10-9-2013

Comparison of performance on static and dynamic language tasks among typically developing preschoolers with bilingual experience

Mireya Hernandez

Mireya Hernandez
Candidate

Speech and Hearing Sciences
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Janet Patterson, Ph.D., CCC-SLP, Chairperson

Dr. Barbara Rodriguez, Ph.D., CCC-SLP

Dr. Philip Dale, Ph.D

COMPARISON OF PERFORMANCE ON STATIC AND DYNAMIC LANGUAGE TASKS AMONG TYPICALLY DEVELOPING PRESCHOOLERS WITH BILINGUAL EXPERIENCE

by

MIREYA HERNANDEZ

BACHELOR OF ARTS IN JOURNALISM AND MASS COMMUNICATION

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

Speech-Language Pathology

The University of New Mexico
Albuquerque, New Mexico

July, 2013

ACKNOWLEDGEMENTS

I whole-heartedly thank Dr. Janet Patterson, my thesis advisor and committee chair, for her counsel and expertise throughout this project. Her mentorship, support, perspective, and critical eye gently challenged me to achieve more than I could have imagined. I also thank my committee members, Drs. Philip Dale and Barbara Rodriguez, for their advice and flexibility throughout this endeavor. Their wisdom helped steer me in the right direction when I most needed it. Jennifer Romero was also an invaluable assistant whose attention to detail and independent work ethic was an enormous help in managing participant data and other logistics. I also thank City of Albuquerque Child Development Centers and Youth Development Inc. Head Start Centers for their support of this study.

COMPARISON OF PERFORMANCE ON STATIC AND DYNAMIC LANGUAGE TASKS AMONG TYPICALLY DEVELOPING PRESCHOOLERS WITH BILINGUAL EXPERIENCE

By

Mireya Hernandez

B.A. Journalism and Mass Communication, University of New Mexico, 2007
M.S. Speech-Language Pathology, University of New Mexico, 2013

ABSTRACT

The purpose of this study was to evaluate the potential of a graduated prompting format for measuring modifiability in a screening context among preschool children with bilingual experience. The dynamic assessment framework, including graduated prompting, is an alternative to traditional, static assessment of language impairment. Performance on four dynamic language tasks by 16 typically developing 4-year-olds with more English than Spanish experience was compared to performance of 16 matched children on static versions of the same tasks. These included novel adjective learning, semantic comparison, phonological awareness, and false-belief theory of mind tasks. When prompted responses were credited, the dynamic group performed significantly higher than the static group on three out of four language tasks. Unlike the static group, the dynamic group showed significant within-task improvement on the semantic comparison and false-belief tasks. These findings suggest graduated prompting is a viable format for measuring modifiability compared to traditional static screening.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

Demographic changes in recent decades have led to more culturally and linguistically diverse caseloads for speech-language pathologists (SLPs). There is a pressing need to develop more appropriate and updated language assessment tools to meet the needs of a changing population. The need is especially urgent for SLPs working in public schools, where changes in federal legislation have increased the emphasis on early literacy and language development. It is estimated that nearly 20 percent of school-age children speak a language other than English at home (U.S. Department of Education, 2004). Between 2000 and 2010, half the U.S. population growth was made up of Hispanics, 75% of which spoke a language other than English at home (Ennis, Rios-Vargas, & Albert, 2011; Johnson, Rios, Drewery, Ennis, & Kim, 2010). Among children entering Head Start programs who are learning English as a second language, 80% come from Spanish-speaking homes (Fortuny, Hernandez, & Chaudry, 2010). Today's school-based SLPs face the challenge of large caseloads with growing numbers of Spanish-speaking children. These children are at high risk for misdiagnosis of language impairment due to the widespread use of culturally and linguistically inappropriate assessments. In order to better identify candidates for services and judiciously allocate resources, SLPs require updated screening and assessment tools that better distinguish language difference from disorder among Spanish-speaking school children.

Research has highlighted the inherent bias in universal application of assessment procedures based on mainstream cultural and linguistic norms. Culturally and linguistically diverse (CLD) children's experiences with multiple languages contribute to

language development that differs from monolinguals (Hammer et al., 2012; Patterson & Rodriguez, 2005). When different language environments are associated with different language content, neither language represents the developing child's total linguistic knowledge. For example, home language typically centers on household and family activities, such as cooking, chores, and household objects. School language includes more academic concepts, such as numbers and colors. For bilingual children, this means developing vocabulary and concept knowledge is distributed between two languages (Pearson, Fernandez, & Oller, 1993). As a result, monolingual language measures may underestimate total language ability.

Historically, standardized tests have been inadequate in parsing out language difference from disorder among CLD children. Tests that are normed on monolingual, English-speaking populations lack diagnostic validity when applied toward bilingual populations. Translations of tests based on English developmental patterns, which are distinctive from both Spanish and bilingual developmental patterns, are not valid for bilingual children and are associated with a high risk for over-diagnosis (Restrepo & Silverman, 2001).

Bilingual children's variable language histories further complicate accurate assessment of their language skills. There are two broad profiles for bilingual children entering the school system, simultaneous learners and sequential learners (McLaughlin, 1984). Simultaneous learners are exposed to two languages prior to entry into preschool and may follow a similar developmental trajectory as monolingual learners. Sequential language learners are essentially monolingual children until their first major exposure to English in preschool. By 2030, they are expected to make up 40 percent of the school-age

population (Thomas & Collier, 2002). Their language profiles are highly variable based individual differences and language experiences. Because sequential learners have started learning the basic rules of their first language, their acquisition of a second language follows a different developmental path than either monolingual or simultaneous language learners (Tabors & Snow, 1994). For example, many sequential learners may have normal receptive language and low expressive language scores (Brown, 2004). This pattern is also characteristic of monolingual children with reading disabilities (Barrera, 1995; Gunderson & Siegel, 2001). Misapplication of tests normed on monolingual children places sequential language learners at risk for misdiagnosis of language impairment. In fact, "...many preschool-age bilingual children who demonstrate some characteristics of language disorder, such as expressive language disorder, do achieve normal speech as they grow older" (Brown, 2004, p. 228).

Clearly, accurate assessment of this rapidly growing population must take into account the inherent complexities of bilingual language development. Some recommendations for improved classification of bilingual children include assessment of skills in both languages and use of multiple measures within natural settings (Langdon, 1989; National Association for the Education of Young Children [NAEYC], 2005). However, the extensive time and resources required for such assessments present a barrier to application in the school system. Such barriers highlight the need for more culturally and linguistically appropriate tools with clinical utility throughout the assessment process. This includes the development of initial language screening tools to better identify candidates for more comprehensive language assessments.

Chapter 2

Literature Review

Static Assessment

Traditionally, static tests have been the primary means of measuring language and learning skills of schoolchildren. Although the purported goals of these static assessments include the identification of specific education needs and effective differentiation in instruction, the overarching emphasis has been classification of children by aptitude (Haywood, Brown, & Wingenfeld, 1990). However, when applied to CLD populations, the extent to which results can be interpreted as diagnostic is limited.

Static tests measure a child's independent performance at a given point in time. In this way, they are measures of previously learned knowledge, experience, and strategies and the ability to recall and apply them to different tasks. The term 'static' is apt in that it assesses performance at fixed points in time. The manner of administration is also fixed, as examiners typically withhold feedback and follow strict guidelines to maintain test validity (Brown & Ferrara, 1985). Even the results may be interpreted as a "more-or-less fixed" measure of ability (Campione & Brown, 1987; Tzuriel & Haywood, 1992, p. 3). Standardized static tests typically employ a normative assessment approach to determine if present levels of performance are within the normal range for peers. When variables such as language, education, and cultural backgrounds are controlled, deviations in performance may be reasonably interpreted as indicators of delay or disorder.

However, when CLD children are administered static tests, deviations may also reflect a lack of experience with test expectations, tasks, and procedures. CLD children's home cultures may differ in emphasis on school-readiness and interaction style between

children and adults. These factors may negatively affect a child's performance on tests in which familiarity with a specific adult-child dynamic is assumed. In such instances, poor performance on static tests may result from differences in language experience rather than disorder. Taken by themselves, such measures can lead to misdiagnosis and inappropriate referral for special education services (Rogoff, 1991). Conversely, children may be under-referred for special services because the lack of culturally appropriate tests leads to cautious interpretation (Winter, 1999).

Dynamic Assessment

Dynamic assessment has emerged as promising approach to better distinguish experience-based language difference from disorder (Gutierrez-Clellen & Peña, 2001; Patterson & Rodriguez, 2005). Whereas static assessments measure the product of learning experiences, dynamic assessments measure the learning process itself (Campione & Brown, 1987). Under this method, some of the examiner's focus shifts away from final results to the observation of different behaviors during the learning process. These include the use of problem-solving strategies, responsiveness to different types of help, and transfer of skills toward new problems. Examining a child's ability to learn rather than what knowledge has already been acquired reduces the potential for experience-based bias.

This approach is based on the sociocultural theory of learning and Vygotsky's proposed Zone of Proximal Development, or ZPD (Vygotsky, 1978). His theory operationalized the concept of learning potential and offered a rough outline for its assessment. Vygotsky proposed that one's learning ability is not readily apparent, but rather, could be facilitated through instruction. Therefore, assessment requires analysis at

two levels, the actual and potential developmental levels. Traditional static assessments measure actual developmental levels—that is, the final product of learning that has already taken place, as reflected by independent performance. Potential developmental levels include what a child can perform when guided by a more knowledgeable adult or peer. The difference between actual and potential developmental levels constitutes the ZPD.

According to Vygotsky (1978), "The ZPD permits us to delineate the child's immediate future and his dynamic mental state…what is in the course of developing" (p. 87). Of central importance is the facilitative role of a more knowledgeable individual in guiding the interaction and eliciting the child's learning potential. Under this model, the interaction and the learning process are not static, but constantly changing. As learning occurs, the responsibility is shifted from the more knowledgeable individual to the learner, who gradually internalizes and develops the skill to carry out activities independently.

Feuerstein expanded on the socio-cultural model of learning and introduced the theory of cognitive modifiability. This theory is based on the premise that human cognitive ability is not fixed and static, but modifiable and adaptable through "mediated learning experiences" (Feuerstein, Rand, & Hoffman, 1979, p. 71). Differences in performance are attributed to dissimilar experiences. Children who lack the experiences required to achieve certain tasks need guided instruction to demonstrate their true learning potential. This support compensates for differences in experience and modifies the child's performance. In contrast with static assessments, the more knowledgeable

individual bears the primary responsibility to provide mediation and determine what environmental changes may improve learning.

The most widely studied form of dynamic assessment is Feuerstein's Mediated Learning Experience (MLE). It represents a comprehensive assessment approach in that it includes a static and dynamic component. This approach employs a test-teach-retest format in which participants are taught the principles needed to perform specific tasks during a short intervention period, or MLE. Data includes a comparison of pre- and post-intervention performance and also describes the child's responsiveness during treatment (Peña, Iglesias & Lidz, 2001).

The graduated prompting approach is an alternative form of dynamic assessment. Pioneered by Campione and Brown (1987), it is characterized by a predetermined set of prompts, each containing more specific information, designed to teach the target skill as needed. The amount of prompting needed is interpreted as a measure of the child's ZPD (Campione & Brown, 1987). Campione and Brown also emphasized the *transfer* of learning, that is, the number of prompts needed to complete a different version of the same task. In their studies, they found this parameter of modifiability to be the most strongly correlated to ability measures. Relative to MLE approaches, graduated prompting does not yield as much prescriptive data about child learning strategies and responsiveness to different types of support. However, it is generally more efficient and structured, as it measures learning that occurs within the test rather than over an extended intervention period (Grigorenko & Sternberg, 1998).

Applications in Language Assessment

One application of dynamic assessment is to distinguish children who are typically developing from those with language impairments. Specifically, research has examined responsiveness to instruction, or modifiability, based on the assumption that it indicates learning potential and language ability (Lidz, 1983; Peña, 2000). Differences in modifiability are central to distinguishing language difference from disorder. Typically developing children are expected to show high modifiability in response to learning experiences that compensate for differences affecting initial task performance. However, children with language impairments are expected to show low modifiability (Gutierrez-Clellen & Peña, 2001).

The extent to which modifiability differentiates CLD children has been examined through mediation of different language skills, including narrative skills, categorization, and novel-word learning (Kapantzoglou, Restrepo, & Thompson, 2012; Peña, Resendiz, & Gillam, 2007; Ukrainetz, Harpell,Walsh, & Coyle, 2000). Using the MLE approach, these studies of CLD children identified modifiability scores as the best indicators of language impairment. For example, when teaching word-learning skills among CLD preschoolers, dynamic assessment proved effective for predicting change and differentiating typically developing from "low-language ability" groups (Peña, Iglesias & Lidz, 2001). Although both groups performed similarly on a standardized pre-test, the combined modifiability ratings and post-test scores more accurately differentiated the two groups. The repeated link between modifiability measures and differentiation of groups indicates their utility in identifying language impairment in CLD children better than traditional standardized measures alone.

The graduated prompting approach has also been useful for predicting change and identifying children with language impairments. In a study of late-talking toddlers, the amount of graduating prompting needed to elicit two-word combinations was positively correlated to the degree of language growth that followed a short intervention period (Olswang, Bain, & Johnson, 1992). When graduated prompting was used to screen phonological awareness skills in a group of kindergartners, it yielded higher predictive data regarding later reading achievement than a standardized screening test (Bridges & Catts, 2011). Both studies demonstrate the graduated prompting format's ability to predict future performance for language skills in specific domains. Hasson, Camilleri, Jones, Smith, and Dodd (2012) also used graduated prompts on bilingual 3-5 year olds within a brief test-teach-retest format. They found that the number of graduated prompts needed to teach vocabulary and sentence structure during the teaching phase accurately differentiated children with language impairments from typically developing children (Hasson et al., 2012). These results suggest the method's potential to identify language impairment better than traditional static screening tools.

Language Screening and Graduated Prompting

Dynamic assessment's emphasis on individual performance along multiple parameters has also led to several criticisms. It has been described as long, labor intensive, and lacking in reliability and validity soundness (Haywood & Tzuriel, 2002; Jitendra & Kameenui, 1993). The MLE format, in particular, is a time and expertise intensive method best suited for comprehensive assessments (Olswang et al.,1992; Peña, et al., 2001). Screening tools to better identify candidates for such evaluations must be developed in order to use time and resources more efficiently. Given its narrowed focus,

the graduated prompting approach may be better suited for this purpose. Its structured design presents a more psychometric orientation and focuses on specific aspects of modifiability that best differentiate children: learning (responsivity) and transfer (Campione, 1989; Lantolf & Poehner, 2004; Peña, 2000).

The widespread adoption of the response to intervention (RTI) model, which requires universal screening, further underscores the need for effective screening tools (Bridges & Catts, 2011). RTI is used for early identification and prediction of learning impairments and subsequent allocation of services. The development of screening tools that are sufficiently sensitive for initial identification of language impairment is an essential step in appropriate distribution of interventions. As is the case with comprehensive assessments, screenings should be developed with consideration of the diverse linguistic profiles of the school-age population (NAEYC, 2005). Evidence of graduated prompting's potential for distinguishing language difference from disorder suggests the format may be appropriate for early identification of impairment.

In a recent study, Patterson, Rodriguez, and Dale (2013) used a graduated prompting framework to examine the modifiability of typically developing preschoolers with bilingual language experience. They developed three brief language tasks in English and Spanish for the purpose of observing changes in performance within a format with potential for use in a screening context. Task items were scored based on the amount of graduated prompting, or assistance, the child needed. For children who required any level of prompting on initial trials, the results indicated improved performance from the first two items to the last two items on 2 out of 3 language tasks. This suggests graduated prompting is a viable format for assessing modifiability for at least some language tasks.

It is the focus the present study to further examine graduated prompting as a measure of modifiability within the same tasks. The broader purpose is to explore the framework's potential to distinguish language difference from disorder in a screening context. By comparing outcomes of the dynamic versus static assessment approaches, the study aims to determine the effect of graduated prompts on performance for typically developing children with bilingual language experience.

Assessment Tasks

Three dynamic language tasks were included in the original Patterson et al. (2013) study: Novel Adjective Learning, Similarity of Function, and Phonological Awareness. The inclusion of a variety of tasks that require both expressive and receptive responses allows for variation in bilingual children's language abilities (Peña & Halle, 2011). This minimizes test bias by presenting the opportunity for children to demonstrate their knowledge in multiple ways (Peña, Bedore, & Rappazzo, 2003). The tasks in this study were developed based on their potential to distinguish typically developing from language-disordered CLD children.

The Novel Adjective Learning task is a receptive measure of semantic learning ability in which participants are taught adjectives in a new language. The task requires them to identify which objects represent the newly learned adjective. Several considerations guided the rationale for this task. Typically developing African-American children from lower SES backgrounds have been found to perform lower on traditional standardized vocabulary tests than children from middle SES homes. However, they performed similarly on dynamic assessments of novel word learning (Burton & Watkins, 2007). Also, children with specific language impairment do not significantly differ from

typically developing peers in learning new object names (Dollaghan, 1987). Instead they are differentiated by their difficulty identifying semantic features of new words, such as color or pattern (Alt, Plante, & Creusere, 2004). When taught in a dynamic assessment framework, novel word learning has accurately differentiated typically developing from language impaired Spanish-speaking preschoolers (Kapatnzoglou et al., 2012).

During the Similarity in Function task, the child is shown a picture of two objects with similar functions and asked how they are alike. Preschool children with language impairment have been known to perform below typically developing peers on similar inferential tasks and show increased risk for future reading comprehension problems (Blank, Rose, & Berlin, 2003; van Kleeck, Vander Woude, & Hammett, 2006). This task was deemed a good fit for Latino preschoolers because of language socialization practices that tend to focus on object functions rather than labels (Peña & Quinn, 1997).

The Phonological Awareness task is a receptive initial phoneme-matching task. For both English-and Spanish-speaking children, phonological awareness skills are critical for reading and predictive of later decoding and word recognition ability (Leafstedt & Gerber, 2005). Deficits in phonological awareness are common among children with language impairments (Catts, 1993; Catts, Adlof, Hogan, & Ellis Weismer, 2005); therefore, this task may be sensitive to language impairment.

Although not included in the final publication, Patterson et al. (2013) also administered a fourth language task, Theory of Mind. They developed a false belief task to assess a child's understanding that others may have different beliefs and perspectives. It was included in the original task set because language-impaired preschool children have demonstrated difficulty with similar tasks relative to typically developing peers

(Miller, 2004). Furthermore, performance on false belief tasks is related to grammatical and receptive vocabulary development (Farrar et al., 2009; Rakhlin et al., 2011). The language comprehension demands of false belief tasks may render them sensitive to language impairment. However, since bilingual children tend to perform better than their monolingual peers on similar false belief tasks, the task may carry less risk of over-identification of language impairment (Goetz, 2003; Kovacs, 2009). It is also included in this study based on its contribution to task diversity, which is recommended in assessments of bilingual children (Pena & Halle, 2011).

Purpose of the present study

This study further examines graduated prompting as a measure of modifiability by comparing the performance of the original sample on four dynamic tasks to a matched group of peers on static versions of the same tasks. More specifically, measurement of differences in performance on dynamic and static tasks aimed to delineate the extent to which improvement on the dynamic language tasks is due to the graduated prompting framework versus task exposure.

The study will address the following questions:

Q1. How does overall performance of children with Spanish and English bilingual experience on dynamic language tasks compare to the performance of children with similar backgrounds on static versions of the same tasks?

Q2. For the individual static language tasks, how does performance on the first two items compare to performance on the last two items?

It is hypothesized that graduated prompting provides a structured learning experience that accounts for the improved within task performance, or modifiability, in

the original study. It is expected that matched peers will perform significantly lower on the static versions, with little change within tasks.

The degree to which the static language tasks correlate with traditional, standardized assessments is also of interest. Positive correlations will provide criterion validity data for content of the language tasks. Overall performance on the static language tasks are expected to correlate with composite language scores on the Preschool Language Scale—4th edition (PLS-4), as they both probe more global language skills. Additionally, the Novel Adjective Learning and Similarity in Function tasks are expected to positively correlate with the Woodcock-Munoz Language Survey—Revised Picture Vocabulary subtest, as they each assess semantic knowledge.

The following research question stems from this secondary line of investigation:

Q3. How do overall and individual static language task performance relate to performance on standardized language assessments?

The original study included 32 children: 16 with more Spanish experience and 16 with more English experience. Because children with more English experience are the focus of present study, comparisons will only be made with the latter group.

Chapter 3

Methods

Participants

Recruitment. All participants for the present study were recruited from Head Start and City of Albuquerque preschool programs in which English was the primary language of instruction. On-site recruitment and screening for participant eligibility took place during preschool drop-off times. In some instances, teachers distributed pamphlets with attached screeners to parents of potential candidates. After the initial screening, parents were administered a complete questionnaire either on site or over the phone.

All children in both groups met the following criteria:

- 4 years old.
- Both English and Spanish experience in the home.
- Total English language exposure greater than Spanish language exposure.
- No history of intervention for language delay/disorder.
- No parent or teacher concern regarding hearing, speech, or language.

The emphasis of both studies is the measurement of modifiability within a graduated prompting framework among typically developing children with bilingual language experience. Therefore, participants were recruited only if there were no current parent or teacher concerns about speech, language, or hearing. Only one child had a history of intervention, which focused on speech articulation rather than language. The child had been discharged and passed a speech and language screening at the beginning of the school year. Because neither parent nor teacher had any current concerns about the child's speech or language development, the child was included in the sample.

Matching. Participants were matched to the children in the original study who were administered the dynamic language tasks in English. In the following discussion, the portion of original sample who were administered the dynamic tasks in English will be referred to as the dynamic group. Participants in the present study who completed the static tasks will be referred to as the static group. The groups were matched based on language experience, age, and gender. Each was comprised of 16 typically developing Hispanic/Latino 4-year-olds (4;0-4;11) with bilingual (English/Spanish) language experience. The average age of the dynamic group was 53.31 months and the static group was 54.44 months. Additionally, the male-to-female ratio (5 males, 11 females) was maintained across both groups, as shown in Table 1. (See Appendix A for complete matching profile.)

Table 1

Matching Characteristics of Static and Dynamic Groups

	Static group ($N = 16$)	Dynamic group ($N = 16$)
Gender	11 females; 5 males	11 females; 5 males
Age (*M*)	54.44 mos.	53.31 mos.
Home Language Exposure Index (*M*)	2.38	2.31

The primary matching criterion was language experience. Total home language experience was determined on the basis of a parent questionnaire. It included selected questions from a questionnaire used in several large-scale studies of bilingual preschoolers to characterize the child's English and Spanish experience in the home (Hammer, Miccio, & Wagstaff, 2003; Hammer et al., 2012). All questionnaires were administered in English, as each parent spoke English fluently. The brief questionnaire used in this study includes seven questions about the child's language experience, and a

5-point scale to quantify the answers: 1=All English, 2=More English than Spanish, 3=Same amount of English and Spanish, 4=More Spanish than English, 5=All Spanish.

Two questions used to describe home language exposure in the original study (Patterson et al., 2013) were used as matching criteria for the children in the present study:

1. When thinking about the languages spoken by everyone in your home, which of the following describes how much English and Spanish are used?
2. Which language(s) do you use at home?

For the purposes of matching groups, each participant's average score for the above questions was used to derive a Home Language Exposure Index. Each participant in the static group was matched within ±0.5 of another participant in the dynamic group. For example, a dynamic group participant with more English spoken in the home (Question 1 = 2) and the same amount of English and Spanish spoken by the responding parent (Question 2 = 3) would have a Home Language Exposure Index of 2.5. The matched participant in the static group would have an index range from 2 to 3. Given the present study's emphasis on matching participants on the basis of equal or higher English-language experience, scores were primarily in the 2 to 3 range (Home Language Exposure Index mean for the dynamic group = 2.31; static group = 2.38).

In the original study, descriptive data about classroom language exposure was also considered in determining language experience. As a result, two participants in the dynamic group had Home Language Exposure Index scores of 3.5. However, they had greater overall English language experience when classroom exposure was factored in. We matched these two participants exactly in the static group to avoid having children in

this study with Home Language Experience Index scores above 3.5. For these two matched participants, additional information about classroom experience indicated they had greater English language experience overall. Additionally, matches for the dynamic group participants with indexes of 1.5 were either exact or +0.5, to ensure some Spanish experience in the home. Home Language Exposure Indexes of 1.0 were not accepted, as this would mean there was no Spanish spoken in the home.

Data Collection Procedures

Parent interviews were used to determine eligibility for the study and to obtain relevant demographic and language profile information. For children who met the matching criteria, the researcher arranged to administer tests at the child's school. Total testing time ranged from 45-90 minutes, with at least one break to participate in lunch, recess, and other classroom activities. Most sessions took place over a single day, with two participants tested over two days. The author, who grew up in a Spanish-English bilingual home, administered all tests. Based on self-ratings, the author's expressive Spanish skills are moderate, which is characterized as functional, with stronger receptive than expressive language (Blumenfeld & Kaushanskaya, 2007). Testing sessions were all administered in the following order: WMLS-RPV English version, WMLS-RPV Spanish version, Static Language Tasks, PLS-4 (English). The order of task presentation for the Static Language Tasks was counterbalanced to include three variations. The Similarity in Function task was not presented first due its more extensive expressive language demands. Teacher questionnaires regarding child language use in the classroom were also completed to gain a more complete language profile.

Measures

Standardized tests. The standardized language assessments administered were the Preschool Language Scale—Fourth Edition and Woodcock-Munoz Language Survey—Revised Picture Vocabulary subtest in English and Spanish. Both tests include parallel versions that are normed on English and Spanish speaking populations. The Preschool Language Scale—Fourth Edition (PLS-4) includes Auditory Comprehension and Expressive Communication subscales, which assess receptive and expressive language, respectively (Zimmerman, Steiner, & Pond, 2002). The test-retest reliability coefficient is .93 for children ages 4;0-4;5 and .96 for ages 4;6-4;11, and internal consistency (Cronbach's alpha) for 4-year-olds is .95 to .96. The Auditory Comprehension subscale assesses comprehension of basic vocabulary, concepts and grammatical markers. The Expressive Communication subscale includes naming and describing objects, and use of specific prepositions, grammatical markers, and sentence structures. Scores from both subtests are used to derive an overall Total Language Score.

The Woodcock-Munoz Language Survey—Revised Picture Vocabulary subtest (WMLS-RPV) is a vocabulary assessment that includes receptive and expressive items (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005). The split half reliability coefficient for this subtest for 4-year-olds in the standardization sample was .91. Both versions begin with several receptive tasks, but primarily consist of expressive picture naming items. The items are of increasing difficulty with ceiling rules for discontinuing testing. In the original study, the WMLS-RPV was administered in the language in which the child was most proficient, as determined by parent report. The children in the present study were administered the English version followed by the Spanish version of the

WMLS-RPV to better characterize proficiency in both languages. Thirteen static group participants performed higher on the English version than the Spanish version of the WMLS-RPV and three children received higher scores on the Spanish WMLS-RPV, although their overall language profile indicated greater English language experience. The remaining 13 participants performed within the average range for English version but not the Spanish version. Performance on all standardized tests is reported is the *Results* section.

Language tasks. The Novel Adjective Learning (NAL) task is a receptive task in which children are taught adjectives in Hawaiian, a language with consonants and vowels that occur in English and Spanish phonological inventories. The first plate introducing each adjective depicted items that represented the word (e.g., yellow objects) and other contrastive objects. The examiner pointed to the objects representing the adjective, saying, "Look at these things. This is melemele, this is melemele, and this is melemele." Participants were shown the next test plate depicting three untrained items, two of which were foils, and asked to point to the object representing the adjective ("Which one is melemele?")

During the Similarity in Function (SiF) task, children were shown pictures of two objects with similar functions (e.g., fork and spoon) and asked how they were alike. This task required participants to identify and express a specific semantic association (i.e., object function) for each task item.

The Phonological Awareness (PA) task measures the ability to identify words with the same initial sounds. During this task, the examiner named a picture with the target initial sound, as well as three other pictures, one of which started with the same

target phoneme. The child was asked to identify the picture starting with the same sound as the target sound (e.g., target word: *mouse*; choices: *lock*, *match*, and *hat*).

The Theory of Mind (ToM) task is a false belief task in which children were shown a picture of a container with unusual contents (e.g., an egg carton filled with rubber bands). The administrator next showed the child a picture of a doll looking at the closed container and asked what the doll thought was in it. For example, "Look! Eva just got here. What does Eva think is inside the carton?" Acceptable responses included pointing to the picture on the container or verbalizing.

Consistent with the original study, each of the four language tasks was introduced with a single demonstration item. These included a trial task item followed by feedback on accuracy and explanations. For example, if the child answered the phonological awareness demonstration item incorrectly, the examiner said, "Hmmm. 'Sock' and 'soup' start with the same sound! They both start with 'ssss.'" Such task familiarization techniques do not change the nature of the test administration as similar methods are typically used in traditional, standardized tests. After the first demonstration item, the examiner did not provide explanations or models for any task items. The static task administration also differed from the dynamic version in that the administrator did not provide feedback regarding accuracy of answers.

Scoring Procedures

Static tasks. Research has identified the limitations of strict adherence to a single language when testing bilingual populations (Patterson, 1998; Patterson & Rodriguez, 2005). Because context-dependent linguistic knowledge is distributed across two languages, neither language encompasses the child's total knowledge. The evidence

indicates that vocabularies of both languages overlap, with vocabulary knowledge in each language tending to align with the different contexts in which they are learned (Umbel, Pearson, Fernandez, & Oller, 1992). Bilingual scoring accounts for distributed knowledge and credits answers given in either language (Bedore, Peña, Garcia, & Cortez, 2005). In a study of typically developing bilingual 5-6 year olds, Bedore et al. (2005) found that applying bilingual scoring to both English and Spanish tests increased classification accuracy compared to traditional single language scoring. Since the language tasks in the present study were designed to evaluate responsiveness rather than language proficiency, it was most appropriate to use bilingual scoring for these tasks.

Child responses to static language tasks were scored as follows: A correct answer earned 1 point and an incorrect answer earned 0 points. Each of the four tasks contained 6 items, for a maximum of 24 points across all tasks.

Dynamic tasks. Because the original study specifically examined the number of prompts provided rather than response accuracy, dynamic group scores included a range of values (0-3) relative to the number of prompts provided. The four tasks were introduced with the same demonstration procedures as the static tasks. Each task item included a maximum of three scripted prompts to provide increasingly explicit verbal and visual support. For example, for one Similarity in Function task item, the child was shown a picture of a fork and spoon and asked, "How are these two things the same?" An unprompted correct response received the maximum score of 3. If the child did not answer correctly, Prompt 1 verbally directed the child to consider function ("Let's think about what people usually *do* with these two things."). If the child did not respond correctly following the first prompt, Prompt 2 added visual support (pictures of people

using the objects appropriately) and again verbally prompted the child to consider object function. If the child did not respond correctly, Prompt 3 modeled and explained the correct response ("We can use them both to eat. That's how they're the same") and the child received a score of 0.

In order to analyze performance across comparable data sets in the present study, dynamic group scores were converted to binary scores, which were scored as either correct or incorrect. However, responsiveness to prompts remained the primary variable of interest. Therefore, two distinct methods of score conversion were used to examine dynamic group responses and compare to the static group.

The traditional scoring method held responses to dynamic tasks to the same standard for accuracy as the static tasks. Dynamic items on which the child answered correctly with no prompts were scored as correct (0 prompts =1); all other responses were scored as incorrect (1, 2, or 3 prompts = 0). However, this method fails to distinguish between children who answered correctly when prompted and those who answered incorrectly, regardless of prompts. Therefore, an alternate dynamic binary score conversion was devised to account for the graduated prompt framework and credit correct responses, even if prompting was required. The third prompt (or model) indicated that the child did not answer correctly after two prompts. Under the dynamic binary scoring method, 0, 1, or 2 prompts = 1 and 3 prompts = 0. This scoring method may be more sensitive to modifiability as it distinguishes between children who are responsive to support from those who are not.

Reliability

All sessions of the static language task administration were audio and video-recorded. An undergraduate student with Spanish and English language skills received one session of training in scoring two of the language tasks: SiF and ToM. These tasks were the only two with open-ended answers, allowing for possible variation in interpretation. For example, when asked how a fork and spoon are the same on the SiF task, responses included "we eat with them" or "you pick up stuff…and put it in your mouth." The second rater independently reviewed the video and rescored the two tasks for a randomly selected 25% (4/16) of the sample. Inter-rater agreement for both tasks was 100%. The PA and NAL tasks were not reviewed because each item had only a single correct answer selected from a closed set.

Data Analyses

The first research question compared static and dynamic group performance on the four language tasks through a series of *t* tests. Another series of *t* tests addressed the second research question by examining the degree of within-task change in performance by the static group. Finally, several Pearson correlations were run to analyze relationships in performance between the static language tasks and standardized tests.

Chapter 4

Results

Research Question 1

The first research question aimed to determine if the groups performed differently on the static versus dynamic versions of the same language tasks. Comparisons were made using the traditional scoring method for the dynamic tasks (i.e., unprompted correct responses were scored as "1"; prompted correct responses and incorrect responses were scored as "0"). Total scores on the four combined language tasks were compared across groups using a paired samples t-test. Total mean scores for the dynamic group (M = 12.50, SD = 3.86) and static group (M = 12.56, SD = 3.76) were not significantly different, as shown in Table 2. Between-group comparisons of performance on the four individual language tasks also were not significantly different.

Table 2

Comparison of Language Task Performance Between Groups: Traditional Scoring

	Static group (N = 16)		Dynamic group (N = 16)			
Task	*M*	*SD*	*M*	*SD*	*t*(15)	1- tailed *p*-value
PA	1.5	.89	2.06	1.69	1.32	.104
SiF	3.38	2.25	2.56	2.42	0.99	.170
NAL	3.31	1.35	2.94	1.48	0.86	.202
ToM	4.38	2.15	4.94	0.93	1.92	.187
Total	12.56	3.76	12.5	3.86	0.04	.485

In order to determine if this lack of difference between groups was consistent across scoring methods, the same analyses were performed using the dynamic binary scoring method for the dynamic tasks. This method assigned credit to all correct answers, regardless of prompts (i.e., 0, 1, 2 prompts were scored as "1"; 3 prompts were scored as "0"). When this scoring method was used, the dynamic group's total combined scores

increased (M = 20.75, SD = 3.15), resulting in significantly different scores from the static group, $t(15) = 5.41$, $p < .001$. The dynamic group performed better than the static group on all four language tasks, although differences were significant for only three of the tasks, as shown in Table 3. Whereas the traditional scoring method indicated no differences in performance between groups, the dynamic binary scoring method showed a clear trend of higher dynamic over static group performance.

Table 3

Comparison of Language Task Performance Between Groups: Dynamic Binary Scoring

	Static group (N = 16)		Dynamic group (N = 16)			
Task	*M*	*SD*	*M*	*SD*	*t*(15)	1- tailed *p*-value
PA	1.50	0.89	4.94	1.61	8.65	<.001*
SiF	3.38	2.25	4.63	1.82	1.67	.058
NAL	3.31	1.35	5.38	0.72	4.66	<.001*
TOM	4.38	2.15	5.81	0.40	2.52	.01*
Total	12.56	3.76	20.75	3.15	5.41	<.001*

*$p < .01$

Research Question 2

Static group. We also used paired samples *t* tests to examine changes in performance across items for each static language task. There were no significant differences in performance between the first two items and the last two items for any of the four language tasks, as shown in Table 4. Because children who answered the first item of a task incorrectly had more opportunities to show change, we examined this subset's performance more closely. Among these children, a paired samples *t* test found significant differences in performance on two out of the four tasks: PA and NAL (see Table 5).

Table 4

Changes in Within Task Performance for Static Group

Static task	First 2 items (*M*)	Last 2 items (*M*)	*t*(15)	1-tailed *p*-value
PA	0.38	0.63	1.0	.17
SiF	1.19	1.19	0.0	.50
NAL	1.06	1.44	1.57	.07
TOM	1.50	1.44	0.32	.38

N =16

Table 5

Changes in Within Task Performance for Static Group Subset Who Answered First Item Incorrectly

Static task	*n*	First 2 items (*M*)	Last 2 items (*M*)	*t*-value	1-tailed *p*-value
PA	12	.00	0.67	t(11) = 3.55	.003*
SiF	8	.39	0.63	t(7) = 0.61	.282
NAL	7	.29	1.43	t(6) = 4.38	.003*
TOM	4	.75	1.25	t(3) = 0.78	.248

*$p < .01$

Dynamic group. In the original study of 32 participants, Patterson et al. (2013) reported a significant difference in performance between the first two and last two items for the SiF task, but not for the PA or NAL tasks. This trend held true when the same analyses were applied to only the 16 children of the dynamic group who were given the tasks in English in the original study. Although the findings for the ToM task were not reported in Patterson et al., performance on the last two items was significantly better than on the first two items, $t(15) = 3.17$, $p = .003$.

Patterson et al. (2013) also examined the role of graduated prompts for children given either the Spanish or English version of the tasks and who required at least one prompt on the first item of a task. In other words, they did not independently answer the

first task item correctly and required some level of support. For these children, performance on the last two items was significantly better than performance on the first two items for SiF and NAL. For the subset of children given the English version of the dynamic tasks and needing at least one prompt on the first item, there was significant improvement on the SiF and ToM tasks using the original scoring. Traditional scoring also resulted in significant differences for SiF and ToM, as shown in Table 6. We found the same pattern of results when we used the dynamic scoring method, SiF $t(12) = 3.49$, $p = .002$ and ToM, $t(7) = 1.53$, $p = .085$. However, this was only a trend for the few children who needed prompting on the ToM task.

Table 6

Changes in Within Task Performance for Dynamic Group Subset Who Needed at Least One Prompt on First Item: Traditional and Dynamic Binary Scoring

		Traditional scoring			
Dynamic task	*n*	First 2 items (*M*)	Last 2 items (*M*)	*t*-value	1-tailed *p*-value
PA	12	0.42	0.50	t(11) = 0.43	.337
SIF	13	0.15	0.85	t(12) = 2.64	.011*
NAL	9	0.56	0.78	t(8) = 0.69	.256
TOM	8	0.63	1.75	t(7) = 3.81	.004*
		Dynamic binary scoring			
Dynamic task	*n*	First 2 items (*M*)	Last 2 items (*M*)	*t*-value	1-tailed *p*-value
PA	12	1.50	1.42	0.32	.380
SiF	13	0.92	1.85	3.49	.002*
NAL	9	1.67	1.89	0.80	.224
TOM	8	1.75	2.00	1.53	.085

*$p < .01$

Although the subset of the original 32 children who needed prompting showed significant improvement on the NAL task, this finding was not replicated among the more English-experienced dynamic group subset who needed prompting. Instead,

children in the dynamic group who needed prompting on the first NAL item showed no significant differences between the first two and last two items using any of the three scoring methods. As in the original study, the dynamic group did not significantly improve on the PA task using any scoring method. To summarize, the dynamic group and its subset trended toward improvement within the ToM and SiF tasks and the static group did not. This trend was strongest for the SiF task, which consistently showed significant improvement by the entire dynamic group and its subset using each scoring method.

Table 7

Static and Dynamic Group Performance on Standardized Measures

	Static group (*N*=16)		Dynamic group (*N*=16)			
Standardized measure	*M*	*SD*	*M*	*SD*	*t*(15)	2-tailed *p*-value
WMLS-RPV English SS	87.63	18.25	94.44	17.01	1.10	.288
Total PLS-4 SS	93.44	10.99	93.94	8.14	0.14	.889

Research question 3.

We used Pearson's *r* to answer the final research question regarding the relationship between performance on the static language tasks and standardized language tests. The possible range of total static language scores was 0-24 (1 point for each correct item, 6 items per task), with higher scores indicating higher performance. Standard scores for the PLS-4 are based on a mean of 100 (*SD* = 15). The WMLS-RPV subtest is based on a mean of 100 (*SD* = 15). The means and standard deviations for the children's performance in the static and dynamic groups are shown in Table 7. There was no significant difference between groups on the PLS-4 and the WMLS English vocabulary test, confirming that the static and dynamic groups were well matched. The static group

was also administered the Spanish version of the WMLS-RPV to gain a more complete language profile ($M = 43.9$, $SD = 26$).

First, relationships between total language task performance and performance on the PLS-4 and the WMLS-RPV English version for the static group were examined. Using one-tailed probability levels, we found a significant positive correlation between the static group's overall task scores and PLS-4 Total Language Scores ($r = .57$, $p = .01$). In contrast, using the original 0-3 scoring range, the English dynamic group's overall task scores were only moderately correlated to PLS-4 Total Language Scores, $r = .39$, $p = .069$. The difference between the static and dynamic group correlations was significant at a two-tailed probability level, $t(13) = 2.19$, $p < .05$. Using one-tailed probability levels, there were no statistically significant correlations between the WMLS-RPV English version and overall performance on either static or dynamic language tasks.

Performance on the English version of the WMLS-RPV relative to individual static tasks was also examined. Specifically, the Similarity in Function and Novel Adjective Learning tasks were expected to correlate with the WMLS-RPV-English because they each measure semantic skills. Separate Pearson correlations were run to determine the relationship between scores on Similarity in Function and Novel Adjective Learning tasks (both individually and together as a single measure of lexical skills) and the WMLS-RPV English version. Because positive correlations were expected, one-tailed probability levels were used. There were no significant relationships between the WMLS-RPV English version and the aforementioned language tasks. As shown in Table 8, there were no significant positive correlations between individual static tasks and standardized measures. However, there was strong positive correlation between static group scores on

the combined language tasks and PLS-4. Furthermore, the correlation between the static tasks and the PLS-4 was significantly stronger than the correlation between the combined dynamic tasks and the PLS-4.

Table 8

Correlations Between Static Tasks and Standardized Measures

Standardized measure	Static task(s)	*r*	1-tailed *p*-value
WMLS-English (monolingual SS)	SiF Task	.20	.234
	NAL Task	.15	.288
	SiF +NAL	.28	.144
	Combined Task Total	-.07	.405
Total PLS SS	Combined Task Total	.57	.011*

*$p < .01$

Chapter 5

Discussion

This study examined the viability of a graduated prompting approach for measuring modifiability among typically developing preschool children with Spanish-English bilingual experience. Graduated prompts are of particular interest because of their potential to be administered within a brief language-screening context. Differences in performance between static and dynamic language tasks groups were not found when both data sets were held to the same standard for accuracy. However, when the dynamic group's scaffolded correct responses were accepted, differences were significant on three out of four language tasks and overall task scores. This suggests that, compared to static assessment, graduated prompting provided additional information about children's modifiability on language tasks.

The second hypothesis, predicting little if any static group improvement for each language task, was also confirmed. In contrast to the dynamic group, which improved on the SiF and ToM tasks, the static group showed no significant differences in performance between the first two items and the last two items of any language task. This suggests that changes by the dynamic group were attributable to the graduated prompts. The trend of improvement on the dynamic version of SiF was replicated when we analyzed only the English-experienced dynamic group in this study. The subset of the dynamic group who needed at least one prompt on the first item of a task performed similarly to the whole group by improving on the SiF and ToM tasks. However, the subset of the static group who answered the first task item incorrectly improved on PA and NAL tasks, in contrast with the whole group. Although Patterson et al. (2013) reported a significant difference in

performance on NAL for the subset of 32 children who needed at least one prompt, the same results were not replicated on the more English-experienced dynamic group subset.

Finally, the predicted positive correlations between overall performance on the four static language tasks and PLS-4 were confirmed. However, neither combined nor individual language tasks correlated with other standardized measures. These findings complement the aforementioned data about modifiability within specific language tasks by supporting their validity as a combined global measure. They also indicate the combined tasks' potential for initial assessment of overall language skills in a screening context.

Static and Dynamic Group Performance

Comparisons between the static and dynamic groups aimed to isolate the effects of graduated prompting on performance. The first hypothesis predicted higher performance on the dynamic versus static versions of the four language tasks. In order to allow for direct comparisons with the static group, the traditional method for the dynamic language tasks credited only unprompted correct answers. When this method was used, the expected differences in performance were not confirmed. Rather, both groups performed similarly on the individual and combined language tasks. One explanation for this similarity is the brief nature of each task. The prompts were designed to provide short, structured lessons that could be reasonably administered in a screening context. Although the tasks were brief, improved performance on the latter items of some tasks suggests participants were able to transfer knowledge acquired through graduated prompts to later items. However, the tasks may have been too brief to effect completely

independent transfer to new items. By crediting only unprompted correct responses, the traditional scoring method may be less sensitive to different degrees of correctness.

Although traditional scoring, which credits only unscaffolded responses, allows for more direct comparison of dynamic and static group performance, it also removes some of their key differences. The case for dynamic assessment rests largely on the argument that it yields more process-oriented data about "how modifiable the child is" (Lidz, 1983, p. 60). Equating a scaffolded correct response with an incorrect response reverts back to a product-oriented approach by dismissing data about responsiveness to support. Therefore, dynamic binary scoring was also applied to the dynamic group, which expanded the criterion for credit to include prompted correct responses. This allowed for comparisons between the groups while maintaining their core distinguishing characteristics. When the dynamic binary scoring method was used, there was a clear trend of higher performance on the dynamic over the static tasks. The dynamic group performed higher than the static group on the overall language task measure. It also outperformed the static group on each individual language task, with significant between group differences on PA, NAL, and ToM.

Application of both scoring methods resulted in two distinct patterns of performance—difference versus no difference—that reflect the core differences between static and dynamic assessments. When independent performance was the standard for accuracy on the dynamic tasks, results were similar to the static group. The dynamic binary method's inclusion of supported responses allowed for crediting responsiveness to support in addition to response accuracy. Moreover, the improved scores when scaffolded responses were credited highlight the potential of the graduated prompt format to yield

more specific data about responsiveness to support. For example, dynamic group participant #7 did not complete any NAL task items independently, as shown in Table 9. When the first scoring method was applied, her task score was 0. However, when the dynamic binary scoring was applied, her score was 5. The higher score achieved using the dynamic binary scoring method indicates the child's potential to complete the task when supported. This suggests that dynamic binary scoring has higher diagnostic sensitivity in distinguishing between children who are responsive to support from those who are not. This is consistent with other findings that modifiability, particularly responsivity, is a useful metric for distinguishing typically developing from language-impaired CLD children (Peña, 2000; Ukrainetz et al., 2000).

Table 9

Dynamic Group Participant #7 Novel Adjective Learning Task Responses

NAL item	# of prompts	Traditional Score	Dynamic Binary Score
1	2	0	1
2	2	0	1
3	1	0	1
4	3	0	0
5	1	0	1
6	1	0	1
Total	10	0	5

The dynamic group's better performance on each language task is one indicator that graduated prompts facilitated learning within those tasks. This finding suggests that graduated prompts may be a measure of responsivity, or learning potential. In an analysis of the ratings scales used in several MLE studies, Peña (2000) defined modifiability as the "combination of planning, attention to task/discrimination, motivation, transfer, responsivity, and examiner effort" (p. 87). With the exception of motivation, all parameters were significantly different in the typically developing and low-language

ability groups. Specifically, the latter three parameters of modifiability best differentiated the groups. In other studies, graduated prompts have accurately predicted later performance (Bridges & Catts, 2011) and distinguished typically developing from language-impaired children (Hasson et al., 2012).

Still, the exact relationship between prompts and modifiability has been unclear. According to Lidz (1991), the greatest concern regarding the graduated prompt framework is the open question about "the nature and meaning of the metric generated by the prompting procedures" (p. 30). The quantifiable difference in performance between children presented with static tasks versus graduated prompts directly addresses this concern. This study describes responsiveness to graduated prompts among typically developing children with bilingual experience for future comparison with language-impaired children. If the number of prompts needed indicates learning potential, it would be expected that less modifiable, language-impaired children would need a consistently high number of prompts.

The dynamic group's improvement on the last two items of some language tasks further suggests that graduated prompts provide a viable format for observing modifiability within those tasks. Comparisons in performance between earlier and later task items demonstrate the children's ability to transfer newly learned information to new contexts. Although the dynamic group improved on SiF and ToM tasks, repeated but unscaffolded exposure for the static group did not result in improved performance on later items for any task. This confirms the second hypothesis predicting no significant improvement on later task items for the static group. Closer analyses of performance among children in both groups who did not complete the first task item without

prompting further supported this finding. Although the static group subset unexpectedly improved on PA and NAL tasks, it did not improve on SiF and ToM tasks. Conversely, the dynamic group subset improved only the SiF and ToM tasks. The pattern of improvement on SiF and ToM tasks by the dynamic group and its subset signifies a consistent trend. This pattern of change versus no change confirms the role of graduated prompts in dynamic group improvement for those two tasks.

Language Tasks

Regardless of which scoring method was applied, the dynamic group and its subset of children who needed prompting on the first item consistently improved on the last two SiF task items. The lack of improvement by the static group subset supports the original study's conclusion that modifiability within semantic tasks may be observed and measured within a relatively abbreviated graduated prompt framework. The particular suitability of semantic tasks for observing modifiability among CLD preschoolers is consistent with previous studies. For example, Ukrainetz et al. (2000) found responsiveness to support during MLE and post-test measures of semantic categorization skills effectively differentiated Native-American preschool children with stronger and weaker language skills. Although we used the graduated prompt framework within brief tasks, children's consistent modifiability within SiF suggests the task's potential for differentiating children within a screening context.

Our findings are novel in establishing that dynamic assessment of theory of mind is viable. Unlike the static group, the dynamic group showed a strong trend of improvement on the ToM task, particularly among the subset who needed prompting on the first item. This indicates that dynamic group participants who didn't initially

understand the task were modifiable when provided prompts. In other words, the ToM task was within their ZPD. This outcome is not surprising, as comprehension of false belief usually occurs by 5 years of age. There is a well-established link between language and theory of mind development, particularly for false-belief tasks (Astington & Baird, 2005). Among children with SLI, grammar and vocabulary skills are especially correlated to performance on theory of mind tasks (Farrar et al., 2009). Our results show that modifiability of theory of mind skills may be observed within a graduated prompt format. Furthermore, assessment of modifiability within theory of mind tasks contributes to a well-rounded, diverse set of language tasks that is recommended in distinguishing language difference from disorder.

Improvement on the Novel Adjective Learning and Phonological Awareness tasks by the subset of the static group who answered the first item wrong was an unexpected finding. Because the subset consisted of participants on the extreme end of the scoring scale, improvement on the tasks may be due to regression to the mean. That is, low scores may have increased due to the tendency of extreme scores to regress toward the average upon repeated measurement.

Another possible explanation for static group improvement on NAL is the role of task exposure in facilitating learning. For the typically developing children in the study, repeated exposure to the same task may have bridged the familiarity gap for semantic learning tasks. A 2004 study by Alt et al. found that children with language impairments had poorer ability than typically developing peers in receptively fast-mapping the semantic features of novel words. Research suggests that children with language impairment require more exposures to a word to comprehend or produce a new word than

their typically developing peers (Ellis Weismer & Hesketh, 1998; Gray, 2004; Rice, Oetting, Marquis, Bode, & Pae, 1994). Given the structured nature of the NAL task, which highlighted a specific semantic feature, the typically developing children in the study may have been able to map the semantic features of novel words in receptive tasks after a single exposure.

The more English-experienced subset of the dynamic group who needed at least one prompt on the first NAL item did not improve on the last two items. This finding was surprising because the subset of the 32 children in the original study showed significant change. The original study included two groups: one with more English and the other with more Spanish experience. Therefore, the different results when only the English subgroup was analyzed indicates that the majority of improvement on the NAL task came from the more Spanish-experienced group. The more English-experienced subset's slight improvement may have been related to task exposure or graduated prompts. However, the improvement was only significant when the more Spanish-experienced subset was included. Although language exposure was the primary selection criterion for the study, different language usage profiles between groups may also have affected task performance (Hammer, et al., 2012; Peña, Gillam, Bedore, & Bohman, 2011). It is likely that participants who had more Spanish-language experience while attending a primarily English-speaking preschool used both languages regularly. Bilingualism acquired in early childhood has been shown to facilitate novel word learning as an adult (Kaushanskaya, 2012). The improvement on NAL by the more-Spanish experienced group suggests this "bilingual advantage" may possibly be observed earlier than previously thought.

The static group subset's improvement on the PA task is surprising because the supported dynamic group did not improve. Patterson et al. (2013) posited that the tasks were too difficult to show change, i.e., not within the children's ZPD. The static group's low performance on the last two task items supports the argument that tasks were too difficult ($N = 16$, $M = 0.63$ out of 2.0). The static subgroup's improved performance likely indicates regression to the mean, as scores on the first two items were extreme ($M = 0$). Although the subgroup improved on the last two items, the low scores ($n = 12$, $M = 0.68$ out of 2.0) were similar to the entire group.

Low performance on the phonological awareness task by both static and dynamic groups may also indicate a poor match between the task, testing format, and participant profiles. Although Kantor, Wagner, Torgesen, and Rashotte (2011) reported dynamic assessment of phonological awareness skills among 4-year olds did not provide more information than static assessment, the participants in their study were middle to upper-middle class. Their participants' ability to perform the tasks using both static and dynamic assessment formats suggests prior experience with the phonological awareness tasks. The static and dynamic groups in this study came from lower socioeconomic backgrounds and may have lacked prior exposure to the specific phonological awareness skill being assessed. Considering the hierarchical nature of phonological awareness development, the brief, graduated prompt format may not be suitable for assessing modifiability within discrete phonological awareness skills. For example, a child with emerging skill at an earlier developing stage of phonological awareness, such as rhyming or syllable segmentation, may be unresponsive to graduated prompts targeting a later developing skill. An alternative approach may be to assess a child's responsiveness along

different developmental levels of phonological awareness. For the children in this study, an earlier developing phonological awareness task may have been more appropriate.

Validating Dynamic Assessment

One of the primary criticisms of dynamic assessment approaches is the lack of "satisfactory metric characteristics" (Haywood & Tzuriel, 2002, p. 43). Research must address concerns about reliability and validity if dynamic assessment approaches are to become widely used. The issue is especially challenging for the MLE approach, which requires individualized mediation that is incompatible with standardization. In previous studies of MLE, modifiability measures have accurately identified children with language impairment (Peña, 2000). But their dependence on Likert scales to measure modifiability required extensive calibration to achieve reliability. This presents obvious challenges to reliability when applied on a larger scale. The graduated prompting approach, however, is sufficiently structured to allow a high level of standardization. The high inter-rater reliability achieved during static task administration further suggests the tasks are well suited for standardization.

We also examined the validity of the four tasks relative to other standardized language measures. The strong positive correlation between the combined static task and PLS-4 Total Language Scores is one indicator that performance on the tasks reflects overall language skills. In contrast, total dynamic group scores were only moderately correlated to PLS-4 Total Language Scores. This pattern aligns well with the theoretical bases of both static and dynamic approaches. Dynamic assessment aims to gain more detailed information about learning potential than static assessment. Therefore, only modest correlations between the two types of assessments would be expected. In this

case, the graduated prompts provided scaffolding when needed, in direct contrast to the unsupported PLS-4 administration. The positive correlation between static language tasks and the PLS-4 is an important step toward validating the combined dynamic tasks as an initial indicator of language ability.

The tasks may also be particularly suitable for bilingual children. Because both studies focus on children with bilingual experience, parallel tasks were developed with consideration of variable linguistic abilities. Similar performance levels between English and Spanish task administration groups in the Patterson et al. study (2013) suggest potential for use among bilingual children across a broad range of acquisition patterns (Peña et al., 2003).

Limitations

There were some limitations in interpreting the results of this study. Since language experience was the emphasis of the original study, it was the primary matching criterion. Although the groups were well matched in that they performed similarly on the standardized measures, we did not control for differences in language usage. Any differences in usage may have influenced performance on the four static language tasks. Performance on language screening tasks among bilingual children has been linked to language usage and dominance profiles (Peña et al., 2011). Of the three participants who performed higher on the Spanish version of the WMLS-RPV, two performed similarly to the rest of the group on the static tasks and one did not. This unpredictable relationship between language experience and performance underscores the need to assess a bilingual child's skills in both languages. Also, we may reconsider inclusion of the PA task in our refinement of the language screener. Although it contributed to a well-rounded task

variety for overall language assessment, the specific task may not be appropriate for measuring modifiability for this age group. Whether administered in a dynamic or static format, children tended to perform poorly on the task. Future revisions of the language screener may include a different phonological awareness skill or set of skills.

Conclusion and Implications

Our results indicate graduated prompts may be used to observe and measure modifiability within a brief initial screening context. The findings of this study support earlier conclusions by Patterson et al. that a brief six-item task was sufficient to observe modifiability among typically developing preschoolers with bilingual language experience. The dynamic group's better performance on all four tasks compared to the static group suggests graduated prompts facilitated learning potential on each task. Graduated prompts also allowed for observation of modifiability within individual tasks. Whereas the dynamic group consistently improved on the last two SiF and ToM items, the static group did not. For those tasks, the prompts appear to have facilitated the children's transfer of newly learned skills to new problems. Given that modifiability measures have been shown to accurately distinguish language difference from disorder, graduated prompts show promise for early identification of language impairment among bilingual children.

The positive correlation between the combined static language tasks and the PLS-4 provides evidence of content validity for the combined tasks as an initial measure of overall language skill. This relationship demonstrates the combined tasks' potential to accurately assess overall language skill within an abbreviated screening context. A strong

trend of dynamic group improvement on the SiF and ToM tasks indicates these tasks are especially well suited for observing modifiability among CLD children.

These findings are consistent with arguments that graduated prompting is not a comprehensive evaluation approach for guiding intervention plans (Lidz, 1991). Instead, its distinctly narrow focus serves a different purpose. Our results suggest this limited scope may effectively evaluate two critical components of modifiability: responsivity and transfer. Although there is some debate whether these are related or distinctive elements of the modifiability paradigm (Kozulin, 2011), their role in identifying language impairment is clear (Peña, 2000). The graduated prompting framework's potential for evaluating these two parameters holds high promise for screening language skills. After identifying candidates for evaluation, more comprehensive measures, such as MLEs, would serve the separate purpose of diagnosing impairments and describing strengths and weaknesses.

In a position statement on screening and assessing bilingual children, The National Association for the Education of Young Children (2005) recommends regular screenings "using linguistically and culturally appropriate screening tools" (p.6). Federal legislation also mandates racially and culturally nondiscriminatory assessments of all school children (IDEA, 2004). However, the complexities of bilingual language development and the scarcity of appropriate assessment tools are significant barriers to the realization of these goals. This study is an important contribution toward the creation of screening tools that are developed and tested on CLD children, particularly for the large percentage who come from Spanish-speaking homes. Still, more research is warranted to further clarify the relationship between modifiability, the number of

graduated prompts needed during language tasks, and the identification of language impairment. Specifically, future studies may examine if measures obtained from graduated prompts during language screenings can accurately differentiate typically developing from language-impaired children with bilingual experience.

APPENDIX A

Complete Matching Profile for Dynamic and Static Groups

Dynamic group		Static group	
Participant #	HLEI	HLEI	Participant #
P01ENG	2	2 (±0)	MP02ENG
P02ENG	3	3 (±0)	MP01ENG
P03ENG	2.5	3 (+0.5)	MP14ENG
P04ENG	2	2 (±0)	MP04ENG
P05ENG	2.5	3 (+0.5)	MP15ENG
P06ENG	3.5*	3.5 (±0)	MP05ENG
P07ENG	2	2 (±0)	MP06ENG
P08ENG	1.5	1.5 (±0)	MP03ENG
P09ENG	1.5	1.5 (±0)	MP07ENG
P10ENG	2.5	3 (+0.5)	MP16ENG
P11ENG	3	3 (±0)	MP08ENG
P12ENG	2.5	2.5 (±0)	MP12ENG
P13ENG	1.5	1.5 (±0)	MP10ENG
P14ENG	1.5	2 (±0)	MP09ENG
P15ENG	3.5*	3.5 (±0)	MP11ENG
P16ENG	2	2 (±0)	MP13ENG
N=16; 11 females; 5 males		*N*=16; 11 females; 5 males	
Avg. age: 53.31 mos.		Avg. age: 54.44 mos.	
Avg. HLEI: 2.31		Avg. HLEI: 2.38	

Note. HLEI = Home Language Exposure Index, average of Questions 1 and 2.
*Matched exactly

REFERENCES

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language & Hearing Research, 47*, 407-420.

Astington, J., & Baird, J. (Eds.). (2005). Why language matters for theory of mind. New York: Oxford University Press.

Barrera, I. (1995). To refer or not to refer: Untangling the web of diversity, "deficit," and disability. *New York State Association for Bilingual Educational Journal, 10*, 54-66.

Bedore, L., Peña, E., Garcia, M. & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools, 36*, 188-200.

Blank, M., Rose, S. A., & Berlin, L. J. (2003). *Preschool language assessment instrument, second edition*. Austin, TX: Pro-Ed.

Blumenfeld, M. & Kaushanskaya (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research, 50*, 940-967.

Bridges, M.S., & Catts, H.W. (2011). The use of dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*(4), 330-338.

Brown, C.L. (2004). Reducing the over-referral of culturally and linguistically diverse students (CLD) for language disabilities. *NABE Journal of Research and Practice*, 2:1, Winter, 225-243.

Brown, A.L. & Ferrara, R.A. (1985). Diagnosing zones of proximal development. In J. Wertsch

(Ed.). *Culture, communication, and cognition: Vygotskian perspectives* (pp. 272-305). New York: Cambridge University Press.

Burton V. & Watkins, R. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders, 40*, 335-356.

Campione, J.C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22*(3), 151-165.

Campione, J.C. & Brown, A.L. (1987). Linking dynamic assessment with school achievement. In C.S. Lidz (Ed.). *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-115). New York: Guilford.

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36,* 948-958.

Catts, H., Adlof, S., Hogan, T., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, 48, 1378-1396.

Dollaghan, C. (1987). Fast mapping in normal and language- impaired children. *The Journal of Speech and Hearing Disorders, 52*(3), 218–222.

Ellis Weismer, S., & Hesketh, L. J. (1998). The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1444-1458.

Ennis, S.R., Rios-Vargas, M. & Albert, N.G. (2011). *The Hispanic population: 2010 (*C2010BR-04). Washington, D.C.: U.S. Census Bureau.

Farrar, M.J., Johnson, B., Tompkins, V., Easters, M., Zilisi-Medus, A., & Benigno, J.P. (2009). Language and theory of mind in preschool children with specific language impairment. *Journal of Communication Disorders, 42,* 428-441.

Feuerstein, R., Rand, Y. and Hoffman, M.B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques.* Baltimore: University Park Press.

Fortuny, K., Hernandez, D. J., & Chaudry, A. (2010). *Young children of immigrants: The leading edge of America's future.* Washington, DC: The Urban Institute.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6,* 1-15.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47*, 1117–1132.

Grigorenko, E.L. & Sternberg, R.J., 1998. Dynamic assessment. *Psychological Bulletin, 124,* 75-111.

Gunderson, L., & Siegel, L. (2001). The evils of the use of IQ tests to define learning disabilities in first- and second- language learners. *The Reading Teacher, 55*(1), 48-55.

Gutierrez-Clellen, V.F. & Peña, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools. 32*, 212-224.

Hammer, C.S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. A. (2012). Predicting Spanish-English bilingual children's language abilities. *Journal of Speech Language Hearing Research, 55,* 1251-1264.

Hammer, C. S., Miccio, A. W., & Wagstaff, D. A. (2003). Home literacy experiences and their relationship to bilingual preschoolers' developing English literacy abilities: An initial investigation. *Language, Speech, and Hearing Services in Schools, 34*, 20-30.

Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2012). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language and Teaching Therapy, 29*, 57-75.

Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psychoeducational assessment. *School Psychology Review, 19*(4), 411-422.

Haywood, H.C. & Tzuriel, D. (2002) Applications and challenges in dynamic assessment. *Peabody Journal of Education*, *77*, 40-63.

Individuals with Disabilities Education Improvement Act of 2004, P.L. 108-446 (2004). Retrieved from http://idea.ed.gov/download/statute.html

Jitendra, A.K. & Kaneenui, E.J. (1993). Dynamic assessment as a compensatory assessment approach: A description and analysis. *RASE: Remedial & Special Education, 14*(5), 6-18.

Johnson, T., Rios, M., Drewery, M. Ennis, S., & Kim, M.O. (2010). *People who spoke a language other than English at home by Hispanic origin and race: 2009 (*ACSBR/09-19). Washington, D.C.: U.S Census Bureau.

Kantor, P., Wagner, R., Torgesen, J., & Rashotte, C. (2011). Comparing two forms of dynamic assessment and traditional assessment of preschool phonological awareness. *Journal of Learning Disabilities*, *44*, 313–321.

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*, 81-96.

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition, 15*(03), 470-489.

Kovacs, A.M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science 12*(1), 48-54.

Kozulin, A. (2011). Learning potential and cognitive modifiability. *Assessment In Education: Principles, Policy & Practice, 18*(2), 169-181.

Langdon, H. W. (1989). Language disorder or difference? Assessing the language skills of Hispanic students. *Exceptional Children*, *56*(2), 160–167.

Lantolf, J. P., & Poehner, M. E. (2004). *Dynamic assessment in the language classroom* (CALPER Professional Development Document CPDD-0411). University Park, PA: The Pennsylvania State University, Center for Advanced Language Proficiency Education and Research.

Leafstedt, J. M. & Gerber, M. M. (2005). Crossover of phonological processing skills: A study of Spanish-speaking students in two instructional settings. *Remedial and Special Education,* 26, 226-235.

Lidz, C. S. (1983). Dynamic assessment and the preschool child. *Journal of Psychoeducational Assessment, 1*(1), 59-72.

Lidz, C. S. (1991). *Practitioners' guide to dynamic assessment.* New York: Guilford.

McLaughlin, B. (1984). *Second language acquisition in childhood: Preschool children. Vol.1.* Hillsdale, NJ: Erlbaum Associates.

Miller, C. A. (2004). False belief and sentence complement performance in children with specific language impairment. *International Journal of Language and Communication*

Disorders, 39(2), 191–213.

National Association for the Education of Young Children (2005). *Screening and assessment of young English-Language Learners: Position statement.* Retrieved from http://www.naeyc.org/positionstatements

Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech & Hearing Research, 39*(2), 414-423.

Olswang, L., Bain, B. & Johnson, G. (1992). Using dynamic assessment with children with language disorders. In Warren, S. & Reichle, J. (Eds.), *Causes and effects in communication and language intervention* (pp.187–215). Baltimore: Paul H. Brookes.

Patterson, J. (1998). Expressive vocabulary development and word combinations of Spanish-English bilingual toddlers. *American Journal of Speech-Language Pathology, 7*, 46–56.

Patterson, J. L. & Rodriguez, B.L. (2005). Designing assessments for multilingual children. In M. Ball (Ed.), *Clinical Sociolinguistics* (pp. 230-241). Oxford: Blackwell.

Patterson, J. L., Rodriguez, B. L., & Dale, P. S. (2013). Response to dynamic language tasks among typically developing Latino preschool children with bilingual experience. *American Journal of Speech-Language Pathology, 22,* 103-112.

Pearson, B. Z., Fernandez, S., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*, 93-120.

Peña, E. D. (1996). Dynamic assessment: The model and language applications. In K. Cole, P. Dale, & D. Thal (Eds.). *Assessment of communication and language* (pp. 281-307). Baltimore: Paul H. Brookes.

Peña, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly, 21*(2), 87-97.

Peña, E. D., Bedore, L.M., & Rappazzo, C. (2003). Comparison of Spanish, English, and bilingual children's performance across semantic tasks. *Language, Speech, and Hearing in Schools, 34,* 5-16.

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

Peña, E. D., & Halle, T.G. (2011). Assessing preschool dual language learners: Traveling a multiforked road. *Child Development Perspectives, 5*, 28-32.

Peña, E. D., Iglesias, A., & Lidz, C.S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138-154.

Peña, E. D., & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech & Hearing Services In Schools, 28*, 323-332.

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.

Peña, E.D., Reséndiz, & Gillam, R.B. (2007). The role of clinical judgments of modifiability in the diagnosis of language impairment. *Advances in Speech-Language Pathology, 9*(4), 332-345.

Rakhlin, N., Kornilov, S.A., Reich, J., Babyonyshev, M., Koposov, R.A., & Grigorenko, E.L. (2011). The relationship between syntactic development and theory of mind: Evidence from a small-population study of a developmental language disorder. *Journal of Neurolinguistics, 14,* 476-496

Restrepo, M. A., & Silverman, S. (2001). The validity of the Spanish Preschool Language Scale for use with bilingual children. *American Journal of Speech-Language Pathology, 10*, 382–393.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37,* 106-122.

Rogoff, B. (1991). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.

Tabors, P. & Snow, C. (1994). English as a second language in preschools. In F.Genesee (Ed.). *Educating second language children: The whole child, the whole curriculum, the whole community* (pp.103-125). New York: Cambridge University Press.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz, CA and Washington, DC: Center for Research on Education, Diversity & Excellence.

Tzuriel, D. & Haywood, H. C. (1992) The development of interactive-dynamic approaches to assessment of learning potential. In H.C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 3-37), New York, NY: Springer-Verlag.

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*, 142-154.

Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D .K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63,* 1012–1020.

U.S. Department of Education, National Center for Education Statistics. (2004). *The condition of*

education. Washington, D.C.: NCES.

van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in head start preschoolers with language impairment using scripted book-sharing discussions. *American Journal of Speech-Language Pathology, 12*, 1-11.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Winter K. (1999). Speech and language therapy provision for bilingual children: Aspects of the current service. *International Journal of Language and Communication Disorders 34*, 85–98.

Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey-Revised*. Itasca, IL: Riverside

Zimmerman, I., Steiner, V. & Pond, R. (2002). *Preschool Language Scale, Fourth Edition.* San Antonio, TX: Harcourt Assessment/The Psychological Corporation.