



Theses and Dissertations

---

2023-06-06

## Validating a Predictive Early Kindergarten Dynamic Assessment of Word Level Reading and Language

Taylor Jordan Cena  
*Brigham Young University*

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



Part of the [Communication Commons](#)

---

### BYU ScholarsArchive Citation

Cena, Taylor Jordan, "Validating a Predictive Early Kindergarten Dynamic Assessment of Word Level Reading and Language" (2023). *Theses and Dissertations*. 9957.  
<https://scholarsarchive.byu.edu/etd/9957>

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 [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

Validating a Predictive Early Kindergarten Dynamic Assessment  
of Word Level Reading and Language

Taylor Jordan Cena

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

Shawn Nissen, Chair  
Douglas B. Petersen, Co-Chair  
Kathryn Cabbage  
Ellie Young

Department of Communication Disorders  
Brigham Young University

Copyright © 2023 Taylor Jordan Cena

All Rights Reserved

## ABSTRACT

### Validating a Predictive Early Kindergarten Dynamic Assessment of Word Level Reading and Language

Taylor Jordan Cena  
Department of Communication Disorders, BYU  
Master of Science

The purpose of this study was to examine the preliminary psychometrics to revisions to the decoding subtest of the Predictive Early Assessment of Reading and Language (PEARL) dynamic assessment and to further examine the predictive validity of the language subtest of the PEARL. Specific aims were to (a) determine if kindergarten students made gains on the PEARL decoding subtest from pretest to a delayed standard posttest and to a delayed novel posttest, indicating that floor effects were mitigated, (b) to examine if there was variance in the delayed standard posttest and novel posttest scores, (c) to determine if there was adequate fidelity of administration and adequate inter-rater reliability of a delayed standard posttest and a delayed novel decoding posttest, and (d) to examine the sensitivity of the language subtest of the PEARL dynamic assessment.

The PEARL Dynamic Assessment was administered to 34 kindergarten students from one elementary school who were identified as at-risk from the DIBELS (Dynamic Indicators of Basic Early Literacy Skills) composite score (Good et al., 2004). Of those 34 students, the PEARL identified 7 students at risk for decoding and 11 students at risk for language. The PEARL dynamic assessment contains two brief subtests, a decoding subtest, and a language subtest. Both subtests use a pretest, teach, posttest format.

Results indicated that the delayed standard posttest correct sounds were significantly different from the pretest correct sounds. The delayed novel posttest correct sounds were significantly different from the pretest correct sounds. For the delayed standard posttest correct words there was a significant difference when compared to the pretest correct words. The delayed posttest novel words were not significantly different from the pretest words. Fidelity of administration was calculated from 10 separate administrators, with 100% fidelity of administration for nine of those administrations. The point-to-point interrater reliability of the delayed posttest sounds, and novel sounds was 94% and the point-to-point interrater reliability of the total words read correctly was 94%. Results indicated that the PEARL correctly identified 9/11 students as having language disorder, yielding a sensitivity index of 82%. The findings suggest that the modification to the decoding subtest of the PEARL has preliminary evidence of validity and reliability, and that the language subtest of the PEARL may be a valid tool to identify kindergarten students at risk for language disorder.

Keywords: dynamic assessment, decoding, language, response to intervention, diagnosis, kindergarten

## ACKNOWLEDGMENTS

I would like to thank the following committee members for their meaningful contribution to the current study: Dr. Douglas Petersen, Dr. Katy Cabbage, Dr. Ellie Young, and Dr. Shawn Nissen. I could not have made it to this point without their guidance and expertise. I would like to personally thank Dr. Petersen for allowing me to work on this research project. Through the process of research and writing, he showed me so much patience, kindness, wisdom, and passion. It was contagious and made the work so much more meaningful to me. I would also like to thank Dr. Shawn Nissen for stepping in as my thesis chair when circumstances changed, and the McKay School of Education at Brigham Young University for providing the internal grant that funded and made this project possible. I am especially grateful for the team of research assistants on the PEARL team for their time and effort in collecting data that was used in this study. I also want to thank my family and friends for showing me continuous love and support throughout the completion of this thesis and my graduate education.

## TABLE OF CONTENTS

TITLE PAGE .....	i
ABSTRACT.....	ii
ACKNOWLEDGMENTS .....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
DESCRIPTION OF THESIS STRUCTURE AND CONTENT .....	vii
Introduction.....	1
Statement of the Problem.....	3
Floor Effects.....	4
Language Difference vs. Language Disorder .....	5
Sensitivity and Specificity .....	6
Dynamic Assessment .....	7
Dynamic Assessment of Decoding.....	8
Dynamic Assessment of Language.....	11
Statement of Purpose .....	13
Research Questions.....	14
Method.....	14
Participants.....	14
Measures .....	15
Predictive Early Assessment of Reading and Language (PEARL) .....	15
Tier-2 Oral Narrative and Expository Language Intervention.....	17
Reference Standard .....	18

Fidelity .....	19
Inter-Rater Reliability .....	20
Results.....	20
Research Question 1 Results.....	20
Research Question 2 Results.....	21
Fidelity .....	21
Inter-Rater Reliability .....	22
Research Question 3 Results.....	22
Discussion.....	22
Pretest to Posttest Gains.....	23
Fidelity and Inter-Rater Reliability.....	25
Sensitivity of the Language Subtest of the PEARL.....	25
Limitations .....	27
Implications for Future Research.....	27
Conclusion .....	28
References.....	29
APPENDIX A: Review of the Literature.....	35
APPENDIX B: Institutional Review Board Approval Letter.....	53
APPENDIX C: Child Assent Form .....	59

## LIST OF TABLES

Table 1	<i>Descriptive Data</i> .....	15
Table 2	<i>PEARL Decoding Subtest Statistics</i> .....	21

## DESCRIPTION OF THESIS STRUCTURE AND CONTENT

This thesis, *Validating a Predictive Early Kindergarten Dynamic Assessment of Word Level Reading and Language*, was written in a hybrid format to adhere to journal publication formats as well as traditional thesis requirements. The initial pages of this thesis fulfill university requirements while the thesis itself is presented in a journal article format. Appendix A consists of the annotated bibliography. The Institutional Review Board (IRB) approval to conduct this study is included in Appendix B.



## Introduction

The foundational skill that serves as the base for all academic and school-based learning is the ability to read (Lyon, 2003). Decoding and language are two fundamental components of reading ability. According to Gough and Tunmer (1986); decoding is the ability to read isolated words quickly and accurately based off a knowledge of letter-sound correspondence rules. Language comprehension is the capacity to understand what is being read or “decoded,” built on the foundation of background knowledge, vocabulary, language structures (semantics, syntax, etc.), verbal reasoning, and literacy knowledge (Ehri et al., 2001; Scarborough, 2001). This supports the idea that even if a child can accurately decode all the sounds in a string of words, it does not mean they will be able to understand the meaning of the words and then apply the background knowledge they have, to make inferences and to fully understand what they read. Conversely, if a child has strong language skills, but poor decoding ability; they will not be able to understand what they read. The simple view of reading suggests that measures of language ability can help reveal problems with reading comprehension. Studies have shown that, more specifically, oral language ability in the early school years is a predictive measure of future reading comprehension (Catts et al., 2016; Lyon, 2003). The key to comprehension is the ability to quickly and accurately read the words on the page. Lyon (2003) states that children need to learn how to use and understand language because it is crucial for reading, and consequently, further academic success. Thus, student performance in both decoding and language are important measures for reading and should be monitored over time.

Schools across the country have adopted a multi-tiered system of support (MTSS) framework to help identify students with reading difficulty, monitor progress, and to provide necessary intensities of reading instruction according to student needs. A key component of this

MTSS is the use of quick and valid universal screeners, which are used to identify students who are at risk for academic difficulties so they can receive the level of intervention needed to achieve academic success (Fuchs & Deno, 1981a; Greenwood et al., 2011; Jenkins & Johnson, 2008; Thomas & January, 2021; VanMeveren et al., 2020). Universal screening is a critical prerequisite for providing early school-based prevention and intervention services. Typically, school-based universal screening is conducted with all students in each classroom, school, or district with three specific purposes; (a) to identify those at risk for future academic deficiencies, (b) identify current difficulties a student is having, and (c) to provide those students with the appropriate instruction to help them succeed (Glover & Albers, 2007; Greenwood et al., 2011). To increase efficiency in the school system, educators must adopt universal screeners which are contextually appropriate, technically sound, and usable; meaning they align with the curriculum, they are supported by local needs, they fit the population of interest, and have high reliability and validity (Glover & Albers, 2007). Universal screeners usually function as general outcome measures (GOMs) which have very specific parameters. For example, GOMs (universal screeners) need to be brief, efficient, easy to administer and score, reliable, valid, sensitive, standardized, have parallel forms, and measure child performance related to long-term outcomes (Fuchs & Deno, 1981b; McConnell & Wackerle-Hollman, 2016).

The administration of universal screeners for both language and decoding are crucial for young children in the school system and can provide substantial information on reading ability. Research has shown that oral language and decoding assessments in early elementary school often predict reading problems in later grades (Catts et al., 2016; Petersen et al., 2018). Furthermore, when reading difficulty is identified at an early age, there is a much greater likelihood of remediating that problem, and a higher probability for stronger academic

achievement. In fact, research suggests that if identified early, and given the proper instruction, reading difficulties can be remediated in the early school years (Lyon, 2003).

### **Statement of the Problem**

Most universal screeners for reading that are administered to young students have two major flaws: First, these screeners exclusively focus on word reading (decoding), rather than language abilities. Presumably this emphasis on decoding in the early grades is because decoding can be measured in a way that meets GOM requirements. Another reason is that oral language is often overlooked as a proxy measure for reading comprehension, thus, comprehension is only measured in students after they have learned to decode with fluency. There is usually no explicit oral language instruction in school, so there is no perceived need to assess and monitor oral language growth over time. Most children enter kindergarten with letter name and letter sound recognition knowledge, even if it is minimal. In fact, Petersen et al. (2018) suggests that many of these students enter the school system unprepared, having not experienced any formal reading instruction. These skills often see much improvement because students receive direct instruction regarding phonemic and phonological awareness in relation to decoding. It is important to note that this is not the same for language abilities. Most language acquisition happens in the preschool years, and before children receive any formal schooling. While knowledge of language and language complexity increases through the kindergarten years; it is not directly measured based on static measures and instruction received in the classroom (Catts et al., 2016). Current research indicates that early explicit instruction focused on oral language and literacy with young children promotes the development of phonological awareness and language complexity. This early instruction facilitates the child's ability to comprehend and understand what they listen to, which is an important foundation, preparing them for the explicit reading instruction that will be

taught in the kindergarten years (Lyon, 2003). Although very little research has been done in the field of early identification of problems with oral language comprehension related to reading abilities, Catts et al. (2016) determined that language measures are indicative of reading comprehension outcomes in later grades and children who are most likely to have problems learning to read can be identified in the early stages of their literacy development. They showed evidence of this in a longitudinal study by taking 263 kindergarten students who were identified as “at risk” for reading disorder based off beginning-of-year school-based assessments (Letter Name Fluency and Initial Sound Fluency subtests of the DIBELS (Dynamic Indicators of Basic Early Literacy Skills)). They administered the students identified as at risk in kindergarten a battery of reading and language assessments that were previously reported to be predictive of word reading outcomes. The assessments tested vocabulary, grammar, and narration. At the end of third grade, the children were given measures of reading comprehension. Results showed that the measures of vocabulary, grammar, and comprehension abilities administered in kindergarten had strong predictive validity for reading comprehension in later years.

The second major flaw with universal screeners for reading ability is that they are static measures. Three specific problems arise when using only static assessments to screen students and evaluate current and future reading ability; (a) They yield high floor effects (the scores are low and there is limited variance in scores across students), (b) They do not accurately identify a language difference from a disorder (poor sensitivity and specificity), and (c) often lack construct validity when administered to young students.

### **Floor Effects**

Static reading measures consistently yield floor effects with low scores and minimal score variance, especially in young children and with children who are culturally and

linguistically diverse (CLD). High floor effects arise because static assessments only test a child's current performance of skill or knowledge. As a result, young children are often misidentified as having a reading disorder when they perform poorly on a one-time static measure of prereading ability (Petersen et al., 2018). For example, children who have limited literacy experiences or who are from CLD backgrounds are at risk for misidentification of reading difficulties using static assessments because they only measure current levels of reading ability, which could be confounded by cultural or linguistic factors. Research shows that static assessment methods used in the United States consistently identify children who are CLD as having higher risk for reading problems, which leads to costly interventions provided to more students than necessary (Denman et al., 2017; Petersen et al., 2016, 2018). These static measures also yield high floor effects because kindergarten students who are being assessed come into the education system with various levels of language and literacy skills and formal reading instruction. Students are also often unfamiliar with testing procedures and school expectations are unclear or different from their individual backgrounds (Laing & Kamhi, 2003; Petersen et al., 2018). These factors can negatively influence the student's performance on testing, and lead to inappropriate classification of special education.

### **Language Difference vs. Language Disorder**

Static measures of oral language ability cannot validly lead to a diagnosis of reading impairment because they often fail to differentiate a language difference from a language disorder. Laing and Kamhi (2003), state that "it is well accepted that the accurate assessment of spoken and written language skills in CLD children should not depend solely on the use of standardized norm-referenced test procedures." Most of the popular norm-referenced language assessments in the field today are static measures used to qualify students for speech-language

services (Petersen & Gillam, 2015; Shenoy, 2014; Spaulding et al., 2006). Norm-referenced, static measures of language contain three significant deficits: content bias, linguistic bias, and disproportionate representation in the normative samples (Laing & Kamhi, 2003). Content bias is present when test stimuli and procedures reflect on mainstream culture and patterns, assuming that all children have been exposed to the same language and life experiences. Linguistic bias includes dialectal disparity between the examiner, the child, and the expectations of the child's responses on the test. Many static measures fail to include CLD populations in their normative sample, and therefore have disproportionate representation of student performance based on individual backgrounds. These factors have also led to an increase in the over- and under-identification of impairments in children, not due to a language disability, but a language difference.

### **Sensitivity and Specificity**

To understand another flaw found in static measures of prereading ability, the factors of sensitivity and specificity must be considered. A test with high sensitivity means that it will accurately identify a person with a disorder. Conversely, a test with high specificity will correctly find and identify those without a disorder. Although static measures are commonly used in the field of education, research has found that most of these assessments yield sensitivity and specificity in the poor to fair range (< 70% -80%; Petersen et al., 2018) at the beginning of the school year, especially with CLD students. If sensitivity and specificity are in the poor range, then the test is not accurately measuring what it is intended to. As indicated by Petersen et al. (2018), evaluating kindergarten students using static measures leads to very poor specificity (< 70%); which results in overidentifying students who are at risk for future reading difficulty. This

can be a costly mistake by using limited resources for students who do not need it, while conversely having students who need the support go undetected or underserved.

### **Dynamic Assessment**

To reduce the negative impact that static measures have on identifying oral language and pre-reading ability, we need an alternative approach. Dynamic assessment can measure not only what a child can understand, but also how well a child learns a new concept. Thus, contrary to static assessment, dynamic assessment measures a child's potential to learn, rather than perform, a specific task. Dynamic assessment includes three specific phases; a pretest, a period to teach the desired skill, and a posttest. According to Petersen et al. (2018), this model, in theory, should be able to measure a child's ability to read before the child has had any formal reading instruction. Dynamic assessment adopts modifiability scoring, which uses the judgement of the examiner and the efforts spent on the child to facilitate learning during the teaching phase of the dynamic assessment process. Modifiability scores have been shown to produce high sensitivity (correctly identifying a problem) and specificity (correctly identifying no problem) and are often measured using Likert scales that ask the examiner to reflect on the teaching experience with the student (Petersen et al., 2018). Sensitivity and specificity are often improved when adopting a dynamic assessment method because learning potential is being measured. Examiners have the option to look at the student's performance on the pretest, the posttest, the length of the teaching phase, and how much support (modifiability) was needed during the teaching phase. Thus, while static assessments use only one assessment point, dynamic testing uses multiple variables that focus on learning to come to a classification decision (Camilleri & Botting, 2013).

One of the hallmark features of dynamic assessment is the teaching phase. This instructional period has a specific purpose which attends to three variables: (a) intentionality and

reciprocity, (b) transcendence, and (c) meaning. *Intentionality and reciprocity* are when the examiner and student are both engaged in the learning process by focusing on the interaction and teaching, rather than treating the process as an assessment. *Transcendence* is the ability to take a newly learned behavior and apply it in different contexts. *Meaning* explains the “why” behind the assessment and teaching experiences with the students. Incorporating these three variables by giving feedback, encouragement, and helping the student understand the reason and purpose of testing, means dynamic assessment can help mitigate negative impacts of static measures (Clark, 2019).

Another major benefit to using dynamic assessment is the ability to mitigate testing bias. Measuring a child’s ability to learn reduces the potential for one-time static measures to be skewed by confounding factors including socioeconomic status, cultural and linguistically diverse backgrounds, and prior exposure to testing materials. However, there is limited research using dynamic assessments to identify reading difficulties. In order for dynamic assessment to be used more widely in school settings, there needs to be strong evidence to support the predictive validity of the testing measures for decoding and language (Petersen et al., 2018).

### **Dynamic Assessment of Decoding**

There has been an additional focus in the research on examining the predictive validity of dynamic assessment of decoding. For example, Fuchs et al. (2011), conducted a study to examine the construct and predictive validity of a dynamic assessment of decoding. Their study assessed 318 students in the fall of first grade on a wide variety of testing measures in hopes to predict student response to reading instruction. The testing battery included dynamic assessments as well as static measures of early alphabetic knowledge, rapid automatized naming, phonemic awareness, oral vocabulary, listening comprehension, attentive behavior, and hyperactive or



impulsive behavior. Results of their study found that when static measures of decoding and reading ability were combined with their dynamic assessment of decoding and word identification, there was strong predictive validity of later reading abilities.

To mitigate the influence and effects that prior reading instruction has on statically assessed decoding skills, Cho and Compton (2015) introduced a new orthographic system as part of their dynamic assessment. Using domain-specific dynamic assessment tasks (decoding), they sought to examine whether learning potential is generalized across domains such as reading and arithmetic outcomes. This study assessed 112 native English speakers in first grade. Each student was given a dynamic assessment that measured learning novel-symbol sound correspondence, blending sounds, and inferring decoding rules. Students in this study were also given a battery of static reading and arithmetic measures to see if the dynamic assessment of decoding better represented the target construct of early reading learning potential. Results of their study found that constructs of the dynamic assessment of decoding were influenced by factors such as attention, motivation, and core phonological processing abilities. The results indicated that dynamic assessment of decoding provides important information of individual differences in decoding ability and can be used as a helpful tool for professionals to correctly identify students with reading difficulties, but that it does not generalize across domains (mathematics).

Although the Cho and Compton (2015) and Fuchs et al. (2011) studies on dynamic assessment of decoding reported promising results, they also had several limitations, including no report on sensitivity and specificity, low variance in student demographics, and no modifiability scoring (i.e., an examiner's judgement of a student's learning potential). They also used dynamic procedures that took considerable time and examined students only during the first-grade year after they had already received formal decoding instruction in kindergarten. To

address these weaknesses, Petersen et al. (2016) compared the predictive validity of current static reading measures used in public schools in the state of Utah to the predictive validity of a dynamic assessment of decoding. The researchers included 600 diverse participants (300 Caucasian and 300 Hispanic, with 75% of them being first language Spanish-speaking English language learners) to introduce a greater variety of student demographics. All participants were administered two static reading measures at the beginning of kindergarten, as well as one dynamic assessment of decoding which took approximately 3 minutes to administer. After the initial assessment, the students were split into two separate groups for the teaching phase. One group was taught decoding skills using an onset-rime strategy, while the second was taught using a sound-by-sound strategy. Both used measures of modifiability after the teaching phase was concluded. At the end of first grade, all students were administered norm-referenced tests of reading. Results of this study found that the dynamic assessment yielded over 80% sensitivity and specificity for the White English-speaking students, with 85% sensitivity and 70% or higher specificity for the Hispanic students; indicating that the dynamic assessment did in fact classify all students more accurately than the combined two static measures of reading ability.

Petersen et al. (2018) sought to increase the evidence for the classification accuracy of their dynamic assessment of decoding by conducting a longitudinal study that screened early kindergarten students for decoding ability and then followed them into later elementary school years. After the initial assessment in kindergarten, the researchers assessed reading ability in second through fifth grades using the DIBELS Next Oral Reading Fluency (DORF) test, which calculates the number of correct words read per minute minus any word errors made. Students were classified as having decoding difficulty at the end of second through fifth grade if they received DORF scores at or below the 7th percentile. Their study concluded that using dynamic

assessments of decoding as a measure for future reading ability in kindergarten is a good predictor of reading ability up to 6 years into the future. They found that combining static and dynamic measures accounted for higher variance than the dynamic assessment alone but combining the two types of assessments did not improve the classification accuracy.

### **Dynamic Assessment of Language**

Various studies have examined the role of using a dynamic assessment of language in classifying language disorder. For example, Kapantzoglou et al. (2012), Peña et al. (1992), and Ukrainetz et al. (2000), sought to investigate the classification accuracy of varying dynamic assessments of vocabulary. Narrative language has been one of the primary foci of research conducted for dynamic assessments of language (Orellana et al., 2019). Examining narrative discourse is a useful language assessment approach because it is highly effective in measuring complex language ability. Narratives use story grammar elements including setting, problem, feeling, attempt, consequence, and resolution, which give rich context that integrates academic language in a naturalistic manner (Petersen & Gillam, 2015; Ukrainetz et al., 2000).

Further research has been conducted to help examine the diagnostic accuracy of dynamic assessments of language in children with CLD backgrounds. Orellana et al. (2019) completed a systematic review of dynamic assessment of language research with bilingual children. They used a meta-analysis structure to examine six studies including Kapantzoglou et al. (2012), Kramer et al. (2009), Peña et al. (2014), Peña et al. (2001), Petersen et al. (2018), and Roseberry and Connell (1991). Two key factors were identified as being the most effective measure for diagnostic accuracy of dynamic assessments of language in CLD populations: (a) children with language disorder (LD) performed substantially lower than their typically developing (TD) peers on both the pre-and post-tests of language and (b) modifiability scores given by clinicians were

significantly lower in the students with LD, than their TD control groups. While this meta-analysis concluded that using dynamic assessment of language can increase the diagnostic accuracy and identification of language disorder in CLD populations, the studies presented with weaknesses. Only half of the studies included an adequate sample size (30 or more). Peña et al. (2001) included 55 participants, Peña et al. (2014) had 54 participants, and Petersen et al. (2018) included 42 participants. Only five studies had high replicability of the procedures used in their research, and no studies included a representative ratio of students with and without language disorder.

Future research should address some of the weaknesses identified in previous studies of dynamic assessment, and in the systematic review completed by Orellana et al. (2019). Research conducted should identify which variables of dynamic assessment are the most effective in predicting language ability, they should include a larger sample size, have better representation of students both with and without language disorder, blind researchers to a child's language ability before testing, and increase cross-validation by replicating studies to support previous findings. Furthermore, it is imperative that language disorder is diagnosed accurately prior to investigating sensitivity and specificity. This reference standard must be valid because language disorder is a learning disability (Sun & Wallach, 2014), and a student's response to evidence-based instruction could be considered a gold standard diagnostic approach for language disorder.

There are also limitations in previous research that focused on the predictive validity and classification accuracy of dynamic assessments of decoding. In the research conducted by Petersen et al. (2018), sensitivity and specificity were in the fair to good range, which was a result of the combination of both dynamic and static measures which yield high floor effects. Cho and Compton (2015) sought to reduce the effects of static measures by creating a dynamic

assessment that introduced novel-symbol sound correspondence but included only a small sample size. Future research should cross-validate dynamic assessment of decoding research and introduce innovative approaches that might yield greater sensitivity and specificity. One possible method to yield greater sensitivity and specificity to validate dynamic assessment of decoding is to administer the standard posttest again after a period of time, and to also administer a delayed novel posttest. Doing so requires the student to recall and use the skills taught during the teaching phase of the dynamic assessment to decode unfamiliar words, which may help predict future decoding problems. Such an assessment could yield greater sensitivity and specificity by focusing on the ability for a student to generalize the skill necessary to decode instead of remembering specific examples from pre- to posttest.

These dynamic approaches to measuring decoding must yield results that are significantly different from the pretest, must have preliminary evidence that they do not yield considerable floor effects (i.e., there is variance in the scores), and must have evidence that they can be administered and scored with fidelity and reliability prior to investigating other aspects of validity such as sensitivity and specificity. Further, no longitudinal studies of dynamic assessment of oral language have been conducted with young students. Thus, only concurrent evidence of validity is available for the dynamic assessment of language. However, a dynamic assessment can also be used to predict future difficulty, and future research should explore predictive validity.

### **Statement of Purpose**

The purpose of this study was to (a) examine whether students made gains from the pretest to a delayed standard posttest and a delayed novel posttest of the Predictive Early Assessment of Reading and Language (PEARL) decoding subtest indicating that pretest floor

effects were mitigated, (b) to determine whether there was variance in the delayed standard and novel decoding posttest scores, (c) examine whether there was adequate fidelity of administration and adequate inter-rater reliability of the delayed standard and novel decoding subtests, and (d) examine the predictive validity of the language subtest of the PEARL dynamic assessment when administered to kindergarten students in the middle of the school year and after the students have received oral narrative language intervention.

### **Research Questions**

1. For kindergarten students, (a) are the scores from a delayed standard posttest and delayed novel decoding posttest of the PEARL dynamic assessment significantly higher from the scores from the pretest (indicating that students made gains from pretest to posttest and that floor effects were reduced)? (b) Was there variance in the delayed standard and novel decoding posttest scores (floor effects were mitigated)?
2. Do delayed standard and a delayed novel decoding posttests of the PEARL have adequate fidelity of administration and adequate inter-rater reliability?
3. What is the classification accuracy of the language subtest of the PEARL dynamic assessment for kindergarten students with language disorder?

### **Method**

#### **Participants**

This research study was approved by the Brigham Young University Institutional Review Board (IRB). We administered a dynamic assessment of language and decoding (the PEARL) to 34 kindergarten students from one elementary school who were identified as at-risk in the fall from the DIBELS composite score (Good et al., 2004), which was a combination of letter naming and phonemic awareness tasks. Of those 34 students, the PEARL identified 7 students at

risk for decoding and 11 students at risk for language. Demographic information of the participants is displayed in Table 1.

**Table 1**

*Descriptive Data*

Demographic Variable	n (%)
<b>Gender</b>	
Male	16 (47%)
Female	18 (53%)
<b>Ethnicity</b>	
White non-Hispanic	18 (53%)
Hispanic	16 (47%)
<b>SES (Free/Reduced Lunch)</b>	
Yes	17 (50%)
No	17 (50%)
<b>ELL</b>	
Yes	15 (44%)
No	19 (56%)

**Measures**

***Predictive Early Assessment of Reading and Language (PEARL)***

In the middle of the school year, each kindergarten student was administered the Predictive Early Assessment of Reading and Language (PEARL). The PEARL dynamic assessment contains two brief subtests. Subtest 1 examines a child's ability to learn to decode

using consonant-vowel-consonant (CVC) nonsense words. During the pretest phase, four nonsense words are given to the student, and they are asked to read the words. The test administrator is allowed to deliver neutral prompts if encouragement is needed for the student. If the child fails to read two or more words accurately, the examiner will move to the teaching phase. Using a script, the child was taught how to read the four CVC words by repeating each sound in the word and then learning how to blend the sounds together. After the teaching phase was completed, the examiner filled out a responsiveness rating scale in real time, which rates the student's errors, confidence, disruptions, rate of learning, and the examiners overall judgement of a student's ability to learn to decode. This learning score was on a scale from 0-4, where 0 = difficult and 4 = easy to learn. The post-test phase began once the responsiveness scale was completed by the examiner. During the post-test measure, the student was asked to read the same four CVC words in a different order and was given a score based on the number of correct sounds and correct words read. Modifications of the decoding subtest of the PEARL, for this study, included (a) a delayed standard posttest where after approximately 7 minutes the participants were asked to decode the same posttest again and (b) a delayed novel set of consonant-vowel-consonant nonsense words with the same vowel ('a') and the same consonants from the standard posttest, but in a different order were presented to the student to decode.

Subtest 2 of the PEARL is the Dynamic Assessment of Language, which measures a child's oral language comprehension. In this subtest, the child is read a short narrative and is then asked to retell the story independently. While scoring in real-time, the examiner awarded points on the retell based on the number of story grammar elements that were included (e.g., character, problem, feeling, action, and ending), the language complexity (the inclusion of temporal, adverbial, and relative clauses), and the episode complexity (groupings of story grammar



elements, e.g., problem, consequence, and ending). Once the pretest was completed and scored, the examiner began the teaching phase by showing the student how to retell the same story using pictures and icons. After the brief teaching phase, the examiner completed a responsiveness scale similar to the one completed in the decoding subtest, to measure the child's language learning potential. The posttest portion of the language subtest used the same format as the pretest, but the child listened to a different story. Previous research conducted to validate the use of the PEARL language subtest yielded 92-100% sensitivity and 84-89% specificity, while the PEARL decoding subtest reports over 80% sensitivity and specificity.

### ***Tier-2 Oral Narrative and Expository Language Intervention***

The 11 students identified as at-risk for language learning difficulty from the PEARL language subtest, received small group Story Champs language intervention that focused on oral narrative retelling, personal story generation, and informational expository text retelling. The oral narrative-based language intervention used personal themed stories from Story Champs that increased in complexity over time. The expository passages used during the expository intervention sessions were chosen from a database of materials aligned with Common Core State Standards. According to a systematic review of 24 research articles published from 1993 to 2018 regarding narrative language intervention, Favot et al. (2021) concluded that oral narrative intervention is effective for children with language disorders. When intervention utilizes icons, visuals, requires students to generate narratives, and allows for clinician modeling (which are all aspects of Story Champs) effectiveness of intervention increases.

Treatment was delivered in small groups ranging from two to four children with one or two research assistants. Research assistants were trained prior to Tier-2 intervention across two 1-hour training sessions. Treatment sessions were completed in the students' school, most often

in a large room with several desks and chairs. Each intervention session lasted approximately 15-20 minutes. The Story Champs small group procedures are outlined in detail in the Story Champs manual (Spencer & Petersen, 2012). The interventionists were also asked to complete a student modifiability rating form at the end of each intervention session. This modifiability form was similar to the one used in the PEARL (more information provided in the PEARL manual). Five specific behaviors such as rate of learning and disruptions were scored on a scale from 0-4 which yielded a modifiability score ranging from 0-20. Examiners also completed an overall learning rating scale ranging from 0-4, reflecting how difficult it was for the child to learn. This score was referred to as the final modifiability rating. The modifiability rating that ranged from 0-20 and the final modifiability rating that ranged from 0-4 was combined to generate a total modifiability score. The mean intervention modifiability scores were calculated for the final modifiability ratings and the total modifiability ratings each student was given by the individual interventionists during each session.

### **Reference Standard**

In order to examine the predictive validity of the language subtest of the PEARL, students identified as having weak language from the PEARL dynamic assessment received small group oral narrative and expository language intervention two times per week for approximately 6 weeks. In order to determine the sensitivity of the language subtest of the PEARL dynamic assessment, first, a specific reference standard was set to establish whether the students had a language disorder or not. To establish the reference standard, researchers examined the posttest expository retell, posttest narrative retell, and posttest personal story scores at the end of the 6 weeks of intervention. They also examined the mean intervention modifiability scores, qualitative modifiability comments from interventionists, and whether or

not the student had an existing language-related Individualized Education Program. To be identified as having a language disorder, students needed to meet at least three of the following criteria: (a) posttest scores that were 4 or lower for the expository retell, narrative retell, or personal story generation, (b) a final modifiability score below 3, (c) a total modifiability score below 19, and (d) a qualitative rating score less than 3.

For students who are English language learners (ELL), posttest expository, narrative, and personal story scores were not taken into account due to the biased nature of such static assessments, and determination of language disorder was based on the modifiability ratings and the qualitative notes. In order to establish language disorder, ELL students also had to meet at least one of the following criteria: (a) a final modifiability score below 3, (b) a total modifiability score below 19, and (c) a qualitative rating score less than 3.

### **Fidelity**

A team of undergraduate students in the communication disorders program at Brigham Young University were trained on the administration and scoring of the PEARL dynamic assessment. Four research assistants who were team leaders received extensive training over several hours. The other research assistants of the assessment team were given a training which lasted approximately 1 hour and were certified and approved by a team leader, suggesting that the individual was ready to administer the assessments on their own. Research assistants administered five practice sessions including using the modifiability rating form on fellow research assistants playing the role of children with and without language disorder. Each team member was required to demonstrate competence and the ability to adhere to and carry out the testing procedures consistently and independently with 100% accuracy. There was always a team leader and/or an experienced, trained individual onsite to monitor adherence to testing

procedures. The researchers observed and provided feedback until each examiner was able to deliver the dynamic assessment procedures independently and accurately. Fidelity was monitored by team leaders while examiners were administering the dynamic assessment.

### **Inter-Rater Reliability**

Inter-rater reliability was examined for approximately 25% (9/34) of the children's scores. The tests rescored were randomly selected using a random number generator. The first author independently performed the re-scoring and was blind to the children's language status (i.e., typically developing, disorder) to eliminate researcher bias. This independent examiner completed the same training requirements as the primary research assistants and demonstrated the ability to administer and score the dynamic assessment accurately. This individual listened to the audio files corresponding to the selected tests and scored the delayed standard decoding posttest and the delayed novel decoding posttest in real-time. The scores from the independent examiner were compared to the total scores given by the initial examiner. The percent agreement and the range of agreement were analyzed.

## **Results**

### **Research Question 1 Results**

In order to answer research question 1a, we examined whether the scores from the delayed standard posttest and the delayed novel decoding posttest of the PEARL were significantly different from the scores of the pretest, we conducted a series of independent samples t-tests (see Table 2). Results indicated that the delayed standard posttest correct sounds ( $M = 6.94$ ,  $SD = 4.77$ ) were significantly different from the pretest correct sounds ( $M = 5.35$ ,  $SD = 5.06$ ,  $t(33) = -3.04$ ,  $p = < 0.01$ , Cohen's  $d = 3.05$ ). Also, the delayed novel posttest correct sounds ( $M = 7.00$ ,  $SD = 4.72$ ) were significantly different from the pretest correct sounds ( $M =$

5.35,  $SD = 5.06$ ,  $t(33) = -2.91$ ,  $p = < 0.01$ , Cohen's  $d = 3.30$ ). For the delayed standard posttest correct words ( $M = 1.56$ ,  $SD = 1.58$ ) there was a significant difference when compared to the pretest correct words ( $M = 1.00$ ,  $SD = 1.61$ ,  $t(33) = -2.64$ ,  $p = < 0.01$ , Cohen's  $d = 1.24$ ). The delayed posttest novel words ( $M = 1.41$ ,  $SD = 1.65$ ) were not significantly different from the pretest words ( $M = 1.00$ ,  $SD = 1.61$ ,  $t(33) = -1.72$ ,  $p = < 0.10$ , Cohen's  $d = 1.40$ ). In order to answer question 1b, we examined the means and standard deviations of the delayed standard and novel posttest scores. The means and standard deviations indicated that there were limited floor effects and there was considerable variance in the delayed standard and novel posttest scores.

**Table 2**

*PEARL Decoding Subtest Statistics*

	Pre-Correct Sounds	Pre-Correct Words	Sec. Post Correct Sounds	Sec. Post Correct Words	Mod. Post Correct Sounds	Mod. Post Correct Words
Mean	5.35	1.00	6.94	1.56	7.00	1.41
SD	5.06	1.61	4.76	1.58	4.71	1.65

**Research Question 2 Results**

***Fidelity***

Fidelity of administration was calculated from ten separate administrators, with 100% fidelity of administration for nine of those administrations. One examiner prompted a student to read the sounds and then also prompted the student to read the words, which was not in line with the standardized administrative procedures.

### ***Inter-Rater Reliability***

The point-to-point inter-rater reliability of the delayed standard posttest sounds, and the delayed novel posttest sounds was 94% and the point-to-point inter-rater reliability of the total words read correctly was 94%.

### **Research Question 3 Results**

Using the reference standard for a language disorder, 11 students were confirmed to have language disorder. We then compared the results of the language subtest of the PEARL dynamic assessment to determine whether the PEARL correctly identified these 11 students as having language disorder. Results indicated that the PEARL correctly identified 9 out of 11 students as having language disorder, yielding a sensitivity index of 82%.

### **Discussion**

The purpose of this study was to (a) determine if students made gains on the PEARL decoding subtest from pretest to a delayed standard and novel posttest, indicating that floor effects were mitigated, (b) to examine if there was variance in the delayed standard and novel decoding posttest scores, (c) to determine if there was adequate fidelity of administration and adequate inter-rater reliability of the delayed standard and novel decoding subtest, and (d) to examine the predictive validity of the language subtest of the PEARL dynamic assessment. Independent samples *t*-tests indicated that the results of the PEARL dynamic assessment of decoding for correct sounds and correct words from pretest to a delayed standard and delayed novel posttest were significantly different, and that there was variance in the scores, indicating that students did in fact make gains from pretest to posttest and that floor effects were reduced. Inter-rater reliability was examined and calculated for approximately 25% (9 out of 34) of the children's scores. The point-to-point inter-rater reliability of the delayed standard posttest

sounds, novel sounds, and total words read correctly was 94%. An analysis of the data indicated that the dynamic assessment of language had good sensitivity.

### **Pretest to Posttest Gains**

In 2018, Petersen et al. found that using a pretest, teach, posttest dynamic assessment approach to predict kindergarten students' decoding ability out to fifth grade yielded higher and more balanced sensitivity and specificity indexes (approximately 80% or above) when compared to kindergarten static measures. A long-term aim of the current research is to determine whether a delayed standard posttest and a modification of the posttest of the dynamic assessment of decoding will yield even better sensitivity and specificity. The first step of this project was to design a novel decoding subtest of the PEARL. It was important to obtain evidence that students could maintain and transfer some of what was taught during the teaching phase of the dynamic assessment to these delayed standard and novel posttests. These new posttests of the dynamic assessment cannot have floor effects with scores demonstrating limited variance. Obtaining evidence that there is variance in the new posttest scores is important because there would be no way to differentiate and identify which students' may or may not be at risk for decoding difficulty if all students perform similarly. Showing gains from pretest to posttest is also crucial for dynamic assessment because it is important to know how well children respond to the instruction in the dynamic assessment which can identify those who are at risk for decoding difficulty.

The gains noted in the current study from pretest to posttest provides evidence that the new decoding posttests of the PEARL didn't have floor effects. For example, mean decoding scores at pretest were 5.35 for sounds and 1.00 for words, whereas mean decoding scores for the delayed standard posttest correct sounds were 6.94 and 1.56 for correct words. The delayed

novel posttest means for correct sounds was 7.00 and 1.41 for correct words. Performance on the pretest indicates that many of the students could not decode most of the nonsense words or even the letter sound correspondence, indicating a floor effect on the static pretest measure. Measures of reading/decoding ability consistently over- and under-identify students who may be struggling with decoding skills because there are children who will perform similarly on a static assessment but will require different amounts of support to reach the target level of understanding and competency (Cho & Compton, 2015). This problem leads to various negative consequences within the school system. It often results in all students who struggled on the one-time static measure to receive extra (costly) intervention. This persistent problem violates a student's legal right to receive an education in the least restrictive environment. Fuchs et al. (2011) state that floor effects arise when using static measures because they (a) devote too few of items on a test to sample elementary level skills and (b) they only assess unaided success or unaided failure. They concluded that to mitigate these issues, dynamic assessment should be used in conjunction with current static measures to assess pre-reading and reading skills in students because dynamic assessment shows a child's readiness for change and potential to learn the skill.

Gains made from pretest to posttest are crucial to classify decoding difficulty. Pretest to posttest gains tell us that students who were ready and able to learn the skill, had the ability to remember and recall the instruction received during the teaching phase of the dynamic assessment, and were able to draw upon that knowledge in the delayed standard and novel posttests. This tells us that the means were in fact higher, and floor effects were mitigated after the teaching phase. Because all the students in the current study were at risk for decoding difficulty, it will be important to administer the delayed standard and novel decoding posttests to students not at-risk to determine whether their gains from pretest to posttest are even greater.



### **Fidelity and Inter-Rater Reliability**

Previous research has indicated that dynamic assessments can have high fidelity and inter-rater reliability. The current research adds to the findings from Cho and Compton (2015), Fuchs et al. (2011), Miller et al. (2001), Orellana et al. (2019), Petersen et al. (2016), and Petersen et al. (2018) that a domain specific dynamic assessment of decoding and language (such as the PEARL) can be administered by adequately trained professionals. It is crucial that all assessments have clear instruction and administrative guidelines to promote high inter-rater reliability and fidelity among all users when administering and scoring the test to improve reliability. Petersen et al. (2016) sought to ensure high fidelity and inter-rater reliability in their study by training all research examiners on how to administer and score the dynamic assessment before administering the test to students. Examiners were required to perform practice examinations and pass off administration and scoring tasks with 100% fidelity through the first and second authors of the study. Likewise, the current study sought to maintain high inter-rater reliability on the new decoding posttest of the PEARL by completing various trainings with all research examiners on administration and scoring guidelines. The point-to-point inter-rater reliability of the delayed standard posttest sounds, and delayed novel sounds of the PEARL dynamic assessment was 94% and the point-to-point inter-rater reliability of the total words read correctly on the delayed standard and novel posttests of the PEARL was 94%. These results support previous findings that dynamic assessments can achieve high inter-rater reliability.

### **Sensitivity of the Language Subtest of the PEARL**

Peña et al. (2006) and Peña et al. (2014) found that the classification accuracy of language disorder in culturally and linguistically diverse school-age children when using dynamic assessment of storytelling measures and response to intervention (modifiability) ratings

was excellent. Similar to Peña et al., the current study required that the kindergarten students had to meet specific criteria in order to be identified as having a language disorder and were administered a dynamic assessment of narrative ability. Previous research supports findings from the current study, that using a dynamic assessment (test, teach, test model) of storytelling and narrative ability, and completing modifiability ratings yields good (75%-82%) classification results. Using only static assessment data can lead to over and under- identification of language impairment (Laing & Kamhi, 2003; Shenoy, 2014) and can result in a lack of resources for students who need it, and intervention provided to those who do not. Dynamic assessment is a potential solution to this problem, and the research evidence to support dynamic assessment is mounting (Orellana et al., 2019).

The current study sought to increase the evidence of using dynamic assessment and modifiability ratings to identify students with language disorder by administering the PEARL dynamic assessment language subtest and by providing narrative language intervention to the students identified as at-risk for language disorder by the PEARL. Students who were identified as having poor language ability consistently demonstrated poor response to the intervention and had low scores at posttest. Using the response to intervention as a reference standard could be considered a gold-standard for identifying language disorder. Language disorder is a language learning disability, thus students who have difficulty learning language are the very same students who have a language disorder. Using response to intervention as a reference standard and finding that the results from this study using the PEARL dynamic assessment language subtest closely aligns with previous findings is promising.

## **Limitations**

There are a few limitations to this study that are worth noting. Each interventionist was required to complete a modifiability rating of every student after the intervention sessions. This rating scale could be subject to bias due to individual differences and background of each examiner such as: experience working with children, experience with students from CLD backgrounds, or experience in schools. These differences could alter their views and ratings of each child. To eliminate potential bias, future research should focus on determining inter-rater reliability of the modifiability rating scale used in this study for the intervention sessions.

Having only a limited number of examiners and interventionists required that the research team only provide intervention to the students identified as at-risk for language disorder. Due to these limited resources, specificity was not reported in the current study. The study also included a small sample size from three separate classrooms from one elementary school within one school district. This small sample limits the generalization of the research findings to the general population.

## **Implications for Future Research**

In order to increase the evidence for external validity, further research should include a larger sample of participants, especially those that are culturally and linguistically diverse. Researchers should also provide narrative language intervention for students identified as typically developing and for those who are identified as at-risk. Doing so will provide response to intervention data to compare the at-risk and typically developing groups to provide both sensitivity and specificity indexes that could add greater evidence to support using dynamic assessment to identify language disorder in kindergarten students.

## **Conclusion**

The purpose of the study was to examine and validate the preliminary psychometrics of two new PEARL decoding subtests and to further investigate the predicative validity of the PEARL language subtest. The results of the study indicated that students did, in fact, make gains from pretest to both the delayed standard and novel posttests and that floor effects were reduced. The point-to-point inter-rater reliability of the delayed standard posttest sounds, delayed novel sounds, and total words read correctly was 94%. The dynamic assessment of language revealed good sensitivity. Results of this study also provide evidence that a dynamic assessment of language reflects a student's response to intervention over time. The findings of this study suggest that the modifications made to the PEARL decoding subtest revealed preliminary evidence of validity and reliability and that the language subtest of the PEARL could be a useful and valid tool to identify kindergarten students at risk for language disorder.

## References

- Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: The dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders, 48*(5), 565–581. <https://doi.org/10.1111/1460-6984.12033>
- Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y.-S. (2016). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities, 49*(5), 451–465. <https://doi.org/10.1177/0022219414556121>
- Cho, E., & Compton, D. L. (2015). Construct and incremental validity of dynamic assessment of decoding within and across domains. *Learning and Individual Differences, 37*, 183–196. <https://doi.org/10.1016/j.lindif.2014.10.004>
- Clark, K. D. (2019). *The cross-validation of the classification accuracy of a dynamic assessment of narrative language for school-age children with and without language disorder* [Master's thesis, Brigham Young University]. Brigham Young University ScholarsArchive. <https://scholarsarchive.byu.edu/etd/7606/>
- Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 Years: A systematic review. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.01515>
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the national reading panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287.
- Favot, K., Carter, M., & Stephenson, J. (2021). The effects of oral narrative intervention on the narratives of children with language disorder: A systematic literature review. *Journal of*

*Developmental and Physical Disabilities*, 33(4), 489-536.

<https://doi.org/10.1007/s10882-020-09763-9>

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read:

Implications for RTI frameworks. *Journal of Learning Disabilities*, 44(4), 339–347.

<https://doi.org/10.1177/0022219411407864>

Fuchs, L. S., & Deno, S. L. (1981a). *A comparison of reading placements based on teacher judgment, standardized testing, and curriculum-based assessment* (ED 211603) ERIC.

<https://files.eric.ed.gov/fulltext/ED211603.pdf>

Fuchs, L. S., & Deno, S. L. (1981b). *The relationship between curriculum-based mastery measures and standardized achievement tests in reading* (ED 212662). ERIC.

<https://files.eric.ed.gov/fulltext/ED212662.pdf>

Glover, T. A., & Albers C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, 45(2), 117-135.

<https://doi.org/10.1016/j.jsp.2006.05.005>

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., & Flindt, N.

(2004). *Technical adequacy of DIBELS: Results of the early childhood research institute on measuring growth and development* (Technical Report No. 7). University of Oregon.

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

Greenwood, C. R., Carta, J. J., & McConnell, S. (2011). Advances in measurement for universal screening and individual progress monitoring of young children. *Journal of Early*

*Intervention*, 33(4), 254–267. <https://doi.org/10.1177/1053815111428467>

- Jenkins, J. R., & Johnson, E. (2008). Universal screening for reading problems: Why and how should we do this? *RTI Action Network*.  
<http://www.rtinetwork.org/essential/assessment/screening/readingproblems#>
- Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81-96. [https://doi.org/10.1044/0161-1461\(2011/10-0095\)](https://doi.org/10.1044/0161-1461(2011/10-0095))
- Kramer, K., Mallett, P., Schneider, P., & Hayward, D. (2009). Dynamic assessment of narratives with grade 3 children in a first nations community. *Canadian Journal of Speech-Language Pathology & Audiology, 33*(3), 119-128.
- Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools, 34*(1), 44-55. [https://doi.org/10.1044/0161-1461\(2003/005\)](https://doi.org/10.1044/0161-1461(2003/005))
- Lyon, G. R. (2003). Reading disabilities: Why do some children have difficulty learning to read? What can be done about it? *Perspectives, 29*(2). Retrieved from <https://www.wrightslaw.com/info/read.disability.lyon.pdf>
- McConnell, S., & Wackerle-Hollman, A. (2016). Can we measure the transition to reading? General outcome measures and early literacy development from preschool to early elementary grades. *AERA Open, 2*(3), 1-15. <https://doi.org/10.1177/2332858416653756>
- Miller, L., Gillam, R. B., & Peña, E. (2001). *Dynamic assessment and intervention: Improving children's narrative abilities*. Pro-Ed.
- Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal*

- of Speech-Language Pathology*, 28(3), 1298-1317. [https://doi.org/10.1044/2019\\_AJSLP-18-0202](https://doi.org/10.1044/2019_AJSLP-18-0202)
- Peña, E. D., Gillam, R. B., & Bedore, L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, 57(6), 2208-2220. [https://doi.org/10.1044/2014\\_JSLHR-L-13-0151](https://doi.org/10.1044/2014_JSLHR-L-13-0151)
- Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, 49(5), 1037-1057. [https://doi.org/10.1044/1092-4388\(2006/074\)](https://doi.org/10.1044/1092-4388(2006/074))
- Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, 10(2), 138-154. [https://doi.org/10.1044/1058-0360\(2001/014\)](https://doi.org/10.1044/1058-0360(2001/014))
- Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A nonbiased procedure. *The Journal of Special Education*, 26(3), 269-280. <http://dx.doi.org/10.1177/002246699202600304>
- Petersen, D. B., Allen, M. M., & Spencer, T. D. (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities*, 49(2), 200-215. <https://doi.org/10.1177/0022219414538518>
- Petersen, D. B., & Gillam, R. B. (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, 48(1), 3-21. <https://doi.org/10.1177/0022219413486930>



- Petersen, D. B., Gragg, S. L., & Spencer, T. D. (2018). Predicting reading problems 6 years into the future: Dynamic assessment reduces bias and increases classification accuracy. *Language, Speech, and Hearing Services in Schools, 49*(4), 875-888. [https://doi.org/10.1044/2018\\_LSHSS-DYSLC-18-0021](https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021)
- Roseberry, C. A., & Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 34*(3), 596-603. <https://doi.org/10.1044/jshr.3403.596>
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook for research in early literacy* (pp. 97–110). Guilford Press.
- Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37*(1), 61-72. [https://doi.org/10.1044/0161-1461\(2006/007\)](https://doi.org/10.1044/0161-1461(2006/007))
- Spencer, T. D., & Petersen, D. B. (2012). *Story Champs: A multi-tiered language intervention program*. Language Dynamics Group.
- Shenoy, S. (2014). Assessment tools to differentiate between language differences and disorders in English language learners. *Berkeley Review of Education, 5*(1), 33–52.
- Sun, L., & Wallach, G. P. (2014). Language disorders are learning disabilities: Challenges on the divergent and diverse paths to language learning disability. *Topics in Language Disorders, 34*(1), 25-38. <https://doi.org/10.1097/TLD.0000000000000005>

- Thomas, A. S., & January, S. A. (2021). Evaluating the criterion validity and classification accuracy of universal screening measures in reading. *Assessment for Effective Intervention, 46*(2), 110–120. <https://doi.org/10.1177/1534508419857232>
- Ukrainetz, T. A., Harpell, S., Walsh, C., & Coyle, C. (2000). A preliminary investigation of dynamic assessment with Native American kindergartners. *Language, Speech, and Hearing Services in Schools, 31*(2), 142-154. <https://doi.org/10.1044/0161-1461.3102.142>
- VanMeveren, K., Hulac, D., & Wollersheim-Shervey, S. (2020). Universal screening methods and models: Diagnostic accuracy of reading assessments. *Assessment for Effective Intervention, 45*(4), 255–265. <https://doi.org/10.1177/1534508418819797>

## APPENDIX A

**Review of the Literature**

Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y.-S. (2016). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities, 49*(5), 451–465.

<https://doi.org/10.1177/0022219414556121>

**Objective:** To determine if measures of language ability would add to the prediction of problems in reading comprehension in addition to the predictors of word reading. To determine if independent measure of language predict reading comprehension difficulties independent of measures related to word reading.

**Method:** This was a longitudinal study that measured 366 kindergarten students until the end of third grade. These kindergarten students were identified as “at risk” for reading difficulties based on screenings from beginning of the school year DIBELS scores. All students underwent a screening battery and a portion of the students considered “at risk” were provided with direct language intervention. Progress was monitored over the kindergarten year, and then at the end of third grade, measures of reading comprehension were conducted, and children were split into two groups with either good or poor outcomes.

**Relevance to work:** Most children enter kindergarten with some letter sound and letter name knowledge but show significant increase in their skills because they receive direct instruction in the kindergarten classroom. This is not the same for language. Language acquisition happens earlier and continues to develop through the school-aged years, but it is not individually measured based on instruction received at school. This

makes it hard to determine what the language levels were for students before, during and after intervention for reading related difficulties.

Cho, E., & Compton, D. L. (2015). Construct and incremental validity of dynamic assessment of decoding within and across domains. *Learning and Individual Differences*, 37, 183–196. <https://doi.org/10.1016/j.lindif.2014.10.004>

**Objective:** This study had two purposes. 1) To test whether a dynamic assessment of decoding could measure early reading learning potential differently or more accurately than static decoding assessments; and 2) to determine the value of a dynamic assessment of decoding for explaining word reading and arithmetic performance beyond what can be measured and explained with static measures.

**Method:** They had 112 native English speaking first grade students from six schools and 20 classrooms participated in the study. They were given a dynamic assessment that measured learning novel-symbol sound correspondence, blending sounds, and inferring decoding rule. They were also given a battery of static reading and arithmetic measures to see if the dynamic assessment of decoding better represented the target construct of early reading learning potential as well as predicting future math performance.

**Relevance to work:** Using an assessment that is more accurate at measuring early reading learning potential in young students could bring a pivotal change for all types of assessment. Dynamic assessment measures a student's ability to learn a concept, which helps decrease testing bias in students from CLD backgrounds. This concept could be transferred over from other academic areas instead of just reading.

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y., & Cordier, R. (2017).

Psychometric properties of language assessments for children aged 4–12 Years: A systematic review. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.01515>

**Objective:** This systematic review’s purpose was to analyze the psychometric properties of available language assessments (more specifically for spoken language assessments) for school-aged children and to identify which tests have the best evidence to support their use of effectiveness.

**Method:** Different language assessments were analyzed for sound psychometric properties using the COSMIN taxonomy and checklist. Only used tests for mono-lingual English-speaking children and used tests from all over the world.

**Relevance to work:** Knowing the psychometric properties of various assessments can help clinicians across the nation choose the types of tests that will be most beneficial for their individual caseloads. Understanding the psychometric deficiencies in certain assessments can help decrease the problem of children being both over-diagnosed and under-diagnosed with a language disorder. Looking at systematic reviews such as this one will help clinicians find more information about the validity of tests so they can implement evidence-based practice even in the evaluation and diagnostic process.

Favot, K., Carter, M., & Stephenson, J. (2021). The effects of oral narrative intervention on the narratives of children with language disorder: A systematic literature review. *Journal of Developmental and Physical Disabilities, 33*(4), 489-536.

<https://doi.org/10.1007/s10882-020-09763-9>

**Objective:** The purpose of this study was to examine the quality, efficacy, and common features of oral narrative interventions on the narratives of children with language disorder.

**Method:** This was a systematic review that analyzed 24 research studies conducted from 1993 to 2018. They reviewed 11 single case research studies, 10 group studies, and one study that included both single case and group data, and two case studies.

**Relevance to work:** This systematic review is helpful in showcasing the importance of narrative intervention on improving oral language abilities in children with language disorder. Since narrative intervention is being implemented in our current research, it is beneficial to know the evidence-base for using this type of intervention to improve oral language skills in children with and without language disorder.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347.  
<https://doi.org/10.1177/0022219411407864>

**Objective:** The purpose of this study was to advance the predictive validity in the study response of reading instruction using dynamic assessment. Their goal was to build on the idea and notion of a response to intervention (RTI) framework used in the school systems to develop a testing battery that was quicker and more efficient in identifying students with reading disability.

**Method:** This study assessed 318 students in the fall of first grade on a wide variety of testing measures in hopes to advance predictive validity in the student response

of reading instruction. The testing battery included dynamic assessments as well as static measures of early alphabetic knowledge, rapid automatized naming, phonemic awareness, oral vocabulary, listening comprehension, attentive behavior, and hyperactive or impulsive behavior.

**Relevance to work:** RTI frameworks are used widely across school systems. It is a chance for educators to identify students who have difficulty learning skills and giving them the support needed to be successful. Developing a testing battery that could give the same type of information quicker will be beneficial to SLP's and educators working to identify students at risk for reading, language, and learning disabilities.

Fuchs, L. S., Deno, S. L. (1981a). *A comparison of reading placements based on teacher judgment, standardized testing, and curriculum-based assessment* (ED 211603) ERIC. <https://files.eric.ed.gov/fulltext/ED211603.pdf>

**Objective:** To investigate the accuracy with which reading instruction placements are made in schools. Reading instruction levels are generally determined based on standardized measures, informal reading inventories and previous teacher reports/judgements. This comparison is useful to determine which placement procedure is most accurate.

**Method:** A group of 91 students were randomly selected from one elementary school. All children spoke English, while 15 received special education, and 23 were enrolled in Title 1 reading programs. Each student was given the Woodcock Reading Mastery Test, the Word Identification Test and the Passage Comprehension. Teachers reported the reading level group each student participated in for reading instruction and

reading passages from the Ginn 720 were given. All students were tested individually during a 45-to-60-minute session.

**Relevance to work:** High level of inaccuracy between test scores and teachers reading placements. Shows the inaccuracy and poor correlations between different testing measures. The effects of reading placement and testing relate to measures of diagnosing language disorders in young children.

Fuchs, L. S., Deno, S. L. (1981b). *The relationship between curriculum-based mastery measures and standardized achievement tests in reading* (ED 212662). ERIC.

<https://files.eric.ed.gov/fulltext/ED212662.pdf>

**Objective:** The objective of this study was to determine if performance on simple curriculum-based measures demonstrate concurrent validity with respect to performance on standardized reading achievement tests, and how results on these measures are dependent upon the instructional criterion employed.

**Method:** 91 randomly selected English speaking students (grades 1-6) were given three measures; standardized achievement tests, teacher judgments, and graded reading passages. Depending on the child's score, they were assigned a different instructional level based on a specific criterion to see if there is a correlation between testing measures and the instruction given to each student.

**Relevance to work:** Standardized measurements are most used to identify reading difficulties in children. The correlation is high between instruction and word identification and reading comprehension regardless of what level of instruction was given.



Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81-96. [https://doi.org/10.1044/0161-1461\(2011/10-0095\)](https://doi.org/10.1044/0161-1461(2011/10-0095))

**Objective:** This study aimed to examine if using a dynamic assessment method of word learning skills is an effective approach to identify bilingual preschool children with primary language impairment.

**Method:** They had 28 preschool participants in the study who were all predominantly Spanish speaking. Each child participated in three sessions of preliminary testing and a three-phase dynamic assessment task of word-learning skills. All testing was done in their native language of Spanish.

**Relevance to work:** Their study found that typically developing bilingual children were able to make phonological and semantic connections to learn the new words faster than the children with language impairment. The dynamic assessment showed that typically developing children often had greater modifiability scores and had an easier time learning the strategies than the bilingual children with language impairment. This is helpful for the current study because it provides evidence that dynamic assessment can be a promising method to identify language impairment in mono-English speaking children, as well as bilingual-Spanish and English-speaking students.

Kramer, K., Mallett, P., Schneider, P., & Hayward, D. (2009). Dynamic assessment of narratives with grade 3 children in a first nations community. *Canadian Journal of Speech-Language Pathology & Audiology, 33*(3), 119-128.

**Objective:** Since many current language assessment methods contain cultural biases, this study sought to use the Dynamic Assessment and Intervention tool as a way to assess language differences from language disorders in a First Nations community. They also sought to compare their findings to test results and classification groups made by school personnel.

**Method:** They had 17 third grade students that were both English and Cree speaking participate in a test, teach, re-test dynamic assessment of narrative skills. Each phase was held three or more days apart from one another, with the teaching phase split into two different intervention days. Examiners rated each child on a scale of 1-5 to describe the amount of effort required to teach the child (1= a lot of effort, 5 = minimal effort).

**Relevance to work:** Storytelling is a skill that is used across cultures. Using this as a measure for language learning disorder can be very effective, especially when it involves measuring how much effort it took for the student to learn the skill. This same dynamic assessment method of narrative skills will be involved in the current study, with an emphasis on the individual examiners recording the modifiability ratings for each child.

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

**Objective:** This article aims to discuss the current problems and recent solutions made in regard to the use of norm-referenced tests with children from CLD backgrounds. The article also suggests two promising procedures to provide unbiased testing measures to

children from CLD backgrounds by implementing processing-dependent measures and dynamic assessment.

**Method:** This article did not have a specific study or research question, but rather discussed the importance of using different methods to measure language abilities in children from CLD backgrounds. They discussed the importance of using criterion-references measures, ethnographic interviewing, language sampling, and dynamic assessment to get a better diagnosis of language impairment with children who speak different languages.

**Relevance to work:** This article helps the current study by laying a foundation for the importance of using more than just norm-referenced, static assessments in the process of identifying and classifying language disorder. Having alternative measures helps improve classification accuracy for all school-aged children, and especially those from CLD backgrounds.

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

**Objective:** The purpose of this article was to complete a meta-analysis of the current use of dynamic assessment and its diagnostic accuracy for language impairment in young children who are bilingual.

**Method:** Using their key word search, they found seven corresponding studies with participants ranging from three to 8 years old. The dynamic assessment procedures used in these studies focused on vocabulary, morpheme rule learning, ability to learn

nonwords, and narrative skills. They analyzed diagnostic accuracy, participant modifiability, and methodological quality.

**Relevance to work:** This article provides strong evidence for the use of dynamic assessment for diagnosing language impairment in children and bilingual children. Some limitations in the studies included in this article provide justification for the current study to mitigate these limitations and to provide more evidence that dynamic assessment can be a helpful tool in diagnosis language impairment in children.

Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research, 49*(5), 1037–1057. [https://doi.org/10.1044/1092-4388\(2006/074\)](https://doi.org/10.1044/1092-4388(2006/074))

**Objective:** This research article had three purposes. The first objective was to evaluate the reliability of using a dynamic assessment of narrative skills for diagnostic purposes. The second focus was to compare pre-to posttest dynamic assessment changes in story creation by children with and without language impairment. The third focus of this research article was to experiment the application of using dynamic assessment of narrative ability for diagnostic purposes in classifying children as having a language impairment.

**Method:** For experiment number one, they had a sample of 58 first and second grade children from central Texas. Participants came from various cultural/linguistically diverse backgrounds including African American (38%), European American (34%), Latino American (28%). They split the first and second grade students into three groups (two groups of typically developing children, and one group with language impairment).

The first two groups were given two wordless stories in different order depending on which TD group they were put in (1 or 2). Both groups of children produced stories that corresponded to two wordless picture books.

In experiment 2, they had 71 participants in first and second grade from central Texas and southern California. These children were split into three groups: one group of typical developing children, one of children with LI, and one no-treatment control group. A pre-posttest control group was designed with children from the TD and LI groups with two mediated learning experiences (MLE) sessions with the goal to improve and increase the length and complexity of children's stories.

**Relevance to work:** This study found that measures of narrative skills are more beneficial with a pre- and posttest, as well as an intervention stage regarding classification accuracy. They found that pre-test data alone led to over and under-identification of language impairment in students. This is important for the current findings of this study because it gives valuable information on the reliability and classification accuracy of using measures of language ability.

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

**Objective:** The purpose of this research study was to use dynamic assessment to identify the effects of testing bias by measuring how well a child can learn a skill instead of measuring their current ability through static measures.

**Method:** They used two groups of preschool-aged children, totaling seventy-nine participants overall. There was an experimental group which received the MLE teaching

method of dynamic assessment, and a no-mediation group. Teaching and testing were done in Spanish and English depending on the needs of the child and their primary language. Researchers in this study also measured language ability from parent reports, classroom observations (peer interactions and play interactions), and teacher judgements.

**Relevance to work:** It is very important to look at all aspects in a child's life when evaluating their language. It simply cannot be done by one static measure of current ability. Assessing a child's ability to learn and looking at how they use their language in different aspects of their life will provide SLP's with a much better picture of the student's overall language ability.

Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A nonbiased procedure. *The Journal of Special Education, 26*(3), 269-280.  
<http://dx.doi.org/10.1177/002246699202600304>

**Objective:** One of the purposes of this study was to explore the differences between home demands and the tasks of labeling and describing given on standardized assessment measures. The other purpose was to show the efficiency of using dynamic assessment as the approach to use in differentiating nondisabled children with language differences and those who have language disorders.

**Method:** The researchers took 60 students ranging 3 to 9 years of age enrolled in a Head Start program. The group of children were all from culturally and linguistically diverse populations. They were administered two to three standardized assessments then went through a period of teaching and instruction then a post test was conducted to compare scores to see if there was improvement in their results. If scores did not

improve, it would be indicative of a language disorder, while the students whose scores did improve would show that a language difference is present.

**Relevance to work:** Dynamic assessment is a very valuable tool for clinicians. It is helpful to stay away from biases that are often present in standardized assessment measures. It shows a child's ability to learn new skills and can help distinguish between a language disorder and a language difference. Having this knowledge will help clinicians avoid over diagnosing children from culturally and linguistically diverse backgrounds.

Petersen, D. B., Allen, M. M., & Spencer, T. D. (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities, 49*(2), 200–215.

<https://doi.org/10.1177/0022219414538518>

**Objective:** The purpose of this study was to compare the classification accuracy of static and dynamic prereading measures in kindergarten students, and to see which method was a better indicator of future reading difficulties in these young students.

**Method:** The participants in the study were kindergarten students and were given two common static prereading measures. After the static assessments were administered, they split the students into two groups who received either a dynamic assessment using an onset-rime decoding strategy, or a dynamic assessment which used a sound-by-sound reading strategy. At the end of first grade, the same students were assessed using multiple reading measures, and it was found that using dynamic assessment resulted in higher classification accuracy.

**Relevance to work:** Using assessment methods that have higher classification accuracy will be a significant help to SLPs, especially those working in a school setting.

This will help avoid the problem of either over or under identification of students for special education. Having higher sensitivity and specificity in these tests can also improve the efficiency of the evaluation and classification process by eliminated the number of steps and assessments needed to identify students who need help.

Petersen, D. B., Gragg, S. L., & Spencer, T. D. (2018). Predicting reading problems 6 years into the future: Dynamic assessment reduces bias and increases classification accuracy. *Language, Speech, and Hearing Services in Schools, 49*(4), 875-888. [https://doi.org/10.1044/2018\\_LSHSS-DYSLC-18-0021](https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021)

**Objective:** The purpose of this study was to use dynamic assessment of decoding to evaluate kindergarten students to see how well it predicts reading difficulty in later elementary grades. They were also looking to see if the dynamic assessment improves the validity of the static reading measures currently used among kindergarten students.

**Method:** Hundreds of kindergarten students (both Caucasian and Hispanic) were given a quick dynamic assessment of decoding ability, as well as static measures of letter/sound identification and phonemic awareness. Once those students reached later grades (second through fifth) their oral reading fluency was assessed to see if there was variance in scores and predictive validity of the static and dynamic measures.

**Relevance to work:** This study was very helpful because it gave a descriptive overview of the current practices school districts are using across the state in regard to identifying students at risk for reading disorders. They used current practices to help support their research in transitioning from using static assessments like the DIBELS, to dynamic assessments that measure a student's ability to learn how to read.



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

<https://doi.org/10.1097/TLD.0000000000000005>

**Objective:** The purpose of this research study was to conduct an evaluation of the current interventions, assessments, and district policies in place to identify English language learners with language learning disabilities. Is there a battery of assessments that is proven to be more useful and accurate than others regarding identifying language differences from language disorders?

**Method:** Participants were professionals in the field of education including speech language pathologists, school psychologists, special educators, and paraprofessionals. All participants were asked to complete an online survey discussing the evaluation measure they use for children at their schools.

**Relevance to work:** It is important for SLPs to be aware of assessments that are bias towards specific populations. Doing so will allow them to be more efficient in the evaluation process. It will also help avoid the error of either over or under qualifying students (especially ESL students) for services based on differences rather than disorder.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37(1), 61-72. [https://doi.org/10.1044/0161-1461\(2006/007\)](https://doi.org/10.1044/0161-1461(2006/007))

**Objective:** The purpose of this article was to look at the available norm-referenced tests to see if their manuals give enough evidence/support as to why they adopted specific cutoff scores to identify language impairment. How likely are they to obtain a low score

on that given test? The second purpose was to determine how many normative tests discuss the sensitivity and specificity within the testing manual.

**Method:** By looking at the effect size in cutoff scores in the testing manuals between the language impaired and the typically developing, as well as the sensitivity and specificity. Using 43 of the most up-to-date commercially available norm-referenced standardized tests, they solely relied on the testing manuals for this information.

**Relevance to work:** Analyzing a test's ability to accurately identify language impairment is an important part of this career. By having the knowledge and skills to look at a testing manual and be able to identify its construct validity can help clinicians avoid the problem of either over or under identifying children with language impairment using arbitrary cutoff scores. This gives clinicians across the nation the power to apply evidence-based practice in the evaluation process.

Thomas, A. S., & January, S.-A. A. (2021). Evaluating the criterion validity and classification accuracy of universal screening measures in reading. *Assessment for Effective Intervention, 46*(2), 110–120. <https://doi.org/10.1177/1534508419857232>

**Objective:** The purpose of this study was to evaluate the validity of different screeners such as the Measures of Academic Progress (MAP), and the Strategic Teaching Evaluation of Progress (STEP) with state assessments, and the ability of each screening method to distinguish between students at risk and not at risk.

**Method:** This study was conducted by using two separate cohorts of students enrolled in public charter schools. MAP and STEP were administered in three cycles in the fall, winter, and spring of each school year.

**Relevance to work:** This study helps to show support and evidence of using a screening system that promotes timely and appropriate intervention to as many students as possible. The screening should be convenient, given by an appropriate administrator, and in the right context to adequately identify students at risk for reading failure using the MAP and STEP procedures in school.

Ukrainetz, T. A., Harpell, S., Walsh, C., & Coyle, C. (2000). A preliminary investigation of dynamic assessment with Native American kindergartners. *Language, Speech, and Hearing Services in Schools, 31*(2), 142-154. <https://doi.org/10.1044/0161-1461.3102.142>

**Objective:** This study sought to examine language learning ability in Native American children using dynamic assessment as a less-biased procedure. Children from CLD backgrounds often and typically score lower on traditional standardized tests, so their objective was to show evidence that dynamic assessment can help mitigate that bias.

**Method:** They identified 23 kindergarten students as having either strong or weak language learners. All students participated in a dynamic assessment of categorization. Examiners filled out modifiability scores that were determined to reflect on the child's ability to attend, plan, self-regulate, and their responses during the intervention sessions. Posttest scores were determined by standardized categorization tests.

**Relevance to work:** This study used both dynamic assessment and kindergarten students, which is like the design of this study. Although their research focused on categorization tasks, the idea can be applied to narrative language-based dynamic assessment to help identify language learning difficulty in young children. This study

helped support the evidence that dynamic assessment can be useful when identifying weak language ability in children from many different backgrounds.

VanMeveren, K., Hulac, D., & Wollersheim-Shervey, S. (2020). Universal screening methods and models: Diagnostic accuracy of reading assessments. *Assessment for Effective Intervention*, 45(4), 255–265. <https://doi.org/10.1177/1534508418819797>

**Objective:** Universal screenings are a valuable measure for identifying at-risk students. The purpose of this study was to investigate the efficiency and effectiveness of different screening measures to determine which is best for identifying students who are at risk for reading problems.

**Method:** Analyzing data from five school districts from the 2014-2015 academic school year. They used fourth and fifth grade students because they had experience with screeners and standardized assessments. They took curriculum-based measures (CBM), measures of academic progress (MAP) reading subtest which is a computer-adaptive testing (CAT), and high-stakes reading test (MCA-III). They used all these assessments to see if they are a good predictor of future risk for reading difficulties if used as a screening tool. The areas of reading assessed were informational text (comprehension), foundational skills/vocabulary, and literature (identifying story elements).

**Relevance to work:** Understanding which screening tools will be the most effective and accurate in identifying future risk for reading difficulties can help schools, teachers, and SLPs guide students to the right tier of intervention needed for them to obtain success. It also eliminates the problem of overidentifying students at risk and under identifying those who need more support.

## APPENDIX B

**Institutional Review Board Approval Letter**

This Institutional Review Board (IRB) Reliance Agreement (“Agreement”) is made by and between Brigham Young University, Provo (“BYU”), and the **University of Wyoming**, (“Institution”) and is effective on the date of the parties’ signatures below.

## Agreement Terms

## 1.0 Purpose

1.1 This IRB Reliance Agreement (Agreement) establishes the authorities, roles, and responsibilities of each party with respect to the following research studies:

**IRB2020-328: Examining the Validity and Reliability of Dynamic Assessments of Reading and Language**

**IRB2021-360: Accurately Identifying and Responding to Oral and Written Language Learning Potential Using Dynamic Assessment and Contextualized Language Intervention**

(Research Projects) described in Attachment A. Those signing below agree that Institution may accept and rely on the review and approval of the Research Project by BYU IRB.

1.2 This Agreement does not preclude either party from entering into or participating in IRB reliance agreements with other entities or taking part in research not covered by this Agreement.

1.3 This document must be kept on file by the parties and be provided to the Federal Drug Administration (FDA), U.S. Department of Health and Human Services Office for Human Research Protections (OHRP), or other applicable regulatory entity upon request.

## 2.0 Responsibilities of BYU IRB

2.1 The BYU IRB will review and approve or disapprove the Research Project, review and approve or disapprove modifications to the Research Project, approve consent forms, collect reports of unanticipated problems and serious or continuing noncompliance, review information that requires reporting, and maintain required IRB records pursuant to applicable laws and regulations. No subjects may be enrolled in research subject to this Agreement prior to the approval of BYU IRB.

2.2 The review performed by the BYU IRB will comply with the U.S. Department of Health and Human Services regulations for the protection of human subjects (45 C.F.R. part 46), applicable FDA regulations (21 C.F.R. parts 50, 56, 312, 812), the terms of the BYU’s OHRP-approved Federalwide Assurance (FWA), and BYU’s policies. BYU IRB will identify, interpret, and comply with the requirements of any additional international, national, state, and local laws and regulations applicable to the Research Project, including, but not limited to, data security, privacy, and reporting requirements.

2.3 BYU IRB has the authority to suspend or terminate approval of research that is not conducted in accordance with its policies, applicable laws and regulations, or that has been associated with unexpected serious harm to participants. BYU IRB will promptly notify

Institution of the suspension or restriction of the Research Project and will copy BYU on communication with the FDA, OHRP, or funding entity on matters relating to the Research Project. Minutes of BYU IRB meetings relating to the Research Project will be made available to Institution upon request.

2.4 The BYU IRB will make determinations regarding the Health Insurance Portability and Accountability Act of 1996 and its implementing regulations (collectively, “HIPAA”) applicable to the Research Project. Institution will comply with HIPAA determinations made by BYU IRB and will use the BYU IRB forms related to HIPAA compliance. If it becomes necessary for the parties to use or disclose personal health information, then the parties will work together to determine the steps necessary to ensure that the required information is used or disclosed in a HIPAA-compliant manner.

2.5 BYU will provide meeting space and sufficient staff to support BYU IRB’s review and record keeping duties.

### 3.0 Institution IRB Responsibilities

3.1 Institution will comply with the U.S. Department of Health and Human Services regulations for the protection of human subjects (45 C.F.R. part 46), applicable FDA regulations (21 C.F.R. parts 50, 56, 312, 812), the terms of its OHRP-approved FWA, the Institution’s IRB policies, Institution’s policies, and any additional international, national, state, and local laws and regulations applicable to the Research Project.

3.2 Institution will conduct a facilitated review to accept and rely on the approval issued by BYU IRB.

3.3 Institution will be responsible for ensuring compliance with BYU IRB’s determinations for research conducted at Institution’s facilities.

3.4 Upon completion of the Research Project, Institution will remove BYU IRB as a designated IRB from Institution’s FWA record maintained by OHRP.

### 4.0 Joint Responsibilities

4.1 Each party is responsible for evaluating the potential financial conflicts of interest of its investigators and research staff associated with the Research Project and for reporting identified financial conflicts of interest to the other party.

4.2 Each party will notify the other party when a regulatory entity has or will conduct an audit or review of the Research Project and will communicate the outcome of the review in writing. If either party determines that it must report the findings of an investigation to a regulatory entity, it will request approval in writing from the non-reporting party in advance, with such approval not being unreasonably withheld. Nothing in this Agreement will be construed to prevent either party from making its own report to regulatory entities in accordance with its written procedures or applicable laws or regulations.

4.3 Each party will cooperate to ensure adequate protection of human research subjects participating in the Research Project and will cooperate to exchange relevant documentation and records when needed.

4.4 Either party may terminate this Agreement with or without cause upon 30 days written advance notice to the other party. The parties may also terminate this Agreement immediately upon written notice that (1) the Research Project is terminated; (2) either party is debarred from participation in federally funded research; or (3) either party is determined to have violated any of the provisions of this Agreement or international, national, state, or local laws or regulations.

4.5 Each party (the “Indemnifying Party”) shall indemnify, hold harmless, and defend the other party, its officers, trustees, employees, investigators, volunteers, and agents (the “Indemnified Parties”) from and against any and all causes of action, liabilities, obligations, judgements, losses, damages, claims, settlement payments, costs and expenses (including reasonable attorney’s fees), interest, awards, judgments, diminution in value, fines, fees, penalties, or other charges arising out of or relating to the Indemnifying Party’s performance of its obligations under this Agreement or the operations conducted by the Indemnifying Party under this Agreement.

4.6 All correspondence and documents relating to this Agreement will be in English.

4.7 This Agreement will be governed and construed in accordance with the laws of the State of Utah without regard to its conflict of law rules. Any dispute arising between the parties will be resolved in the United States District Court for the State of Utah or the Fourth District Court in Provo, Utah, depending on the nature of the claim.

4.8 This Agreement may be executed in any number of counterparts, either in original, emailed, or faxed form.

4.9 No Amendments or changes to this Agreement will be effective unless made in writing and signed by the other party.

4.10 This Agreement constitutes the entire agreement and understanding between the parties and supersedes all prior communications, contracts, or agreements between the parties with respect to the subject matter of this Agreement.

4.11 Any notices to institutional officials or correspondence regarding IRB review and oversight of the Research Project will be addressed as follows:

If to BYU:

Sandee M.P. Aina, MPA  
Associate Director of Human Research Protections (HRPP)  
Brigham Young University  
Research Administration Office  
A-268 ASB Campus Drive  
Provo, UT 84602  
1-801-422-1461  
[sandee.aina@byu.edu](mailto:sandee.aina@byu.edu)

If to Institution:

University of Wyoming  
Nichole Person  
Research Compliance Coordinator  
1000 E. University Ave. Dept. 3355  
[irb@uwyo.edu](mailto:irb@uwyo.edu)  
307-766-5322

University of Wyoming

By:  \_\_\_\_\_

Name: Diana G. Hulme

Title: Associate Vice President for Research

Date:

Brigham Young University

By: \_\_\_\_\_

Name: Larry Howell, Ph.D.

Title: Associate Academic Vice President, Research

Date:



## ATTACHMENT A

## IRB Research Project

Title of Research Project	<p><b>IRB2020-328:</b> Examining the Validity and Reliability of Dynamic Assessments of Reading and Language</p> <p><b>IRB2021-360:</b> Accurately Identifying and Responding to Oral and Written Language Learning Potential Using Dynamic Assessment and Contextualized Language Intervention</p>
Name of Principal Investigator	Kendra Hall Kenyon
Name of Sponsor	

## Name of Organization Providing IRB Review (hereinafter, "BYU"):

BYU's OHRP Federalwide Assurance (FWA) #	FWA00001266
IRB Registration Number	IORG0001302
Street Address	A285 ASB Campus Drive
City	Provo
State (if US)	UT
Zip/Postal Code	84602
Country	USA

Name of Individual Responsible for Administration of the Reliance Agreement	Sandee Aina
Title of Individual	Associate Director, Human Research Protections Program
Phone Number	801-422-1461
Email address	<a href="mailto:Byu.hrpp@byu.edu">Byu.hrpp@byu.edu</a>

## Name of Organization Relying on BYU IRB Review (hereinafter, "University of Wyoming"):

UW's OHRP Federalwide Assurance (FWA) #	00000186
Name of Institutional Official	Diana G. Hulme
Street Address	1000 E. University Ave. Dept. 3355
City	Laramie
State (if US)	Wyoming
Zip/Postal Code	82071

Country	USA
---------	-----

Name of Individual Responsible for Administration of the Authorization Agreement	Nichole Person
Title of Individual	Research Compliance Coordinator
Phone Number	307-766-5322
Email address	<a href="mailto:irb@uwyo.edu">irb@uwyo.edu</a>

## APPENDIX C

## Child Assent Form

Petersen Dynamic Assessment

**Child Assent** (7-14 years old)**What is this research about?**

My name is \_\_\_\_\_. I want to tell you about a research study I am doing. A research study is a special way to find the answers to questions. We are trying to learn about how well students can talk and read.

Your parents have given you permission to be in this study. If you decide you want to be in this study, this is what will happen.

I am going to say some made up words and ask you to repeat them. I am also going to tell you some stories and then ask you to tell them back to me. I am also going to teach you how to tell stories. We will probably work together for about 30 minutes. We will audio record you.

**Can anything bad happen to me?**

You might get tired of talking or reading. That's OK. If you don't want to tell stories or read anymore, just let me know and we will stop.

**Can anything good happen to me?**

We don't know if being in this study will help you. But we hope to learn something that will help other people some day.

**Do I have other choices?**

You can choose not to be in this study.

**Will anyone know I am in the study?**

We won't tell anyone you took part in this study. When we are done with the study, we will write a report about what we learned. We won't use your name in the report.

**What happens if I get hurt?**

Your parents/legal guardians have been given information on what to do if you get injured during the study. But we won't be doing anything dangerous, so you probably won't get hurt.

**What if I do not want to do this?**

You don't have to be in this study. It's up to you. If you say yes now, but change your mind later, that's okay too. All you have to do is tell us. If you do not want to do this with us, you can join the rest of your class in your classroom.

Before you say yes to be in this study; be sure to ask me any questions that you have.

[For Kindergarteners, say "Do you want to be in this study?"]

[For First through Sixth Grade, say "If you want to be in this study, please sign and print your name."]

Name (Printed): \_\_\_\_\_ Signature \_\_\_\_\_ Date: \_\_\_\_\_