English Spelling was correlated with the Spelling subtest from the Woodcock-Johnson IV (WJ IV ACH) test (Schrank, McGrew, and Mather 2014). Spanish Spelling results are forthcoming.

Table 22.7: Concurrent Criterion Validity Correlations for the English and Spanish Spelling Tasks

English	Spanish

	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
G2	212	0.80 [0.75, 0.84]	42	0.79 [0.64, 0.88]	220	0.72 [0.65, 0.78]

23 Sentence Repetition

23.1 Task Description

Children listen to audio-recorded sentences and are prompted to repeat each sentence verbatim.

23.2 Construct

The Sentence Repetition task measures a child's ability to recall and reproduce sentences of varying length and complexity. This task taps into syntactic processing, lexical knowledge, and verbal working memory.

23.3 Item Development

23.3.1 English

Sentences varied by:

- **Vocabulary:** simpler, familiar vocabulary was used for easier items, while less frequent, complex vocabulary was used for harder items.
- **Sentence length:** sentences ranged from 4 to 12 words.
- **Targets:** one or two clauses were targeted by sentence. The targets varied by language. We selected target structures that have been shown to identify struggling readers and/or language difficulties in English. The following targets were used: article-noun, negative, passive, past, plural, possessive, prepositional phrase, question inversions, relative clause, and third person singular.

23.3.2 Spanish

Sentences varied by:

- **Vocabulary:** simpler, familiar vocabulary was used for easier items, while less frequent, complex vocabulary was used for harder items.
- **Sentence length:** sentences ranged from 4 to 12 words.
- **Targets:** one or two clauses were targeted by sentence. The following targets were used: adjective agreement, article, conditional, direct object clitic, imperfect past, negative, preposition, preterite past, progressive, reflexive, relative clause, and subjunctive.

23.4 Scoring

Items were polytomously scored according to the following scoring scheme: 2: The sentence was fully repeated without any error. 1: The sentence was repeated with 2 or less errors. 0: The sentence was repeated with 3 or more errors.

23.5 Calibration Samples

Table 23.1: Demographic Characteristics of Calibration Samples for the English and Spanish Sentence Repetition Tasks

	English			Spanish		
Characteristic	**K** N = 1,275	**G1** N = 1,695	**G2** N = 1,342	**K** N = 1,128	**G1** N = 1,313	**G2** N =
Timepoint						
Fall 2023	608 (48%)	666 (39%)	697 (52%)	0 (0%)	0 (0%)	357 (3
Winter 2024	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	275 (2
Fall 2024	667 (52%)	1,029 (61%)	645 (48%)	443 (100%)	522 (100%)	490 (4
Administration Format						
CAT	667 (52%)	1,029 (61%)	645 (48%)	443 (39%)	522 (40%)	528 (4
Forms	608 (48%)	666 (39%)	697 (52%)	685 (61%)	791 (60%)	632 (5
Race						
American/Alaskan Native	48 (3.8%)	60 (3.6%)	26 (2.0%)	28 (6.3%)	32 (6.2%)	24 (2.
Asian	107 (8.5%)	166 (10.0%)	128 (10.0%)	11 (2.5%)	23 (4.4%)	20 (1.
Black/African American	131 (10%)	189 (11%)	163 (13%)	6 (1.4%)	9 (1.7%)	8 (0.7
Not reported	219 (17%)	276 (17%)	230 (18%)	94 (21%)	142 (27%)	618 (5
Other	315 (25%)	242 (15%)	128 (10.0%)	157 (36%)	91 (18%)	67 (5.
White	441 (35%)	731 (44%)	611 (48%)	146 (33%)	220 (43%)	419 (3
Unknown	14	31	56	686	796	4
Ethnicity						
Hispanic/Latin(o/a)	790 (68%)	1,124 (68%)	820 (64%)	361 (98%)	494 (98%)	1,082 (
Intentional nonreport	13 (1.1%)	5 (0.3%)	3 (0.2%)	3 (0.8%)	2 (0.4%)	2 (0.2
Not Hispanic/Latin(o/a)	359 (31%)	515 (31%)	449 (35%)	6 (1.6%)	10 (2.0%)	53 (4.
Unknown	113	51	70	758	807	23
Gender						
Female	573 (49%)	845 (52%)	622 (49%)	163 (45%)	274 (57%)	600 (5
Male	600 (51%)	772 (48%)	646 (51%)	203 (55%)	206 (43%)	539 (4
Non-binary	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (<0
Unknown	102	78	74	762	833	20
Home Language						
English	655 (53%)	921 (57%)	789 (64%)	10 (2.3%)	26 (5.0%)	115 (1
Spanish	504 (41%)	603 (38%)	355 (29%)	431 (98%)	489 (95%)	1,026 (
Other	77 (6.2%)	84 (5.2%)	92 (7.4%)	1 (0.2%)	0 (0%)	9 (0.8
Unknown	39	87	106	686	798	10
English Proficiency Label						
(Re-)Classified Proficient	63 (5.8%)	113 (7.4%)	105 (8.6%)	23 (6.3%)	46 (9.6%)	184 (1
English Learner	460 (43%)	546 (36%)	340 (28%)	335 (91%)	406 (85%)	821 (7
English-only	559 (52%)	875 (57%)	776 (64%)	10 (2.7%)	25 (5.2%)	110 (9
Unknown	193	161	121	760	836	45
Ever IEP/504	69 (6.9%)	141 (10%)	117 (12%)	33 (9.1%)	39 (8.6%)	89 (10
Unknown	282	315	328	766	859	29
Unknown				685	791	38

23.6 Psychometric Analysis

23.6.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 1 items from the English task and 1 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 23.2 summarizes the basic item characteristics, Figure 23.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 23.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Sentence Repetition Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 180**	**N = 179**	**N = 158**	**N = 157**
No. of Responses	253 (243)	253 (243)	345 (228)	346 (228)
Proportion Correct	1.03 (0.55)	1.03 (0.55)	0.72 (0.48)	0.72 (0.47)
Point-biserial Correlation	0.61 (0.14)	0.61 (0.14)	0.61 (0.14)	0.61 (0.14)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	1 (0.6%)	0 (0%)	1 (0.6%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

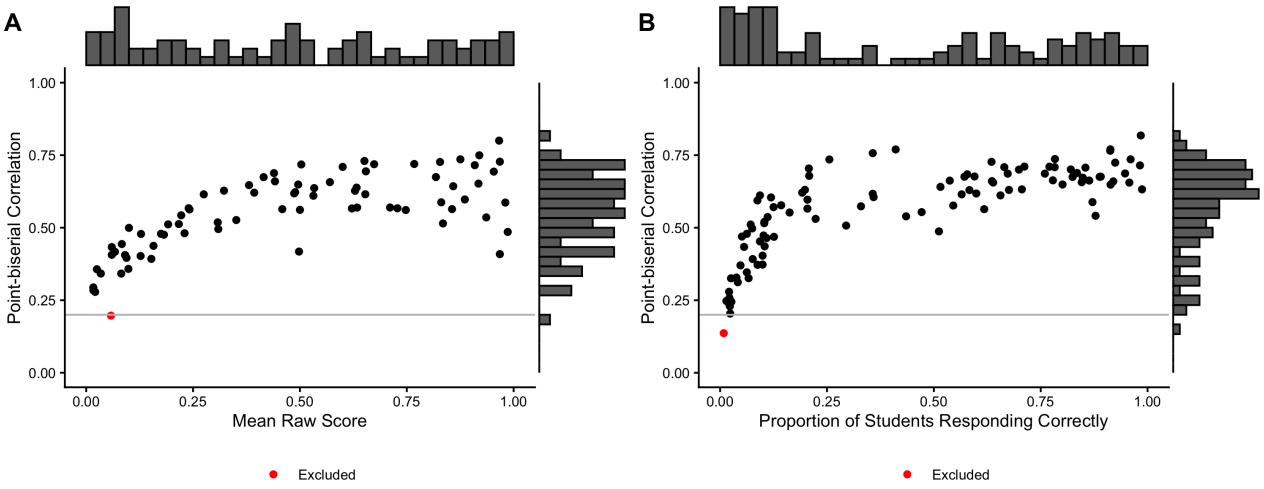

Figure 23.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Sentence Repetition Tasks

23.6.2 Rasch Analysis

23.6.2.1 Item Location Estimates

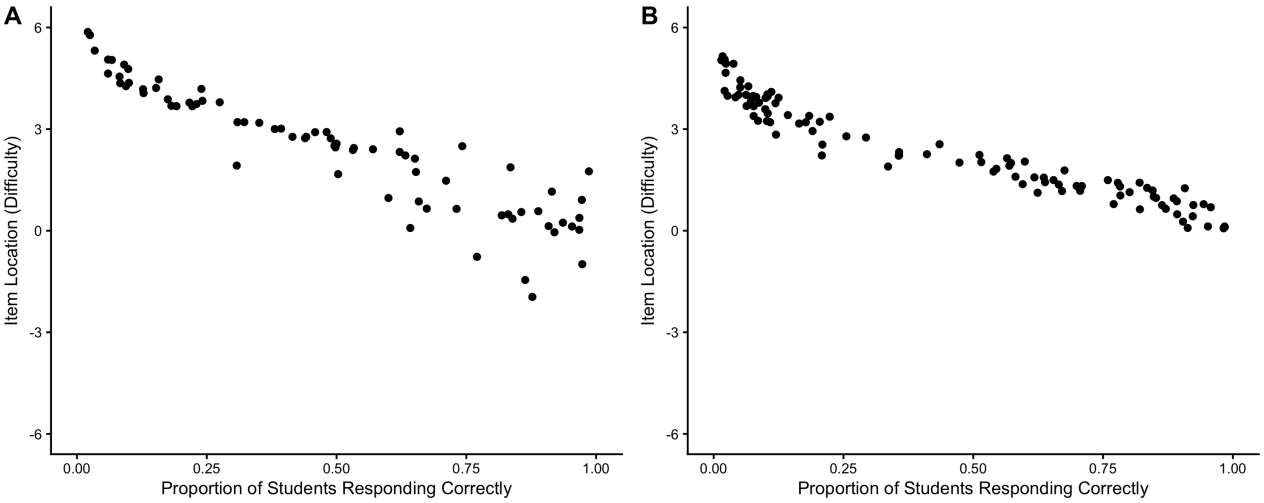

Figure 23.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Sentence Repetition Tasks

23.6.2.2 Item Fit Statistics

Table 23.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Sentence Repetition Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	163	0	0	0	163	122	0	0	0	122
B	9	0	0	0	9	28	0	0	0	28
C	5	0	0	0	5	4	0	1	0	5
D	2	0	0	0	2	2	0	0	0	2
Total	179	0	0	0	179	156	0	1	0	157

23.6.2.3 Person Location Estimates

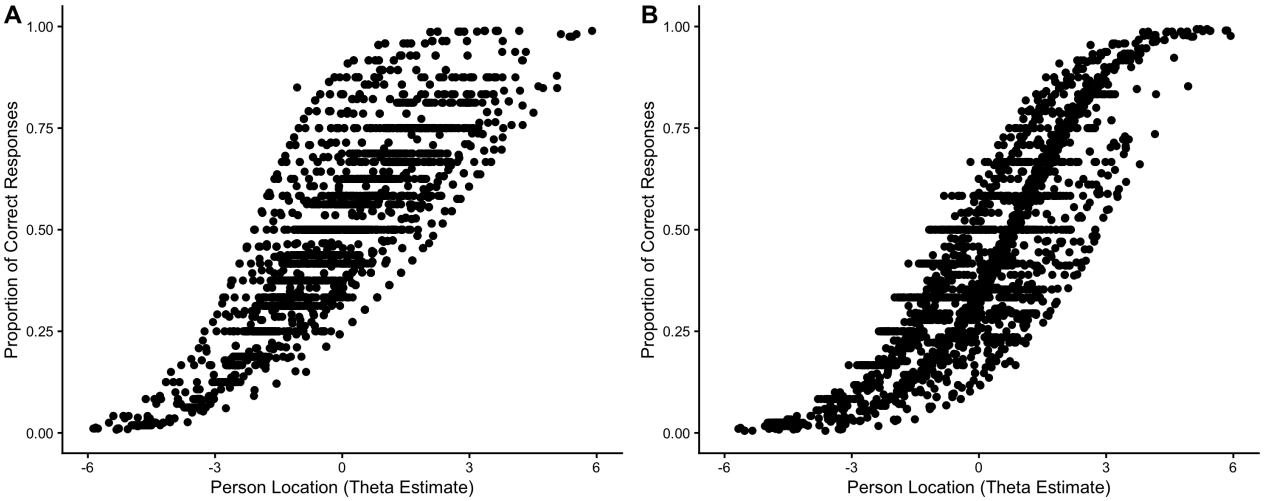

Figure 23.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Sentence Repetition Tasks

23.6.2.4 Person Fit Statistics

Table 23.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Sentence Repetition Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	2,207	37	69	3	2,316	1,652	13	60	9	1,734
B	526	1,036	0	0	1,562	516	945	2	1	1,464
C	64	0	79	14	157	67	0	45	7	119
D	61	0	44	47	152	35	0	38	97	170
Total	2,858	1,073	192	64	4,187	2,270	958	145	114	3,487

23.6.2.5 Distribution of Theta Estimates

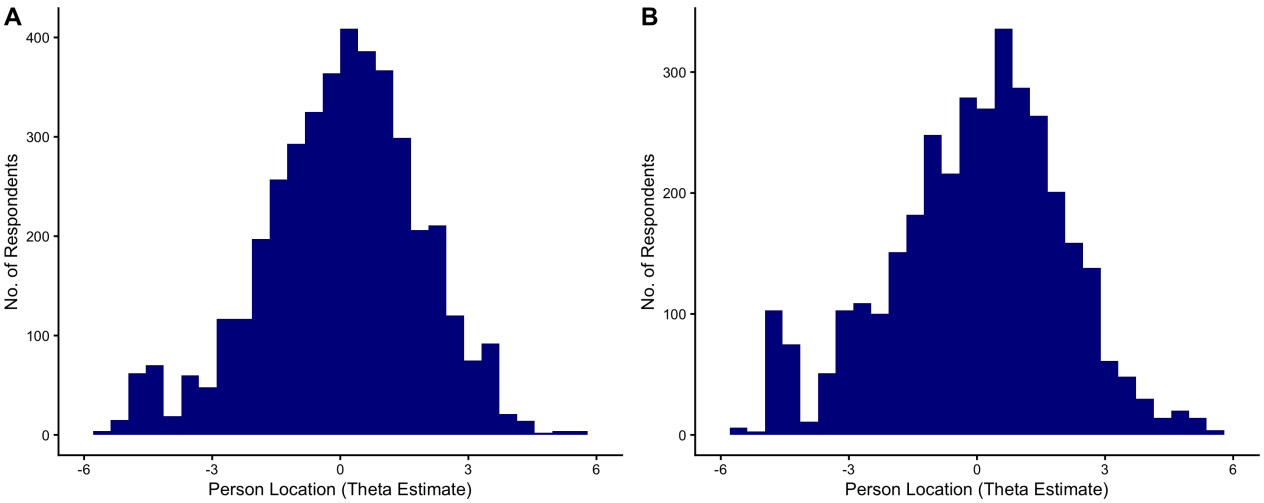

Figure 23.4: Distribution of Theta Estimates for the English and Spanish Sentence Repetition Tasks

23.6.2.6 Wright Maps

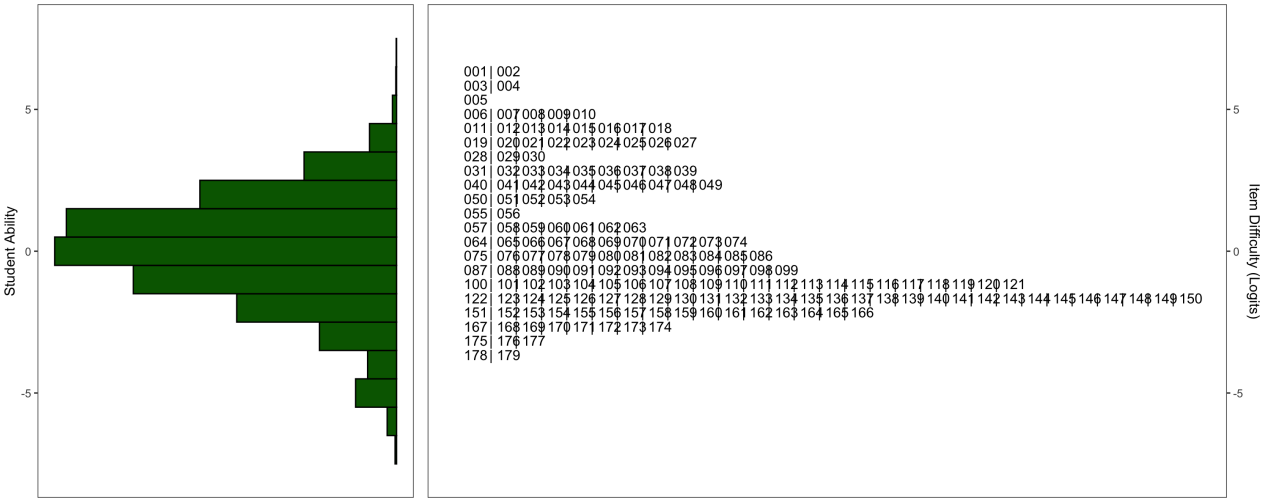

Figure 23.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Sentence Repetition Task

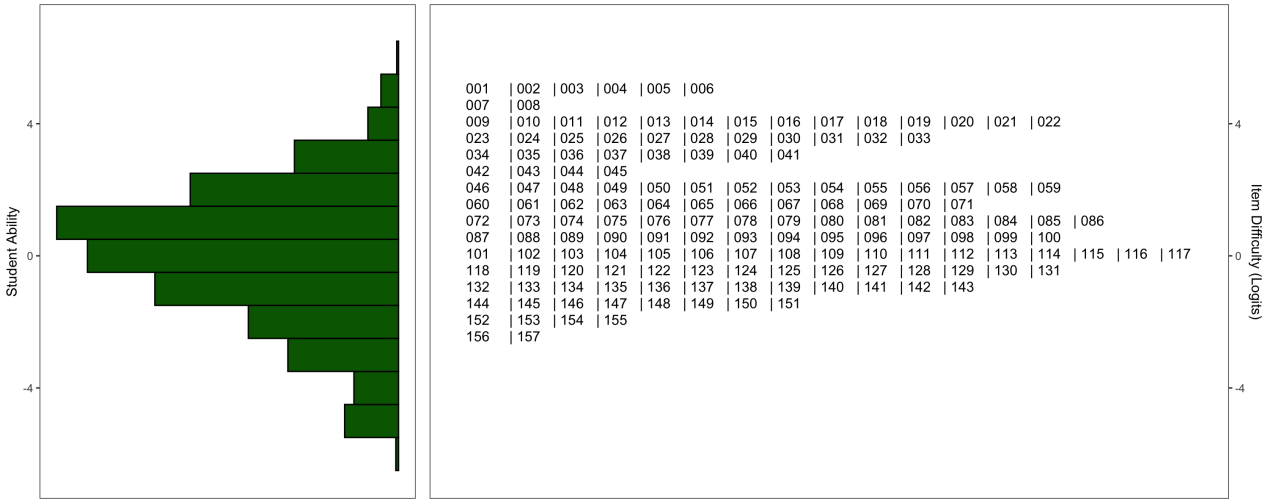

Figure 23.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Sentence Repetition Task

23.6.2.7 Model Summary

Table 23.5: Summary of Rasch Model Statistics for the English and Spanish Sentence Repetition Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 179	N = 4,187	N = 157	N = 3,487
Logit Scale Location	0.26 (2.43)	0.12 (-1.18, 1.26)	1.29 (1.93)	0.15 (-1.29, 1.37)
Outfit	0.99 (0.34)	0.63 (0.38, 0.93)	0.91 (0.63)	0.58 (0.34, 0.90)
Infit	0.98 (0.11)	0.74 (0.49, 1.00)	0.99 (0.14)	0.71 (0.47, 0.99)
Reliability of Separation	0.8898	0.8497	0.9046	0.8590

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 179 and 157 for the English and Spanish task, respectively.

23.7 Criterion Validity Evidence

23.7.1 Sample

Table 23.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Sentence Repetition Tasks

	English			Spanish		
Characteristic	**K** N = 262	**G1** N = 230	**G2** N = 203	**K** N = 238	**G1** N = 216	**G2** N = 265
Timepoint						
Winter 2024	262 (100%)	230 (100%)	203 (100%)	238 (100%)	216 (100%)	265 (100%)
Race						
American/Alaskan Native	5 (1.9%)	3 (1.3%)	1 (0.5%)	2 (0.8%)	4 (1.9%)	4 (1.5%)
Asian	36 (14%)	37 (16%)	8 (4.3%)	7 (3.0%)	2 (0.9%)	0 (0%)
Black/African American	29 (11%)	30 (13%)	34 (18%)	1 (0.4%)	0 (0%)	0 (0%)
Not reported	29 (11%)	32 (14%)	13 (7.0%)	132 (56%)	146 (69%)	172 (65%)
Other	73 (28%)	44 (19%)	3 (1.6%)	41 (17%)	8 (3.8%)	18 (6.8%)
White	87 (34%)	84 (37%)	126 (68%)	53 (22%)	53 (25%)	69 (26%)
Unknown	3	0	18	2	3	2
Ethnicity						
Hispanic/Latin(o/a)	106 (40%)	96 (42%)	120 (59%)	214 (91%)	200 (93%)	248 (98%)
Intentional nonreport	8 (3.1%)	2 (0.9%)	0 (0%)	1 (0.4%)	0 (0%)	2 (0.8%)
Not Hispanic/Latin(o/a)	148 (56%)	132 (57%)	82 (41%)	20 (8.5%)	16 (7.4%)	2 (0.8%)
Unknown	0	0	1	3	0	13
Gender						
Female	130 (50%)	105 (46%)	98 (48%)	126 (53%)	105 (49%)	137 (52%)
Male	132 (50%)	125 (54%)	105 (52%)	112 (47%)	111 (51%)	128 (48%)
Home Language						
English	192 (75%)	172 (75%)	127 (82%)	29 (12%)	23 (11%)	23 (8.7%)
Spanish	32 (12%)	25 (11%)	23 (15%)	204 (87%)	189 (89%)	239 (91%)
Other	33 (13%)	32 (14%)	5 (3.2%)	2 (0.9%)	1 (0.5%)	1 (0.4%)
Unknown	5	1	48	3	3	2
English Proficiency Label						
(Re-)Classified Proficient	11 (5.1%)	18 (8.0%)	11 (7.1%)	31 (14%)	24 (11%)	42 (17%)
English Learner	49 (23%)	40 (18%)	17 (11%)	182 (81%)	167 (80%)	178 (73%)
English-only	157 (72%)	167 (74%)	127 (82%)	11 (4.9%)	19 (9.0%)	23 (9.5%)
Unknown	45	5	48	14	6	22
Ever IEP/504	20 (9.9%)	22 (12%)	17 (11%)	20 (9.5%)	22 (11%)	16 (12%)
Unknown	59	50	48	27	12	127

English Sentence Repetition was correlated with the Sentence Repetition subtest from the Woodcock-Johnson IV (WJ IV OL) test (Schrank, McGrew, and Mather 2014). Spanish Sentence Repetition was correlated with the Recordando Oraciones subtest from the Clinical Evaluation of Language Fundamentals, 4th Edition, Spanish (CELF 4 Spanish) test (Semel et al. 2006b).

Table 23.7: Concurrent Criterion Validity Correlations for the English and Spanish Sentence Repetition Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
K	262	0.69 [0.62, 0.75]	49	0.54 [0.31, 0.71]	238	0.69 [0.62, 0.76]
G1	230	0.69 [0.61, 0.75]	40	0.55 [0.29, 0.74]	216	0.80 [0.75, 0.84]
G2	203	0.77 [0.71, 0.82]	NA	NA	265	0.69 [0.62, 0.75]

24 Word Reading

24.1 Task Description

Children are shown words and are asked to read them.

24.2 Construct

The Word Reading task measures the construct of decoding accuracy, the ability to translate print into speech by correctly pairing graphemes (letters) with their corresponding phonemes (sounds) using pronounceable real words.

24.3 Item Development

24.3.1 English

For the development of the item pool, multiple curricula to create a list of frequent, decodable words, including curricula used in the United States, like McGraw-Hill's "Wonders", Benchmark's "Benchmark Advance", and HMH's "Journey" were reviewed.

From this pool of items, we selected a sample of words, whose semantic meaning was overall easily accessed by the target population, with a variety of word types (e.g., nouns, verbs, adjectives, adverbs, etc.) and with varying orthographic and phonological length.

24.3.2 Spanish

For the development of the item pool, the research team reviewed multiple curricula to build up a list of frequent, decodable words, including curricula used in dual language programs in the United States, like the McGraw-Hill Maravillas, Estrellita, Houghton Mifflin Lectura, in addition to existing kindergarten and first-grade materials from Mexico, Panama, and Chile.

From the pool of items, we selected a sample of words, whose semantic meaning was overall easily accessed by the target population, with a variety of word types (e.g., nouns, verbs, adjectives, adverbs, etc.) and with varying orthographic and phonological length. Concepts represented by multiple words –that is, reflecting dialectic variability based on the cultural and/or geographic background of the respondent–were excluded to avoid benefiting certain cultural groups (e.g., "pig": puerco, cerdo, chancho; "shirt": polera, polo, remera; "avocado": aguacate, palta). In addition to the selected pool

of words, we included cognates to explore the possible interference or advancement of cognates in word reading for students in dual language programs.

24.3.3 Scoring

Dichotomous fixed response format of 0 points for incorrect responses or non-responses and 1 point for correct ones.

24.4 Calibration Samples

Table 24.1: Demographic Characteristics of Calibration Samples for the English and Spanish Word Reading Tasks

	English		Spanish	
Characteristic	**G1** N = 3,249	**G2** N = 3,251	**G1** N = 1,354	**G2** N = 1,060
Timepoint				
Fall 2023	605 (19%)	648 (20%)	0 (0%)	0 (0%)
Winter 2024	0 (0%)	0 (0%)	0 (0%)	432 (52%)
Fall 2024	2,644 (81%)	2,603 (80%)	627 (100%)	396 (48%)
Administration Format				
CAT	2,644 (81%)	2,603 (80%)	827 (61%)	628 (59%)
Forms	605 (19%)	648 (20%)	527 (39%)	432 (41%)
Race				
American/Alaskan Native	117 (3.8%)	63 (2.1%)	40 (4.9%)	15 (1.4%)
Asian	247 (8.1%)	231 (7.6%)	32 (3.9%)	25 (2.4%)
Black/African American	307 (10%)	343 (11%)	19 (2.3%)	14 (1.3%)
Not reported	393 (13%)	366 (12%)	231 (28%)	443 (42%)
Other	441 (14%)	376 (12%)	115 (14%)	68 (6.5%)
White	1,543 (51%)	1,653 (55%)	379 (46%)	479 (46%)
Unknown	201	219	538	16
Ethnicity				
Hispanic/Latin(o/a)	2,134 (70%)	2,105 (70%)	764 (96%)	951 (92%)
Intentional nonreport	13 (0.4%)	5 (0.2%)	2 (0.3%)	4 (0.4%)
Not Hispanic/Latin(o/a)	898 (29%)	892 (30%)	34 (4.3%)	75 (7.3%)
Unknown	204	249	554	30
Gender				
Female	1,512 (50%)	1,479 (49%)	428 (55%)	549 (53%)
Male	1,507 (50%)	1,509 (51%)	345 (45%)	489 (47%)
Non-binary	0 (0%)	0 (0%)	0 (0%)	1 (<0.1%)
Unknown	230	263	581	21
Home Language				
English	1,825 (64%)	1,821 (64%)	101 (12%)	148 (15%)
Spanish	901 (32%)	884 (31%)	709 (87%)	855 (85%)
Other	107 (3.8%)	144 (5.1%)	2 (0.2%)	7 (0.7%)
Unknown	416	402	542	50
English Proficiency Label				
(Re-)Classified Proficient	159 (5.8%)	294 (10%)	84 (11%)	129 (13%)
English Learner	835 (30%)	728 (26%)	599 (78%)	708 (73%)
English-only	1,763 (64%)	1,800 (64%)	88 (11%)	138 (14%)
Unknown	492	429	583	85
Ever IEP/504	229 (9.5%)	241 (10%)	63 (8.9%)	81 (10%)
Unknown	829	918	649	255
Unknown			727	232

24.5 Psychometric Analysis

24.5.1 Basic Item Statistics

We excluded 0 items from the English task and 0 items from the Spanish task based on low response counts ($n < 90$). 0 items were excluded because they had no variance in the Spanish task, and 0 items in the English task. Additionally, we excluded 2 items from the English task and 0 items from the Spanish task based on low point-biserial correlations ($r < 0.2$). Table 24.2 summarizes the basic item characteristics, Figure 24.1 shows the relationship between point-biserial correlations and the proportion of correct responses for each item.

Table 24.2: Basic Item Statistics Before and After Application of Exclusion Criteria, for the English and Spanish Word Reading Tasks

	English		Spanish	
	Before Excl.	After Excl.	Before Excl.	After Excl.
Characteristic	**N = 187**	**N = 185**	**N = 209**	**N = 209**
No. of Responses	363 (325)	365 (326)	214 (146)	214 (146)
Proportion Correct	0.54 (0.23)	0.55 (0.22)	0.51 (0.16)	0.51 (0.16)
Point-biserial Correlation	0.64 (0.12)	0.65 (0.11)	0.69 (0.09)	0.69 (0.09)
Excluded (n < 90)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Excluded (pbis < .2)	2 (1.1%)	0 (0%)	0 (0%)	0 (0%)
Excluded (no variation)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

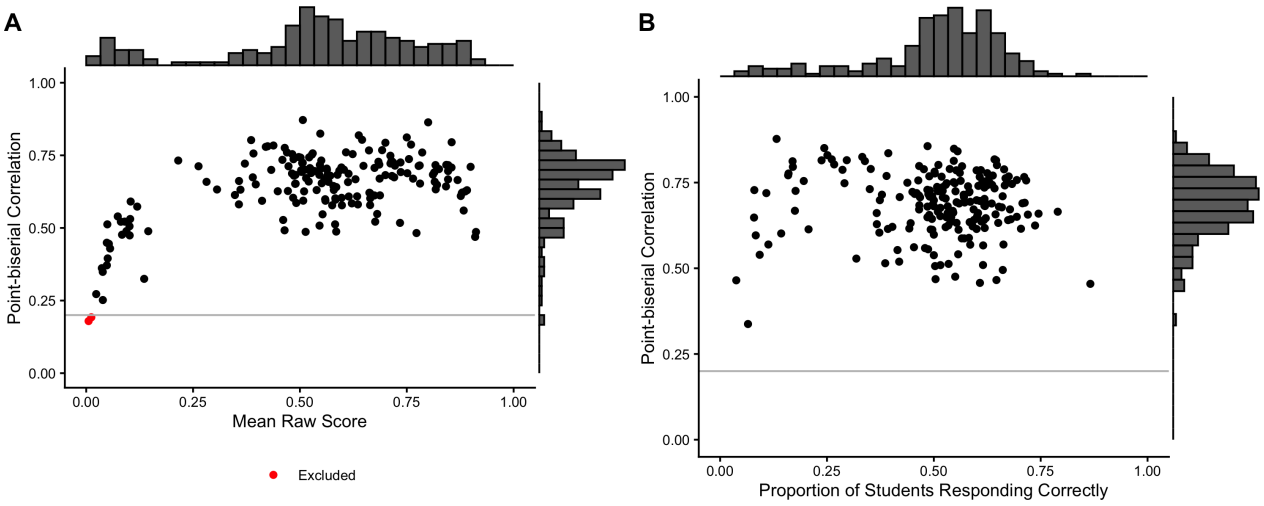

Figure 24.1: Scatterplot Showing Point-biserial (Item-total) Correlations and Proportion of Correct Responses for the English (Panel A) and Spanish (Panel B) Word Reading Tasks

24.5.2 Rasch Analysis

24.5.2.1 Item Location Estimates

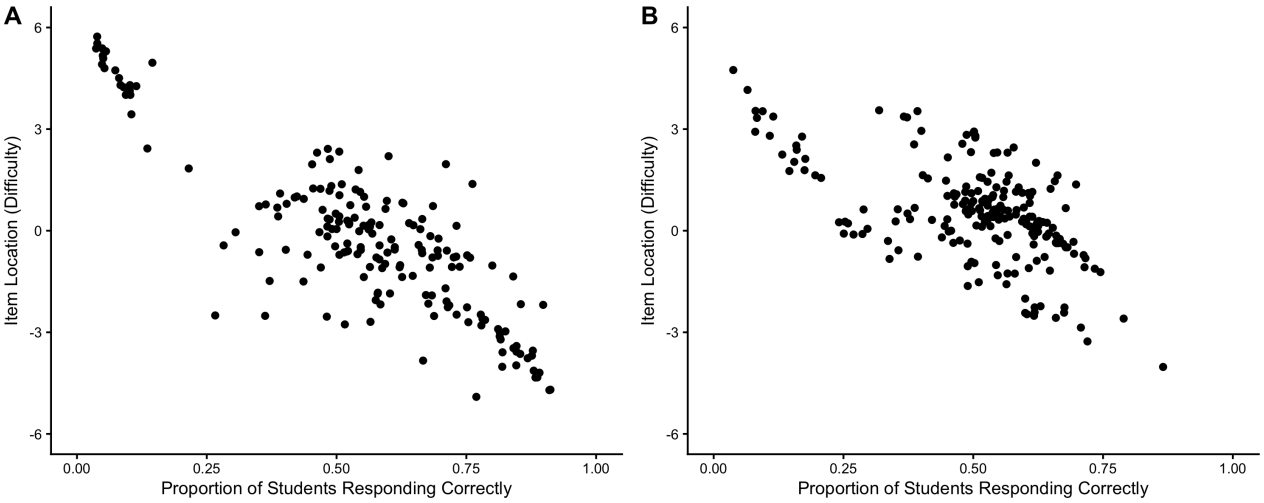

Figure 24.2: Scatterplot Showing Item Location and Proportion of Correct Response for the English (Panel A) and Spanish (Panel B) Word Reading Tasks

24.5.2.2 Item Fit Statistics

Table 24.3: Frequencies of Item Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Word Reading Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	145	0	0	0	145	173	0	0	0	173
B	27	0	0	0	27	24	0	0	0	24
C	8	0	0	0	8	7	0	0	0	7
D	3	0	2	0	5	3	0	2	0	5
Total	183	0	2	0	185	207	0	2	0	209

24.5.2.3 Person Location Estimates

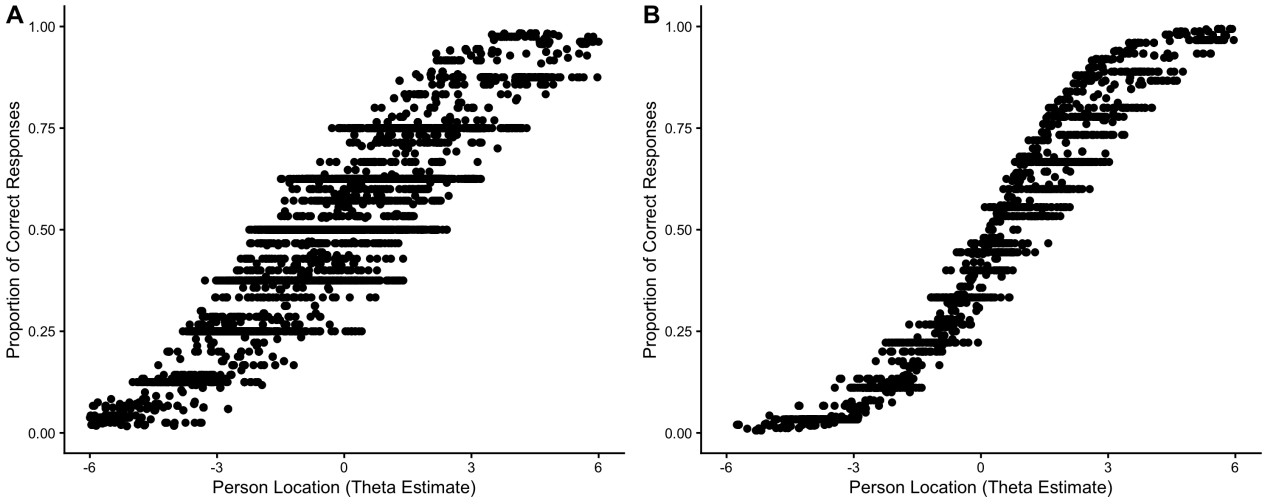

Figure 24.3: Scatterplot Showing Person Location Estimates (Obtained using the MLE method) and the Proportion of Correct Responses for English and Spanish Word Reading Tasks

24.5.2.4 Person Fit Statistics

Table 24.4: Frequencies of Person Misfit Categories Based on Infit/Outfit MSE Values for the English and Spanish Word Reading Tasks

	English					Spanish				
	Infit MSE									
	A	B	C	D	Total	A	B	C	D	Total
Outfit MSE										
A	2,718	0	36	2	2,756	1,363	0	2	2	1,367
B	1,581	1,794	3	0	3,378	172	745	0	0	917
C	90	0	30	7	127	26	0	3	0	29
D	80	0	44	18	142	15	0	11	0	26
Total	4,469	1,794	113	27	6,403	1,576	745	16	2	2,339

24.5.2.5 Distribution of Theta Estimates

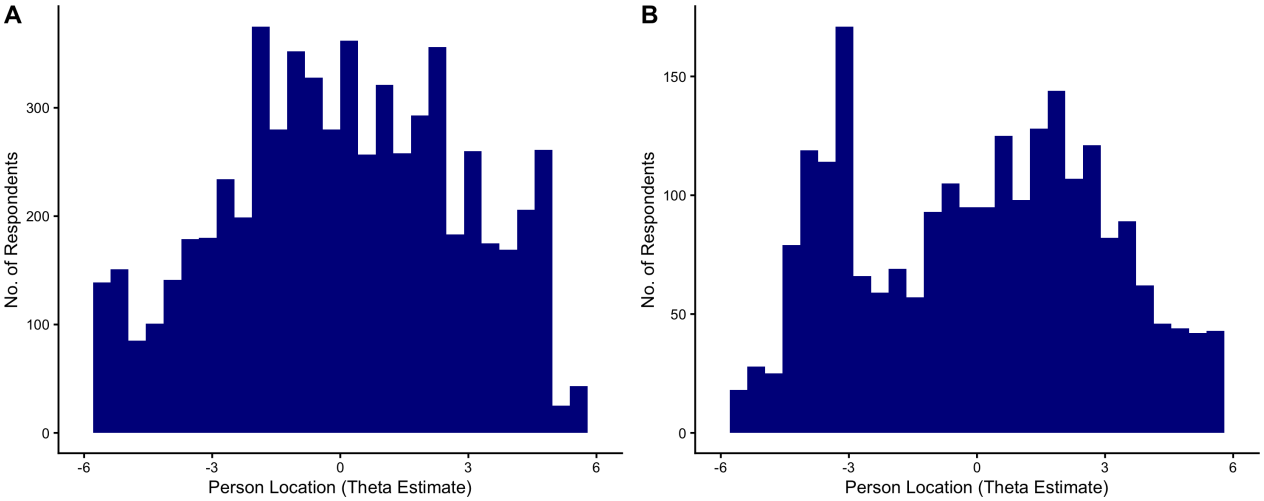

Figure 24.4: Distribution of Theta Estimates for the English and Spanish Word Reading Tasks

24.5.2.6 Wright Maps

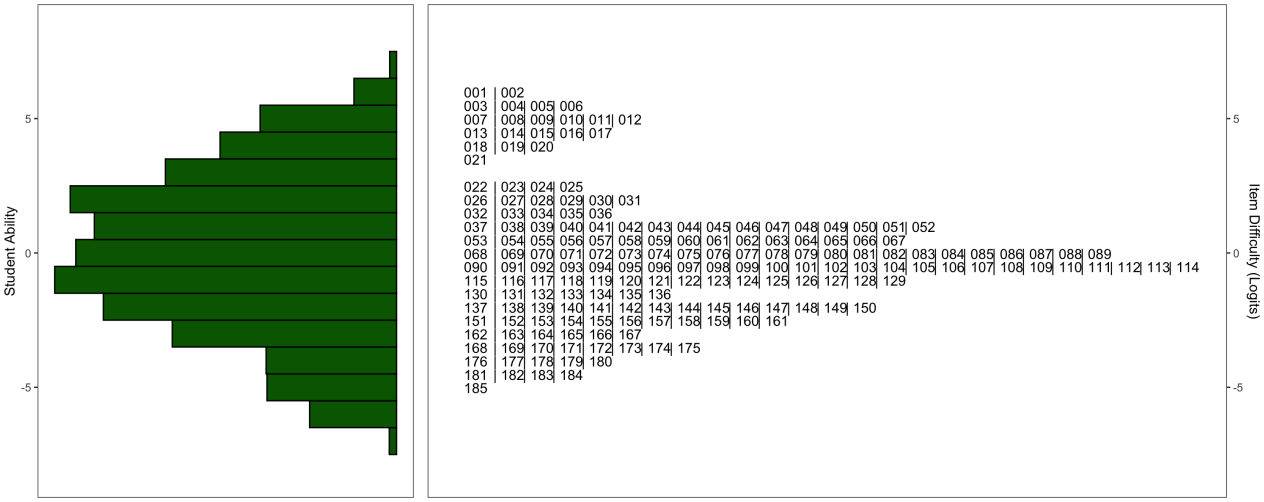

Figure 24.5: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the English Word Reading Task

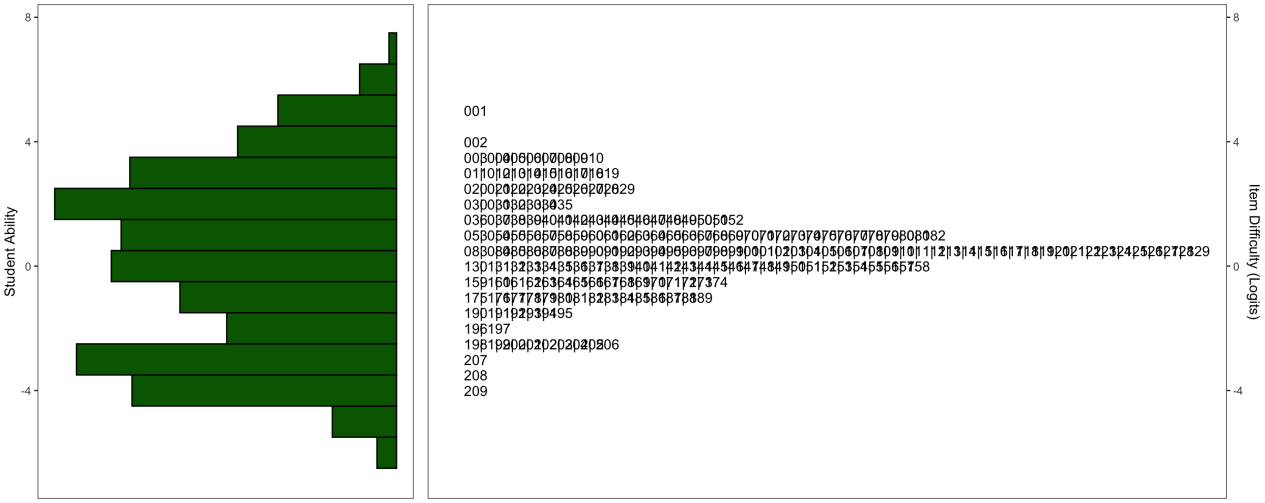

Figure 24.6: Wright Maps Showing the Relationship Between Item and Person Location Estimates for the Spanish Word Reading Task

24.5.2.7 Model Summary

Table 24.5: Summary of Rasch Model Statistics for the English and Spanish Word Reading Tasks

	English		Spanish	
	Item	Person	Item	Person
Characteristic	N = 185	N = 6,403	N = 209	N = 2,339
Logit Scale Location	-0.19 (2.41)	0.05 (-2.02, 2.22)	0.53 (1.46)	0.33 (-2.68, 2.36)
Outfit	0.87 (0.72)	0.48 (0.28, 0.70)	0.89 (0.48)	0.65 (0.05, 0.90)
Infit	0.92 (0.15)	0.70 (0.43, 0.90)	0.96 (0.16)	0.80 (0.19, 0.94)
Reliability of Separation	0.8997	0.8540	0.9130	0.8429

Final Number of Items

Following the exclusion of items with point-biserial correlations $< .20$ and items with poor fit statistics, the final versions of the task contain 185 and 209 for the English and Spanish task, respectively.

24.6 Criterion Validity Evidence

24.6.1 Sample

Table 24.6: Demographic Characteristics of the Concurrent Criterion Validity Evidence Samples for the English and Spanish Word Reading Tasks

	English		Spanish	
Characteristic	**G1** N = 221	**G2** N = 259	**G1** N = 191	**G2** N = 227
Timepoint				
Spring 2024	221 (100%)	259 (100%)	191 (100%)	227 (100%)
Race				
American/Alaskan Native	5 (2.3%)	1 (0.4%)	4 (2.1%)	4 (1.8%)
Asian	25 (11%)	34 (13%)	6 (3.2%)	3 (1.3%)
Black/African American	27 (12%)	32 (12%)	4 (2.1%)	4 (1.8%)
Not reported	55 (25%)	68 (26%)	73 (39%)	96 (43%)
Other	34 (15%)	26 (10%)	10 (5.3%)	14 (6.2%)
White	75 (34%)	98 (38%)	92 (49%)	104 (46%)
Ethnicity				
Hispanic/Latin(o/a)	102 (46%)	140 (54%)	172 (90%)	198 (87%)
Intentional nonreport	2 (0.9%)	0 (0%)	0 (0%)	2 (0.9%)
Not Hispanic/Latin(o/a)	117 (53%)	119 (46%)	19 (9.9%)	27 (12%)
Gender				
Female	97 (44%)	127 (49%)	100 (52%)	128 (56%)
Male	124 (56%)	132 (51%)	91 (48%)	99 (44%)
Home Language				
English	159 (73%)	177 (69%)	43 (23%)	57 (26%)
Spanish	37 (17%)	41 (16%)	144 (76%)	164 (74%)
Other	23 (11%)	40 (16%)	2 (1.1%)	1 (0.5%)
Unknown	2	1	2	5
English Proficiency Label				
(Re-)Classified Proficient	21 (9.7%)	23 (9.0%)	33 (17%)	37 (17%)
English Learner	47 (22%)	60 (23%)	120 (63%)	129 (59%)
English-only	148 (69%)	173 (68%)	36 (19%)	54 (25%)
Unknown	5	3	2	7
Ever IEP/504	22 (11%)	29 (14%)	16 (10%)	17 (9.3%)
Unknown	23	47	34	45
Unknown			2	2

English Word Reading was correlated with the Letter-Word Identification subtest from the Woodcock-Johnson IV (WJ IV ACH) test (Schrank, McGrew, and Mather 2014). Spanish Word Reading was correlated with the Identificación de Letras y Palabras subtest from the Batería IV Woodcock-Muñoz (Batería IV APROV) test (Woodcock et al. 2019).

Table 24.7: Concurrent Criterion Validity Correlations for the English and Spanish Word Reading Tasks

	English				Spanish	
	All		EL		All	
Grade	n	r [CI]	n	r [CI]	n	r [CI]
G1	220	0.90 [0.87, 0.92]	47	0.90 [0.83, 0.94]	191	0.85 [0.80, 0.88]
G2	259	0.77 [0.72, 0.82]	60	0.80 [0.69, 0.88]	227	0.80 [0.75, 0.84]

Part III

Universal Screening

25 Introduction

The Multitudes universal screener uses multiple measures to estimate a student's risk of reading difficulty by the end of the academic year. Performance on Multitudes does not diagnose disability, but instead indicates a need for more targeted instruction and possibly further evaluation. Administration in the first half of the year, following initial instruction of at least eight weeks for kindergarteners and four to six weeks for first and second graders, supports accurate identification while leaving opportunity for instruction to address identified needs.

Screening is only effective if it leads to the accurate identification of students in need of intervention. If the screening threshold is too low, students may be missed, and if the threshold is too high, students or groups of students may be over-identified. Therefore, in evaluating the performance of the screener, decisions are guided by maximizing classification accuracy. In setting these parameters for screening it is important to maximize sensitivity to reduce the number of children that may be missed by the screening. However, specificity is also important to reduce the number of false positives, or the number of children that are inaccurately identified by the screening. Ultimately, screening prioritizes sensitivity over specificity to support early identification and increase the likelihood that children will receive the support they need before reading problems become entrenched.

This chapter heavily draws on and, in part, reproduces analyses discussed in more detail in (Siebert et al. 2025).

26 Definition of 'Support Needed'

We defined the *support needed* category by means of their performance on language-appropriate standardized reading outcome measures from the *Woodcock-Johnson Tests of Achievement IV* (WJ; Schrank, McGrew, and Mather 2014) and the parallel *Batería IV Woodcock-Muñoz* (WM; Woodcock et al. 2019). Specifically, we used the *Basic Reading/Destrezas Básicas en Lectura* cluster scores in kindergarten and the *Broad Reading/Lectura Amplia* cluster scores in grades 1 and 2. The *Basic Reading* cluster taps into foundational reading skills and comprises a weighted average of the *Letter-Word Identification/Identificación de Letras y Palabras* and *Word Attack/Análisis de Palabras* tests. The *Broad Reading* cluster combines the *Letter-Word Identification/Identificación de Letras y Palabras*, *Passage Comprehension/Comprensión de Textos*, and *Sentence Reading Fluency/Fluidez en Lectura de Frases* tests to provide an overall measure of reading achievement, thereby assessing decoding, fluency, and comprehension skills.

Importantly, we classified students as needing support when scoring at or below the 20th percentile *within* our large and diverse sample. Our choice to use a within-sample percentile cut-off was motivated by our sample's relatively low performance in relation to the published national norms for the U.S (Figure 26.1). Several factors may contribute to the observed lower performance of our sample compared to WJ/WM norms for the U.S., including higher child poverty rates (Reyes-Vallarde, 2023), higher rates of multilingualism (CDE, 2023b), variability in effectiveness of reading instruction in public schools (Langreo, 2024), and possible bias in regional norms.

We chose to use the *20th* percentile cutoff so that our screener could identify students before they fall behind so far that they become eligible for special educational services, often at or below the 7th percentile (National Center for Education Statistics 2024; Abu-Hamour et al. 2012).

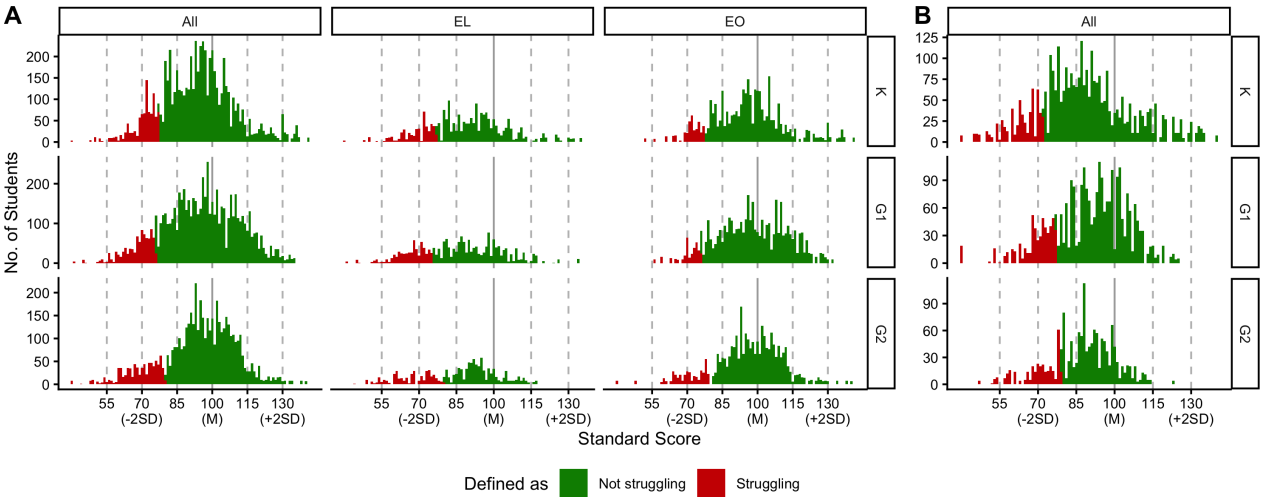

Figure 26.1: Distribution of Woodcock-Johnson (Panel A) and Woodcock-Muñoz (Panel B) Basic Reading Skills Cluster (for Kindergarten) Broad Reading Cluster (for Grades 1 and 2) by Grade, with Struggling Readers ($\leq$ 20th Percentile in Sample) Highlighted in Red.

27 Sample

The overall sample (Table 27.1) comprises 3130 students from 39 schools in 18 rural and urban school districts across California. We sampled purposefully such that (i) our sample is broadly representative of the public school population in California and (ii) we have sufficiently large group sizes for subgroup analyses.

Table 27.1: Demographic Characteristics of the Sample (N = 3,130) by Grade and Screening Language

	English			Spanish		
Characteristic	**K** N = 889	**G1** N = 800	**G2** N = 818	**K** N = 392	**G1** N = 297	**G2** N = 255
Academic Year						
AY 23/24	650 (73%)	515 (64%)	663 (81%)	278 (71%)	174 (59%)	174 (68%)
AY 24/25	239 (27%)	285 (36%)	155 (19%)	114 (29%)	123 (41%)	81 (32%)
Gender						
Female	425 (49%)	393 (50%)	406 (50%)	181 (49%)	148 (53%)	140 (55%)
Male	445 (51%)	390 (50%)	410 (50%)	189 (51%)	131 (47%)	115 (45%)
Unknown	19	17	2	22	18	0
English Proficiency Label						
English Learner	358 (44%)	323 (41%)	236 (30%)	321 (84%)	236 (80%)	186 (75%)
(Re)classified Proficient	56 (6.9%)	80 (10%)	91 (12%)	48 (13%)	46 (16%)	38 (15%)
English-only	395 (49%)	379 (48%)	451 (58%)	15 (3.9%)	13 (4.4%)	25 (10%)
Unknown	80	18	40	8	2	6
Home Language						
English	463 (53%)	396 (50%)	451 (58%)	33 (8.5%)	18 (6.1%)	29 (12%)
Spanish	342 (39%)	334 (42%)	251 (32%)	350 (90%)	276 (94%)	217 (88%)
Other	74 (8.4%)	64 (8.1%)	75 (9.7%)	4 (1.0%)	1 (0.3%)	2 (0.8%)
Unknown	10	6	41	5	2	7
Ever IEP/504	65 (8.5%)	55 (7.9%)	64 (10%)	31 (9.0%)	21 (8.8%)	13 (6.0%)
Unknown	128	107	191	46	57	37
Struggling Readers (our definition)	166 (19%)	165 (21%)	148 (18%)	78 (20%)	71 (24%)	63 (25%)
Struggling Readers (normative)	357 (40%)	318 (40%)	241 (29%)	212 (54%)	125 (42%)	119 (47%)

28 Approach to Screening Model Building and Evaluation

28.1 Logistic Regression with LOGO Cross-validation

We opted for logistic regression models due to their interpretability, ease of implementation, and computational efficiency. We built a different prediction models for each grade in each language. All models have the following general form:

$$Pr(Y_i = 1) = \sigma(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_n X_{in})$$

where $P(Y_i = 1)$ denotes the probability of student i experiencing reading difficulty, the βs denote the coefficients (weights) of student i's scores on the n different grade-specific screening tasks (i.e., X_{1i}, X_{2i}, ..., X_{ni}), and $\sigma()$ is the standard logistic sigmoid function $f(x) = (1 + \exp(-x))^{-1}$). All models were implemented using the *glm* function from the *stats* package for *R*.

28.2 Selection of Predictor Tasks

We selected tasks for inclusion in the prediction model (**?@tbl-screener-tasks**)based on the following theoretical, empirical, and pragmatic criteria:

- We chose to include tasks from different domains to increase coverage of the broad range of skills that contribute to reading in English and Spanish (i.e., language, phonological awareness, processing speed.
- We selected tasks that fit the model and were found to have strong correlations with the outcome measures for both English-only and multilingual students.
- We included measures in English and Spanish that are known to predict future reading difficulty and dyslexia in each language.
- We minimized the time of administration of the overall battery.

Table 28.1: Tasks in the Multitudes Universal Screener Step 1 (both English and Spanish)

Grade	Tasks
Kindergarten	- Letter Naming Fluency
	- Elision Receptive
	- Rapid Automatized Naming - Objects
Grade 1	- Word Reading
	- Letter Sound Fluency
	- Expressive Vocabulary (Spanish Screening Only)
	- Rapid Automatized Naming - Objects
Grade 2	- Word Reading
	- Expressive Vocabulary
	- Rapid Automatized Naming - Letters

28.3 Evaluating Screener Performance

To evaluate each model's performance and classification accuracy, we computed–among other metrics–sensitivity and specificity (Table 29.1 to Table 29.4). We targeted commonly used thresholds of $> .80$ and specificity $> .70$ (Johnson et al. 2009). We also provide receiver-operating characteristic curves and precision-recall curves for all models (Figure 29.1 and Figure 29.2). To assess the linguistic fairness of our English prediction model, we evaluated predictions not only for the entire sample, but also separately by English proficiency designation.

28.4 Other Notes

It is important to note that the English and Spanish universal screening batteries include the same measures across all grades. This decision, based on data and implementation considerations, was made to ensure accuracy, fairness, and efficiency. Different measures for the two languages would have required different testing times, additional teacher training, and made performance comparisons more difficult for children who need to be screened in both languages. The four assessments included for each grade and each language require less than 10 minutes of testing.

29 Final Screening Models

29.1 English Screener

Table 29.1: English-to-English Prediction Models Evaluated for the Entire Sample.

	All					
Grade	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	877	0.7283	0.7486	0.7811	0.7162	0.186	0.376
G1	791	0.7983	0.8316	0.8895	0.7736	0.213	0.368
G2	807	0.8929	0.9009	0.9133	0.8884	0.182	0.258

Note. green: >= .8; black: >= .7; yellow: >= .6; red otherwise

Table 29.2: English-to-English Prediction Models Evaluated for the Students Classified as English Learners Only.

	All					
Grade	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	376	0.6615	0.723	0.8544	0.5915	0.266	0.527
G1	325	0.7803	0.8002	0.8986	0.7019	0.399	0.538
G2	244	0.8381	0.864	0.9375	0.7904	0.324	0.445

Note. green: >= .8; black: >= .7; yellow: >= .6; red otherwise

Table 29.3: English-to-English Prediction Models Evaluated for English-only Students Only

	All					
Grade	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	387	0.7804	0.7321	0.6667	0.7976	0.132	0.264
G1	388	0.7938	0.8049	0.8182	0.7915	0.085	0.260
G2	445	0.9124	0.9221	0.9355	0.9086	0.139	0.209

Note. green: >= .8; black: >= .7; yellow: >= .6; red otherwise

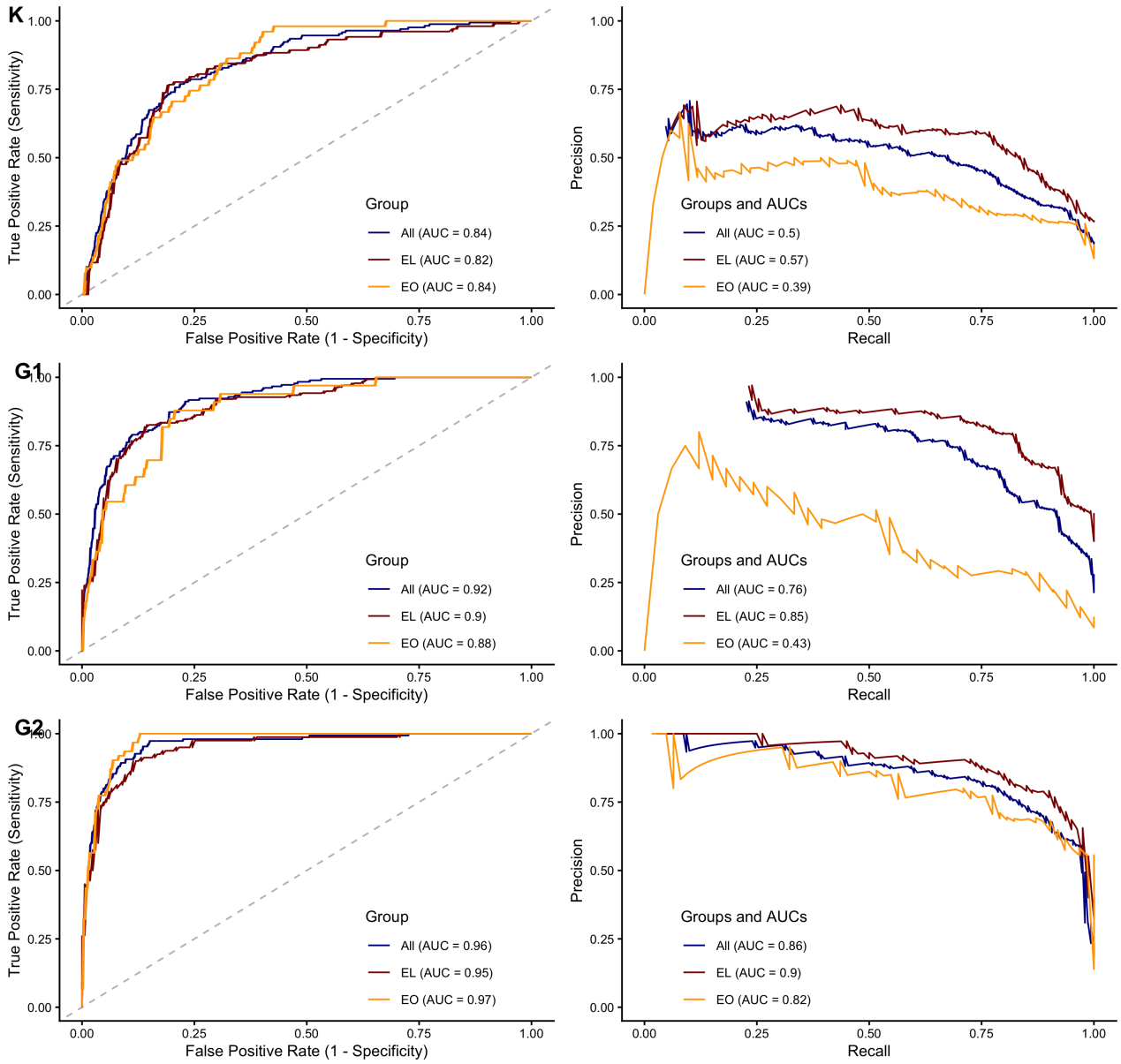

Figure 29.1: Receiver-operating Characteristic Curves (Left Panels) and Precision-Recall Curves (Right Panels) for the English Prediction Models

29.2 Spanish Screener

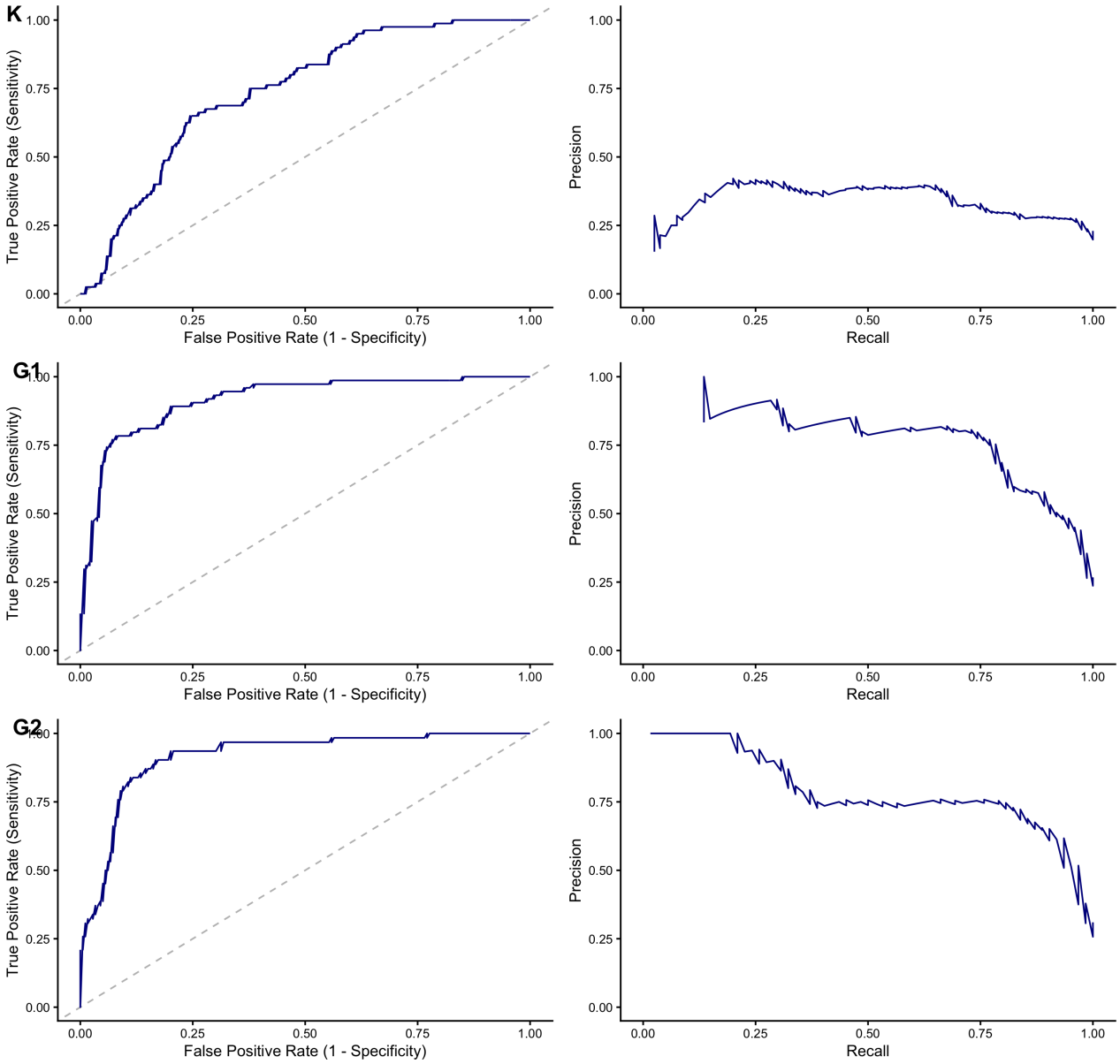

Figure 29.2: Receiver-operating Characteristic Curves (Left Panels) and Precision-Recall Curves (Right Panels) for the Spanish Prediction Models

Table 29.4: Spanish-to-Spanish Prediction Models.

	All						
Grade	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	392	0.5644	0.6248	0.725	0.5247	0.198	0.525
G1	297	0.8754	0.8391	0.7703	0.9079	0.236	0.252
G2	239	0.8672	0.879	0.9032	0.8547	0.257	0.340

Note. green: >= .8; black: >= .7; yellow: >= .6; red otherwise

References

Abu-Hamour, B., H. Al Hmouz, J. Mattar, and M. Muhaidat. 2012. "The Use of Woodcock Johnson Tests for Identifying Students with Special Needs—a Comprehensive Literature Review." *Procedia - Social and Behavioral Sciences* 47: 665–73. https://doi.org/10.1016/j.sbspro.2012.06.714.

Adams, Marilyn J. 1990. *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA: MIT Press.

Adlof, Suzanne M., Hugh W. Catts, and Jaehoon Lee. 2010. "Kindergarten Predictors of Second Versus Eighth Grade Reading Comprehension Impairments." *Journal of Learning Disabilities* 43 (4): 332–45.

Adlof, Suzanne M., and Hannah Patten. 2017. "Nonword Repetition and Vocabulary Knowledge as Predictors of Children's Phonological and Semantic Word Learning." *Journal of Speech, Language, and Hearing Research* 60 (3): 682–93. https://doi.org/10.1044/2016_jslhr-l-15-0441.

Alonso, M. Á., E. Díez, and Á. Fernández. 2016. "Subjective Age-of-Acquisition Norms for 4,640 Verbs in Spanish." *Behavior Research Methods* 48 (4): 1337–42. https://doi.org/10.3758/s13428-015-0675-z.

Alonso, M. Á., Á. Fernández, and E. Díez. 2015. "Subjective Age-of-Acquisition Norms for 7,039 Spanish Words." *Behavior Research Methods* 47 (1): 268–74. https://doi.org/10.3758/s13428-014-0454-2.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Andrich, David. 1988. "The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude." *Applied Psychological Measurement* 12 (1): 33–51. https://doi.org/10.1177/014662168801200105.

Anthony, Jason L., Jeffrey M. Williams, Rachel G. Aghara, Martha Dunkelberger, Barbara Novak, and Anuja Divatia Mukherjee. 2009. "Assessment of Individual Differences in Phonological Representation." *Reading and Writing* 23 (8): 969–94. https://doi.org/10.1007/s11145-009-9185-7.

Anthony, Jason L., Jeffrey M. Williams, Lillian K. Durán, Sandra Laing Gillam, Lan Liang, Rachel Aghara, Paul R. Swank, Mike A. Assel, and Susan H. Landry. 2011. "Spanish Phonological Awareness: Dimensionality and Sequence of Development During the Preschool and Kindergarten Years." *Journal of Educational Psychology* 103 (4): 857–76. https://doi.org/10.1037/a0025024.

Anthony, Jason L., Jeffrey M. Williams, Renee McDonald, Deborah Corbitt-Shindler, Coleen D. Carlson, and David J. Francis. 2006. "Phonological Processing and Emergent Literacy in Spanish-Speaking Preschool Children." *Annals of Dyslexia* 56 (2): 239–70. https://doi.org/10.1007/s11881-006-0011-5.

Araújo, Susana, and Luís Faísca. 2019. "A Meta-Analytic Review of Naming-Speed Deficits in Developmental Dyslexia." *Scientific Studies of Reading* 23 (5): 349–68. https://doi.org/10.1080/10888438.2019.1572758.

Baddeley, Alan. 2003. "Working Memory and Language: An Overview." *Journal of Communication Disorders* 36 (3): 189–208. https://doi.org/10.1016/s0021-9924(03)00019-4.

Baddeley, Alan D., and Graham Hitch. 1974. "Working Memory." In, 47–89. Elsevier. https://doi.org/10.1016/s0079-7421(08)60452-1.

Baker, Diana L., Young-Suk Park, and Scott K. Baker. 2010. "Effect of Initial Status and Growth in Pseudoword Reading on Spanish Reading Comprehension at the End of First Grade." *Psicothema* 22 (4): 955–62.

Baker, Doris Luft, Yonghan Park, and Scott K. Baker. 2010. "The Reading Performance of English Learners in Grades 1–3: The Role of Initial Status and Growth on Reading Fluency in Spanish and English." *Reading and Writing* 25 (1): 251–81. https://doi.org/10.1007/s11145-010-9261-z.

Biemiller, Andrew, and Naomi Slonim. 2001. "Estimating Root Word Vocabulary Growth in Normative and Advantaged Populations: Evidence for a Common Sequence of Vocabulary Acquisition." *Journal of Educational Psychology* 93 (3): 498–520. https://doi.org/10.1037/0022-0663.93.3.498.

Bosse, M.-L., M.-J. Tainturier, and S. Valdois. 2007. "Developmental Dyslexia: The Visual Attention Span Deficit Hypothesis." *Cognition* 104 (2): 198–230. https://doi.org/10.1016/j.cognition.2006.05.009.

Bosse, M.-L., and S. Valdois. 2009. "Influence of the Visual Attention Span on Child Reading Performance: A Cross-Sectional Study." *Journal of Research in Reading* 32 (2): 230–53. https://doi.org/10.1111/j.1467-9817.2008.01387.x.

Brady, Susan. 1986. "Short-Term Memory, Phonological Processing, and Reading Ability." *Annals of Dyslexia* 36 (1): 138–53. https://doi.org/10.1007/bf02648026.

Brady, Susan, Donald Shankweiler, and Virginia Mann. 1983. "Speech Perception and Memory Coding in Relation to Reading Ability." *Journal of Experimental Child Psychology* 35 (2): 345–67. https://doi.org/10.1016/0022-0965(83)90087-5.

Brysbaert, M., and A. Biemiller. 2017. "Test-Based Age-of-Acquisition Norms for 44 Thousand English Word Meanings." *Behavior Research Methods* 49: 1520–23. https://doi.org/10.3758/s13428-016-0811-4.

Caravolas, M., and Anna Samara. 2015. "Learning to Read and Spell Words in Different Writing Systems." In.

Catts, Hugh W., Marc E. Fey, Xuyang Zhang, and J. Bruce Tomblin. 2001. "Estimating the Risk of Future Reading Difficulties in Kindergarten Children." *Language, Speech, and Hearing Services in Schools* 32 (1): 38–50. https://doi.org/10.1044/0161-1461(2001/004).

Catts, Hugh W., Tiffany P. Hogan, and Suzanne M. Adlof. 2005. "Developmental Changes in Reading and Reading Disabilities." In *The Connections Between Language and Reading Disabilities*, edited by Hugh W. Catts and Alan G. Kamhi, 25–40. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Chen, Yi-Jui Iva, Christopher G. Thompson, Zhihong Xu, Robin C. Irey, and George K. Georgiou. 2021. "Rapid Automatized Naming and Spelling Performance in Alphabetic Languages: A Meta-Analysis." *Reading and Writing* 34 (10): 2559–80. https://doi.org/10.1007/s11145-021-10160-7.

Chua, Shi Min, Susan J. Rickard Liow, and Stephanie H. M. Yeong. 2014. "Using Spelling to Screen Bilingual Kindergarteners At Risk for Reading Difficulties." *Journal of Learning Disabilities* 49 (3): 227–39. https://doi.org/10.1177/0022219414538519.

Clayton, Francina J., Gillian West, Claire Sears, Charles Hulme, and Arne Lervåg. 2019. "A Longitudinal Study of Early Reading Development: Letter-Sound Knowledge, Phoneme Awareness and RAN, but Not Letter-Sound Integration, Predict Variations in Reading Development."

Scientific Studies of Reading 24 (2): 91–107. https://doi.org/10.1080/10888438.2019.1622546.
Coady, Jeffry A., and Julia L. Evans. 2008. "Uses and Interpretations of Non-Word Repetition Tasks in Children with and Without Specific Language Impairments (SLI)." *International Journal of Language & Communication Disorders* 43 (1): 1–40. https://doi.org/10.1080/13682820601116485.
Crosson, Amy C., and Nonie K. Lesaux. 2009. "Revisiting Assumptions about the Relationship of Fluent Reading to Comprehension: Spanish-Speakers' Text-Reading Fluency in English." *Reading and Writing* 23 (5): 475–94. https://doi.org/10.1007/s11145-009-9168-8.
Cui, Zhongmin. 2020. "A Seed Usage Issue on Using catR for Simulation and the Solution." *Applied Psychological Measurement* 44 (5): 409–12. https://doi.org/10.1177/0146621620920934.
Cummine, Jacqueline, Eszter Szepesvari, Brea Chouinard, Wahab Hanif, and George K. Georgiou. 2014. "A Functional Investigation of RAN Letters, Digits, and Objects: How Similar Are They?" *Behavioural Brain Research* 275 (December): 157–65. https://doi.org/10.1016/j.bbr.2014.08.038.
Cunningham, Anna J., Adrian P. Burgess, Caroline Witton, Joel B. Talcott, and Laura R. Shapiro. 2020. "Dynamic Relationships Between Phonological Memory and Reading: A Five Year Longitudinal Study from Age 4 to 9." *Developmental Science* 24 (1). https://doi.org/10.1111/desc.12986.
Cutting, Laurie E., and Hollis S. Scarborough. 2006. "Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured." *Scientific Studies of Reading* 10 (3): 277–99. https://doi.org/10.1207/s1532799xssr1003_5.
De la Calle, Guzmán-Simón, A. M. 2018. "Letter Knowledge and Learning Sequence of Graphemes in Spanish: Precursors of Early Reading." *Revista de Psicodidáctica (English Ed.)* 23 (2): 128–36.
De Ramírez, Romilia Domínguez, and Edward S. Shapiro. 2007. "Cross-Language Relationship Between Spanish and English Oral Reading Fluency Among Spanish-Speaking English Language Learners in Bilingual Education Classrooms." *Psychology in the Schools* 44 (8): 795–806. https://doi.org/10.1002/pits.20266.
Deacon, S. Hélène, and Michael Kieffer. 2018. "Understanding How Syntactic Awareness Contributes to Reading Comprehension: Evidence from Mediation and Longitudinal Models." *Journal of Educational Psychology* 110 (1): 72–86. https://doi.org/10.1037/edu0000198.
Defior, Sylvia, and Francisca Serrano. 2005. "The Initial Development of Spelling in Spanish: From Global to Analytical." *Reading and Writing* 18 (1): 81–98. https://doi.org/10.1007/s11145-004-5893-1.
Denckla, Martha Bridge, and Rita Rudel. 1974. "Rapid "Automatized" Naming of Pictured Objects, Colors, Letters and Numbers by Normal Children." *Cortex* 10 (2): 186–202. https://doi.org/10.1016/s0010-9452(74)80009-2.
Denckla, Martha Bridge, and Rita G. Rudel. 1976. "Rapid 'Automatized' Naming (R.A.N.): Dyslexia Differentiated from Other Learning Disabilities." *Neuropsychologia* 14 (4): 471–79. https://doi.org/10.1016/0028-3932(76)90075-0.
Dickinson, David K., Allyssa McCabe, Nancy ClarkChiarelli, and Anne Wolf. 2004. "Cross-Language Transfer of Phonological Awareness in Low-Income Spanish and English Bilingual Preschool Children." *Applied Psycholinguistics* 25 (3): 323–47. https://doi.org/10.1017/s0142716404001158.
Durgunoğlu, Aydin Y., William E. Nagy, and Barbara J. Hancin-Bhatt. 1993. "Cross-Language Transfer of Phonological Awareness." *Journal of Educational Psychology* 85 (3): 453–65. https://doi.org/10.1037/0022-0663.85.3.453.
Ehri, Linnea C. 2020. "The Science of Learning to Read Words: A Case for Systematic Phonics Instruction." *Reading Research Quarterly* 55 (S1). https://doi.org/10.1002/rrq.334.

Ellis, Nick. 1994. "Longitudinal Studies of Spelling Development." In *Handbook of Spelling: Theory, Process and Intervention*, edited by Gordon D. A. Brown and Nick C. Ellis, 155–77. Chichester, UK: John Wiley & Sons.

Ellis, Nick, and Suzanne Cataldo. 1990. "The Role of Spelling in Learning to Read." *Language and Education* 4 (1): 1–28. https://doi.org/10.1080/09500789009541270.

Engelhard, George Jr., and Jue Wang. 2020. *Rasch Models for Solving Measurement Problems: Invariant Measurement in the Social Sciences*. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: SAGE Publications, Inc.

Engelhard, George, Jue Wang, and Stefanie A. Wind. 2018. "A Tale of Two Models: Psychometric and Cognitive Perspectives on Rater-Mediated Assessments Using Accuracy Ratings." *Psychological Test and Assessment Modeling* 60 (1): 33–52.

Ferreiro, Emilia, and Ana Teberosky. 1982. *Literacy Before Schooling*. Heinemann Educational Books Inc.

Fiestas, Christine E., and Elizabeth D. Penã. 2004. "Narrative Discourse in Bilingual Children." *Language, Speech, and Hearing Services in Schools* 35 (2): 155–68. https://doi.org/10.1044/0161-1461(2004/016).

Foorman, Barbara. 2023. "Learning the Code." In *Handbook on the Science of Early Literacy*, edited by Sonia Q. Cabell, Susan B. Neuman, and Nicole P. Terry, 73–82. New York, NY: Guilford Press.

Foorman, Barbara R., Sarah Herrera, Yaacov Petscher, Alison Mitchell, and Adrea Truckenmiller. 2015. "The Structure of Oral Language and Reading and Their Relation to Comprehension in Kindergarten Through Grade 2." *Reading and Writing* 28 (5): 655–81. https://doi.org/10.1007/s11145-015-9544-5.

Fuchs, Lynn S., Douglas Fuchs, Michelle K. Hosp, and Joseph R. Jenkins. 2001. "Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis." *Scientific Studies of Reading* 5 (3): 239–56. https://doi.org/10.1207/s1532799xssr0503_3.

Garcia, Gilbert E., Gail McKoon, Diane August, and Timothy Shanahan. 2006. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

García, J. Ricardo, and Kate Cain. 2014. "Decoding and Reading Comprehension." *Review of Educational Research* 84 (1): 74–111. https://doi.org/10.3102/0034654313499616.

Gathercole, Susan E. 2006. "Nonword Repetition and Word Learning: The Nature of the Relationship." *Applied Psycholinguistics* 27 (4): 513–43. https://doi.org/10.1017/s0142716406060383.

Gathercole, Susan E., Catherine S. Willis, Alan D. Baddeley, and Hazel Emslie. 1994. "The Children's Test of Nonword Repetition: A Test of Phonological Working Memory." *Memory* 2 (2): 103–27. https://doi.org/10.1080/09658219408258940.

Genesee, Fred, Esther Geva, Cheryl Dressler, and Michael Kamil. 2006. "Synthesis: Cross-Linguistic Relationships." In *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, 153–74. Mahwah, NJ: Lawrence Erlbaum Associates.

Gentry, J. Richard. 1982. "An Analysis of Developmental Spelling in "GNYS AT WRK"." *The Reading Teacher* 36 (2): 192–200. http://www.jstor.org/stable/20198182.

———. 2000. "A Retrospective on Invented Spelling and a Look Forward." *The Reading Teacher* 54 (3): 318–32. http://www.jstor.org/stable/20204910.

Georgiou, George K., and Rauno Parrila. 2020. "What Mechanism Underlies the Rapid Automatized

Naming–reading Relation?" *Journal of Experimental Child Psychology* 194 (June): 104840. https://doi.org/10.1016/j.jecp.2020.104840.

Georgiou, George K., Rauno Parrila, Ying Cui, and Timothy C. Papadopoulos. 2013. "Why Is Rapid Automatized Naming Related to Reading?" *Journal of Experimental Child Psychology* 115 (1): 218–25. https://doi.org/10.1016/j.jecp.2012.10.015.

Gershon, Richard C. 2005. "Computer Adaptive Testing." *Journal of Applied Measurement* 6 (1): 109–27.

Geva, Esther, and Linda S. Siegel. 2000. *Reading and Writing* 12 (1/2): 1–30. https://doi.org/10.1023/a:1008017710115.

Goodwin, Amanda P., Diane August, and Margarita Calderon. 2015. "Reading in Multiple Orthographies: Differences and Similarities in Reading in Spanish and English for English Learners." *Language Learning* 65 (3): 596–630. https://doi.org/10.1111/lang.12127.

Gottardo, Alexandra. 2002. "The Relationship Between Language and Reading Skills in Bilingual Spanish-English Speakers." *Topics in Language Disorders* 22 (5): 46–70. https://doi.org/10.1097/00011363-200211000-00008.

Gottardo, Alexandra, Xi Chen, and Michelle Ru Yun Huo. 2021. "Understanding Within- and Cross-Language Relations Among Language, Preliteracy Skills, and Word Reading in Bilingual Learners: Evidence From the Science of Reading." *Reading Research Quarterly* 56 (S1). https://doi.org/10.1002/rrq.410.

Gottardo, Alexandra, and Julie Mueller. 2009. "Are First- and Second-Language Factors Related in Predicting Second-Language Reading Comprehension? A Study of Spanish-Speaking Children Acquiring English as a Second Language from First to Second Grade." *Journal of Educational Psychology* 101 (2): 330–44. https://doi.org/10.1037/a0014320.

Gutierrez-Clellen, Vera F., Elizabeth Peña, and Rosemary Quinn. 1995. "Accommodating Cultural Differences in Narrative Style." *Topics in Language Disorders* 15 (4): 54–67. https://doi.org/10.1097/00011363-199508000-00006.

Heikkilä, Riikka, Vesa Närhi, Mikko Aro, and Timo Ahonen. 2009. "Rapid Automatized Naming and Learning Disabilities: Does RAN Have a Specific Connection to Reading or Not?" *Child Neuropsychology* 15 (4): 343–58. https://doi.org/10.1080/09297040802537653.

Heilmann, John, Jon F. Miller, Ann Nockerts, and Claudia Dunaway. 2010. "Properties of the Narrative Scoring Scheme Using Narrative Retells in Young School-Age Children." *American Journal of Speech-Language Pathology* 19 (2): 154–66. https://doi.org/10.1044/1058-0360(2009/08-0024).

Hogan, Tiffany P., Suzanne M. Adlof, and Crystle N. Alonzo. 2014. "On the Importance of Listening Comprehension." *International Journal of Speech-Language Pathology* 16 (3): 199–207. https://doi.org/10.3109/17549507.2014.904441.

Hogan, Tiffany P., Hugh W. Catts, and Todd D. Little. 2005. "The Relationship Between Phonological Awareness and Reading." *Language, Speech, and Hearing Services in Schools* 36 (4): 285–93. https://doi.org/10.1044/0161-1461(2005/029).

Hoover, Wesley A., and Philip B. Gough. 1990. "The Simple View of Reading." *Reading and Writing* 2 (2): 127–60. https://doi.org/10.1007/bf00401799.

Hulme, C. 1988. "The Implausibility of Low-Level Visual Deficits as a Cause of Children's Reading Difficulties." *Cognitive Neuropsychology* 5 (3): 369–74.

Ives Wiley, Hilda, and Stanley L. Deno. 2005. "Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment." *Remedial and Special Education* 26 (4): 207–14. https://doi.org/10.1177/07419325050260040301.

Jasso, Javier, Stephanie McMillen, Jissel B. Anaya, Lisa M. Bedore, and Elizabeth D. Peña. 2020. "The Utility of an English Semantics Measure for Identifying Developmental Language Disorder in Spanish–English Bilinguals." *American Journal of Speech-Language Pathology* 29 (2): 776–88. https://doi.org/10.1044/2020_ajslp-19-00202.

Jasso, J., S. McMillen, J. B. Anaya, L. M. Bedore, and E. D. Peña. 2020. "The Utility of an English Semantics Measure for Identifying Developmental Language Disorder in Spanish–English Bilinguals." *American Journal of Speech-Language Pathology* 29 (2): 776–88.

Jeon, Eun Hee, and Junko Yamashita. 2014. "L2 Reading Comprehension and Its Correlates: A Meta-Analysis." *Language Learning* 64 (1): 160–212. https://doi.org/10.1111/lang.12034.

Johnson, E. S., J. R. Jenkins, Y. Petscher, and H. W. Catts. 2009. "How Can We Improve the Accuracy of Screening Instruments?" *Learning Disabilities Research & Practice* 24 (4): 174–85.

Katzir, Tami, Youngsuk Kim, Maryanne Wolf, Beth O'Brien, Becky Kennedy, Maureen Lovett, and Robin Morris. 2006. "Reading Fluency: The Whole Is More Than the Parts." *Annals of Dyslexia* 56 (1): 51–82. https://doi.org/10.1007/s11881-006-0003-5.

Kieffer, Michael J., and Nonie K. Lesaux. 2007. "The Role of Derivational Morphology in the Reading Comprehension of Spanish-Speaking English Language Learners." *Reading and Writing* 21 (8): 783–804. https://doi.org/10.1007/s11145-007-9092-8.

Kilpatrick, David A. 2012. "Phonological Segmentation Assessment Is Not Enough." *Canadian Journal of School Psychology* 27 (2): 150–65. https://doi.org/10.1177/0829573512438635.

Kim, Young-Suk. 2012. "The Relations Among L1 (Spanish) Literacy Skills, L2 (English) Language, L2 Text Reading Fluency, and L2 Reading Comprehension for Spanish-Speaking ELL First Grade Students." *Learning and Individual Differences* 22 (6): 690–700. https://doi.org/10.1016/j.lindif.2012.06.009.

Kim, Young-Suk Grace. 2020. "Hierarchical and Dynamic Relations of Language and Cognitive Skills to Reading Comprehension: Testing the Direct and Indirect Effects Model of Reading (DIER)." *Journal of Educational Psychology* 112 (4): 667–84. https://doi.org/10.1037/edu0000407.

———. 2023. "Oral Discourse Skills: Dimensionality of Comprehension and Retell of Narrative and Expository Texts, and the Relations of Language and Cognitive Skills to Identified Dimensions." *Child Development* 94 (5). https://doi.org/10.1111/cdev.13935.

Kim, Young-Suk, and Daniel Pallante. 2010. "Predictors of Reading Skills for Kindergartners and First Grade Students in Spanish: A Longitudinal Study." *Reading and Writing* 25 (1): 1–22. https://doi.org/10.1007/s11145-010-9244-0.

Kim, Y.-S. 2023. "Dimensionality of Comprehension and Retell of Narrative and Informational Texts, and the Relations of Language and Cognitive Skills to Identified Dimensions." *Child Development* 94: e246–63.

Kirby, John R., Alain Desrochers, Leah Roth, and Sandy S. V. Lai. 2008. "Longitudinal Predictors of Word Reading Development." *Canadian Psychology / Psychologie Canadienne* 49 (2): 103–10. https://doi.org/10.1037/0708-5591.49.2.103.

Kirby, John R., George K. Georgiou, Rhonda Martinussen, and Rauno Parrila. 2010. "Naming Speed and Reading: From Prediction to Instruction." *Reading Research Quarterly* 45 (3): 341–62. https://doi.org/10.1598/rrq.45.3.4.

Kremin, Lena V., Maria M. Arredondo, Lucy Shih-Ju Hsu, Teresa Satterfield, and Ioulia Kovelman. 2016. "The Effects of Spanish Heritage Language Literacy on English Reading for Spanish–English Bilingual Children in the US." *International Journal of Bilingual Education and Bilingualism* 22 (2): 192–206. https://doi.org/10.1080/13670050.2016.1239692.

Landerl, Karin, H. Harald Freudenthaler, Moritz Heene, Peter F. De Jong, Alain Desrochers,

George Manolitsis, Rauno Parrila, and George K. Georgiou. 2018. "Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency." *Scientific Studies of Reading* 23 (3): 220–34. https://doi.org/10.1080/10888438.2018.1510936.

Leafstedt, Jill M., and Michael M. Gerber. 2005. "Crossover of Phonological Processing Skills." *Remedial and Special Education* 26 (4): 226–35. https://doi.org/10.1177/07419325050260040501.

Legible. 2024. "Analizador de Legibilidad de Texto." https://legible.es/.

Lexile Text Analyzer. 2024. "Lexile Text Analyzer [Software]." https://hub.lexile.com/analyzer.

Linacre, John M. 2006. "Data Variance Explained by Rasch Measure." *Rasch Measurement Transactions* 20 (1): 1045. https://www.rasch.org/rmt/rmt201k.htm.

Lindsey, Kim A., Franklin R. Manis, and Caroline E. Bailey. 2003. "Prediction of First-Grade Reading in Spanish-Speaking English-Language Learners." *Journal of Educational Psychology* 95 (3): 482–94. https://doi.org/10.1037/0022-0663.95.3.482.

Lobier, M., R. Zoubrinetzky, and S. Valdois. 2012. "The Visual Attention Span Deficit in Dyslexia Is Visual and Not Verbal." *Cortex* 48 (6): 768–73.

Lonigan, Christopher J. 2006. "Development, Assessment, and Promotion of Preliteracy Skills." *Early Education & Development* 17 (1): 91–114. https://doi.org/10.1207/s15566935eed1701_5.

Magis, David, and Gilles Raîche. 2012. "Random Generation of Response Patterns Under Computerized Adaptive Testing with the*R*Package**catR**." *Journal of Statistical Software* 48 (8). https://doi.org/10.18637/jss.v048.i08.

Majerus, Steve, and Nelson Cowan. 2016. "The Nature of Verbal Short-Term Impairment in Dyslexia: The Importance of Serial Order." *Frontiers in Psychology* 7 (October). https://doi.org/10.3389/fpsyg.2016.01522.

Mancilla-Martinez, Jeannette, Jin Kyoung Hwang, Min Hyun Oh, and Janna Brown McClain. 2020. "Early Elementary Grade Dual Language Learners from Spanish-Speaking Homes Struggling with English Reading Comprehension: The Dormant Role of Language Skills." *Journal of Educational Psychology* 112 (5): 880–94. https://doi.org/10.1037/edu0000402.

Mandler, George. 1980. "Recognizing: The Judgment of Previous Occurrence." *Psychological Review* 87 (3): 252–71. https://doi.org/10.1037/0033-295x.87.3.252.

Marian, V., J. Bartolotti, S. Chabal, and A. Shook. 2012. "CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities." *PLOS ONE* 7 (8): e43230.

Marinis, Theodoros, and Sharon Armon-Lotem. 2015. "Sentence Repetition." In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, edited by Sharon Armon-Lotem, Jan de Jong, and Nirit Meir, 13:95–124. Multilingual Matters.

Martin, N. A. 2013. *Expressive One-Word Picture Vocabulary Test (4th Ed., Spanish-Bilingual Ed.)*. Academic Therapy Publications.

Martin, N. A., and R. Brownell. 2011. *Expressive One-Word Picture Vocabulary Test (4th Ed.)*. Academic Therapy Publications.

Martínez, Naroa, and Edurne Goikoetxea. 2019. "Predictors of Reading and Spelling Words Change as a Function of Syllabic Structure in Spanish." *Psicología Educativa* 26 (1): 37–48. https://doi.org/10.5093/psed2019a20.

McBride-Chang, Catherine. 1995. "What Is Phonological Awareness?" *Journal of Educational Psychology* 87 (2): 179–92. https://doi.org/10.1037/0022-0663.87.2.179.

McDougall, Siné, Charles Hulme, Andrew Ellis, and Andrew Monk. 1994. "Learning to Read: The Role of Short-Term Memory and Phonological Skills." *Journal of Experimental Child Psychology*

58 (1): 112–33. https://doi.org/10.1006/jecp.1994.1028.

McGregor, K. K., J. Oleson, A. Bahnsen, and D. Duff. 2013. "Children with Developmental Language Impairment Have Vocabulary Deficits Characterized by Limited Breadth and Depth." *International Journal of Language & Communication Disorders* 48 (3): 307–19.

McGregor, Karla K., Jacob Oleson, Alison Bahnsen, and Dawna Duff. 2013. "Children with Developmental Language Impairment Have Vocabulary Deficits Characterized by Limited Breadth and Depth." *International Journal of Language & Communication Disorders* 48 (3): 307–19. https://doi.org/10.1111/1460-6984.12008.

McWeeny, Sean, Soujin Choi, June Choe, Alexander LaTourrette, Megan Y. Roberts, and Elizabeth S. Norton. 2022. "Rapid Automatized Naming (RAN) as a Kindergarten Predictor of Future Reading in English: A Systematic Review and Meta-Analysis." *Reading Research Quarterly* 57 (4): 1187–1211. https://doi.org/10.1002/rrq.467.

Meijer, Rob R., and Michael L. Nering. 1999. "Computerized Adaptive Testing: Overview and Introduction." *Applied Psychological Measurement* 23 (3): 187–94. https://doi.org/10.1177/01466219922031310.

Melby-Lervåg, Monica, and Arne Lervåg. 2011. "Cross-Linguistic Transfer of Oral Language, Decoding, Phonological Awareness and Reading Comprehension: A Meta-Analysis of the Correlational Evidence." *Journal of Research in Reading* 34 (1): 114–35. https://doi.org/10.1111/j.1467-9817.2010.01477.x.

Melzi, Gigliana. 2000. "Cultural Variations in the Construction of Personal Narratives: Central American and European American Mothers' Elicitation Styles." *Discourse Processes* 30 (2): 153–77. https://doi.org/10.1207/s15326950dp3002_04.

Míguez-Álvarez, Carla, Miguel Cuevas-Alonso, and Ángeles Saavedra. 2021. "Relationships Between Phonological Awareness and Reading in Spanish: A Meta-Analysis." *Language Learning* 72 (1): 113–57. https://doi.org/10.1111/lang.12471.

Miller, Jon F., John Heilmann, Ann Nockerts, Aquiles Iglesias, Leah Fabiano, and David J. Francis. 2006a. "Oral Language and Reading in Bilingual Children." *Learning Disabilities Research & Practice* 21 (1): 30–43. https://doi.org/10.1111/j.1540-5826.2006.00205.x.

———. 2006b. "Oral Language and Reading in Bilingual Children." *Learning Disabilities Research & Practice* 21 (1): 30–43. https://doi.org/10.1111/j.1540-5826.2006.00205.x.

Misra, Maya, Tamar Katzir, Maryanne Wolf, and Russell A. Poldrack. 2004. "Neural Systems for Rapid Automatized Naming in Skilled Readers: Unraveling the RAN-Reading Relationship." *Scientific Studies of Reading* 8 (3): 241–56. https://doi.org/10.1207/s1532799xssr0803_4.

MOLL, KRISTINA, CHARLES HULME, SONALI NAG, and MARGARET J. SNOWLING. 2013. "Sentence Repetition as a Marker of Language Skills in Children with Dyslexia." *Applied Psycholinguistics* 36 (2): 203–21. https://doi.org/10.1017/s0142716413000209.

Nagy, W. E., and J. A. Scott. 2000. "Vocabulary Processes." In *Handbook of Reading Research*, edited by M. L. Kamil, P. B. Mosenthal, P. D. Pearson, and R. Barr, 3:269–84. Mahwah, NJ: Lawrence Erlbaum Associates.

Nakamoto, Jonathan, Kim A. Lindsey, and Franklin R. Manis. 2006. "A Longitudinal Analysis of English Language Learners' Word Decoding and Reading Comprehension." *Reading and Writing* 20 (7): 691–719. https://doi.org/10.1007/s11145-006-9045-7.

———. 2008. "A Cross-Linguistic Investigation of English Language Learners' Reading Comprehension in English and Spanish." *Scientific Studies of Reading* 12 (4): 351–71. https://doi.org/10.1080/10888430802378526.

Nation, Kate, Paula Clarke, Catherine M. Marshall, and Marianne Durand. 2004. "Hidden Language

Impairments in Children." *Journal of Speech, Language, and Hearing Research* 47 (1): 199–211. https://doi.org/10.1044/1092-4388(2004/017).

Nation, Kate, Joanne Cocksey, Jo S. H. Taylor, and Dorothy V. M. Bishop. 2010. "A Longitudinal Investigation of Early Reading and Language Skills in Children with Poor Reading Comprehension." *Journal of Child Psychology and Psychiatry* 51 (9): 1031–39. https://doi.org/10.1111/j.1469-7610.2010.02254.x.

Nation, Kate, and Charles Hulme. 2010. "Learning to Read Changes Children's Phonological Skills: Evidence from a Latent Variable Longitudinal Study of Reading and Nonword Repetition." *Developmental Science* 14 (4): 649–59. https://doi.org/10.1111/j.1467-7687.2010.01008.x.

National Center for Education Statistics. 2024. "Students with Disabilities. Condition of Education." U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cgg.

National Reading Panel (US). 2000. "Report of the National Reading Panel: Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups." Technical Report. Washington, DC: National Institute of Child Health; Human Development, National Institutes of Health.

National Research Council. 1998. *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Academies Press.

Naveh-Benjamin, Moshe, and Thomas J. Ayres. 1986. "Digit Span, Reading Rate, and Linguistic Relativity." *The Quarterly Journal of Experimental Psychology Section A* 38 (4): 739–51. https://doi.org/10.1080/14640748608401623.

Norton, Elizabeth S., and Maryanne Wolf. 2012. "Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities." *Annual Review of Psychology* 63 (1): 427–52. https://doi.org/10.1146/annurev-psych-120710-100431.

O'Brien, G., and J. D. Yeatman. 2021. "Bridging Sensory and Language Theories of Dyslexia: Toward a Multifactorial Model." *Developmental Science* 24 (3): e13039.

O'Connor, Rollanda E., and Joseph R. Jenkins. 1999. "Prediction of Reading Disabilities in Kindergarten and First Grade." *Scientific Studies of Reading* 3 (2): 159–97. https://doi.org/10.1207/s1532799xssr0302_4.

Oh, Min Hyun, Jeannette Mancilla-Martinez, and Jin Kyoung Hwang. 2023. "Revisiting the Traditional Conceptualizations of Vocabulary Knowledge as Predictors of Dual Language Learners' English Reading Achievement in a New Destination State." *Applied Psycholinguistics* 44 (1): 51–75. https://doi.org/10.1017/s0142716422000479.

Ozernov-Palchik, O., E. S. Norton, G. Sideridis, S. D. Beach, M. Wolf, J. D. E. Gabrieli, and N. Gaab. 2017. "Longitudinal Stability of Pre-Reading Skill Profiles of Kindergarten Children: Implications for Early Screening and Theories of Reading." *Developmental Science* 20 (5).

Parrila, Rauno, John R. Kirby, and Lynn McQuarrie. 2004. "Articulation Rate, Naming Speed, Verbal Short-Term Memory, and Phonological Awareness: Longitudinal Predictors of Early Reading Development?" *Scientific Studies of Reading* 8 (1): 3–26. https://doi.org/10.1207/s1532799xssr0801_2.

Pasquarella, Adrian, Xi Chen, Alexandra Gottardo, and Esther Geva. 2015. "Cross-Language Transfer of Word Reading Accuracy and Word Reading Fluency in Spanish-English and Chinese-English Bilinguals: Script-Universal and Script-Specific Processes." *Journal of Educational Psychology* 107 (1): 96–110. https://doi.org/10.1037/a0036966.

Pelli, D. G., C. W. Burns, B. Farell, and D. C. Moore-Page. 2006. "Feature Detection and Letter Identification." *Vision Research* 46 (28): 4646–74.

Pennington, B. F. 2011. "Controversial Therapies for Dyslexia." *Perspectives on Language and Literacy* 37 (1): 7–8.

Petersen, Douglas B., and Trina D. Spencer. 2012. "The Narrative Language Measures: Tools for Language Screening, Progress Monitoring, and Intervention Planning." *Perspectives on Language Learning and Education* 19 (4): 119–29. https://doi.org/10.1044/lle19.4.119.

Piasta, Shayne B., David J. Purpura, and Richard K. Wagner. 2009. "Fostering Alphabet Knowledge Development: A Comparison of Two Instructional Approaches." *Reading and Writing* 23 (6): 607–26. https://doi.org/10.1007/s11145-009-9174-x.

Polišenská, Kamila, Shula Chiat, and Penny Roy. 2014. "Sentence Repetition: What Does the Task Measure?" *International Journal of Language & Communication Disorders* 50 (1): 106–18. https://doi.org/10.1111/1460-6984.12126.

Powell-Smith, K. A., R. A. Kaminski, R. H. Good, M. Abbott, S. Stollar, J. Wallin, and C. E. Wheeler. 2020a. *Acadience RAN Assessment Manual*. Acadience Learning Inc.

———. 2020b. *Acadience RAN Assessment Manual*. Acadience Learning Inc.

Pratt, Amy S., Elizabeth D. Peña, and Lisa M. Bedore. 2020. "Sentence Repetition with Bilinguals with and Without DLD: Differential Effects of Memory, Vocabulary, and Exposure." *Bilingualism: Language and Cognition* 24 (2): 305–18. https://doi.org/10.1017/s1366728920000498.

Proctor, C. Patrick, Diane August, María S. Carlo, and Catherine Snow. 2006. "The Intriguing Role of Spanish Language Vocabulary Knowledge in Predicting English Reading Comprehension." *Journal of Educational Psychology* 98 (1): 159–69. https://doi.org/10.1037/0022-0663.98.1.159.

Proctor, C. Patrick, Jeffrey R. Harring, and Rebecca D. Silverman. 2017. "Linguistic Interdependence Between Spanish Language and English Language and Reading: A Longitudinal Exploration from Second Through Fifth Grade." *Bilingual Research Journal* 40 (4): 372–91. https://doi.org/10.1080/15235882.2017.1383949.

Qi, Ting, Maria Luisa Mandelli, Christa L. Watson Pereira, Emma Wellman, Rian Bogley, Abigail E. Licata, Edward F. Chang, Yulia Oganian, and Maria Luisa Gorno-Tempini. 2023. "Anatomical and Behavioral Correlates of Auditory Perception in Developmental Dyslexia." http://dx.doi.org/10.1101/2023.05.09.539936.

Ramamurthy, M., A. L. White, and J. D. Yeatman. 2024. "Children with Dyslexia Show No Deficit in Exogenous Spatial Attention but Show Differences in Visual Encoding." *Developmental Science* 27 (3): e13458. https://doi.org/10.1111/desc.13458.

Randall, Jennifer, Mya Poe, David Slomp, and Maria Elena Oliveri. 2023. "Our Validity Looks Like Justice. Does Yours?" *Language Testing* 41 (1): 203–19. https://doi.org/10.1177/02655322231202947.

Reckase, Mark D. 1979. "Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications." *Journal of Educational Statistics* 4 (3): 207–30. https://doi.org/10.3102/10769986004003207.

Reese, Elaine, Sebastian Suggate, Jennifer Long, and Elizabeth Schaughency. 2009. "Children's Oral Narrative and Reading Skills in the First 3 Years of Reading Instruction." *Reading and Writing* 23 (6): 627–44. https://doi.org/10.1007/s11145-009-9175-9.

Reschly, Amy L., Todd W. Busch, Joseph Betts, Stanley L. Deno, and Jeffrey D. Long. 2009. "Curriculum-Based Measurement Oral Reading as an Indicator of Reading Achievement: A Meta-Analysis of the Correlational Evidence." *Journal of School Psychology* 47 (6): 427–69. https://doi.org/10.1016/j.jsp.2009.07.001.

Riccio, C. A., B. Imhoff, J. E. Hasbrouck, and G. N. Davis. 2004b. *Test of Phonological Awareness in Spanish*. Pro-Ed.

———. 2004a. *Test of Phonological Awareness in Spanish*. Pro-Ed.

Riedel, Brant W.RIEDEL. 2007. "The Relation Between DIBELS, Reading Comprehension, and Vocabulary in Urban First-Grade Students." *Reading Research Quarterly* 42 (4): 546–67. https://doi.org/10.1598/rrq.42.4.5.

Riva, Anna, Alessandro Musetti, Monica Bomba, Lorenzo Milani, Valentina Montrasi, and Renata Nacinovich. 2021. "Language-Related Skills in Bilingual Children with Specific Learning Disorders." *Frontiers in Psychology* 11 (January). https://doi.org/10.3389/fpsyg.2020.564047.

Roberts, Theresa A. 2005. "Articulation Accuracy and Vocabulary Size Contributions to Phonemic Awareness and Word Reading in English Language Learners." *Journal of Educational Psychology* 97 (4): 601–16. https://doi.org/10.1037/0022-0663.97.4.601.

Rubin, Renée, and Verónica Galván Carlan. 2005. "Using Writing to Understand Bilingual Children's Literacy Development." *The Reading Teacher* 58 (8): 728–39. https://doi.org/10.1598/rt.58.8.3.

Scarborough, Hollis S. 1998. "Predicting the Future Achievement of Second Graders with Reading Disabilities: Contributions of Phonemic Awareness, Verbal Memory, Rapid Naming, and IQ." *Annals of Dyslexia* 48 (1): 115–36. https://doi.org/10.1007/s11881-998-0006-5.

Scarborough, Hollis S., Susan B. Neuman, and David K. Dickinson. 2001. "Connecting Early Language and Literacy to Later Reading (Dis) Abilities: Evidence, Theory, and Practice." In *Handbook of Early Literacy Research*, 1:97–110. New York, NY: Guilford Press.

Schatschneider, Christopher, Jack M. Fletcher, David J. Francis, Coleen D. Carlson, and Barbara R. Foorman. 2004. "Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis." *Journal of Educational Psychology* 96 (2): 265–82. https://doi.org/10.1037/0022-0663.96.2.265.

Schrank, F. A., K. S. McGrew, and N. Mather. 2014. *Woodcock-Johnson IV*. Riverside Assessments, LLC.

Semel, E., E. H. Wiig, W. A. Secord, and H. W. Langdon. 2006b. *Clinical Evaluation of Language Fundamentals (4th Ed., Spanish)*. Pearson Education Inc.

———. 2006a. *Clinical Evaluation of Language Fundamentals (4th Ed., Spanish)*. Pearson Education Inc.

Serrano, Francisca, and Sylvia Defior. 2008. "Dyslexia Speed Problems in a Transparent Orthography." *Annals of Dyslexia* 58 (1): 81–95. https://doi.org/10.1007/s11881-008-0013-6.

———. 2010. "Spanish Dyslexic Spelling Abilities: The Case of Consonant Clusters." *Journal of Research in Reading* 35 (2): 169–82. https://doi.org/10.1111/j.1467-9817.2010.01454.x.

Seymour, Philip H. K., Mikko Aro, and Jane M. Erskine. 2003. "Foundation Literacy Acquisition in European Orthographies." *British Journal of Psychology* 94 (2): 143–74. https://doi.org/10.1348/000712603321661859.

Siebert, Julian M, Phaedra Bell, Nuria Gutiérrez, Mónica Zegers, Eric Falke, Benjamin W Domingue, Yaacov Petscher, et al. 2025. "Differences in Reading Screening Accuracy by Percentile Cutoff and English Proficiency: Feature Selection and Group-Wise Prediction Evaluation." *Reading Research Quarterly*, e70074. https://doi.org/10.1002/rrq.70074.

Signorini, Angela. 1997. "Word Reading in Spanish: A Comparison Between Skilled and Less Skilled Beginning Readers." *Applied Psycholinguistics* 18 (3): 319–44. https://doi.org/10.1017/s014271640001050x.

Silverman, Rebecca D., C. Patrick Proctor, Jeffrey R. Harring, Anna M. Hartranft, Brie Doyle, and Sarah B. Zelinke. 2015. "Language Skills and Reading Comprehension in English Monolingual and Spanish–English Bilingual Children in Grades 2–5." *Reading and Writing* 28 (9): 1381–1405. https://doi.org/10.1007/s11145-015-9575-y.

Simon-Cereijido, Gabriela, and Vera F. Gutiérrez-Clellen. 2017. "Bilingual Education for All: Latino Dual Language Learners with Language Disabilities." In *Immersion Education in the Early Years*, 117–36. Routledge.

Sireci, Stephen G., and Jennifer Randall. 2021. "Evolving Notions of Fairness in Testing in the United States." In *The History of Educational Measurement*, 111–35. Routledge.

Snowling, Margaret J., Alison Gallagher, and Uta Frith. 2003. "Family Risk of Dyslexia Is Continuous: Individual Differences in the Precursors of Reading Skill." *Child Development* 74 (2): 358–73. https://doi.org/10.1111/1467-8624.7402003.

Snowling, Margaret J., and Charles Hulme. 2020. "Annual Research Review: Reading Disorders Revisited – the Critical Importance of Oral Language." *Journal of Child Psychology and Psychiatry* 62 (5): 635–53. https://doi.org/10.1111/jcpp.13324.

Snowling, Margaret, Kate Nation, Philippa Moxham, Alison Gallagher, and Uta Frith. 1997. "Phonological Processing Skills of Dyslexic Students in Higher Education: A Preliminary Report." *Journal of Research in Reading* 20 (1): 31–41. https://doi.org/10.1111/1467-9817.00018.

Solano-Flores, G. 2023. "How Serious are We About Fairness in Testing and How Far are We Willing to Go? A Response to Randall and Bennett with Reflections About the Standards for Educational and Psychological Testing." *Educational Assessment* 28 (2): 105–17. https://doi.org/10.1080/10627197.2023.2226388.

Solari, Emily J., Terese C. Aceves, Ignacio Higareda, Cara Richards-Tutor, Alexis L. Filippini, Michael M. Gerber, and Jill Leafstedt. 2013. "LONGITUDINAL PREDICTION OF 1ST AND 2ND GRADE ENGLISH ORAL READING FLUENCY IN ENGLISH LANGUAGE LEARNERS: WHICH EARLY READING AND LANGUAGE SKILLS ARE BETTER PREDICTORS?" *Psychology in the Schools* 51 (2): 126–42. https://doi.org/10.1002/pits.21743.

Sperling, G. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1.

Stahl, S. A., and W. E. Nagy. 2007. *Teaching Word Meanings*. New York, NY: Routledge.

Stein, J., and V. Walsh. 1997. "To See but Not to Read; the Magnocellular Theory of Dyslexia." *Trends in Neurosciences* 20 (4): 147–52.

Storch, Stacey A., and Grover J. Whitehurst. 2002a. "Oral Language and Code-Related Precursors to Reading: Evidence from a Longitudinal Structural Model." *Developmental Psychology* 38 (6): 934–47. https://doi.org/10.1037/0012-1649.38.6.934.

———. 2002b. "Oral Language and Code-Related Precursors to Reading: Evidence from a Longitudinal Structural Model." *Developmental Psychology* 38 (6): 934–47. https://doi.org/10.1037/0012-1649.38.6.934.

Sun-Alperin, M. Kendra, and Min Wang. 2009. "Cross-Language Transfer of Phonological and Orthographic Processing Skills from Spanish L1 to English L2." *Reading and Writing* 24 (5): 591–614. https://doi.org/10.1007/s11145-009-9221-7.

Swank, Linda K., and Hugh W. Catts. 1994. "Phonological Awareness and Written Word Decoding." *Language, Speech, and Hearing Services in Schools* 25 (1): 9–14. https://doi.org/10.1044/0161-1461.2501.09.

Swanson, H. Lee, and Linda Siegel. 2011. "Learning Disabilities as a Working Memory Deficit." *Experimental Psychology* 49 (1): 5–28.

Swanson, H. Lee, Xinhua Zheng, and Olga Jerman. 2009. "Working Memory, Short-Term Memory, and Reading Disabilities." *Journal of Learning Disabilities* 42 (3): 260–87. https://doi.org/10.1177/0022219409331958.

Talcott, J. B., C. Witton, G. S. Hebb, C. J. Stoodley, E. A. Westwood, S. J. France, P. C. Hansen,

and J. F. Stein. 2002. "On the Relationship Between Dynamic Visual and Auditory Processing and Literacy Skills; Results from a Large Primary-School Study." *Dyslexia* 8 (4): 204–25.

Taran, Nikolay, Rola Farah, Mark DiFrancesco, Mekibib Altaye, Jennifer Vannest, Scott Holland, Keri Rosch, Bradley L. Schlaggar, and Tzipi Horowitz-Kraus. 2022. "The Role of Visual Attention in Dyslexia: Behavioral and Neurobiological Evidence." *Human Brain Mapping* 43 (5): 1720–37. https://doi.org/10.1002/hbm.25753.

Torgesen, Joseph K., and D. Griffith Houck. 1980. "Processing Deficiencies of Learning-Disabled Children Who Perform Poorly on the Digit Span Test." *Journal of Educational Psychology* 72 (2): 141–60. https://doi.org/10.1037/0022-0663.72.2.141.

Torgesen, Joseph, and Tina Goldman. 1977. "Verbal Rehearsal and Short-Term Memory in Reading-Disabled Children." *Child Development* 48 (1): 56. https://doi.org/10.2307/1128881.

Torppa, Minna, Anna-Maija Poikkeus, Marja-Leena Laakso, Kenneth Eklund, and Heikki Lyytinen. 2006. "Predicting Delayed Letter Knowledge Development and Its Relation to Grade 1 Reading Achievement Among Children with and Without Familial Risk for Dyslexia." *Developmental Psychology* 42 (6): 1128–42. https://doi.org/10.1037/0012-1649.42.6.1128.

Treiman, Rebecca, and Brett Kessler. 2021. "Statistical Learning in Word Reading and Spelling Across Languages and Writing Systems." *Scientific Studies of Reading* 26 (2): 139–49. https://doi.org/10.1080/10888438.2021.1920951.

Uccelli, Paola, and Mariela M. Paéz. 2007. "Narrative and Vocabulary Development of Bilingual Children From Kindergarten to First Grade: Developmental Changes and Associations Among English and Spanish Skills." *Language, Speech, and Hearing Services in Schools* 38 (3): 225–36. https://doi.org/10.1044/0161-1461(2007/024).

Valdois, S., M.-L. Bosse, and M.-J. Tainturier. 2004. "The Cognitive Deficits Responsible for Developmental Dyslexia: Review of Evidence for a Selective Visual Attentional Disorder." *Dyslexia* 10 (4): 339–63.

Valdois, S., C. Reilhac, E. Ginestet, and M.-L. Bosse. 2021. "Varieties of Cognitive Profiles in Poor Readers: Evidence for a VAS-Impaired Subtype." *Journal of Learning Disabilities* 54 (3): 221–33.

Van Den Boer, M., E. Van Bergen, and P. F. de Jong. 2015. "The Specific Relation of Visual Attention Span with Reading and Spelling in Dutch." *Learning and Individual Differences* 39: 141–49.

Vellutino, Frank R., Jack M. Fletcher, Margaret J. Snowling, and Donna M. Scanlon. 2004. "Specific Reading Disability (Dyslexia): What Have We Learned in the Past Four Decades?" *Journal of Child Psychology and Psychiatry* 45 (1): 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x.

Vellutino, Frank R., William E. Tunmer, James J. Jaccard, and RuSan Chen. 2007. "Components of Reading Ability: Multivariate Evidence for a Convergent Skills Model of Reading Development." *Scientific Studies of Reading* 11 (1): 3–32. https://doi.org/10.1080/10888430709336632.

Verhagen, Josje, Jan Boom, Hanna Mulder, Elise de Bree, and Paul Leseman. 2019. "Reciprocal Relationships Between Nonword Repetition and Vocabulary During the Preschool Years." *Developmental Psychology* 55 (6): 1125–37. https://doi.org/10.1037/dev0000702.

Verhoeven, Ludo, and Jos Keuning. 2017. "The Nature of Developmental Dyslexia in a Transparent Orthography." *Scientific Studies of Reading* 22 (1): 7–23. https://doi.org/10.1080/10888438.2017.1317780.

Vettori, Giulia, Oriana Incognito, Lucia Bigozzi, and Giuliana Pinto. 2023. "Relationship Between Lexical, Reading and Spelling Skills in Bilingual Language Minority Children and Their Monolingual Peers." *Frontiers in Psychology* 14 (August). https://doi.org/10.3389/fpsyg.2023.1121505.

Wagner, R. K., J. K. Torgesen, C. A. Rashotte, and N. A. Pearson. 2013. *Comprehensive Test of*

Phonological Processing (2nd Ed.). Pro-Ed.

Wainer, Howard, Neil J. Dorans, Ronald Flaugher, Bert F. Green, and Robert J. Mislevy. 2000. *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Routledge.

Washington, J. A., and M. S. Seidenberg. 2021. "Teaching Reading to African American Children: When Home and School Differ." *American Educator* 45 (2): 26–40.

Washington, Julie A. et al., eds. 2020. *Dyslexia: Revisiting Etiology, Diagnosis, Treatment, and Policy*. Baltimore, MD: Paul H. Brookes Publishing.

Weiss, David J. 1985. "Adaptive Testing by Computer." *Journal of Consulting and Clinical Psychology* 53 (6): 774–89. https://doi.org/10.1037/0022-006x.53.6.774.

Whitehurst, Graver J., and Christopher J. Lonigan. 1998. "Child Development and Emergent Literacy." *Child Development* 69 (3): 848–72. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x.

Whitehurst, Grover J., and Christopher J. Lonigan. 2001. *Get Ready to Read! Revised*. National Center for Learning Disabilities; Pearson Education Inc.

Wiig, E. H., E. Semel, and W. A. Secord. 2013. *Clinical Evaluation of Language Fundamentals (5th Ed.)*. Pearson Education Inc.

Willburger, Edith, Barbara Fussenegger, Kristina Moll, Guilherme Wood, and Karin Landerl. 2008. "Naming Speed in Dyslexia and Dyscalculia." *Learning and Individual Differences* 18 (2): 224–36. https://doi.org/10.1016/j.lindif.2008.01.003.

Wilson, Mark. 2023. *Constructing Measures: An Item Response Modeling Approach*. New York, NY: Routledge.

Winters, Katherine L., Javier Jasso, James E. Pustejovsky, and Courtney T. Byrd. 2022. "Investigating Narrative Performance in Children With Developmental Language Disorder: A Systematic Review and Meta-Analysis." *Journal of Speech, Language, and Hearing Research* 65 (10): 3908–29. https://doi.org/10.1044/2022_jslhr-22-00017.

Wolf, Maryanne, and Patricia Greig Bowers. 1999. "The Double-Deficit Hypothesis for the Developmental Dyslexias." *Journal of Educational Psychology* 91 (3): 415–38. https://doi.org/10.1037/0022-0663.91.3.415.

Wolf, Maryanne, Alyssa Goldberg O'Rourke, Calvin Gidney, Maureen Lovett, Paul Cirino, and Robin Morris. 2002. *Reading and Writing* 15 (1/2): 43–72. https://doi.org/10.1023/a:1013816320290.

Wolf, M., and M. B. Denckla. 2003a. *RAN/RAS Rapid Automatized Naming and Rapid Alternating Stimulus Tests*. Pro-Ed.

———. 2003b. *RAN/RAS Rapid Automatized Naming and Rapid Alternating Stimulus Tests*. Pro-Ed.

Wolf, M., R. J. Gotlieb, S. A. Kim, V. Pedroza, L. V. Rhinehart, M. L. G. Tempini, and S. Sears. 2024. "Towards a Dynamic, Comprehensive Conceptualization of Dyslexia." *Annals of Dyslexia*, 1–22.

Woodcock, R. W., C. G. Alvarado, F. A. Schrank, N. Mather, K. S. McGrew, and A. F. Muñoz-Sandoval. 2019. *Batería IV Woodcock-Muñoz*. Riverside Assessments, LLC.

Wren, Yasmin, and Socorro Herrera. 2021. "Early Literacy Predictors of Spanish Reading Achievement in Bilingual Children." *Bilingual Research Journal* 44 (1): 58–73.

Yeo, Seungsoo. 2009. "Predicting Performance on State Achievement Tests Using Curriculum-Based Measurement in Reading: A Multilevel Meta-Analysis." *Remedial and Special Education* 31 (6): 412–22. https://doi.org/10.1177/0741932508327463.

Yesil-Dagli, Ummuhan. 2011. "Predicting ELL Students' Beginning First Grade English Oral Reading Fluency from Initial Kindergarten Vocabulary, Letter Naming, and Phonological Awareness Skills."

Early Childhood Research Quarterly 26 (1): 15–29. https://doi.org/10.1016/j.ecresq.2010.06.001.

Yue, Qiuhai, and Randi C. Martin. 2021. "Maintaining Verbal Short-Term Memory Representations in Non-Perceptual Parietal Regions." *Cortex* 138 (May): 72–89. https://doi.org/10.1016/j.cortex.2021.01.020.

Ziegler, Johannes C., and Usha Goswami. 2005. "Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory." *Psychological Bulletin* 131 (1): 3–29. https://doi.org/10.1037/0033-2909.131.1.3.