The Development and Preliminary Evidence of Validity and Reliability of a Spanish Static and Dynamic Assessment of Decoding

Kayla Brooke Ditty
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Communication Sciences and Disorders Commons

BYU ScholarsArchive Citation
Ditty, Kayla Brooke, "The Development and Preliminary Evidence of Validity and Reliability of a Spanish Static and Dynamic Assessment of Decoding" (2020). Theses and Dissertations. 8481. https://scholarsarchive.byu.edu/etd/8481

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
The Development and Preliminary Evidence of Validity and Reliability
of a Spanish Static and Dynamic Assessment of Decoding

Kayla Brooke Ditty

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

Douglas B. Petersen, Chair
Connie Summers
Shawn L. Nissen

Department of Communication Disorders
Brigham Young University

Copyright © 2020 Kayla Brooke Ditty
All Rights Reserved
ABSTRACT

The Development and Preliminary Evidence of Validity and Reliability of a Spanish Static and Dynamic Assessment of Decoding

Kayla Brooke Ditty
Department of Communication Disorders, BYU
Master of Science

The purpose of the current study was to develop and administer a Spanish static and dynamic assessment of decoding and other important reading measures. Using the best available resources regarding Spanish literacy acquisition and Spanish language development, a preliminary Spanish dynamic and static screener was developed. The resulting test was named the Spanish Dynamic Decoding Measure (S-DDM) and is a counterpart to the original DDM available in English. This test contains three subtests: decoding, phonemic awareness, and word identification. The S-DDM was administered to eight bilingual Spanish-English speaking students with varying levels of Spanish language ability and reading proficiency in grades ranging from second to fourth grade. The results from all three subtests of the S-DDM were analyzed to report the levels of validity, reliability, and variance observed in the administration of the test and performance of the participants. There was a wide range of performance on the pretest sections of the dynamic assessment decoding subtest, and performance at or near ceiling on the posttest sounds sections. Modifiability scores were strong for the students who were not receiving special education services while the two participants receiving special education both received weaker modifiability ratings. There was a wide range of variance observed on all subtests of both the phonemic awareness and word identification subtests, with the exception of the phoneme blending target of the phonemic awareness subtest. Interrater reliability and administration fidelity were found to be within acceptable measures. Results indicated that the test is sensitive to varying degrees of Spanish reading proficiency and reading ability. Future research is needed to determine the effectiveness of this screener in identifying Spanish-speaking students who are at-risk for reading difficulties.

Keywords: reading, dynamic assessment, Spanish, decoding
ACKNOWLEDGEMENTS

I express my appreciation to my committee members: Dr. Petersen, Dr. Summers, and Dr. Nissen who provided support and feedback throughout the process of writing this thesis. I especially want to express my gratitude to Dr. Petersen for taking me on as a thesis student and spending countless hours revising and editing this thesis with me while at home and abroad. His dedication and flexibility have enabled me to reach my goals regarding the timing and quality of this research project. I also want to express my gratitude to my family who have constantly supported me in my academic endeavors. Finally, I wish to express my love and gratitude for my husband who has never doubted in my abilities and who has given so much of himself to enable me to pursue my academic dreams.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>DESCRIPTION OF THESIS STRUCTURE AND CONTENT</td>
<td>viii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Traditional Assessments</td>
<td>4</td>
</tr>
<tr>
<td>Dynamic Assessment</td>
<td>8</td>
</tr>
<tr>
<td>Phase One</td>
<td>14</td>
</tr>
<tr>
<td>Phase Two</td>
<td>14</td>
</tr>
<tr>
<td>Method</td>
<td>14</td>
</tr>
<tr>
<td>Phase One: Test Development</td>
<td>14</td>
</tr>
<tr>
<td>Part Two: Preliminary Investigation</td>
<td>16</td>
</tr>
<tr>
<td>Participants</td>
<td>16</td>
</tr>
<tr>
<td>Procedures</td>
<td>18</td>
</tr>
<tr>
<td>Results</td>
<td>19</td>
</tr>
<tr>
<td>Phase One: Test Development</td>
<td>19</td>
</tr>
<tr>
<td>Decoding</td>
<td>19</td>
</tr>
<tr>
<td>Phonemic Awareness</td>
<td>22</td>
</tr>
<tr>
<td>Word Identification</td>
<td>23</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1  Demographic Information for Participants ................................................................. 17
Table 2  Language Testing in English Results ......................................................................... 18
Table 3  Decoding Subtest Performance Results ................................................................. 26
Table 4  Phonemic Awareness Subtest Performance Results ................................................. 27
Table 5  Word Identification Subtest Performance Results ................................................... 28
LIST OF FIGURES

Figure 1  Decoding Subtest Histograms with the Normal Curve Delineated.........................29

Figure 2  Phonemic Awareness Subtest Histograms with the Normal Curve Delineated .........30

Figure 3  Word Identification Subtest Histograms with the Normal Curve Delineated......... 30
DESCRIPTION OF THESIS STRUCTURE AND CONTENT

To adhere to traditional thesis requirements and journal publication formats, this thesis, *The Development and Preliminary Evidence of Validity and Reliability of a Spanish Static and Dynamic Assessment of Decoding*, is written in a hybrid format. The initial pages of the thesis adhere to university requirements while the thesis report is presented in journal article format.

The annotated bibliography is included in Appendix A. Appendix B contains the S-DDM Decoding subtest, Appendix C contains the S-DDM Phonemic Awareness subtest, and Appendix D contains the S-DDM Word Identification subtest. Finally, Appendix E contains the Fidelity checklist used for this study.
**Introduction**

Spanish is one of the most spoken languages in the world, consisting of over 650 million speakers worldwide (Sawe, 2016). Furthermore, Spanish is the second most spoken language in the United States (U.S. Census Bureau, 2015). This is due in large part to the continuously growing Hispanic population in the United States. Between the years of 2000 and 2010, more than half of the population growth experienced by the United States was Hispanic (Bauman, 2017). The Hispanic population is the largest ethnic minority group in the United States.

Unsurprisingly, the extensive growth seen in the Hispanic population has led to an increase in the number of Hispanic students attending U. S. schools in all regions. Hispanic students make up 22.7 percent of all students across the country (Bauman, 2017). Whereas the proportion of White and African American children was projected to continue to decrease between 2013 and 2025, the Hispanic population of children was expected to increase (Musu-Gillette et al., 2017). These numbers and projections indicate an extensive Spanish influence in schools nationwide.

With such large numbers of Hispanic students enrolled in U. S. schools, the English language abilities of these students vary across a wide spectrum of proficiency and use (Petersen et al., 2018). This in part is because many Hispanic students are exposed to differing levels of Spanish and English in the home, creating cultural and linguistic differences observed in many classrooms. For example, 70 percent of Hispanic students report speaking Spanish at home (Mancilla-Martinez & Lesaux, 2011). Because of the rising number of diverse students, it is likely that all education professionals will be faced with more culturally and ethnically diverse students in need of their services (Laing & Kamhi, 2003). However, professionals have expressed concern regarding the capacity of the U. S. educational system to serve the needs of
the ever-expanding diverse, multilingual student population (Geva et al., 2000). Furthermore, there has been no clear consensus on when and how to evaluate culturally and linguistically diverse (CLD) children for potential learning and reading disabilities. Determining whether a child is struggling academically simply because of a linguistic difference or because of a true disorder has been a pervasive dilemma for educators for many years. Historically, CLD children have been overrepresented in special education programs as well as overidentified for disorders and disabilities (Durgunoğlu, 2002; Peña, 2000; Petersen et al., 2016). In order to properly serve these children, it is necessary that educators be able to accurately identify those who need special education services because of real disorders, not differences in language exposure.

Concerning the Hispanic population specifically, educators are concerned with the number of Latino students that struggle with reading (Petersen & Gillam, 2015). According to a National Assessment of Educational Progress (NAEP) report regarding reading performance, the fourth grade White-Hispanic achievement gap for reading in 2017 was not measurably different from the gap measured in 1992 (Musu-Gillette et al., 2017). This trend is especially concerning because reading ability is an integral part of academic success. Researchers have cited that 75 percent of first graders who manifest reading difficulties will continue having problems with reading well into adulthood (Petersen et al., 2016; Petersen & Gillam, 2015). Additionally, delayed development of reading abilities adversely affects vocabulary growth and leads to the augmentation of negative feelings towards reading which decreases a child's motivation to read, dramatically reducing their exposure to literacy-building opportunities (Torgesen, 2002). Torgesen cited the results of multiple studies which indicated that children who are poor readers at the end of first grade almost never catch up to grade-level reading abilities by the end of elementary school. Furthermore, it has also been reported that when intervention is delayed until
third grade or later, the proportion of children with reading disabilities who actually obtain successful remediation decreases dramatically (Chiappe et al., 2002). Thus, the seriousness of accurate, early identification of children at risk for reading disability is imperative in providing early remediation to promote later academic success.

Increasing awareness of the pervasiveness and adverse effects of reading difficulty has led professionals and researchers to advocate for the improvement of early identification of children who are at-risk for reading disability (Bridges & Catts, 2011; Chiappe et al., 2002; Gellert & Elbro, 2018; Petersen & Gillam 2015; Petersen et al., 2016; Torgesen, 2002). Children who are identified early on as at-risk for reading difficulty will be more likely to receive the necessary remediation which in turn will decrease the chance for academic failure and other ramifications associated with reading disorders. Due to the limited nature of educational resources, it is important to allocate these resources in a systematic way to children who are genuinely in need of such services. To accomplish this design, accurate identification is just as crucial as early identification (Gellert & Elbro, 2018). Because of the pervasive nature of reading disabilities across all cultures as well as the observed shifts in the U. S. demographics, it is prudent to acknowledge the necessity to include CLD groups in our objective to improve early and accurate identification measures (Laing & Kamhi, 2003). For example, the current research available on the development of reading skills has almost exclusively centered on English speakers (Mancilla-Martinez & Lesaux, 2011). There is a lack of research regarding literacy acquisition and patterns of development in word reading in Spanish-speaking Latino children (Geva et al., 2000; Mancilla-Martinez & Lesaux, 2011). Thus, there are severe limitations in the current pool of information and resources to assess minority groups.
Traditional Assessments

Currently, many traditional assessments, screeners, and progress monitoring tools of reading abilities are limited in evaluating CLD children because of well-documented biases towards such groups (Durgunoğlu, 2002; Laing & Kamhi, 2003; Lidz & Peña, 1996; Peña, 2000; Petersen & Gillam, 2015; Petersen et al., 2016). Biases arise because test taking is essentially a cultural phenomenon which would imply that children who are not familiar with test procedures, test content, or the language of the test would naturally perform poorly on such tests (Laing & Kamhi, 2003; Lidz & Peña, 1996). Assessment biases are a serious hindrance to the educational and academic progress of minorities, which merits the careful consideration of appropriate solutions.

One dominant bias of traditional assessments is the consistency with which these tests over identify CLD children as being at greater risk for reading difficulties (Petersen et al., 2016). Overidentification implies that a large number of the CLD children who are misidentified as at-risk may only be experiencing difficulty because of linguistic differences, and not because they are disordered (Lidz & Peña, 1996). As previously mentioned, past statistical trends have shown that CLD children were often overrepresented in special education populations, suggesting the likelihood of associating difficulties with more general learning problems, when none existed (Durgunoğlu, 2002). In contrast, recently, despite unchanged test scores, educators have avoided labelling CLD children as disordered until later on in their schooling when it can be more certain that problems are consistent with a diagnosis of a disability, and not a difference in linguistic proficiency (Durgunoğlu, 2002; Geva et al., 2000). However, this approach is also unacceptable because of the loss of valuable intervention time. Therefore, another important quality of an
appropriate assessment for CLD children is the ability to distinguish between differences versus disorders.

It has been suggested that a way to circumvent certain biases in traditional assessment is to assess a CLD child in their primary language (Durgunoğlu, 2002; Elbro et al., 2012). Early static reading assessments that measure phonemic awareness, sight words, and letter names and letter sounds in a child's native language may provide more accurate and informative data because the child is able to perform to their complete capacity, which would be otherwise impossible in a second language that has yet to be mastered. However, commonly used assessments are not always available in a child's primary language, making it difficult to truly evaluate that child's ability (Geva et al., 2000). Additionally, many tests that are available in minority languages tend to be simply translated versions of their English counterparts that fail to consider specific information regarding the language (Peña et al., 2015; Restrepo & Silverman, 2001). Conversely, even popular tests with a Spanish version that have been developed after considering specific language considerations may not prove to be effective because of lack of supporting evidence and research for their specific use. For example, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) developed a Spanish version called Indicadores Dinámicos del Éxito en la Lectura (IDEL) that was published over 10 years ago. The IDEL measures first sound fluency (phonemic awareness) and letter naming fluency in Spanish. However, the developers only published one technical report regarding its efficiency, in comparison to the 20 articles available for the support of their primary test, DIBELS (University of Oregon, Center on Teaching and Learning, 2019). Furthermore, the report on IDEL states that the estimates for instructional recommendations derived from IDEL may not be generalizable because their sample only consisted of 15% of the original students' across-year data (Baker et
al., 2007). Moreover, the report does not report the predictive validity or sensitivity of the IDEL test, both of which are important psychometric measures for choosing a trustworthy assessment. Thus, this assessment lacks supportive evidence to indicate its use clinically. More research is warranted before such a determination can be made.

Another limitation of traditional assessments is that they are normed on populations that do not adequately represent minority children which precludes reliable results when used on CLD children. Some developers have sought to mitigate this limitation by renorming their tests with little to no change in the actual test material. This was the case with the *Woodcock Reading Mastery Test - Revised/ Normative Update* (WRMT-R/NU), a popular test used to evaluate reading in practice and research. The original version of this test was normed on a sample designed to match the demographics of the 1980 census. However, the 1980 census was limited to "white" or "non-white" indicators of ethnicity (Rose, 1999). In contrast, the updated norms were based on a sample designed to coincide with the 1990 Census, which had more information regarding minority sub-groups (e.g., White, Black, Hispanic, and Other). However, simply including a more diverse population in a sample may do nothing more than decrease the mean distribution of the normative sample (Laing & Kamhi, 2003). A group of researchers evaluated the new WRMT- R/NU to see if such an effect had taken place. They tested a group of 899 monolingual, English participants using the same test and then compared the original norms against the new norms. Using the original WRMT-R norms, 534 (59.4%) were classified as having reading disabilities. Comparatively, 357 (39.7%) were classified as reading disabled using the new norms (Pae et al., 2005). These results indicate that renorming a test can significantly change the classification of children with a disorder, which does not necessarily mean the results are more accurate. Moreover, shifted results from a CLD population may simply
change the classification of the children to be within normal age limits, but still below the mean. This could mean that some may be disqualified from receiving services whereas previous norms would have qualified them for the same services. The fact still remains that these children may still be misclassified despite their "new" classification on a test whose material was not made to match their needs. The possibility of over- and under-identification is not eliminated simply by including more diverse individuals in a normative sample for a test that was not originally designed with the CLD population in mind (Laing & Kamhi, 2003).

Another crucial limitation of current assessments is that many traditional tests are static, meaning the results of an assessment only provide information about the child at a single point in time. Inclusively, the majority of research has only utilized tests that possess static qualities, even when assessing CLD children. Static measures can provide valuable information about what a child currently can do. However, with very young children or with children who have had limited reading instruction, the construct of reading cannot be assessed because they have not yet received sufficient instruction to learn to read. Because of this, peripheral or foundational static measures such as phonemic awareness and letter identification are assessed. Unfortunately, static assessments have significant deficits in predicting future achievement in reading due to "floor effects" when assessing young children (Cho et al., 2017; Fuchs et al., 2011). Floor effects can significantly skew the data making accurate assessment improbable. Floor effects have especially been observed in the results of CLD populations (Petersen et al., 2016). Furthermore, static assessments only reflect a child's current level of performance which entails that all decisions regarding that child are based on results from a one-time administration, which can lead to incomplete, erroneous conclusions (Durgunoğlu, 2002; Petersen et al., 2016).
Additionally, there are important concerns regarding the psychometric properties of static measures. Namely, many static measures have not been proven to meet acceptable standards of classification accuracy and predictive validity, especially in CLD subgroups (Petersen & Gillam, 2015). These psychometric standards are imperative to the accurate identification of reading disabilities. For example, Petersen and Gillam (2015) found that the DIBELS static measures classified almost every participant (95%) in their study of bilingual students as at risk. In another study, Petersen et al. (2016), found that the DIBELS test was again shown to yield unsatisfactory results. The statistical analysis revealed that while the DIBELS test had good sensitivity (87%-88%) of two Hispanic subgroups, the specificity was far from acceptable (33%-36%; Petersen et al., 2016). These data convey the need to develop tests that are better equipped to accurately identify CLD children with and without reading disabilities.

**Dynamic Assessment**

Because of the aforementioned reasons, there is a continuous effort being made to design tests that accurately identify CLD children that are at-risk and in need of intervention (Lidz & Peña, 1996). One such alternative to traditional, static assessment that has demonstrated promising results is dynamic assessment (DA). The concept of DA was derived from the ideology of Vygotsky and his principle of a *zone of proximal development* (ZPD), which is the gap between what a child is capable of doing independently and what they can do with the support and assistance from a competent adult (Laing & Kamhi, 2003; Lidz, 1995; Peña, 2000). Vygotsky himself was very critical of traditional assessments that evaluated children based on the observations of independent performance without taking into consideration untapped potential. The ZPD promotes the evaluation of a child's ability by engaging and interacting with the child through assisting them to realize latent potential. This idea of interaction is an essential,
defining characteristic of DA because the focus of the assessment turns from not only the product of a task being performed, but the learning behavior that represents underlying cognitive development (Fuchs et al., 2011; Elbro et al., 2012; Lidz, 1995; Lidz & Peña, 1996; Petersen & Gillam, 2015). By analyzing a child's learning behavior, an examiner is able to capture a more authentic picture of the child's learning potential, providing insight into whether a disability exists because of a difference in learning experience or a disorder in developmental processes. Additionally, DA has the ability to mitigate floor effects because the construct being measured is related to learning potential as opposed to current knowledge. Thus, even when students have limited Spanish reading instruction, a DA will not result in floor effects if those students do not have a reading disorder.

A popular form of DA is the test-teach-retest approach. The teaching period allows a clinician to probe a child's modifiability or responsiveness to the interaction. Researchers have asserted that the responsiveness of a child, or lack thereof, during intervention can indicate whether or not a disability is present (Fuchs et al., 2011). For example, Lidz and Peña (1996) conducted a case study of two Hispanic children who were participants in a larger study. The two children evaluated had scored very similarly on a traditional, standard test of expressive vocabulary. However, modifiability ratings taken during the intervention period suggested that these two children were not similar at all. In fact, one child proved to be highly responsive, needed only minimal examiner effort, and had a high degree of transfer in skills. On the other hand, the second child was only slightly responsive, required extreme effort from the examiner, and had medium transfer. The posttest scores of these children showed that while the first child improved by over two standard deviations, the second child showed very little improvement. This DA allowed researchers to determine that the first child was not in need of additional
services while the other child required further attention. If the children had been evaluated solely by their pretest scores on the static assessment, both children might have been diagnosed with a disorder when only one truly required the resources.

The study cited above not only lends evidence to the argument for examining a child's modifiability to indicate disorder, but it also renders support for the claim that DA can help reduce biases against CLD children. Other researchers have purported the same claim that examining a child's responsiveness and learning processes, such as is accomplished in DA, can significantly limit biases in language assessment and produce more reliable diagnoses of disordered children, particularly those who are CLD (Bridges & Catts, 2011; Peña, 2000; Laing & Kamhi, 2003). A few reasons why this can happen is because DA can assist in limiting floor effects that occur when evaluating low scoring students as well as CLD individuals. By limiting floor effects, DA is capable of obtaining better predictive validity beyond that of traditional tests, which is important in identifying those with reading difficulties (Bridges & Catts, 2011; Elbro et al., 2012). Furthermore, DA may also have stronger predictive validity because the actual construct of reading can be measured, and an exclusive reliance on peripheral or foundational measures such as phonemic awareness and letter naming is not necessary. For example, a study conducted regarding the ability of DA of reading to mitigate cultural and linguistic biases reported important conclusions. Petersen and Gillam (2015) evaluated the ability of a DA of reading (decoding) to predict future reading disability in Latino children. Using a sample of 63 Hispanic children at-risk for language impairment, the researchers compared the results of a DA with a traditional, static assessment, DIBELS. The results indicated that DIBELS identified 95% of the entire sample as being at-risk for a reading disability. The authors determined that such a high percentage negated the need to calculate sensitivity and specificity because a measure that
classifies almost all participants as at-risk will undoubtedly yield near-perfect sensitivity with varying imperfect specificity (Petersen & Gillam, 2015). In contrast, the DA had a sensitivity that ranged from 86% to 100% for sensitivity (correctly identify those with disorder) and 80% to 88% specificity (correctly identify those without disorder). This study supports the claim that DA yields better classification accuracy for CLD children than traditional assessments which can be used in future research.

Although DA has been shown to provide better classification accuracy for CLD children than traditional assessments, more research is required to develop tests that can provide comparable classification accuracy to that which we see in tests developed to evaluate mainstream students. Petersen et al. (2016) conducted a study evaluating the decoding skills of a sample of 600 kindergartners from 14 elementary schools in a large urban school district in Utah. They administered both a DA of reading as well as a static assessment, DIBELS, which measured traditional phonemic awareness and letter naming skills. Results proved to be similar to those mentioned in the previous study, namely, the static assessment exhibited significant floor effects, even more so with the Hispanic subgroup. While posttest and modifiability sections of the DA of reading yielded sensitivity and specificity ranging between 81% to 100% for all children, the results of just the Hispanic subgroup had a specificity measure which dropped down to between 70% and 76%. The Hispanic specificity was below the examiner's expectations of 80% for all measures. Nevertheless, the DA's scores were significantly higher than the static assessment for the sample overall as well as for the Hispanic subgroup, being nearly 40% higher than the Hispanic subgroup's specificity measure (33% -36%). The authors concluded that changes to the DA itself could perhaps yield better sensitivity and specificity measures, especially for CLD children (Petersen et al., 2016). One such change that could accomplish this
essential objective is to assess CLD children using a DA of decoding in their primary language in conjunction with static measures.

Several studies have been carried out regarding the evaluation of reading disabilities using a DA of decoding. However, each has had a different research purpose. Some of the studies focused on the benefits of a DA of decoding in improving the Response to Intervention (RTI) model of treatment (Cho et al., 2014; Fuchs et al., 2011). On the other hand, Gellert and Elbro (2018) analyzed the use of DA of decoding in assessing the predictive validity of predicting reading disabilities before and after the onset of reading instruction. They asserted that a DA of decoding could specifically provide more validity than a static assessment of decoding as well as contribute to early prediction measures. Furthermore, dynamic measures of decoding have shown their competence in predicting future reading abilities more effectively than dynamic tests that administer other tasks such as phonological awareness tasks. This can be concluded because decoding reflects the construct of reading more closely than phonological awareness tasks (Petersen et al., 2018). For example, Gellert and Elbro (2017) found that their DA of phonological awareness measures only accounted for predictive variance up through the middle of first grade, limiting its clinical utility in identifying older students as reading disabled. However, a different study that used a DA of decoding found that this DA accounted for variance from kindergarten to fifth grade (Petersen et al., 2018). The contrast between these two studies clearly illustrates the potential for a DA of decoding to accurately identify children with reading disabilities across a greater span of time.

Despite evidence of the clinical utility of a DA of decoding, there is limited research regarding the implications of a DA of decoding in assessing the CLD population specifically for reading disabilities (Petersen et al., 2016; Petersen et al., 2018; Petersen & Gillam, 2015).
Moreover, the majority of research that has been done on this topic has consisted of DAs administered solely to samples of children in kindergarten and first grade. No research has been done to assess the decoding abilities of CLD children in grades higher than the first grade using DA. Additionally, all these studies involving CLD children have been conducted only in English. This is likely due to the limited number of tests available in Spanish, or any other minority language. Given that there are very few assessments available in the Spanish language in general, it is not surprising that a DA aimed at evaluating decoding abilities in Spanish does not exist. Therefore, the purpose of this study is to a) develop a Spanish DA of decoding and traditional Spanish static reading measures that can be administered to Hispanic children and b) examine the preliminary psychometric qualities of the test by administering the test to Spanish-speaking students who have varying levels of Spanish reading proficiency, which will offer information on the extent to which the test is sensitive to a broad range of variance in reading ability. By developing an assessment in the Spanish language, much needed information regarding the literacy development and patterns of difficulties seen in CLD children can be gathered and utilized for future research involving language minorities. It is hypothesized that typically developing children with normal reading abilities will perform at or near ceiling for all subsections of the test, and that typically developing children who have limited Spanish reading ability will perform poorly. It is also hypothesized that typically developing children who do not have a reading disorder and who are fluent Spanish readers will perform well on the pretest, posttest, and modifiability sections of a DA of decoding. Students who are typically developing yet who are not fluent Spanish readers will perform poorly on a DA pretest yet will have good modifiability and strong posttest performance. Children who are receiving special education services will perform poorly on the pretest and modifiability sections of a DA. Because the
Spanish DDM is a criterion referenced test, it is anticipated that typical children will do well. Moreover, we expect the ceiling effects to be moderated and correlated with Spanish-speaking proficiency. This thesis project includes two phases: first, the development of a Spanish literacy screener and second, a preliminary investigation of validity and reliability. The research aims and research questions for each phase were as follows:

**Phase One**

1. Develop a Spanish static and dynamic criterion-referenced screener for emergent literacy and reading for preschool through early elementary school students.

**Phase Two**

1. Do the dynamic and traditional static measures of the Spanish DDM reveal a range of performance levels for bilingual English-Spanish speaking students who have a range in Spanish reading ability?
2. What is the fidelity of administration and the inter-rater reliability of the dynamic and traditional static measures of the Spanish DDM for bilingual English-Spanish speaking students?

**Method**

**Phase One: Test Development**

The framework for the development of the Spanish literacy screener was influenced by the original English DDM, which is a subtest of the CUBED reading assessment (Petersen & Spencer, 2016). The examiner’s and technical manuals for the English DDM were used as a guide for the research and development of the test into the Spanish language. The assessment developed for the purpose of this study is known as the Spanish Dynamic Decoding Measure (S-DDM). It is important to highlight that the S-DDM is not simply a translated version of the
English DDM, rather, the S-DDM was developed using information regarding developmental norms of Spanish-speaking individuals as well as resources that delineate Spanish linguistic structure such as phonology and orthography. Data regarding high-frequency words and frequent syllable patterns in Spanish were also used.

Specific resources included a Spanish phonology fact sheet compiled by the American Speech Language and Hearing Association (ASHA; n.d.). This fact sheet contained a list of the Spanish phonemic inventory and presented multiple facts regarding Spanish phonology (e.g., consonant clusters found in Spanish, consonant possibilities for the word final position). This fact sheet also included Spanish developmental norms for phonemic acquisition. A number of commercially and publicly available lists of Spanish sight words were also used to determine a consensus on the most frequently used words in the Spanish language (Cornejo, 1972; Grover, 2012; Núñez, 2016). An article written by Ford and Palacios (2015) contained valuable information concerning the order in which Spanish orthography is taught to children in schools as well as additional developmental information regarding special rules for silent letters (e.g., /h/ in all contexts and /u/ in certain contexts). Lastly, a dual language immersion program called “Estrellita” was also consulted for information regarding certain patterns in literacy instruction (Estrellita, n.d.).

The S-DDM was designed to contain three sections that screen for essential elements in literacy and reading abilities, including decoding, word identification, and phonemic awareness. Using all of the materials listed above and a review of current literature, a Spanish equivalent for each of these sections was designed with the purpose of creating a Spanish static and DA to screen students from preschool through early elementary grade levels.
Phase Two: Preliminary Investigation

Participants

The Brigham Young University Institutional Review Board approved this study. Participants for this study were selected from a larger pool of students participating in a study on developmental language disorder. Because we were looking for a range of performance on the DDM subtests, we included students who were receiving special education services under specific learning disability with related language services. Because no prior research has included students in second grade or higher in DA research, we purposefully selected second through fourth-grade students. Eight second through fourth-grade Spanish-speaking students were selected to participate in this study. Table 1 provides descriptive information about each participant, including information on bilingual English/Spanish status, ethnicity/race, gender, grade, and whether the child has an Individualized Education Program (IEP) for special education services. Table 2 provides the language testing results from the English tests for each child.

The children’s ability to learn language was assessed using a DA. The DA posttest scores and modifiability scores are listed in Table 2. In order to be identified as language disordered, the child had to have an active IEP, perform poorly on a language assessment, and a non-word repetition task. The language assessment administered to participants was the Narrative Language Measure (NLM) from the CUBED assessment. Furthermore, a language sample was collected in both English and Spanish for each student to be able to classify their Spanish language ability. Reading performance in Spanish was characterized by performance on the experimental reading test used in this study.
Table 1

Demographic Information for Participants

<table>
<thead>
<tr>
<th></th>
<th>Percent of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2 (25.0%)</td>
</tr>
<tr>
<td>Male</td>
<td>6 (75.0%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (100%)</td>
</tr>
<tr>
<td><strong>Home Language</strong></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>English and Spanish</td>
<td>7 (87.5%)</td>
</tr>
<tr>
<td><strong>Special Education</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 (25.0%)</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
</tr>
<tr>
<td>Second</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>Third</td>
<td>6 (75.0%)</td>
</tr>
<tr>
<td>Fourth</td>
<td>1 (12.5%)</td>
</tr>
</tbody>
</table>

*Note. Special education with related language services was determined based on an active individualized education program for specific learning disability and for language. *Significant difference $p < .05.$*
Table 2

*Language Testing in English Results*

<table>
<thead>
<tr>
<th></th>
<th>NLM in English</th>
<th>DA Posttest</th>
<th>DA Total</th>
<th>Grade</th>
<th>Special Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>10</td>
<td>14</td>
<td>21</td>
<td>3rd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 2</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>3rd</td>
<td>Y</td>
</tr>
<tr>
<td>Participant 3</td>
<td>18</td>
<td>18</td>
<td>23</td>
<td>3rd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 4</td>
<td>8</td>
<td>9</td>
<td>19</td>
<td>3rd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 5</td>
<td>6</td>
<td>10</td>
<td>11</td>
<td>3rd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 6</td>
<td>21</td>
<td>21</td>
<td>19</td>
<td>2nd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 7</td>
<td>4</td>
<td>7</td>
<td>13</td>
<td>3rd</td>
<td>N</td>
</tr>
<tr>
<td>Participant 8</td>
<td>13</td>
<td>10</td>
<td>16</td>
<td>4th</td>
<td>Y</td>
</tr>
</tbody>
</table>

*Note.* Test scores are raw scores. NLM = Narrative Language Measure. DA = DYMOND Dynamic Assessment of language.

**Procedures**

An initial form of the S-DDM was developed for this study with the purpose of gathering preliminary psychometric data. The S-DDM was administered to eight bilingual, Spanish-speaking children. These children were selected from a larger pool of students at a Utah elementary school who were receiving language intervention for a separate study. Participation for the current study was determined by the students’ bilingual Spanish-English status. Testing
was performed by two bilingual undergraduate research assistants with near-native fluency. The research assistants received a 90-minute training on the administration procedures for the S-DDM. The training was conducted by one of the principal authors of the original English DDM and the main author of this paper who developed the DDM into the Spanish language.

All the testing took place within a three-week period. The examiners pulled the students from their classrooms and tested each student individually in a separate room in the school. The order of testing included first, the decoding DA subsection, followed by the phonemic awareness static assessment subsection, and then the word identification static assessment subsection. The decoding subsection took an average of four minutes to administer, the phonemic awareness subsection took an average of six minutes to administer, and the word identification section took an average of eight minutes to administer. On average, the total time for administering the S-DDM was 18 minutes. All testing sessions were audio recorded with the purpose of performing an inter-rater reliability and fidelity check.

Results

Phase One: Test Development

The resulting product for each subsection of the S-DDM will be discussed including how each decision in the development process was reached.

Decoding

The purpose of the decoding subtest of the S-DDM was to evaluate the child’s knowledge of phonemes and the phonological rules that govern the use of their native language which aid them in successfully decoding words. The decoding task consists of three targets, which increase in difficulty. Each target consists of four nonsense words that the child is asked to read. If a child is able to complete the first target by reading two of the four words correctly on the pretest, then
they are prompted to move on to the next target. If they are unable to do so, then the rest of the first target is administered, and the subtest is ended. This process continues until the instructions direct the examiner to discontinue the subtest and determine the score. All nonsense words included in the three decoding targets were designed using information regarding which phonemes were typically taught earlier as well as the order of development of phonemes in the Spanish language. Each target in the decoding subtest contains a pretest and posttest which measures correct sounds worth 16 total points and correct words worth a total of four point. There is also a portion in each target where the examiner is to indicate the amount of support required to teach the child (e.g., high, moderate, low). A lower support rating indicates better modifiability.

The first target of the English DDM decoding subtest assesses consonant-vowel-consonant (CVC) nonsense words. However, single syllable words are uncommon in the Spanish language. According to Ford and Palacios (2015), the most commonly occurring syllable pattern is consonant-vowel (CV). Therefore, it was concluded that a CV/CV pattern would be suitable to include as the first and least difficult target. The four nonsense words in this level begin with letters selected from a list of earlier developing Spanish phonemes (e.g. /m/, /n/, /b/, /p/, /s/, /l/, /d/, /t/, and /f/). The letters that compose this list are generally taught first because they blend easily with vowels and also allow children to more readily distinguish between vowels (Ford & Palacios, 2015). Twelve nonsense words were created following these principles, from which, four were chosen for the final test. The ending phonemes of each of the four nonsense words are the same, with “asa” structure (i.e., lasa), resulting in an onset-rime pattern. This structure was chosen specifically with the vowel /a/ due to it being an earlier developing vowel. The remaining eight nonsense words will be used in future alternate forms of the test.
The second target of the decoding subtest is composed of consonant-consonant-vowel-consonant-vowel (CCV/CV) words, with the purpose of evaluating the child’s ability to decode consonant clusters at the beginning of words. The English DDM also evaluates consonant clusters for the second target using a CCVCV structure. Consonant clusters contained in the first syllable were chosen as the second target because the ability to decode consonant clusters is a skill that develops later on in Spanish but is important in being able to decode more complex structures. Additionally, CCV is one of the most common syllable structures in Spanish. In the Spanish language, consonant clusters found in a single syllable only occur with the phonemes /l/ and /r/. A complete list of Spanish consonant clusters compiled by ASHA (n.d.) was used to create 16 possible nonsense words. Four nonsense words were randomly selected to be in the final version of the test. The remaining 12 nonsense words may be utilized in future alternate forms of the S-DDM.

The third target of the decoding subtest was designed to be the most complex. The English DDM’s third target tests participants on the silent /e/ rule. This is not a rule in the Spanish phonetic system. However, the Spanish language has special rules that are similarly complex. Thus, a complex rule in line with a typical teaching sequence specific to the Spanish language was chosen. Originally, three possibilities were evaluated including the silent /h/ at the beginning of words, the silent /u/ when preceded by a “g” or a “q,” and the sound changes to “c” and “g” when followed by /a/, /o/ and /u/ versus /e/ and /i/. This last rule entails that the “c” makes a /k/ sound and the “g” makes a /g/ sound when followed by an /a/, /o/ or /u/. Whereas, when followed by an /e/ or an /i/, the “c” makes a /s/ sound and the “g” makes a /x/ sound (an aspirated “h” sound). A list of nonsense words was created for each rule. After deliberation about the complexity of each rule, the “c” and “g” rule was selected to make up the third target for the
final test because it is generally taught later on to kids and was determined to be the best suited for the most difficult level of evaluating complex decoding abilities.

**Phonemic Awareness**

All the words for the three phonemic awareness targets were chosen after a list of high frequency words in Spanish was compiled. This list was created using five high frequency word lists: Cornejo (1972), Grover (2012), Núñez (2016), a Spanish Language Arts word list for 1st grade, and one commercially available list. The overall list compiled using these five sources yielded 199 words. From this list, 109 words met the criteria of appearing on at least two of the five lists. The 109 words were separated into distinct lists depending on how many sounds each word contained (e.g., two, three, four, and five+). Each of these lists were entered into an automatic randomizing device which numbered the words in each list into a completely novel order. Afterwards, a random number generator was used to select words from each list to be assigned to the specific targets. The words for the last target were randomly selected from a list of words that contained only continuous sounds. Adjustments to the final lists were made afterwards to account for words that had the same beginning sounds for each target to help increase variety.

The phonemic awareness subtest of the S-DDM was designed to assess a child’s ability to be able to recognize and separate the phonemes that make up specific words. This section was made up of three targets that decreased in difficulty. In alignment with the English Phonemic Awareness subtest, it was determined that if the child does not reach a predetermined benchmark during the first target, then the next target is administered until the child can complete a target or the test directs the evaluator to discontinue the test. The first target consists of the most frequently occurring ten words from several word lists: two words with two phonemes, four
words with three phonemes, and four words with four phonemes. This target is a phoneme segmentation task in which the child, upon hearing a word, is required to segment the word into the sounds that make up the word. The second target, which is completed if the first target proves to be too difficult, consists of ten words ranging from two to five phonemes. This target requires the child to provide only the first sound of the word after being told a word. The last target has five words that consist only of letters that are considered to be continuous sounds. This target evaluates a child’s phoneme blending abilities. It requires that the evaluator produce the target words slowly, elongating each sound, and then ask that the student repeat the word back normally, with no elongation. The first target is worth 32 points, the second target is worth 20 points, and the third target is worth 10 points.

**Word Identification**

The word identification subtest of the S-DDM was designed to assess a child’s ability to identify sight words, letter names, and letter sounds easily. Identifying sight words and letters is an important reading skill because it helps increase the speed and fluency with which a person decodes simple, common words. The word identification subtest contains three targets that decrease in difficulty. If a child is unable to meet a certain criterion within a target, the test instructs the examiner to administer the next easiest target. This process continues until the child can successfully complete a target or until the test instructs the examiner to discontinue the test.

The target of the first subsection in the word identification section is a sight word task that contains 54 words. To determine which 54 words should be included on the screener, a list of Spanish sight words was compiled using three commercial Spanish sight word lists. One list contained 377 words, another had 299 words, and the third had 246 words. The three lists were compared to determine which words were common among the three lists. A total of 102 words
were common among the three lists. The words were organized alphabetically and then reduced to 54 words by eliminating words that started with letters whose lists contained the most words. This was done in order to allow the words to begin with a variety of letters. The final words for this task were organized by placing words beginning with earlier taught letters (m, n, b, p, s, l, d, f, and t) near the beginning followed by words starting with the remaining letters taught later on (ll, h, c, v, y, g, and q).

The second target is a letter sound identifying task. This section includes 30 Spanish sounds, each repeated twice, once displayed as an upper-case letter and once as a lower-case letter. The list includes all 27 letters in the Spanish alphabet, as well as other sounds such as “ll,” “rr,” and “ch.” The 60 letter symbols were ordered such that sounds taught earlier on were placed near the beginning, followed by sounds taught later on. According to Ford and Palacios (2015), vowels are taught before consonants, for this reason, the five Spanish vowels were placed in a random order at the beginning of the list. Information for the order of the consonants was taken from Ford and Palacios and a dual immersion program called Estrellita (n.d.). Letters that were common among the two lists as taught earlier included m, p, l, s, t, d, and n. These letters were randomly organized following the vowels. Ford and Palacio’s and Estrellita’s materials also agreed that z, k, x, and w are all letters taught later on, which is why these letters were randomly organized at the end of the list. The order of the remainder of the letters in the middle was determined using a weekly first grade curriculum found in the Estrellita program which provided a weekly sequence of when certain letters and sounds were taught. The only letter sounds that were unmentioned by either list were “q” and “rr.” In summary, using the guidelines found in the research, the organization of the letters suggested the following sequence:

O A I U E/ M P S L T/ D N R C F/ B J G C H Ñ/ V LL Z H/ Y X K W/RR Q
The above sequence provided the parameters by which we ordered the letters for both the second and third targets of the word identification subtest. The final target of the word identification subtest is a letter naming task, composed of all 27 Spanish alphabet letters. These letters are also presented twice in the list, once as a lower-case letter and once as an upper-case letter, creating a total of 54 letters on the target list. The vowels remained at the beginning of the list, followed by consonants that follow the same parameters as described in the second target. The letters were, however, randomized differently so that the lists in target two and target three differ in letter order. The first target of the word identification subtest is worth 54 points, the second target is worth 60 points, and the third target is worth 54 points.

**Phase Two: Preliminary Investigation**

To examine whether the dynamic and traditional static measures of the S-DDM exhibit variance for second through fourth grade bilingual English-Spanish speaking students, descriptive data is provided, including demographic information about the participants and means and standard deviations, range, and skew and kurtosis of the DA and static subtests in Tables 3, 4, and 5, and in Figures 1, 2, and 3. Table 3 lists the participant’s results from the decoding subtest. Scores range from 0 to 16 on the CVC sounds, CCVCV sounds, C-Rule sounds, and posttest sounds. The scores range from 0 to 4 on the CVC words, CCVCV words, C-Rule words, and posttest words. Modifiability is scored with a range of 1 to 3. The results from the phonemic awareness subtest are reported in Table 4. The phoneme segmentation target score ranges from 0 to 32, the first sounds target score ranges from 0 to 20, and the phoneme blending target score ranges from 0 to 10. The results from the word identification subtest are reported in Table 5. The irregular words target has a score range from 0 to 54, the letter sounds target score ranges from 0 to 60, and the letter names target score ranges from 0 to 54.
Table 3

*Decoding Subtest Performance Results*

<table>
<thead>
<tr>
<th></th>
<th>CVC Sounds</th>
<th>CVC Words</th>
<th>CCVCV Sounds</th>
<th>CCVCV Words</th>
<th>C-Rule Sounds</th>
<th>C-Rule Words</th>
<th>Mod.</th>
<th>Post Sounds</th>
<th>Post Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>08.00</td>
<td>08.00</td>
<td>07.00</td>
<td>07.00</td>
<td>04.00</td>
<td>04.00</td>
<td>08.00</td>
<td>07.00</td>
<td>07.00</td>
</tr>
<tr>
<td>Mean</td>
<td>14.00</td>
<td>03.25</td>
<td>13.14</td>
<td>02.57</td>
<td>13.75</td>
<td>02.25</td>
<td>01.25</td>
<td>14.86</td>
<td>02.86</td>
</tr>
<tr>
<td>Mode</td>
<td>16.00</td>
<td>04.00</td>
<td>15.00*</td>
<td>01.00*</td>
<td>14.00</td>
<td>02.00</td>
<td>01.00</td>
<td>14.00</td>
<td>02.00</td>
</tr>
<tr>
<td>SD.</td>
<td>04.28</td>
<td>01.49</td>
<td>04.26</td>
<td>01.27</td>
<td>02.06</td>
<td>01.26</td>
<td>00.46</td>
<td>01.07</td>
<td>01.07</td>
</tr>
<tr>
<td>Skew.</td>
<td>-02.34</td>
<td>-01.95</td>
<td>-02.13</td>
<td>-00.22</td>
<td>-00.71</td>
<td>01.13</td>
<td>01.44</td>
<td>00.37</td>
<td>00.37</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>05.47</td>
<td>03.21</td>
<td>04.73</td>
<td>-01.72</td>
<td>01.79</td>
<td>02.23</td>
<td>00.00</td>
<td>-2.80</td>
<td>-2.80</td>
</tr>
<tr>
<td>Range</td>
<td>12.00</td>
<td>04.00</td>
<td>12.00</td>
<td>03.00</td>
<td>05.00</td>
<td>03.00</td>
<td>01.00</td>
<td>02.00</td>
<td>02.00</td>
</tr>
<tr>
<td>Min.</td>
<td>04.00</td>
<td>00.00</td>
<td>04.00</td>
<td>01.00</td>
<td>11.00</td>
<td>01.00</td>
<td>01.00</td>
<td>14.00</td>
<td>02.00</td>
</tr>
<tr>
<td>Max.</td>
<td>16.00</td>
<td>04.00</td>
<td>16.00</td>
<td>04.00</td>
<td>16.00</td>
<td>04.00</td>
<td>02.00</td>
<td>16.00</td>
<td>04.00</td>
</tr>
</tbody>
</table>

Note. *= more than one mode exists. SD = Standard Deviation. Mod = Modifiability.

Modifiability scores ranged from 1 to 3, with 1 representing good modifiability and 2 or 3 representing weaker modifiability. Test scores are raw scores.
### Table 4

**Phonemic Awareness Subtest Performance Results**

<table>
<thead>
<tr>
<th>Statistics on Phonemic Awareness Targets</th>
<th>Phoneme Segmentation</th>
<th>First Sounds</th>
<th>Phoneme Blending</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>08.00</td>
<td>08.00</td>
<td>08.00</td>
</tr>
<tr>
<td>Mean</td>
<td>20.25</td>
<td>17.25</td>
<td>09.38</td>
</tr>
<tr>
<td>Mode</td>
<td>32.00</td>
<td>20.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>12.40</td>
<td>04.27</td>
<td>00.92</td>
</tr>
<tr>
<td>Skewness</td>
<td>-00.96</td>
<td>-01.82</td>
<td>-00.99</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-00.56</td>
<td>03.02</td>
<td>-01.04</td>
</tr>
<tr>
<td>Range</td>
<td>32.00</td>
<td>12.00</td>
<td>02.00</td>
</tr>
<tr>
<td>Min.</td>
<td>00.00</td>
<td>08.00</td>
<td>08.00</td>
</tr>
<tr>
<td>Max.</td>
<td>32.00</td>
<td>10.00</td>
<td>10.00</td>
</tr>
</tbody>
</table>

*Note.* *= more than one mode exists. Test scores are raw scores.*
<table>
<thead>
<tr>
<th></th>
<th>Sight Words</th>
<th>Letter Sounds</th>
<th>Letter Names</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>08.00</td>
<td>07.00</td>
<td>08.00</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>34.00</td>
<td>31.00</td>
<td>20.13</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>04.00*</td>
<td>02.00*</td>
<td>01.00*</td>
</tr>
<tr>
<td><strong>Std. Dev.</strong></td>
<td>18.66</td>
<td>23.04</td>
<td>18.57</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-00.67</td>
<td>-0.149</td>
<td>00.16</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>-01.32</td>
<td>-01.89</td>
<td>-02.35</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>48.00</td>
<td>57.00</td>
<td>42.00</td>
</tr>
<tr>
<td><strong>Min.</strong></td>
<td>04.00</td>
<td>02.00</td>
<td>01.00</td>
</tr>
<tr>
<td><strong>Max.</strong></td>
<td>52.00</td>
<td>59.00</td>
<td>43.00</td>
</tr>
</tbody>
</table>

*Note.* *= more than one mode exists. Test scores are raw scores.*
Figure 1

Decoding Subtest Histograms with the Normal Curve Delineated
As hypothesized, based on the varying reading abilities of the participants, there was a wide range of performance on the pretest sections of the DA decoding subtest, and performance at or near the ceiling on the posttest sounds sections. Modifiability scores were strong, as predicted, for the students who were not receiving special education services. The two students receiving special education both received a score of 2 for the modifiability rating, indicating that their modifiability was weaker. The first two targets of the S-DDM phonemic awareness portion exhibited a wide range of performance as well. All participants performed well on the final target of the phonemic awareness subtest, phoneme blending, and therefore this target did not exhibit
the same level of variance as the other targets. All three targets from the word identification subtest exhibited a wide degree of variance among the participants.

**Fidelity**

To examine the fidelity of administration and the inter-rater reliability of the dynamic and traditional static measures of the S-DDM for second through fourth grade for bilingual English-Spanish speaking students, 100% of test administration and scoring was evaluated using a fidelity checklist and rescored by independent examiners.

Two undergraduate research assistants administered the S-DDM to the eight research participants. A fidelity checklist was created to evaluate the degree to which the research assistants were able to follow the administration instructions and protocol. Each of the three subtests of the S-DDM had their own specified set of fidelity questions. The fidelity checklist for the decoding subtest included items such as Did the examiner read the pretest instructions correctly? Was each phase of the teaching section followed correctly? The phonemic awareness subtest was evaluated for how well the script was followed and whether or not the examiner provided the corrective prompts as needed. The word identification subtest was reviewed for similar items as the phonemic awareness subtest as well as if the examiner followed the discontinue rule correctly. Fidelity was evaluated for seven out of the eight children’s data (87.5%). The fidelity check yielded scores ranging from 64% to 88% fidelity, with an average of 78.85% fidelity.

**Inter-Rater Reliability**

Point-to-point inter-rater reliability of scoring for the decoding, phonemic awareness, and word identification subtests of the S-DDM was conducted. One hundred percent of both the decoding and phonemic awareness administrations were evaluated for inter-rater reliability.
However, only 87.5% of the data from the word identification section was evaluated for inter-rater reliability because of technical difficulties with the recording of one test administration. Inter-rater reliability was conducted by the two original examiners, each evaluating the other’s test administrations.

The decoding subtest had inter-rater reliability scores ranging from 80.16% to 98.34%, with an average of 92.24%. The phonemic awareness subtest had inter-rater reliability scores ranging from 85.48% to 100% agreement between examiners, with an average of 94.35%. The word identification subtest had inter-rater reliability ranging from 61.3% to 95.23%, with an average of 87.49%. One outlier in the data significantly affected these results, given that 6 of the 7 data sets were above 87% in reliability.

Discussion

The purpose of this study was to develop a Spanish DA of decoding and traditional Spanish static reading measures that can be administered to Hispanic Spanish-speaking children. This study also set out to examine the preliminary psychometric qualities of the test by administering the screener to Spanish speaking students with varying levels of language ability. The test that resulted from the development phase of this study is called the Spanish Dynamic Decoding Measure (S-DDM). The final version of the test contained three subtests: a decoding measure, a phonemic awareness measure, and a word identification measure. These three subtests evaluate a student’s reading ability and important skills which support children in their reading development. It was hypothesized that the participants who were typically developing and who were strong readers in Spanish would perform at or near ceiling across all three subtests, yet poor readers with and without reading disorder would struggle with the pretest sections. It was also hypothesized that for the DA subtest, typically developing students would
all respond well to the instruction, with good modifiability scores and posttest scores. Students receiving special education services, however, would receive poor pretest and weaker modifiability scores. The administration of the test resulted in a wide range of variance in all measures of the decoding subtest except for the posttest sounds results, in all targets of the phonemic awareness subtest except the phoneme blending target, and all three targets of the word identification subtest. Regarding the two targets that did not exhibit variance (e.g., posttest sounds of decoding and phoneme blending of phonemic awareness), most participants reached the ceiling, as expected. Fidelity and inter-rater reliability were found to be within acceptable measures (Fidelity = 78.85%; Inter-rater reliability = 87.49%).

**Phase One: Test Development**

The development of a combined dynamic and static Spanish reading screener is unique. To our knowledge, there is no developed DA of reading in Spanish. There is also a dearth of research regarding the reading acquisition of Spanish-speaking students, subsequently, this reading screener was developed with the best information available regarding Spanish-language development and literacy acquisition. Thus, by developing this test, we can accumulate more data on student performance and normative and developmental data for Spanish-speaking students. This information will be important in adding to the current pool of research surrounding Spanish-literacy development.

Items should be changed before the distribution of this screener is made available. First, there are prompts provided for the examiner to say when introducing each item of the three targets in the phonemic awareness subtest (e.g., “Dime los sonidos en __”; “Cuál es el primer sonido que escuchas en la palabra___?”; “Digo la palabra lento, tú lo dices rápido”). After reviewing the administrations of this subtest, it was determined that repeating the prompt for
each test item was redundant and halted the administration process. The participants were able to understand the response expectation without the frequent repetition of the instructions. Therefore, the instructions for the prompts should be changed to indicate that the prompts only be used as needed.

Another item that requires an adjustment is the exemplar provided to the student in both the second and third targets of the word identification subsection. The exemplar provided in both of these targets remained the same as those found in the English DDM. However, these exemplars were both vowels. To better provide clear examples of the difference between letter naming and letter sounds in Spanish, consonants should have been used as exemplars. This is because of the nature of vowels in the Spanish language. The letter names and letter sounds of Spanish vowels are the same, and therefore, are not good examples. This likely caused confusion among participants about the desired response for these two targets. Therefore, this element should be adjusted to reflect the needs of the Spanish language when providing clear examples.

The resulting screener is a tool that, if all necessary adjustments are made, may prove to be an asset in evaluating Spanish-speaking students both in the United States and Spanish-speaking countries around the world. There is much to be learned about how we can correctly evaluate Spanish-literacy skills in order to provide the necessary intervention measures to those children in need of services.

**Phase Two: Preliminary Investigation**

The pilot administration of the S-DDM was an important first step in evaluating and analyzing the usefulness of this tool when evaluating Spanish-speaking students. We expected to see most typically developing participants who were strong readers in Spanish to reach the ceiling on most measures. Variance, as expected, was observed on most targets, with both
typically developing students with a range of Spanish reading ability and those receiving special education showing variance in their reading performance. The students receiving special education services had weaker modifiability scores than the typically developing students.

As expected, there was a wide range of variance for pretest measures across the three targets of the decoding subtest. More specifically, six children reached the ceiling for both pretest sounds and pretest words on target 3, the easiest target. Two children demonstrated poor performance on this decoding pretest (CVC words). One of these children had a current IEP for special education services, however, the child that received the lowest score on this target was a child classified as typically developing. There was more variance observed in the more advanced decoding target pretests. Furthermore, most participants scored well on the posttest sounds portion across targets and were able to read about 50% of the words on the posttest.

The phonemic awareness subtest also revealed variance in performance. On the first target, only two participants reached ceiling while two other participants exhibited extreme floor effects (e.g., scores of 0 and 3). The participant who received a score of 0 did have an IEP while the other child was classified as typical. The rest of the children scored in the upper-middle portion (e.g., scores from 19 to 28 out of 32). On the second target, 50% of the participants reached the ceiling while the other participants showed moderate performance. Only one participant’s score significantly deviated from the rest of the participants (8 out of 20). Most participants reached ceiling on the third measure (phoneme blending), which was in line with expectations.

The word identification subtest resulted in the greatest amount of variance. This was not surprising considering the tasks that the participants were asked to perform (e.g., read sight words, identify letter sounds, identify letter names), and given the fact that the participants had a
wide range of Spanish reading ability. On the first target, about 50% the participants performed well while the other 50% struggled to complete the task. The second target resulted in two participants being able to complete the task with adequate success. The other participants did not perform well on this target. The third target of this subtest showed the most interesting results. No participants were able to reach ceiling when asked to identify letter names in Spanish. Fifty percent of the children scored between 31 and 42 on this measure (out of 54 points possible). This was particularly surprising that not even one participant could provide more than 80% of the letter names. The other four participants experienced significant difficulty with this task, with scores ranging from 1 to 6.

The two targets with which most participants struggled were the letter names and letter sounds targets. It is possible that the participants of this study had not received adequate instruction regarding Spanish letter names and sounds despite the fact of being identified as bilingual, Spanish-English speaking students. This factor may have played a large role in the variance observed on these measures. It is also possible that the failure to provide an adequate exemplar impacted the participants’ understanding of the expected response (e.g., a vowel was provided instead of a consonant, causing confusion).

The two participants in this study with IEPs received weaker modifiability ratings on the DA of decoding than the participants who were not receiving special education services. One participant receiving special education services struggled with all measures across all three subtests. However, his scores did not deviate meaningfully from the scores of other typical students with the exception of modifiability. The other participant receiving special education services performed consistently well on all measures, except for modifiability. It is interesting to note that there is a discrepancy between the pretest DA and static reading performance of these
two students receiving special education services, yet modifiability for both students was rated lower.

There were also interesting findings in the fidelity of administration of this test. Despite having received training on administering this test, neither examiner administered the second part of the two-part teaching section of the decoding subtest to any of the participants. It is unclear as to why this was the case. The impact that the exclusion of this component of the subtest had on the performance of the participants is also unknown. One examiner also demonstrated confusion on the administration protocol for the word identification subtest. On each target of the word identification subtest, the examiners are instructed to wait three seconds before intervening and providing the correct answer for the child before moving on to the next item. One examiner did not follow this protocol when administering the screener. The confusion observed among examiners perhaps indicates the need for more and/or clearer instructions and training before administration begins.

Limitations and Future Research

One limitation is the small participant sample size. The participants were a convenience sample who likely do not represent the Spanish-English bilingual population in the United States, let alone the monolingual Spanish-speaking population found throughout the world. The sample of this study only represents a certain demographic located in a specific region of the country. Without a larger sample size from varied locations, it is hard to generalize findings to a larger, broader population. An additional limitation that arises from the sample is that the participants are not native Spanish speakers and they may not have received very much explicit reading instruction in the Spanish language. Future studies should consider using a sample size that contains a larger range of Spanish language ability among participants. Additionally, future
research should include more descriptive information regarding the Spanish language and reading proficiency of all participants in order to draw conclusions about the correlation between Spanish language proficiency, reading ability, and performance on this assessment.

Another recommendation for future research is to include a larger set of items for each subtest that have been carefully chosen with Spanish orthographic and phonotactic probability taken into consideration. After administering these larger sets of words, the items should be evaluated to find out which items work best for the students and if some items work better for students based on Spanish language exposure. Based off the results of this evaluation, the targets discovered to work best should be included in future research as well as the final version of the S-DDM.

The performance of the examiners was another limitation. Given the difficulties following the administration protocol, it was determined that a 90-minute training was insufficient to prepare the examiners in accurately administering this test to participants. Future studies should consider providing additional training or providing a more detailed, clear manual and evaluation of the administration abilities of the examiners before beginning data collection in order to ensure higher levels of both fidelity and reliability.

**Conclusion**

This is the first study to develop and evaluate the psychometric qualities of a dynamic and static Spanish reading assessment tool. This study found that the items developed for this assessment were in line with the current information available regarding the development of Spanish decoding and literacy skills in Spanish speakers. This study also discovered that the targets of each of the three subtests all exhibited a wide range of variance with the exception of the posttest sounds portion of the decoding subtest and the phoneme blending subtest of the
phonemic awareness subtest, indicating that the test is sensitive to varying degrees of Spanish reading proficiency and reading ability. Both of these targets resulted in the majority of participants reaching ceiling or near-ceiling static scores resulting in a low amount of variance, yet the DA pretest revealed greater variance. Participants classified as typically developing did not reach ceiling for all measures. Instead, performance varied across a wide spectrum of scores for most measures, particularly the measures in the word identification portion of the assessment.

Additional research is required in order to determine the effectiveness of this screener in identifying Spanish-speaking students who are at risk for reading difficulties. This screener may be a powerful tool for future research in acquiring much needed information about the development of Spanish reading skills among Spanish-speaking students. This screener may also be useful clinically in enabling clinicians to identify Spanish-speaking students in need of reading remediation services. By developing and evaluating new static and dynamic reading assessments in Spanish, we will be able to fill the gap of tools available for assessing CLD children in order to provide the best quality intervention possible.
References


https://www.asha.org/uploadedFiles/Spanish-Phonemic-Inventory.pdf


https://doi.org/10.1017/S014271640200005X


Cornejo, R. (1972). Spanish high-frequency word list. Southwestern Educational Laboratory.


kindergarten through grade 2. *Annals of Dyslexia*, 68, 126-144.

https://doi.org/10.1007/s11881-018-0159-9


https://doi.org/10.1007/s11881-000-0020-8


http://iteachduallanguage.blogspot.com/2012/10/palabras-de-alta-frecuenciahigh.html


University of Oregon, Center on Teaching and Learning. (2019). *Dynamic Indicators of Basic Early Literacy Skills.* Retrieved from https://dibels.uoregon.edu/assessment/idel

APPENDIX A

Annotated Bibliography


**Objective:** The objective of this study was to determine the validity of different phonological processing abilities (PPAs), namely: phonological awareness (PA), phonological memory (PM), and phonological access to lexical storage (also known as RAN). Researchers also aimed to ascertain whether each of these PPAs were distinguishable from each other as well as distinguishable from general cognitive ability, and how the related to emergent literacy skills in general. All of these research questions were specifically made regarding the Spanish-speaking Hispanic population.

**Method:** Using a randomly selected sample of 147 children between the ages of three and five, researches conducted a battery of tests regarding phonological processing abilities (PPAs). All participants were Spanish speaking, Hispanic preschoolers that came from economically disadvantaged families. The majority of the sample were labeled as nonreaders. Each child was given the DIAL-3 at the beginning of the school year to assess their general cognitive ability. Furthermore, at the end of the school year, the children were tested with the battery of tests designed to assess each phonological processing skill as well as emergent literacy. Phonological awareness was assessed using subtests of blending and elision tasks. Phonological memory was assessed by tasks on
memory for sentences, memory for words, and memory for nonwords. Lastly, RAN was tested by rapid size-naming and rapid-object naming tasks.

Results: The statistical analyses of the results of this study showed some important findings regarding the emergent literacy and phonological processing skills of Spanish-speaking preschoolers. First, it was determined that all three of the phonological processing abilities were distinguishable from general cognitive ability. Additionally, it was concluded that each PPA was best described as separate but correlated and that each PPA was significantly related to the emergent literacy skills of the participants. Finally, another important conclusion drawn was that general cognitive ability only indirectly influenced emergent literacy skills, and therefore was not directly predictive of any emerging abilities. Furthermore, the research findings supported the validity of phonological awareness as a predictor of emergent literacy in the Hispanic population, especially in regards to the ability to distinguish between alphabetic and nonalphabetic test.

Relevance to Current Work: This study provides important results regarding the analysis of literacy skills in Hispanic children. The authors acknowledge the importance of conducting research on minority populations that continue to grow within the United States. Additionally, each of the phonological processing abilities, which are similar to the current study's screener for decoding abilities (namely the phonological awareness subsection), are proposed to be a possible integral part in identifying and instructing children who are at-risk for reading challenges.
Objective: This study was conducted with the goal of determining the usefulness and predictive validity of the Dynamic Screening of Phonological Awareness (DSPA) as a screening tool for identifying children at risk for reading disabilities. This article examines the results from two samples to evaluate multiple research questions. The first sample was used to compare the DSPA versus a static version of the DSPA to determine how much variance was accounted for by the DSPA. Furthermore, the second sample was used to explore the differences between the DSPA and a commonly used screening tool, a screener from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test called Initial Sound Fluency.

Method: As described above, this study consisted of two samples. The first sample (n = 90) included kindergarten children from Iowa. This group of children were administered the Static Deletion Task (SDT) and the DSPA. The second sample (n = 96) consisted of kindergarten children from Kansas. This group of children were also given the DSPA as well as the Initial Sound Fluency (ISF) subtest of DIBELS. Furthermore, both samples were given the Word Identification and the Word Attack subtests from the Woodcock Reading Mastery Tests-Revised/NU. Regarding the first group, the DSPA was given immediately following the SDT. Both the SDT and the DSPA consist of the same four sets of word which increased with complexity in the required task. All tasks consisted of words which the student was required to repeat after having deleted a syllable or phoneme from the initial word (i.e. "cowboy" without "cow"). During the
DSPA, the examiner would use a series of prompts until the child produced the correct response. Now considering the second group which was administered the DSPA and the ISF, the DSPA was the same as was given to the first sample. Furthermore, the ISF is a subtest of the DIBELS test and evaluates a child's ability to identify and produce the initial sound in a word presented to them by the examiner. All the aforementioned tests were administered in September. Finally, the Word Identification and Word Attack subtests were administered to all students in both samples at the end of April or beginning of May. These tasks were both untimed and evaluated a student's abilities to read real words in isolation and non-sense words in isolation, respectively.

**Results:** The results of the predictive validity of the DSPA when compared to the SDT in the first sample showed that the DSPA accounted for a significant amount of variance with both end-of-year tasks, namely 4% for the Word Identification (real word) task and 9% for the Word Attack (non-sense word) task. Overall, logistic analyses showed that the DSPA was a significant predictor for both end-of-year tasks when evaluated alone and in combination with the SDT. However, for the Word Attack measure, statistics showed that the DSPA alone or combined with the SDT was a fair predictor of reading ability while they failed to show acceptable prediction for the Word Identification measure. The results of the second sample determined that both the DSPA and ISF were significant predictors of reading ability. However, the predictions of reading outcomes for both end-of-year measures made by the DSPA test were over and above the predictions of the ISF. Additionally, the DSPA accounted for a significant amount of unique variance for each measure as well (e.g. 5% for Word Attack and 2% for Word Identification). The overall better prediction of the DSPA when compared to the
Word Attack results for both groups is expected given that phonological decoding is more related to phonological awareness than sight word recognition.

Relevance to Current Work: The results of this study support the claim that DA can be an effective and efficient means of improving screenings of phonological awareness to determine reading disabilities. Additionally, the authors claim that they are unsure as to whether a DA would be more effective as a primary measure of phonological awareness or as a supplementary measure. Thus, the current work will add to the pool of research that aims to answer this uncertainty. Furthermore, this study supports the idea that the earlier a child is identified for reading difficulties, the more we can help improve their reading outcome. Lastly, this article states that phonological awareness has been shown to uniquely predict reading outcomes, which supports the design of our decoding screening measure.


https://doi.org/10.1017/S014271640200005X

Objective: The purpose of this study was to determine if the acquisition of basic literacy skills was the same for native English speakers (NS) as well as bilingual children (BL) and children learning English as a second language (ESL). They also sought to determine if a difference existed in which literacy measure was the most effective in predicting reading difficulty for a native speaking child and children who were learning a second language.

Method: A group of 540 native English speakers (NS), 59 bilingual children (BL), and 60 children who began learning English at the beginning of school (ESL) were used
in this study. Researchers administered a variety of tests that examined the children's abilities in the following areas: phonological awareness, alphabet knowledge, naming speed, syntactic awareness, and verbal memory. These tests were administered near the beginning of the kindergarten school year in November and then again at the end of the year in May. Using a variety of statistical measures, the results were analyzed to determine any significant findings.

Results: The results of this study demonstrated that the same literacy skills that were essential for NS participants were also crucial in the English literacy development of BL and ESL children as well, despite their limited proficiency. The study also concluded that although BL and ESL children performed more poorly than NS children, they developed the same skills in a similar manner. Additionally, although no significant difference existed between the three groups and the predictive power of the literacy measures, the most significant predictors for all three groups were alphabetic knowledge and phonological awareness.

Relevance to Current Work: The results of this article support the use of literacy screeners that assess alphabetic knowledge and phonological awareness to evaluate risk of reading difficulty in all languages, such as is used in the current study. This article also upholds the use of early identification to find and treat any children that may be at risk for reading deficits. However, this study fails to acknowledge the fact that by testing BL or ESL children in a language other than their native language, we are unable to appropriately identify those who are at risk for language difficulty because any deficits would be masked by linguistic differences. If we administer screeners in a child's native
language, we will be able to determine whether difficulty is due to difference versus disorder.


**Objective:** The purpose of this test was to assess the effectiveness of a DA of decoding in predicting a child's responsiveness to Tier-2 treatment in a response-to-intervention (RTI) framework. Researchers hoped to determine whether a DA could be used to enhance the predictive validity and classification of student's who would not respond to Tier-2 intervention in order for them to obtain more intensive treatment for reading disability. The researchers also compare the predictive validity of the DA of decoding with two static assessments of decoding to determine any superior qualities of DA over traditional test materials.

**Method:** This study was composed of students from a larger evaluation of RTI success. A final sample of 134 students was analyzed. The students were tutored in small groups receiving Tier 2 intervention for 14 weeks, three days a week, with 45-minute-long sessions. Students were also tested with a battery of tests to determine pre-intervention skills so that this information could be considered. Furthermore, a DA of decoding was administered using a gradual prompt design. Following, two static Das were administered, one time and the other untimed.

**Results:** Results from the statistical analysis concluded that the DA assessment significantly explained small but unique variance in Tier 2 responsiveness, indicating its
usefulness in assessing kids in need of more intense intervention. Another conclusion that was drawn was that the decoding DA contributed to improving the prediction and identification accuracy for children who are unresponsive to Tier 2 beyond what either of the static assessments could provide. However, the researchers did qualify their conclusions stating that their findings were only preliminary and more research is required to corroborate their findings.

Relevance to Current Work: This study provides support for the use of DAs of decoding in identifying children with reading disability. Although this article primarily researched the effectiveness of DA in an RTI framework for intervention, it is possible to extend the findings to support the need for more research on DA of decoding outside of RTI models. While RTI and DA may be useful in concert with each other, the current study's focus is on the use of DA outside of RTI.


Objective: The purpose of this article consisted of two main objectives. First, they sought to evaluate the effectiveness of a DA of decoding in predicting word reading skills of children at the end of first grade, while controlling for autoregressive factors (i.e., present level of word reading skills). The second purpose of this article was to discover if DA of decoding had more predictive validity for students who are at-risk for developing reading disabilities than students who were not classified as such.
**Method:** Using a sample of 105 English speaking students, researchers administered a series of three dynamic measures regarding word reading, which assessed the children's ability to connect new symbols to English sounds, read CVC words composed of these new symbols, and infer a decoding task rule, namely the "silent e" rule for CVCe words. Graduated prompts were administered throughout these tasks to evaluate when mastery was achieved. Static tests were also administered which evaluated phonological awareness, rapid automatic naming (RAN), decoding, and word recognition. All of the dynamic and static tests were administered at the beginning of first grade. At the end of first grade, only the decoding and word recognition measures were given again.

**Results:** The statistical analysis showed that the autoregressive factors explained approximately 50% to 80% of the variance in later word level reading. This means that 20% to 50% was left to be explained by other means. The results indicated that the DA added an additional 2% of the variance. The examiners suggested that although apparently small, the 2% of variance is significant given that the study spanned a short amount of time. They concluded that the DA of word reading allows for the evaluation of later word reading skills to be predicted by the actualizing skills assessed by examining learning potential. Concerning the second objective of this study, it was concluded that DA was more predictive for those students who were at-risk for reading disabilities, likely because children not at-risk were already performing to their full capacity on both standardized and DAs. However, the DA was a significant predictor of word reading skills for children who were at-risk.
Relevance to Current Work: This study adds to the pool of research being conducted regarding DAs of decoding. The results add insight into the usefulness of DA for children who are at-risk for reading disability beyond the usefulness of static assessments with this population. The author's also state that in future research, if more DA tasks are used that align better with actual reading processes, such as decoding, then higher predictive validity may be achieved in evaluating young children than has currently been found when assessing phonological skills.


Objective: The purpose of this article is to summarize the literature regarding evidence to support the existence of several elements of language that transfer across languages. The author states that a significant problem educators face is assessing whether a language learner (LL) struggles with literacy because of insufficient exposure to a second language or because of a learning or cognitive disorder. Historically, LLs have been overrepresented in the special education population, signifying that educators thought that these children had cognitive impairments which interfered with their ability. More recently, educators have been more hesitant to place LLs in resource classes in hopes that they will either catch up in their skill level or show more concrete signs of problems that are not related to linguistic differences. However, this approach is inefficient because of the loss of what could be valuable intervention time. To help resolve this issue, the author suggests using elements of cross-linguistic transfer to determine the cause of the literacy challenges for each child.
Results: Using the results of a plethora of existing studies, several elements are suggested as potential elements that exhibit cross-language transfer and could be beneficial in assessing the source of a child's literacy challenges. These elements include phonological awareness, syntactic awareness, functional awareness, decoding, use of formal definitions and decontextualized language, knowledge of writing conventions and story grammar, and finally, the use of meaning-making strategies to comprehend text. Essentially, the idea behind each element is that if a child exhibits proficient skills in each of the areas listed above in their native language, then we can expect that these elements will transfer over to the new language they are learning. If such transfer has not yet occurred, then we can assume that the child is delayed because of insufficient time with and exposure to the new language. In contrast, if the child does not demonstrate sufficient ability in these areas in their first language and it is expected that they should be, then we can be surer that the problem is likely rooted in a cognitive or learning deficit. Additionally, these children should be observed further and provided with much needed intervention.

Relevance to Current Work: The research summarized in this article supports the claim that evaluation of a child in their native language is effective in determining the root of literacy challenges. The author supports the claim that language learning children are in need of more efficient methods of evaluation. Moreover, in order to apply the principle of cross-linguistic transfer, evaluating a child in their native language is necessary. Additionally, many of the elements discussed in this article, such as phonological awareness and decoding, are also elements that we have attempted to incorporate in our decoding screener for Spanish speaking children.

**Objective:** This study aimed to develop a dynamic measure of decoding acquisition that could be used with both native speakers and language learners of the Danish language. The dynamic test the researchers hoped to develop would specifically be able to overcome language biases which lead to overidentification of language learners as disordered or dyslexic. Overidentification of language learners as dyslexic tends to occur when traditional, static measures of reading ability are used for assessment.

**Method:** This study consisted of a sample of a total of 159 adults. Among this group were dyslexic native speakers, non-dyslexic native speakers, possible dyslexic language learners, and non-dyslexic language learners. All instructions were provided nonverbally, using gestures and demonstrations as well as practice trials to ensure comprehension. Following instructions, the researchers administered the tasks to each participant individually in the same order, as follows: non-word reading, dynamic reading, letter knowledge, receptive vocabulary, non-word repetition, phoneme awareness, and word reading. Specifically, the dynamic reading test contained three parts: learning novel letter shapes and sounds, learning to read two-letter non-words, and independently reading 12 new words made up of combinations of the three letters from the first part.

**Results:** The researcher's claim that besides one exception, all the measures were reliable, measuring at Cronbach's Alpha or above 0.86. The letter knowledge task was the only measure that resulted in ceiling effects in all four participant groups. Furthermore,
the results did support the concern that standard reading measures overidentify dyslexia in second language learners. This claim was illustrated in the differing percent of incidences of dyslexia found between the two types of assessment. The standard word reading test found the incidence of dyslexia to 88% in the sample of second language learners while the DA concluded that only 48% were dyslexic. Unfortunately, the study could not tell for sure which participants were truly dyslexic among this group. However, according to teacher reports, 60% might have been dyslexic in their professional opinion. Overall, the DA was determined to be sensitive to dyslexia in that it identified those with difficulties in learning new words and phonological processing. The dynamic measure was also less sensitive to second language proficiency, allowing it to overcome typical biases that occur in standard assessments.

Relevance to Current Work: This study supports the claim that standard assessments of reading and decoding abilities overidentify individuals as having dyslexia or other disorders regarding reading. Similarly, the research also lends evidence to the fact that DAs are more capable of overcoming biases that accompany linguistic and cultural differences. It is also worth mentioning that the researchers acknowledged that assessing individuals in their native language is an efficient way of circumventing these biases as well. However, the researchers qualified this statement by stating that such assessment is not always the most practical or possible method, which is the reason these researchers conducted the study that they did, testing all participants in one language. The current study that we are conducting will modify an existing dynamic measure to make it available in the Spanish language, allowing for the assessment of Hispanic children in their native language.

**Objective:** This study's purpose was to examine the construct and predictive validity of a DA of decoding. Determining these psychometric values concerning DA is important in the consideration of including a DA as part of a test battery to predict which children will be in need of special services. If DAs could successfully predict academic outcome, then its use in an RTI framework would greatly enhance the effectiveness of such a program. RTI itself is designed so that a child might be receiving weeks of inefficient tiered instruction before their need for more intense intervention is determined. Therefore, the more rapid nature of DA would be helpful in cutting down the time it takes to identify a child that is at-risk for poor achievement.

**Method:** A sample of 318 first grade children was used to conduct this study. Each child was given a prediction battery at the beginning of first grade. The battery included measures of alphabetic knowledge, rapid automatized naming, phonemic awareness, oral vocabulary, DA of decoding, listening comprehension, and teacher ratings regarding attentive behavior and hyperactivity or impulsivity. The DA included pseudoword reading and instruction which increased in the level of difficulty and included five levels of scaffolded instruction. At the end of second grade, the students were given an IQ test which was added to the prediction battery. In the spring of first grade, the students were administered the outcome battery consisting of an untimed
measure of work identification, a timed measure of sight word reading, and an evaluation of reading comprehension.

Results: First of all, the results indicate the DA could play an important role in improving test batteries in identifying children who need intensive instruction more quickly than traditional RTI procedures. The evaluation of the construct validity of the DA suggests that DA is correlated with language ability and IQ level. Regarding the predictive validity, in one model, it was concluded that the correlation between the DA and reading outcomes was high. However, the word attack (WA) skill, administered as part of the alphabetic knowledge section by having children read lists of pseudowords, was even more correlated with reading outcomes than DA. In a second model of analysis, results concluded that DA accounted for unique variance of word identification and passage comprehension abilities (1.0% and 2.3%, respectively). These findings suggest that DA can be described as a significant predictor of responsiveness to reading instruction beyond the other predictors of reading used. However, it can also be said that these proportions are small.

Relevance to Current Work: The researchers of this article support the use of integrating DA into a predictive test battery to identify children at-risk for academic difficulties. However, the results of this article are not as impressive in supporting this claim. Moreover, the researchers were not satisfied with their DA. Indeed, it is necessary that more research be conducted to develop a better DA that can offer quantitative evidence to support the above stance on assessment utilizing DA. For this purpose, the current study has developed what we believe to be a more effective and efficient DA that is more capable in accomplishing the aforementioned goals.
Objective: The purpose of this study was to evaluate the predictive validity of a dynamic screening measure while controlling for letter knowledge and phonological awareness. The authors also made it clear that it was important that the measures have no floor effects. Due to the fact that early intervention can help minimize reading difficulties, the authors claim that there is an increasing need to improve early screening measurements to more accurately identify children in need of such intervention. Additionally, the authors also hoped to validate the use of dynamic testing with children who have not had formal instruction of how to read.

Method: Using a longitudinal study design, the researchers administered a dynamic test of phonological awareness as well as traditional tests of phonological awareness and letter knowledge to a total of 160 Danish children at the beginning of kindergarten. Reading abilities were tested at three points: at the end of kindergarten, during the first half of first grade, and at the end of first grade.

Results: Results of this study indicated that the dynamic screening measure did account for significant variance in the child's reading skills at the end of kindergarten and during the first half of first grade. The dynamic test accounted for 7% of the variance during the test taken at the end of kindergarten. The results remained the same after controlling for letter knowledge as well as individual differences on the static test of phonological awareness. The dynamic test accounted for a significant 4% of the variance from the tests conducted during November of the first-grade year. This went down to 3%
after controlling for letter knowledge and static phonological awareness. However, the research also concluded that the variance previously found at the first two test points did not continue into the third test point at the end of first grade. In May of first grade, the dynamic test ceased accounting for the unique variance to reading.

Relevance to Current Work: This study contributes to the current study by calling attention to the fact that more research is needed regarding the effectiveness of dynamic screeners to predict reading ability past the beginning of first grade. The current study hopes to find significant results regarding the reading abilities of a population of Hispanic students in grade levels from kindergarten to second grade.


Objective: The objective of this article was to evaluate the ability of a DA of decoding to predict reading difficulties before and after reading instruction was administered. Furthermore, secondary objectives included evaluating whether the DA could accurately predict difficulties in both reading accuracy and fluency, or whether it was only effective in predicting reading accuracy as has been demonstrated by other studies. Furthermore, the researchers also hoped to determine if DA could take the place of traditional predictors in assessing reading difficulties.

Method: This study was conducted using 158 children from 19 different kindergarten classes at six Danish schools. The children were seen on three separate occasions spanning the time between the end of kindergarten and the end of 2nd grade.
At the end of kindergarten, the children were given a form A of the dynamic test of decoding as well as tests that evaluated word reading accuracy, letter knowledge, phoneme identification, and rapid automatized naming of objects. Near the beginning of grade 1, the children were given form B of the DA of decoding in conjunction with tests evaluating word reading accuracy, letter knowledge, phoneme synthesis, and rapid automatized naming of digits. Finally, at the end of second grade, children were tested for their reading accuracy and their reading fluency.

Results: Statistic results of this study indicated that the dynamic decoding assessment contributed to the prediction of reading difficulties both when given before and after formal reading instruction. This result supports the claim that DAs evaluates both how easily a child can grasp the idea of an alphabetic writing system and how easily the child can then use that system. Furthermore, regarding the secondary objectives of the study, results indicated that the DA predicted problems with both reading accuracy and fluency. However, the prediction value for fluency was very limited and weak in comparison to the prediction of reading accuracy. The researchers determined that further research is needed regarding this topic. Lastly, researchers concluded that a DA of decoding cannot fully replace the use of traditional predictors such as letter knowledge and RAN, as was the case in the results of this study. However, the DA of decoding did account for a substantial amount of the variance in the results indicating that it was an important component of the prediction measures.

Relevance to Current Work: This study supports the claim that DA is a promising measure in accurately identifying children who have difficulties in reading. This study also provides valuable knowledge regarding the need for measures other than just
decoding when identifying such difficulties. This study made mention of the importance of letter knowledge and RAN abilities as well as the decoding measures in predicting reading challenges. This claim supports the design of the current study's decoding screener which includes sections that assess not only decoding, but word identification (which has a letter knowledge subcomponent) and phonemic awareness as well. Additionally, this study also supports the notion that many current screeners fall short of acceptable accuracy standards and it is therefore necessary to create screeners that meet acceptable and accurate prediction standards.


**Objective:** The purpose of this study was to determine to what extent ESL and native English speakers differ in word recognition abilities and if it is necessary to evaluate them using different measures. Namely, measures that these researchers evaluate include analyses of phonological processing skills and rapid naming tasks. The researchers were also interested in determining whether differences existed between typically developing students and the at-risk students of each group regarding their scores on the measures mentioned.

**Method:** After a certain amount of attrition to the sample size, this study ended up with results from 70 native English children and 200 ESL children, namely of Asian origin. Researchers administered tasks regarding nonverbal intelligence, receptive vocabulary, phonological awareness, phonological memory, and rapid automatized naming. These measures were taken and then compared with word recognition abilities
which was the dependent measure of the study. The results were used to determine if differences existed between the ESL group of children and the native English speakers.

Results: Results of this study indicated that ESL and native English speakers did not differ in their word recognition skills, although they did differ on overall language proficiency. There were no significant differences within the typical developing students and the at-risk students of both the ESL and native English-speaking groups. Namely, both at-risk groups had low phonological awareness tests and also scored slow results on the rapid naming measures. The opposite is true of the typically developing kids in each group, who all scored within normal parameters. Additionally, it was determined that phonological awareness was a good measure of word recognition skills for both native English speakers as well as those who spoke English as a second language. However, phonological awareness proved to be more important for accounting for the variance in the individual differences of the word recognition scores for the ESL students whereas the rapid naming was more important in this same regard for the native English group.

Relevance to Current Work: This study supports the claim that phonological awareness and rapid naming are both excellent predictors in evaluating the reading skills of ESL children which in turn is helpful for early identification and treatment. This study also acknowledges that assessment completed in a child's native language can be more accurate and informative when it comes to evaluating literacy skills. The authors also state that this type of assessment is generally unavailable because of lack of trained professionals and assessment tools. Another important point made is that ESL students may have difficulty reading because of difficulties with decoding and not because of linguistic differences, however, we lack research regarding the acquisition of literacy in
ESL children. This statement supports the work being done in the current study to develop an assessment tool to be administered to children whose L1 is Spanish.


Objective: This article sought to describe two alternative assessment methods that can be used in place of traditional norm-referenced and criterion-referenced tests when testing culturally and linguistically diverse (CLD) populations. The authors purport that such traditional tests have well documented problems in presenting biases when used on CLD students. Such biases include content bias, linguistic bias, and normative samples that are not representative of CLD populations. Two alternative assessments proposed include the use of processing-dependent measures, such as nonword repetition and working memory tasks, and DA to evaluate the language and literacy abilities of CLD populations.

Results: The researchers cite multiple studies that provide evidence of their claims. The results that are of interest to the purpose of the current study include the conclusions regarding dDA. The researchers report that DA is an excellent, bias-free method of distinguishing between language disorders and language differences in CLD children as well as predicting early reading abilities. The use of DA will help solve the problem of over- and underdiagnoses of language and learning disabilities in CLD children.

Relevance to Current Work: This article provides evidence for the use of DA in evaluating language and reading abilities in children who are from culturally and linguistically distinct backgrounds. This article also enumerates the biases that are
inherent in traditional, standardized assessment of language and literacy when testing CLD children. The researcher also asserts that a way to eliminate biases would be to develop tests specially designed for CLD groups, which is what the current study hopes to accomplish by designing a decoding screener for Latino children.


**Objective:** The purpose of this article is to expound on the relationship between the ideology of Vygotsky and his notion of the Zone of Proximal Development (ZPD) and the currently developing research of DA as an evaluation approach.

**Summary:** The author explains that although Vygotsky never used the term "dynamic assessment" to describe his ideas, researchers have based the foundation of their work in developing dynamic assessment approaches in the ideas of Vygotsky. The author lists three major definitive features of DA: its interactive nature, its focus on the learning process, and the type of information that results. The interactive nature of DA allows evaluators to take an active part in the assessment and allows observations of the learner-evaluator interactions to be used in the diagnostic process. The focus of DA on the learning process allows clinicians to gather information regarding the ability of the student to apply cognitive skills to the task being assessed. Lastly, DA allows us to obtain results concerning the modifiability of the student. These characteristics relate to the central idea of the ZPD, which asserts that just as important as it is to know where a child performs independently is to also know where a child can perform when provided with adult scaffolding and assistance.
Relevance to Current Work: This article will provide insightful background information into the nature and development of DA as an evaluation tool. By understanding the background of DA in relation to its roots in Vygotsky's ZPD, we are better able to understand the clinical purpose of using intervention and interaction to more completely assess a child's capabilities to learn and perform.


Objective: The objective of this article was to illustrate that DA can be used to add insight into the results of traditional assessment. The researchers supported their claim using a case study of two participants of a larger study. DA has been purported to reduce the biases that are generally present in traditional, static assessments when evaluating minority children. Specifically, many minority children are not familiar with test procedures and also have less exposure to test content, putting them at a disadvantage when compared to their non-minority peers.

Method: The authors presented a case study to support their claims that DA could uncover important information regarding a child's performance that could not be evaluated using more typical assessments. The two Latino participants evaluated in this cast study were among those assessed in the study using a test-teach-retest method to assess vocabulary knowledge. In the cited research, the Expressive One Word Picture Vocabulary Test (EQWPVT) served as the pre and post-test material used in comparing the scores for normal groups versus those with language disorders. During the teaching
period, each child received two intervention sessions designed using principles of Mediated Learning Experiences (MLE) such as intentionality, planning, meaning, and change. The modifiability exhibited by each child was documented using three Likert-type scales. The scales evaluated responsiveness of child, effort required from the evaluator, and indication of ability to transfer learned skills.

**Results:** Both children had scored similarly on the pretest measure of the EQWPVT and had the results of the test been static in nature, each most likely would have been diagnosed with a language impairment. However, because of the intervention and post-test measures, researchers were able to determine that one child seemed to demonstrate results consistent with a language disorder while the other child seemed to only be exhibiting poor results due to a language difference. For instance, the disorder child was only slightly responsive during the intervention and required extreme effort from the examiner to achieve success. She also only presented with medium transfer of learned skills during the post-test. In contrast, the other child was highly responsive, only needed minimal examiner effort, and exhibited a high degree of transfer. Finally, the first child's post-test scores showed low levels of improvement while the other child improved his score by two standard deviations.

**Relevance to Current Work:** The details of this case study support the claim that DA can be used to differentiate between minority children who are language different versus language disordered. Furthermore, this article adds evidence to the effectiveness of dynamic assessments in analyzing results with more insight than traditional assessments are capable of performing. For instance, the results of this DA provided
more insight into the development of the children than the static scores would have done if analyzed without the mediated intervention efforts.


**Objective:** The purpose of this article was to explore how the results of modifiability from two studies helped differentiate between children who were typically developing and others who had low language ability (LLA). Specifically, the author wanted to explore this hypothesis regarding the results from culturally and linguistically diverse children. The author purports that DA, of which modifiability can play an essential role, can be used to differentiate between those with a language difference versus a language disorder. Furthermore, DA can also help limit certain biases that are common during assessment, especially with culturally diverse children.

**Method:** The author analyzed results from two previous studies to determine the efficiency of modifiability in obtaining the goals mentioned above. Both studies involved culturally and linguistically diverse elementary school children, namely African American children and Latino children. The first study administered DA in the form of test-mediation-retest model to the participants. Components of the study consisted of a baseline testing period, a mediated language experience (MLE) intervention period, and a posttest all of which spanned over a period of 12 weeks. A modifiability scale was also completed for each child following the second session of MLE. The second study followed a similar design in which a pretest-mediation-posttest approach was administered. After the MLE sessions in this study, examiners completed a Learning
Strategies Checklist as well as a Modifiability Scale after the second MLE session. The Learning Strategies Checklist consists of items that evaluate child responsivity such as attention, planning, self-regulation, and motivation. It also looks at transfer of skills as well.

**Results:** The analysis of the results from the first study concluded that the Modifiability Scale differentiated children with LLA from the typically developing (TD) children with high efficiency. Specifically, TD children consistently scored higher on measures of effort, responsivity, and transfer. Results from the second study indicated similar results. Typically developing children scored higher on the Learning Strategies Checklist as well as the Modifiability Scale, although there was some indication of slight overlap. Overall, the results from the analysis of these two studies show that measures modifiability can be used to differentiate TD children from LLA children, even those who are culturally and linguistically diverse.

**Relevance to Current Work:** This work supports the claim that DA can be used to differentiate between children who have language differences versus those who have language disorders. Additionally, this article supports the application of modifiability into the assessment of children who are culturally diverse. The author supports the claim that it is important to determine the need of children who come from different language backgrounds for special education services. This article cites the research of Vygotsky and the Zone of Proximal Development to uphold the use of measures that evaluate the amount of support a child requires to accomplish a task. The present study does apply such a scale to evaluate the level of help a child requires to accomplish decoding tasks in one of the sections contained on the dynamic screener.
Objective: The purpose of this article was to compare the classification accuracy between common static measurements and DA measurements in evaluating pre-reading abilities in kindergarten children. Additionally, the researchers sought to compare the sensitivity and specificity yielded by each of the measures in identifying children at-risk for reading difficulties by using data gathered at the end of first grade. A secondary objective was to compare two types of instruction during DA, namely a sound-by-sound method and an onset-rime method. The researchers also made special mention of results regarding a Hispanic subset.

Method: The participants of this study included 600 participants from 14 schools in a school district in Utah. These children were followed from kindergarten until the end of first grade. Each child was administered typical static measurements (e.g. DIBELS) at the beginning of kindergarten by their school district to determine which children were at risk for reading difficulties. Within three weeks after the static measurements were administered, the researchers divided the children into two groups and administered the DA to evaluate their reading capabilities. One group was assigned to the sound-by-sound implementation and the other was assigned to the onset-rime method. Each of the two groups participated in three phases of the DA: the pretest phase, the teaching phase, and the posttest phase. The only difference was found in the teaching phase style as described
above. The data were analyzed to determine whether any significant differences existed between the static and dynamic measures as well as the two different dynamic styles.

**Results:** The researchers concluded that the DA’s results were overall significantly better than the static measurements. The classification accuracy was better for the dynamic measures for both groups, sound-by-sound and onset-rime, by a significant amount (e.g., 72% and 70%, respectively for static measures; 85% and 83%, respectively for dynamic assessments). Likewise, the DA also proved superior in the sensitivity and specificity measures as well (e.g., sound-by-sound for static: 79% and 50%, respectively; onset-rime for static: 69% and 51%, respectively; sound-by-sound for dynamic: 92% and 83%, respectively; onset-rime for dynamic: 81% for each measure). DA proved more effective in each measure when compared to the static measurements. However, there was no significant difference between the sound-by-sound method versus the onset-rime method. The results for the Hispanic subgroup, although not as effective, were still found to be within more acceptable parameters using the DA’s over the static measures (88% sensitivity and 70% specificity for the DA; 88% sensitivity and 33% specificity for the static measures).

**Relevance to Current Work:** This article is relevant to the current work because it supports the assumption that DA is an effective way of determining whether a child is at risk for reading problems, especially with populations that are considered culturally and linguistically diverse (CLD). The researchers report that more Hispanic children may be misidentified as having future difficulty than desired; however, the benefits of DA still outweigh static measures when evaluating minority groups. This article also supports the
claim that early identification is an important step in helping to prevent and treat reading difficulty.


**Objective:** This study sought to evaluate the predictive validity of a DA to identify bilingual Latino children that are at risk for reading difficulty. Additionally, the authors aimed to provide evidence for the benefits of DA in mitigating biases that are typical when evaluating culturally and linguistically diverse children through normal, static measures.

**Method:** This research study was conducted using a sample of 63 Latino children identified through a kindergarten screener as at-risk for future language impairment. The researchers gathered many data regarding the participants using parent interviews and information provided from the school (i.e. SES information, activities in which the child engaged, and which language was used and with whom during these activities). Data was collected regarding the children's performance on a nonsense word recoding DA administered by the researchers, as well as DIBELS subtests during kindergarten. These tests will provide comparison between DA and static measures. Later, three types of first-grade English reading measures were administered by the schools (e.g., oral reading fluency, word identification, and nonword fluency) to serve as criterion measures. Three scores were derived from the DA: sounds gain, residuum gain, and modifiability score. The latter two were significantly correlated with the first-grade reading measures while the modifiability score proved to be the most informative.
Results: Results indicated that the modifiability score from the kindergarten DA yielded a 100% sensitivity and 80% specificity for first-grade oral reading fluency, 100% sensitivity and 88% specificity for first-grade word identification, and 86% sensitivity and 85% specificity for first-grade nonword fluency. These significant measures indicate that the use of a nonsense word recoding DA for predicting reading difficulty is effective. The results of the DIBELS test presented evidence against the efficiency of static measures when accurately attempting early identification, namely that the DIBELS subtests overidentified students as at-risk. Additionally, the modifiability score proved to be the most predictive measure for future reading ability. Specifically, the modifiability scores of the children indicated that those who were identified as at-risk were those who failed to implement the strategies taught during the DA as well as those who were labeled as less responsive to instruction by the examiners.

Relevance to Current Work: This article supports the claim that Latino children are at a disadvantage when it comes to early, accurate identification of risk for reading difficulty because static measures do not compensate for biases against culturally and linguistically diverse children. Furthermore, the research results support the claim that DA is an effective way in mitigating such biases against ELL children. Additionally, this article provides support for DA in describing the effectiveness of a test that takes a relatively short time to administer while also producing significantly relative data for assessment.

Objective: The purpose of this study was to evaluate the results of a kindergarten DA of decoding to verify whether the test predicted future reading difficulty at 2nd, 3rd, 4th, and 5th grades for both Caucasian and Hispanic students. These researchers also sought to establish whether DAs have more predictive validity and classification accuracy than conventional static measures.

Method: This current study builds upon the work of a study conducted in 2016 by the main researcher and others. In the 2016 study, participants consisted of 300 Caucasian and 300 Hispanic students. These participants were administered two static reading measures and one of two DAs at the beginning of kindergarten and a norm-referenced assessment of reading at the end of first grade. The current study's sample consisted of all students from the 2016 study that still resided in the same Utah school district. The remaining students (over 365 for each grade level) were administered the Oral Reading Fluency subtest of DIBELS Next assessment at the end of each school year from 2nd to 5th grade. The students were subsequently classified as having decoding difficulties or not based on the results of the evaluation. The results from the initial kindergarten evaluation and the tests administered from 2nd to 5th grade were compared to determine how much variance the kindergarten DA of decoding accounted for second through fifth grade decoding skills, as well as how accurately the kindergarten DA classified those with and without decoding difficulty.

Results: The results of this study indicated that the kindergarten DA was a fair to good predictor of future reading ability for both Caucasian and Hispanic students up
through 5th grade. The results also determined that the DA did account for variance in reading outcomes that static measures were not able to explain, particularly for the Hispanic subgroup. Additionally, the DA also had significantly better classification accuracy (fair to good) than static measurements (poor to fair). Static measures tend to over identify subjects as at risk, especially culturally diverse subjects. This fact makes DA a more effective way of identifying students at risk for reading difficulties.

Relevance to current work: This study supports the claim that DAs are effective and efficient in measuring a child's ability to decode and can be used to identify individuals who need early intervention better than static measures. Additionally, this article acknowledges and seeks to mitigate biases that are inherent when cultural and linguistic diversity is not accounted for in assessments. It is worth mentioning that the researchers acknowledge that the classification accuracy of their study did not meet the typically preferred benchmark of ≥ 80% and therefore their results are limited when it comes to a more diverse population, which our study will seek to resolve.


TESL Canada Journal, 16(2), 86-93.

Objective: This review's purpose is to take a close look into the Woodcock Reading Mastery Test--Review (WRMT-R). Specifically, the author hopes to establish whether the WRMT-R is appropriate to use when evaluating the reading ability of an ESL learner. The author describes the sections of the test in detail. Additionally, she provides the norming data concerning the original test as well as the updated norms that took place 12 years after the test had been in use.
Results: The author concludes that the WRMT-R is unfit in any regard to be used on ESL learners. This conclusion is drawn because of the analysis of the norming data of the test. The WRMT-R was originally normed on a sample whose structure was designed to reflect the 1980 US Census. The author highlights the fact that the 1980 US Census was limited to a white/non-white racial classification and therefore the original norms could not have been accurate in representing minorities in the US. The WRMT-R was renormed in 1995 and was designed to reflect the ratios of the 1990 US census. However, this census was also limited in its racial categories, only including White, Black, Hispanic, and other. Furthermore, the manual states that individuals who were not proficient in English were not included in the norm sample. The author also expresses concerns over the questionable psychometric properties of the test, namely the lack of evidence of content validity. Construct validity is not discussed in the original manual, and little is said regarding it in the new manual.

Relevance to Current Work: This review provides useful information in evaluating current standardized procedures commonly used in evaluating mainstream students as well as minorities in current practice. Subtests of the WRMT-R such as the Word Identification and the Word Attack subtest are commonly used in research today when evaluating reading disability. This review casts doubt on the use of these subtests in assessment practices.


Objective: The purpose of this article is to summarize facts and methods surrounding the prevention and intervention of reading failure. The author used information from two
consensus documents to outline a detailed document that could aid in organizing better techniques to decrease the incidence of reading difficulties in schools across the country. The two consensus reports that the author used include Preventing Reading Difficulties in Young Children (1998) and another titled Teaching Children to Read (2000).

**Results:** The author stated that 37% of fourth-grade school children were not reading at the level needed in order to successfully complete grade-level work. With this statistic in mind, the author advocated for the enhancement of our current reading instruction practices as well as evaluation of those who are at-risk for reading failure. To support this claim, the article declares that children who do not obtain the necessary literacy skills will be disadvantaged not only in academic settings but in their future adult lives as well. Thus, identifying children at-risk for reading difficulties is an essential part of preparing children for their futures. The author states that it is imperative that children receive the necessary intervention before waiting until mid-elementary school years to increase the chance of a child becoming capable of adequate reading skills. The author claims that children who are inefficient readers at the end of first grade almost never ascertain the average-level reading skills that are necessary at the end of elementary school.

**Relevance to Current Work:** Critical elements of this article that are relevant to the current work include the identification of challenges with sight words, phonemic awareness, and decoding skills as indicators of children who are experiencing reading difficulties. The author asserts that a child's ability to identify sight words is a crucial skill needed for success in becoming a fluent, capable reader. Similarly, failure to develop the necessary phonemic decoding skills will negatively affect a child's reading
progress. All of these claims support the current work in the reasoning behind the test items included in the screener of decoding abilities.
APPENDIX B

S-DDM Decoding Measure
## APPENDIX C

**S-DDM Phonemic Awareness Measure**

### TARGET 1: Phoneme Segmentation

**Say:** "Voy a decir una palabra. Después de decirla, digame todos los sonidos de la palabra. Por ejemplo, si digo 'gato', diga /g/ /a/ /t/.

- "Probemos uno diferente: dime todos los sonidos en 'mapa'."
- "Si el estudiante claramente separa cada sonido correctamente, dígale 'Correcto'." (ver la columna 3)
- "Así es, los sonidos en 'mapa' son /m/ /a/ /p/ /a/. Inténtalo de nuevo. Dime todos los sonidos en 'mapa'."
- "Bueno, vamos a hacer algunas más."
- Begin the test, starting with "me".

<table>
<thead>
<tr>
<th>Prompt for each word: &quot;Dime los sonidos en _&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>me   /m/ /a/      0 1 2</td>
</tr>
<tr>
<td>yo   /o/ /a/      0 1 2</td>
</tr>
<tr>
<td>día  /d/ /i/ /a/  0 1 2 3</td>
</tr>
<tr>
<td>hace /a/ /e/ /c/  0 1 2 3</td>
</tr>
<tr>
<td>por  /p/ /o/ /r/  0 1 2 3</td>
</tr>
<tr>
<td>ven  /v/ /e/ /n/  0 1 2 3</td>
</tr>
<tr>
<td>solo /s/ /o/ /l/  0 1 2 3 4</td>
</tr>
<tr>
<td>chico /k/ /i/ /o/ 0 1 2 3 4</td>
</tr>
<tr>
<td>cada /k/ /a/ /d/ 0 1 2 3 4</td>
</tr>
<tr>
<td>ellos /e/ /l/ /o/ 0 1 2 3 4</td>
</tr>
</tbody>
</table>

**TOTAL CORRECT PHONEMES** /32

### TARGET 2: First Sounds

**Say:** "Escúchame decir esta palabra: 'miel'. El primer sonido que escuchas en esa palabra es /m/.

- "¿El primer sonido en 'miel' es /m/?"
- "¿Cuál es el primer sonido que escuchas en la palabra 'miel'? Correcto.
- "Así es, /m/ es el primer sonido en la palabra 'miel'."
- "Ahora voy a decir algunas palabras más. Dime el primer sonido que escuchas en cada palabra."

**Corrective Prompt (as needed):** "Recuerda, el primer sonido de la palabra, no el nombre de la letra."

**TOTAL SCORE** /20

### TARGET 3: Phoneme Blending

**Say:** "Voy a decir una palabra de forma lenta y luego la voy a decir de forma rápida. Escucha, /sssooolll/ /sol/.* (2 seconds for each sound).

- "Ahora voy a decir la palabra de forma lenta y tú la dices de forma rápida. Listo. sssooolll."
- "Ahora voy a decir más palabras de forma lenta. Tú dílas las palabras de forma rápida."

**Corrective Prompt (as needed):** "Correcto." (If necessary)

<table>
<thead>
<tr>
<th>Prompt for each word: &quot;Digo la palabra lenta, tú la dices rápida&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>hola  /o/ /l/      0 1 2</td>
</tr>
<tr>
<td>les   /e/ /s/     0 1 2</td>
</tr>
<tr>
<td>ese   /e/ /s/     0 1 2</td>
</tr>
<tr>
<td>nos   /o/ /s/     0 1 2</td>
</tr>
<tr>
<td>son   /o/ /s/     0 1 2</td>
</tr>
</tbody>
</table>

**TOTAL SCORE** /10
### APPENDIX D

**S-DDM Word Identification Measure**

**TARGET 1: Irregular Words**

Display corresponding Irregular Words from stimulus book. Say:
"Por favor lee estas palabras. Si no sabes una palabra, está bien. Solo continúa."

- **Discontinue Rule:** If student reads five consecutive words incorrectly, STOP and go to Target 2.
- **Next Target:** Administer Target 2 Letter Sounds even if Target 1 benchmark is met. **Exception:** Do not administer Target 2 Letter Sounds if mastery of letter sounds has been previously documented and student has likely retained letter sound knowledge.

**TARGET 2: Letter Sounds**

Display corresponding Letter Sounds from stimulus book. Say:
"Las letras hacen sonidos." (Examiner points to the "A").
Say:
"El sonido que hace esta letra es 'a'... O 'aa'..."
Say:
"Quiero que me digas los sonidos que hacen estas letras. Si no sabe un sonido, está bien. Solo continúa."

- **Point to the first letter - A.**

- **Start the 2-minute timer.**
  - Corrective Prompt (maximum 2 times):
    "Dime el sonido que hace la letra."
- **Discontinue Rule:** Stop the test after 2 minutes, or when the student is finished, or it is clear the student does not know letter sounds.
- **Next Target:** If student’s score is not at benchmark, go to Target 3 (see Benchmark table).

**TARGET 3: Letter Names**

Display corresponding Letter Names from stimulus book. Say:
"Las letras tienen nombres." (Examiner points to the O on the top left corner of the stimulus page).
Say:
"El nombre de esta letra es 'oa'. Di 'oa'."
Say:
"Todas estas letras tienen nombres. Quiero que me digas los nombres de estas letras."
Say:
"Si no sabe un nombre, está bien. Solo continúa."

- **Point to the first letter - O.**

- **Start the 2-minute timer.**
  - Corrective Prompt (maximum 2 times):
    "Dime el nombre de la letra."
- **Discontinue Rule:** Stop the test after 2 minutes, or when the student is finished, or it is clear the student does not know letter names.
Fidelity Checklist

Fidelity for Decoding
- Did they read the pretest instructions at the beginning?
- Did they provide prompts for the pretest appropriately?
- Did they engage the teaching phase appropriately?
- Phase 1 of the teaching phase was followed?
- Phase 2 of the teaching phase was followed?
- Posttest script was followed
- Prompts for the posttest given correctly

Fidelity for phonemic awareness
Target 1
- Was the script followed correctly? Word for word introduction?
- If the student didn’t answer correctly, did the examiner provide the demonstration prompt?
- Did the examiner prompt for each word: “dime los sonidos en...”
- Was the corrective prompt provided correctly and when necessary?

Target 2
- Was demo provided word for word?
- Was the prompt for each word?
- Was the corrective prompt provided when necessary?

Target 3
- Was the demo provided correctly?
- Did they prompt for each word?

Fidelity for Word Identification
Target 1:
- Did the examiner say the script?
- Did they say the word after 3 seconds if the child struggles?
- Did they follow the discontinue rule? (as long as it followed and doesn’t affect score)

Target 2:
- Did the examiner say the script correctly?
- Did they follow the discontinue rule?
- After 3 seconds, did they say the sound for the student

Target 3:
- Did the examiner say the script correctly?
- Did they follow the discontinue rule?
- Did they say the letter after 3 seconds?