Technical Report # 2

On Demonstrating Construct Validity Using Wilson's Model to Create S-IGDI Pilot Measures

Given the current challenges in assessment for Spanish-English bilingual (SEB) children (see Technical Report 1 for a thorough description of these challenges) it is important to approach measure development with conceptually-sound and psychometrically-robust methodology. Currently-available measures of Spanish early literacy demonstrate design characteristics that are less than desirable (Wackerle-Hollman, Durán, Rodriguez, Brunner, Palma & McConnell, under review). More specifically, existing measures are rarely designed with Spanish language development in mind, and rather are translations of current English measures (see Technical Report 1). As such, the conceptual underpinnings of these measures are anchored to English development and do not appropriately model the development of Spanish early literacy skills and how to best assess them.

To alleviate this issue we have approached the design of our assessment tool by using Wilson's model (Wilson, 2005) for measure construction, which allows the measure to represent a manifestation of the construct of interest. Specifically, the Spanish Individual Growth and Development Indicators (S-IGDI) are being developed under the *construct modeling* principles for test development. The construct modeling framework is based on an item response modeling approach and consists of four building blocks: the construct map, item design, outcome space, and the measurement model (Wilson, 2005). This framework is often referred to as the Wilson model. This model provides guidelines for the process of test development and intends to facilitate understanding of how an instrument works by understanding how it is created. The Wilson model also relates to the principles set by the National Research Council Committee on the Foundations of Assessment on how to know what students know and it is typically summarized in the Assessment Triangle (Kennedy, Brown, Draney & Wilson, 2005). The relationship between the Wilson model and the Assessment Triangle is illustrated in Figure 1.
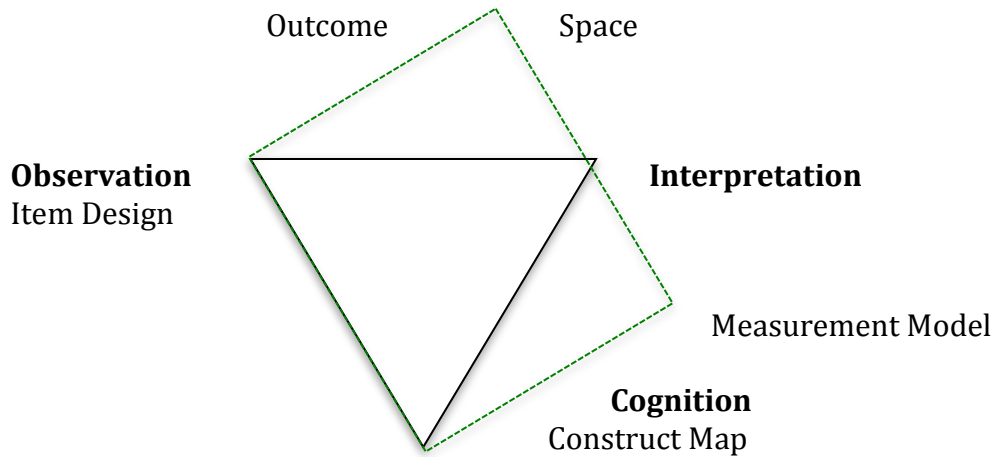
Figure 1. The NRC Assessment Triangle embedded within the Wilson's model.

The four building blocks of Wilson's model offer a unique perspective on measure design because they incorporate validity standards at each set of the process, allowing for inferences to be made about the construct(s) of interest. As such, the model is intended to be cyclical by using the information defined in each building block to further refine the tasks via model fit, reliability evidence, validity evidence, etc.

The first building block is the Construct Map, which is a conceptual representation of an underlying or latent cognitive skill that may be illustrated along a continuum of no skill to complete skill. Construct maps including operational definitions for each domain of interest (phonological awareness, oral language and alphabet knowledge) are provided in Figures 2-4.

Phonological Awareness

Definition: The meta-linguistic ability to understand that spoken words are comprised of small sound units; to detect, discriminate between, and manipulate these structural components; and to perform these skills independent of word meaning (Durgunoglu, Nagy & Hancin-Bhatt, 1993; Branum-Martin, Mehta et al., 2006; Cardenas-Hagan, Carlson & Pollard-Durodola, 2007; Kuo & Anderson, 2010; Gorman & Gillam, 2003; Anthony et al., 2011; Cisero & Royer, 1995).
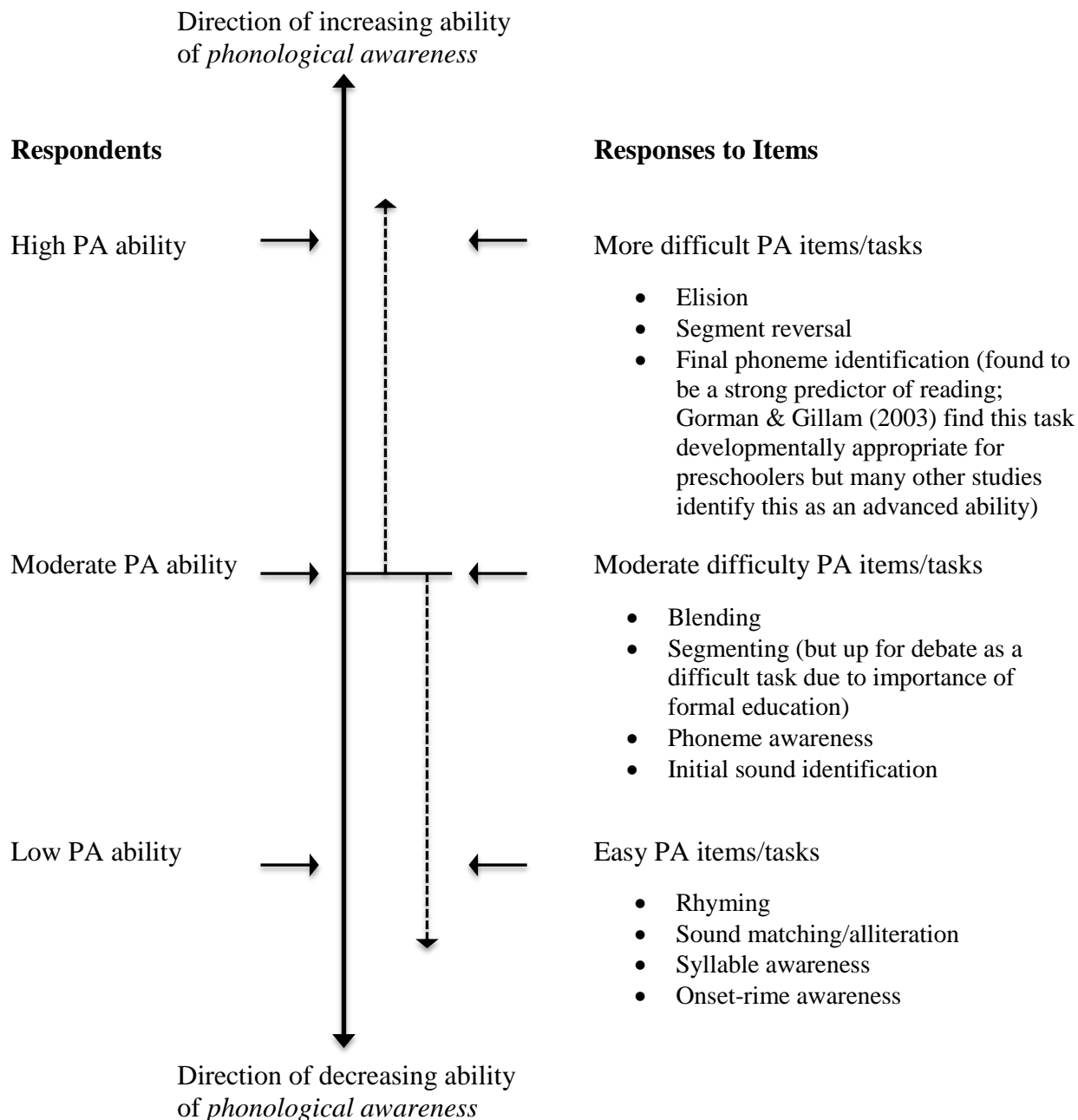
Direction of increasing ability
of *phonological awareness*

Respondents   **Responses to Items**

High PA ability →   ← More difficult PA items/tasks

- Elision
- Segment reversal
- Final phoneme identification (found to be a strong predictor of reading; Gorman & Gillam (2003) find this task developmentally appropriate for preschoolers but many other studies identify this as an advanced ability)

Moderate PA ability →   ← Moderate difficulty PA items/tasks

- Blending
- Segmenting (but up for debate as a difficult task due to importance of formal education)
- Phoneme awareness
- Initial sound identification

Low PA ability →   ← Easy PA items/tasks

- Rhyming
- Sound matching/alliteration
- Syllable awareness
- Onset-rime awareness

Direction of decreasing ability
of *phonological awareness*

Figure 2. Construct map illustrating students' performance and item locations for the Phonological Awareness construct.

Oral Language

Definition: The ability to use words to communicate thoughts and ideas to others, and in turn, understand ideas and thoughts from others (Dunst, Trivette, Masiello, Roper, & Robyak, 2008; Morgan & Meier, 2008).

- Expressive language: the use of words to express meaning.
- Receptive language: the ability to listen, process, and understand the meaning of spoken language.
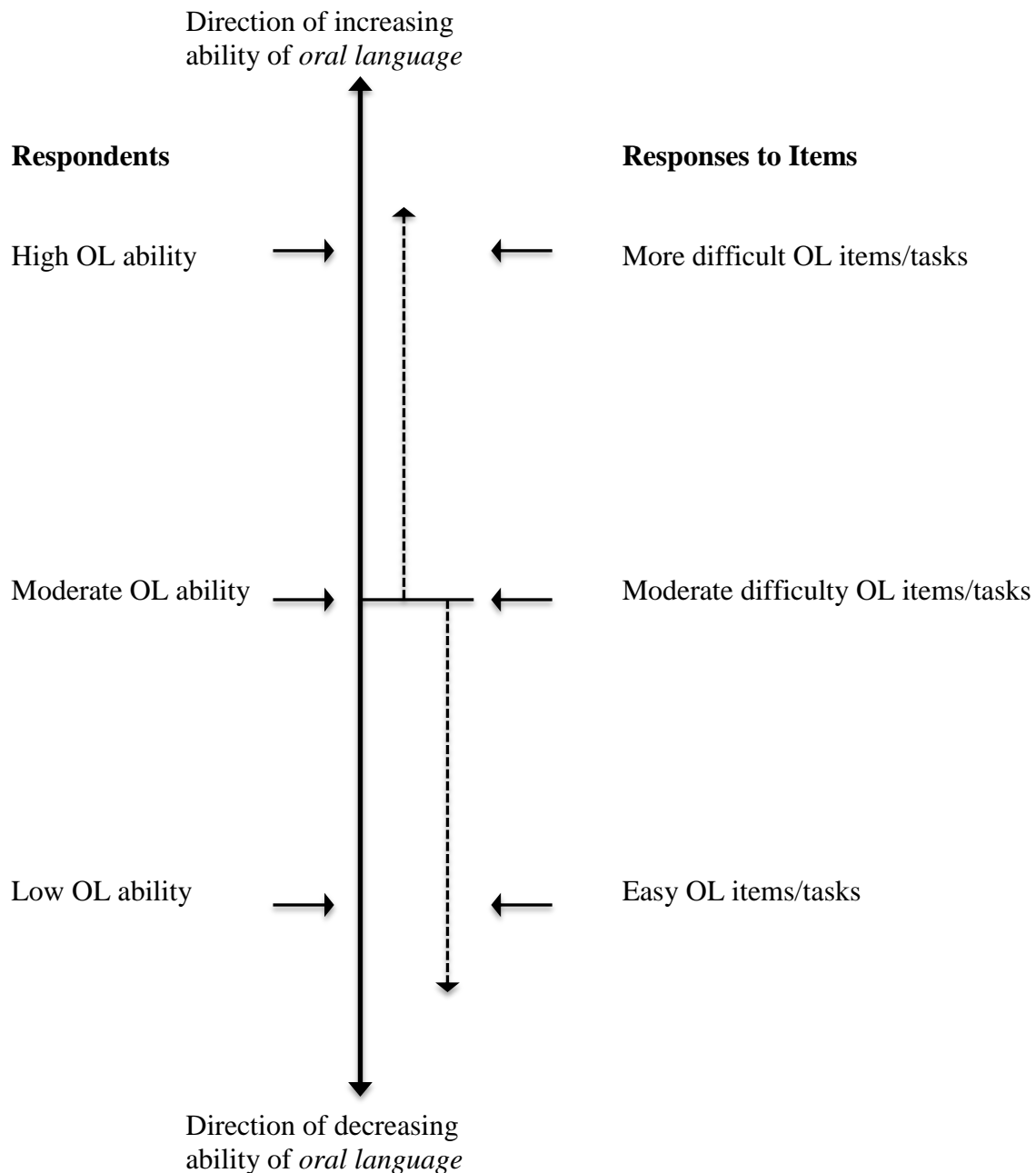
Direction of increasing
ability of *oral language*

Respondents		Responses to Items
High OL ability →	←	More difficult OL items/tasks
Moderate OL ability →	←	Moderate difficulty OL items/tasks
Low OL ability →	←	Easy OL items/tasks

Direction of decreasing
ability of *oral language*

Figure 3. Construct map illustrating students' performance and item locations for the Oral Language construct.

Alphabet Knowledge (AK)
Definition: Knowledge about the names, sounds, and symbolic representation of the 27 letters of the alphabet (McBride-Chang, 1999; Davison & Brea-Spahn, 2012).
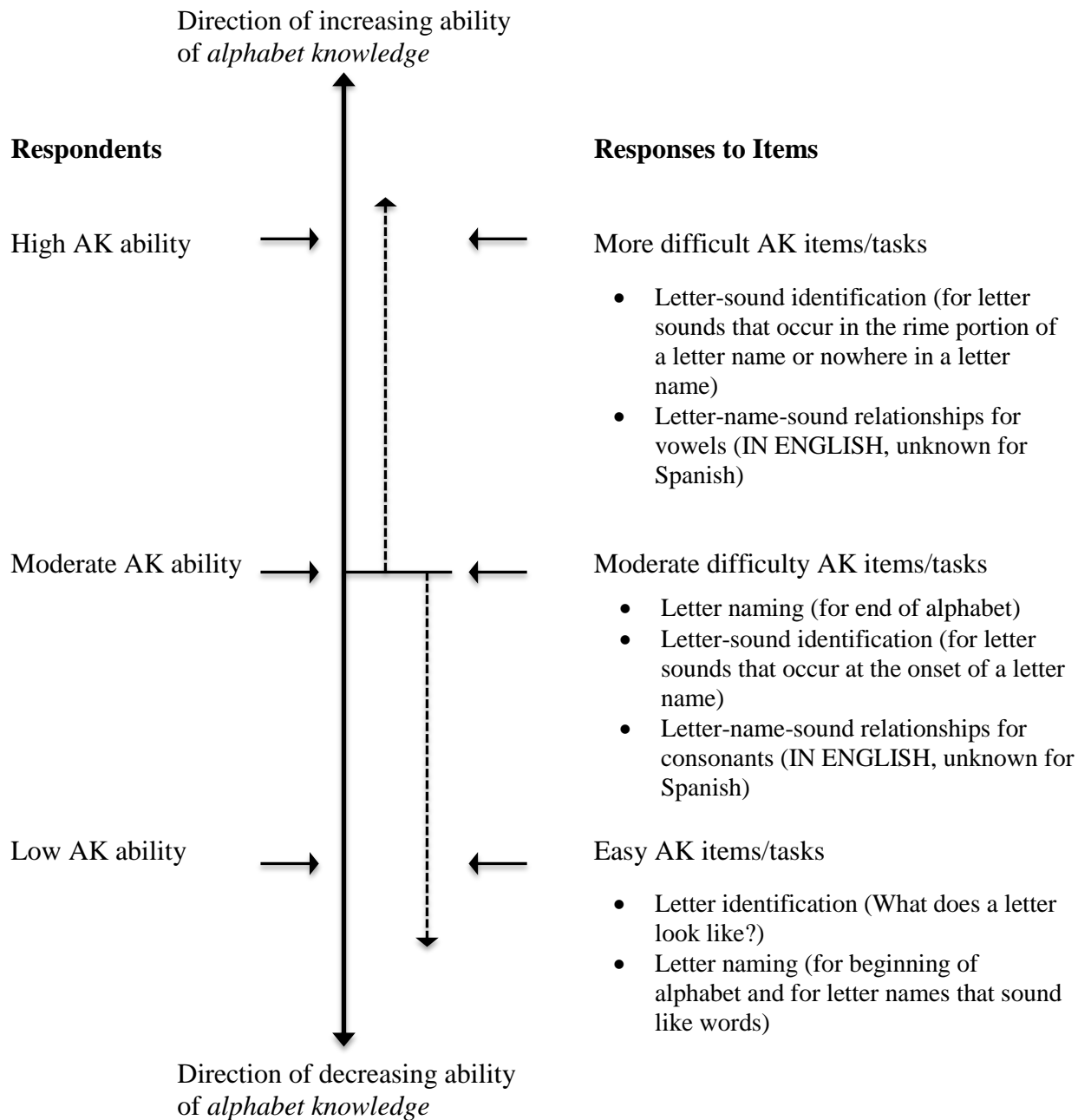
Direction of increasing ability
of *alphabet knowledge*

Respondents

High AK ability

Moderate AK ability

Low AK ability

Responses to Items

More difficult AK items/tasks

- Letter-sound identification (for letter sounds that occur in the rime portion of a letter name or nowhere in a letter name)
- Letter-name-sound relationships for vowels (IN ENGLISH, unknown for Spanish)

Moderate difficulty AK items/tasks

- Letter naming (for end of alphabet)
- Letter-sound identification (for letter sounds that occur at the onset of a letter name)
- Letter-name-sound relationships for consonants (IN ENGLISH, unknown for Spanish)

Easy AK items/tasks

- Letter identification (What does a letter look like?)
- Letter naming (for beginning of alphabet and for letter names that sound like words)

Direction of decreasing ability
of *alphabet knowledge*

Figure 4. Construct map illustrating students' performance and item locations for the Alphabet Knowledge construct.

The second building block is Item Design. The Item Design process is intended to answer the question, "How is this construct manifested into items?" To create items that accurately represent the construct, we first evaluated what information gathered from preschool age children can be considered evidence of knowledge and reasoning. Initially, we considered 23 tasks in which we hypothesized the construct could be manifested (see Technical Report 1). Then, using qualitative information such as observations of child behavior when interacting with the measures and quantitative data such as descriptive statistics, the number of measures was narrowed down to 11. At the beginning of such a process, the relationship between the construct and the items is often only vaguely known. To establish causality between the construct and the items, we assume that the test taker has a particular amount of the latent construct and that amount of the construct is a cause of that test taker's responses to the items (see Figure 5).   The content was determined through a literature review of Spanish language and literacy development as well as a corpus of Spanish words from Spanish children's books, Spanish early childhood curricular reviews and expert input (see Tech Report 1 for more information on developing the content). Once the content was established we then explored how we might solicit such evidence through multiple interactions or tasks. Finally, in creating items we carefully controlled features in items that would distract from the construct of interest, allowing for the inferences made about performance to be as close as possible to the item content. Taken together, this process allowed for the development of 23 different S-IGDI tasks.
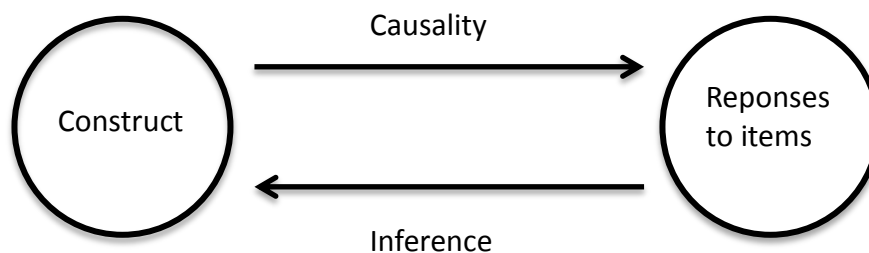

Figure 5. Picture of the relationship between the construct and the item responses.

The third building block is the Outcome Space. Once the items have been created it is important to clearly define how inferences can be made about the construct of interest through scoring. During this step we defined which characteristics of responses were valid

and which were extraneous. Positive scores indicate more knowledge (toward complete knowledge on the construct map). For multiple-choice items, responses were categorized as correct and incorrect (with the scoring being "1" and "0" respectively). Similar guidelines were used for the expressive measures where a valid response was scored as "1" and "0" otherwise. For the performance assessment (storybook), the scoring rubric was constructed such that partial credit options were available, with some additional items including the most complete response as a score of "2", partial credit as a score of "1" and no credit as a score of "0".

The last building block is the Measurement Model. Selecting a robust measurement model allows for empirical support of the inferences made during the outcome space. The measurement model is intended to help us understand and evaluate the scores that come from the item responses and, as a result, informs the construct and practical application of each task. When the measurement model is applied, each item is scaled on the construct of interest, illustrating the degree to which items map onto the level of difficulties represented.

Purpose of the Report

Using the first three building blocks from Wilson's model to support measure design and the conceptual foundations and design guidelines described in Technical Report 1 (i.e., Pena's tenets for measure design), this study examined the validity and utility of the 23 S-IGDI tasks during a pilot study trial. This process was established using qualitative and quantitative metrics to reduce the pool of candidate measures to a more manageable set of two to three tasks per domain, further supporting the final goal of one to two measures per domain. Measures selected within this report will move forward to field testing, where all building blocks will be utilized to scale and select promising measures (see Technical Reports 3 and 4).

<div align="center">

Method

</div>

Measures

To begin the measure creation process, 23 tasks were developed to tap into one of the three research-based early literacy construct definitions: oral language, phonological awareness, or alphabet knowledge. Both expressive and receptive tasks were designed for each construct.

Oral language. For S-IGDIs, oral language is defined as the ability to use words to communicate ideas and thoughts to others (Dunst, Trivette, Masiello, Roper, & Robyak, 2008;

Morgan & Meier, 2008). It includes expressive language, or the use of words to express meaning and receptive language: the ability to listen, process, and understand the meaning of spoken words .

Expressive measures. Expressive tasks for oral language included: (1) Picture Naming/*Denominación de los Dibujos*, (2) Categories/*Categorías*, (3) Functions/*Funciones*, (4) Verbs (Expressive)/*Verbos (Expresivo),* (5) Definitional Vocabulary (Expressive)/*Vocabulario de Definiciones (Expresivo),* and (6) Analogies/*Analogías.*

(1) Picture Naming/*Denominación de los Dibujos* requires children to name images of common and culturally-relevant objects, animals, foods, etc. Thus, this task evaluates children's ability to produce spoken vocabulary words. To administer the task, the child is shown each card in succession and asked "¿Qué es?" (What is this?). If an image has more than one name due to dialectical differences, all possible correct answers are listed on the back of the card.

(2) In Categories/*Categorías*, children must state the category to which the three images on a card belong, or how these images "go together". This task involves the semantic ability to understand group and category membership, as well as the ability to produce spoken language. When giving the task, the administrator names each image and then asks how these items go together. The back of the card contains all possible answers to account for cases in which objects may belong to more than one category or in which there are multiple ways to word the correct answer.

(3) Functions/*Funciones* items provide images of household objects, toys, and everyday nouns and children are asked to identify their function. Ability to describe an object's function was hypothesized to be especially important in this context considering that children acquiring Spanish may learn verbs before nouns (Peña et al., 2003). When displaying each item, the administrator names the image for the child and then asks "¿Para qué sirve?" (What is this object used for?). When an object has multiple functions, or when there are multiple verbs used to describe the same function (i.e., un carro sirve para conducir o manejar), all potential verbs or purposes are and accepted as correct responses.

(4) Expressive Verbs/*Verbos (Expresivo)* involves production of a verb that describes the action being portrayed in a picture. Each card contains one image and the examiner asks the child "¿Qué está pasando?" (What is happening?) Tasks containing verbs were supported

in the literature due to verbs' salience in Spanish language acquisition (Peña et al., 2003). While attempts were made to select images portraying one clear action, multiple possible responses are included for cards whose images solicit multiple verbs.

(5) *Vocabulario de Definiciones (Expresivo)* provides children with common everyday items and asks the children to first name the item or image and then provide the function of that item. As a result, this task offers a partial credit model in which children may respond to one portion of the two questions (vocabulary vs. definition) and receive half the points available. During this task the child is asked "¿Qué es?" (What is this?). If the child responds correctly, the assessor provides the prompt "¿Qué está pasando?"(What is happening?). If the child cannot name the image, the assessor provides the correct name of the item or objects and then prompts, "¿Qué está pasando?" (What is happening?).

(6) *Analogías* is designed to elicit knowledge of relationships between everyday nouns, adjectives and verbs. In this task children are provided with a verbal prompt that compares two words to demonstrate the relationships. The child is then asked to complete a second phrase by responding with the appropriate word (i.e., "Día es a la noche como el invierno es _____.").

Receptive measures. Receptive tasks for oral language included: (7) Which one doesn't belong?/*¿Cuál dibujo es diferente?*, (8) Receptives/*Receptivos*, (9) Verbs (Receptive)/*Verbos (Receptivo),* (10) Definitional Vocabulary (Receptive)/*Vocabulario de Definiciones (Receptivo)*, (11) Story Comprehension: Recall and Prediction/*Comprensión de la historia: el retiro y la predicción*, and (12) Defined Linguistic Interactions/*Interacciones definidas lingüísticas*.

(7) Which one doesn't belong?/*¿Cuál dibujo es diferente?* is a semantic task evaluating children's ability to distinguish between categories. Each card contains three images, two of which belong to the same category. After the administrator names each image on the card, the child may respond by either pointing to or saying the name of the image that does not belong.

(8) Receptives/*Receptivos* requires children to point to the image whose name the administrator says. As children must understand spoken words in order to correctly identify images, this task satisfies the receptive portion of the oral language construct definition. There are three image choices on each card.

(9) Receptive Verbs/*Verbos (Receptivo)* presents children with two or three images per card. Each image portrays an action. Children must use their receptive language skills to match the action said by the administrator to the correct image. This task differs from *Receptivos* only in the content of the images: in *Receptivos*, all images are nouns.

(10) Definitional Vocabulary (Receptive)/*Vocabulario de Definiciones (Receptivo)* presents children with one image per card. The assessor names the item and offers a statement such that a defining feature of the item is required for a response. For example, the assessor would say "Este es el sol, hace calor o frío?" (This is the sun, is it hot or cold?).

(11) Story Comprehension: Recall and Prediction/*Comprensión de la historia: el retiro y la predicción* offers children the opportunity to listen to a brief picture book selected from popular Spanish children's literature. After the story the child is asked questions such as "Que va a pasar despues?" (What will happen next?) (prediction) or "What just happened?"(Qué pasó?) (recall). The administrator also asks children to name nouns on given pages (vocabulary) and to answer general questions such as "Qué est pasando en éste cuento? "/"What happened in this story?".

(12) Defined Linguistic Interactions/*Interacciones definidas lingüísticas* is a task intended to support natural language use between the assessor and the child. For this task the assessor brings a box or bag of everyday items that go together in a primary theme. For example, the assessor might present to a child a small bucket, a shovel, a pair of sunglasses and a small towel. The assessor then asks the child to talk about what these objects are for and uses prompts such as "How do you know?" or "What would you use these for?"

Contextualized measures. Let's go to the store!/*¡Vamos a la tienda!* and Let's Talk!/*¡Vamos a hablar!* are performance-based tasks designed to capture a child's natural communication skills. Research indicates SEB students may be more likely to produce oral language in naturalistic settings (Hammer & Rodriguez, 2012; Peña & Halle, 2011); as a result, these measures were developed to appropriately tap into naturalistic resources for expressive and receptive language.

(13) The *¡Vamos a la tienda!* assessment is situated within a storybook format with all assessment questions centered on the context of a trip to the grocery store. Pictures were taken at a local Latino market to gather images of familiar food items in a setting that is familiar to SEB students. Manipulatives attached to the storybook with Velcro enhance

interaction during this assessment and are designed to encourage active engagement in the assessment process.

(14) *¡Vamos a hablar!* (originally titled *Narativos*, then *¡Vamos a jugar!*) is designed to elicit children's natural language as they describe the events depicted in various scenes. The embedded narrative in the pictures is intentionally designed to be unusual and engaging to encourage children's interest and motivation to talk about what they see in the pictures.

Phonological awareness. For the purposes of the S-IGDIs, we defined phonological awareness as the meta-linguistic ability to understand that spoken words are comprised of small sound units; to detect, discriminate between, and manipulate these structural components; and to perform these skills independent of word meaning (Durgunoglu, Nagy & Hancin-Bhatt, 1993; Branum-Martin, Mehta et al., 2006; Cardenas-Hagan, Carlson & Pollard-Durodola, 2007; Kuo & Anderson, 2010; Gorman & Gillam, 2003; Anthony et al., 2011; Cisero & Royer, 1995).

Expressive measures. Expressive tasks to measure phonological awareness skills included (15) Blending/*Mezclar* and (16) What word is left?/*¿Qué palabra queda?*.

(15) Blending/*Mezclar* is a task involving phonemic awareness understanding. Children listen to two sounds, or phonemes, separately spoken (i.e., bo/ca), then combine them to form a single word (i.e., boca). This allows children to associate individual sounds that when said together create a new word. When giving the task, the administrator says one sound (i.e., par), pauses, then makes the other sound (i.e., aguas). The child must respond by saying the answer. No picture cards are used in this measure.

(16) In What word is left?/*¿Qué palabra queda?*, children first hear a word or sound. Then, part of the word or sound is omitted. This task allows children to identify sound structure. The administrator says the entire word, (i.e., sandía), then takes away part of the word (san) and asks the child what word remains (día). These items are presented both with and without picture scaffolding. For the first half of the task the items are provided on picture cards and involve answers that are real words in the Spanish language. The second half of these items are presented verbally without pictures, as child responses are simply pieces of Spanish words that cannot be imaged on a card. For pictured items, the child can respond by pointing to the picture that corresponds with the answer or by saying the answer; for non-pictured items, the child must respond expressively.

Receptive measures. To measure children's receptive phonological awareness skills, (17) Rhyming/*Rimar*, (18) First Sounds/*Primeros Sonidos*, and (19) Detection/*Detección* were designed.

(17) Rhyming/*Rimar* requires the ability to discriminate between the endings of words independent of word meaning. For this task, children are required to match the ending sound of a target word to the word that rhymes when presented with either two or three word choices. To give this task, the administrator points to and names the target image in pairs with each of the word choices and asks the child "¿Cuáles son las dos palabras que riman?" (Which two words rhyme?).

(18) First Sounds/*Primeros Sonidos* items require detection of and discrimination between the initial sounds of words independent of word meaning. For this task, the administrator names each object on the card and then provides the beginning sound of one object, the target sound. Children must point to the image corresponding to this target sound.

(19) Detection/*Detección* items show children an image of an object and require children to distinguish between the object's complete name and an incomplete word distractor. For example, if the image were a strawberry, the assessor would ask the child: "Es una fres, o una fresa?"

Alphabet knowledge. For the purposes of the S-IGDIs, alphabet knowledge was defined as knowledge about the names, sounds, and symbolic representation of the 27(29) letters of the alphabet (McBride-Chang, 1999; Davison & Brea-Spahn, 2012).

Expressive measures. The only expressive measure of alphabet knowledge was (20) Letter Naming (Expressive)/*Denominación de las Letras (Expresivo)*.

(20) Letter Naming/*Denominación de las Letras (Expresivo)* requires children to name the upper- or lower-case letter of the Spanish alphabet present on a card when asked "¿Cuál es esta letra?"(What is this letter?).

Receptive measures. The tasks designed to measure receptive alphabet knowledge were (21) Letter Naming (Receptive)/*Denominación de las Letras (Receptivo)*, (22) Letter Detection/*Detección de las Letras*, and (23) Sound Identification/*Identificación de los Sonidos*.

(21) Letter Naming/*Denominación de las Letras (Receptivo)* requires children to use their receptive language understanding to point to the correct letter (out of three letters)

when the administrator says the target letter name. Such a task measures children's ability to distinguish between and know the names of the written letters of the alphabet.

(22) Letter Detection/*Detección de las Letras* asks children to correctly select the alphabet letter on a card that includes two other letter-like symbols. This ability to discriminate between letters and other symbols is demonstrated when the child points to the alphabet letter when asked "¿Cuál es una letra? (Which one is a letter?).

(23) Sound Identification/*Identificación de los Sonidos* requires children to correctly identify the target letter once the administrator makes the target letter sound. The child responds to each item by pointing to the correct letter on a card that includes the target letter and two distractor letters.

Sample Population

During the data collection effort each task was administered to a sample of 10 students, with a total sample of 33 students across sites. Basic demographic information is provided in Table 2, including gender, program location and type, and simultaneous or sequential acquisition of language (via parent and teacher report).

All students were four to five years in age, participating in preschool classrooms and school readiness programs with eligibility for kindergarten in the following academic year.

Table 2: Pilot Testing Demographics by Location

Site	Program Name	Sample Size	Program Type	Gender (% male)	Simultaneous/ Sequential Ratio (%)
Minnesota (n=14)	Joyce Preschool-Windom (Urban)	5	Private/ Scholarship	40%	60/40
	Joyce Preschool-Uptown (Urban)	3	Private/Scholarship	33%	33/66
	Spring Lake Park (Suburban)	6	Private /Dual Immersion	33%	50/50
Utah (n=19)	Centro de la Familia de Utah (Rural)	18	Migrant Head Start	44%	N/A
	Bear River Head Start (Rural)	1	Head Start	100%	N/A

Data Collection, Timeline and Fidelity Standards

Data collectors included three fluent Spanish graduate students funded by the S-IGDI project as Graduate Research Assistants (GRA). Prior to pilot data collection, each GRA was observed using a fidelity checklist via video-conference and obtained 100% fidelity on each of the 16 S-IGDI pilot measures. All student interactions with data collectors were recorded for video coding. Following data collection, video recordings were shared across sites. One GRA from Minnesota and one GRA from Utah watched all pilot videos and scored each measure according to the qualitative rubric. Rubric scores across sites were compared and discussed in preparation for the task elimination process.

The pilot testing occurred for 12 of the 16 S-IGDI measures between October and December 2012. The remaining four measures, *Mezclar, ¿Qué palabra queda?, ¡Vamos a la tienda!* and *¡Vamos a hablar!*, required further development and were thus piloted between January and April 2013.

Methods and Procedures

To determine if the measures demonstrated social validity, face validity, robust correlations with other existing and quality measures of Spanish early literacy and have utility in the hands of practitioners, we considered a series of qualitative and quantitative variables.

Qualitative Analysis

    Development standards. When designing the S-IGDI measures, it was critically important to keep the practitioner in mind, as teachers and professionals who interact with SEB children in U.S. preschool classrooms demonstrate dramatic variance in skill, knowledge-base and ability to understand and instruct Spanish speaking children. For the measures to be useful in these various classrooms we paid particular attention to the needs, desires and capabilities of professionals who work with SEB children. As a result, after designing each measure we provided end-users (teachers, para-professionals, support staff in classrooms with SEBs) with a survey that asks for opinions about a selection of tasks from each domain. The teacher survey is provided in Appendix A. Teachers completed these surveys simultaneous to the window of assessment in which children received pilot measure testing.

    Utility standards. Three utility standards were created to limit the burden on student and practitioner use and interpretation. First, the task could not continue on to piloting if it included tedious or overwhelming materials or test kits. Many early childhood measures come with large supply kits, manipulatives, and multiple manuals. These measures often take a significant amount of time to deliver, are difficult to maintain if materials are missing or damaged, and can be confusing for practitioners with many separate parts and protocols. Second, tests must demonstrate a cost-benefit relationship such that producing the task would not outweigh the benefit of the scores achieved. As such, any task that was particularly expensive or time-consuming to produce was eliminated from the pool of potential measures for piloting. Finally, any task that provided redundant information already available in a psychometrically sound format in the field was eliminated. Of the original 23 tasks, 16 were produced for initial pilot testing, with 5 measures (Defined Language Interactions/*Interacciones Definidas Lingüísticas*, Analogies/*Analogías*, Story Comprehension-Recall and Prediction/*Comprensión de la historia: el retiro y la predicción*, Definitional Vocabulary (Receptive)/*Vocabulario de Definiciones (Receptivo)* and Definitional Vocabulary (Expressive)/*Vocabulario de Definiciones (Expresivo)*) removed from the pool of tasks based on utility standards.

    Functionality rubrics. A qualitative rubric was developed to evaluate each measure's functioning in the field and to address each measure's adherence to design and development

standards. The qualitative rubric included the following criteria: (A) active engagement of child, (B) valid response patterns obtained from child, (C) ease of use by administrator, and (D) timeliness of measure administration and scoring. Each of these four components were rated using a 0-3 scale, where 0 represented an unsatisfactory and unresolvable measure that did not achieve its desired outcome, and where 3 represented a superior measure that achieved its desired outcome to the highest standard and was considered for further testing without reservations. The qualitative rubric is provided in Appendix B. Two independent coders reviewed each measure via video-recorded child interactions and used the qualitative rubric to arrive at an overall score for each measure. Overall scores ranged from 0-12.

(A) Active engagement was rated based on coder-observed child attention to the task, the extent to which the child seemed to enjoy the task, and whether or not the child responded to the administrator when asked a question. (B) Valid response patterns were determined by the degree to which meaningful data could be obtained from children. Data were considered meaningful by the coders when children responded thoughtfully and demonstrated understanding of what was being asked of them. A measure produced unreliable or invalid data when children guessed or consistently chose item distractors on the left, center, or right of the card. (C) Ease of administrator use was determined by the success with which an administrator could give the task to a child. This rating was completed with end users in mind: would the procedures allow for successful administration of the task by someone with minimal academic training (i.e., paraprofessionals)? (D) To evaluate the timeliness with which a measure could be delivered and scored, video coders timed the child's interaction with each measure's content. If administration of a measure took more than 5 minutes and/or if scoring took more than 1 minute, a lower rating was given, as according to General Outcome Measures (GOM) standards, S-IGDI measures must be quick and easy to administer (see Tech report 1 for more information about General Outcome Measure qualities).

In addition to the formal qualitative rubric, anecdotal evidence from assessor-child think-aloud interactions completed during pilot testing were used to bolster rubric-based decisions.

Quantitative Analysis

Empirical standards. To support the qualitative findings, we also examine standard empirical criteria for each task including mean, median, standard deviation, minimum value and maximum value. Empirical standards were used to evaluate the distribution of scores and provide a cursory view of the likelihood for a task to be successful with four to five year old children.

Results

Each measure was reviewed using the standards and rubric performance scale after pilot data was collected to determine candidacy for further field testing.

Qualitative Results

Of the 16 measures, functional and development standards were used to determine which tasks demonstrated the most face validity and social validity with end-users. A selection of measure was provided to end-users (teachers) in an online survey format. Results from the teacher survey are presented in Tables 3, 4 and 5. Survey questions were presented in three Likert scale with four degrees of agreement ranging from none/not very to very (none/not very, slightly, somewhat, very; none, some, most, all). It is important to note that the sample size for this survey was small, with varying samples across domains and a total sample of 25 responses. When delivering the survey we noted that each site could participate collectively or individually. This means that a team of teachers at any one site may have answered the survey as a group or teachers may have completed the survey on their own.

Descriptive responses and anecdotal summaries of teacher impressions about the measures were also obtained. Results are summarized here by domain.

For *Alphabet Knowledge,* teachers reported positive impressions: "This task [letter detection] was very clear, easy to use and the print was very clear", "This task [letter naming-expressive] is easy to use and it clearly aligns with what I teach in my classroom," and "this assessment taps something our other assessments don't attend to." Teachers also reported comparative results between tasks, such as "I liked this task better than letter detection because it is more difficult to identify specific letters among other real letters than other symbols." Similarly, some teachers also voiced challenges in using tasks, such as "Some of the Spanish instructions were unclear and scoring should be clarified," and "this task overlaps with other assessments I already use."

17

For *Phonological Awareness* teachers reported that they generally liked the suggested tasks and found the information useful. Teachers showed the strongest interest and satisfaction with the First Sounds task. Specifically, for phonological awareness tasks teachers reported, "It is nice to see and interact with a research-based rhyming task" and "It's [First Sounds] an easy to use format". At the same time, teachers also voiced challenges in their reviews: "some of my students would not use the same labels for the pictures you have listed here" and "it [Que Palabra Queda] would be challenging to give because of the sounds."

For *Oral Language* teachers reported various impressions about the tasks. A significant proportion of responses illustrated teacher's satisfaction with the measures: "The [Which One Doesn't Belong/ Categories] tasks measure critical thinking skills, which I like", "The tasks have easily recognizable and culturally appropriate photos and this would help because we don't currently have an assessment of Spanish oral language." Other teachers reported, "I like how it [verbs] goes beyond picture naming and elicits phrases rather than single words, and the photos are good." Similarly, teachers acknowledge the tasks were designed appropriately. Various responses recorded included statements like, "they are easy to use and the photos are good." Finally, teachers extended their satisfaction with the tasks to their teaching strategies with statements such as, "This assessment [Functions] and the others gives me ideas about what skills are important to teach beyond vocabulary. Verbs are often left out of vocabulary assessments. Good pictures." However, just as with the other domains, teachers also voiced their concerns and challenges in using the tasks. For example one responses noted, "Categories can be interpreted differently. Some of the images may not be the best representations for that word." Similarly, some teachers noted that variability in responses could be impacted by the test design. For example, one responses stated, "the answers could vary greatly, a variety of answers should be considered correct" another wrote, "The correct answers are in Mexican Spanish, a list of acceptable answers is requested," and finally, "Some of the pictures might not be sensitive to cultural differences and some photos are not accurate enough."

Results generally indicated that most measures were appropriate representations of the teacher's knowledge of Spanish early literacy and teachers found them accessible and easy to use. For an oral language set of tasks, Which One Doesn't Belong and Categories, teachers reported the items would be difficult and would not provide much meaningful information in

their interaction with students. Within this domain teachers were most interested in the verbs and picture naming tasks. In the phonological awareness domain teachers were least interested in the Rhyming task and most interested in the First Sounds task. Finally, in the alphabet knowledge domain teachers were least interested in the Letter Detection task and most interested in the Sound Identification and Letter Naming (receptive) tasks.

Table 3. Qualitative Teacher Survey Results: Alphabet Knowledge				
Question	Alphabet Knowledge tasks			
	Letter Naming Receptive	Letter Detection	Letter Naming Expressive	Sound Identification
To what extent would this assessment measure alphabet knowledge?	0% not at all 14% slightly 29% somewhat 57% very much	0% not at all 23% slightly 46% somewhat 31% very much	0% not at all 7% slightly 64% somewhat 29% very much	0% not at all 0% slightly 36% somewhat 64% very much
What portion of Spanish speaking children in your class would be able to answer at least one item on this task?	0% none to a few 29% some 57% most 14% all	7% none to a few 29% some 57% most 7% all	7% none to a few 21% some 64% most 7% all	21% none to a few 29% some 43% most 7% all
How accurately would this assessment measure your students' Spanish early literacy skills?	0% not at all accurately 14% slightly accurately 57% somewhat accurately 29% accurately 0% very accurately	0% not at all accurately 21% slightly accurately 64% somewhat accurately 0% accurately 14% very accurately	0% not at all accurately 7% slightly accurately 64% somewhat accurately 0% accurately 29% very accurately	7% not at all accurately 7% slightly accurately 36% somewhat accurately 0% accurately 50% very accurately
How easy, if at all, is this assessment to use?	0% not at all easy 7% slightly easy 29% somewhat easy 0% easy 64% very easy	0% not at all easy 14% slightly easy 0% somewhat easy 29% easy 57% very easy	0% not at all easy 0% slightly easy 50% somewhat easy 0% easy 50% very easy	0% not at all easy 0% slightly easy 46% somewhat easy 0% easy 54% very easy
How easy, if at all is it to evaluate correct and incorrect responses on this assessment?	0% not at all easy 21% slightly easy 0% somewhat easy 29% easy 50% very easy	0% not at all easy 21% slightly easy, 0% somewhat easy 36% easy 43% very easy	0% not at all easy 21% slightly easy 0% somewhat easy 43% easy 36% very easy	0% not at all easy 7% slightly easy 0% somewhat easy 43% easy 57% very easy
To what extent, if at all, would this assessment measure the skills you teach in your classroom?	0% not at all 7% slightly 43% somewhat 50% very much	7% not at all 14% slightly 71% somewhat 7% very much	8% not at all 0% slightly 46% somewhat 46% very much	0% not at all 7% slightly 36% somewhat 43% very much
How usable, if at all, would this task's score be in planning for instruction?	0% not very usable 7% slightly usable 43% somewhat usable 50% very usable	0% not very usable 43% slightly usable 43% somewhat usable 14% very usable	8% not very usable 8% slightly usable 69% somewhat usable 15% very usable	0% not very usable 7% slightly usable 50% somewhat usable 43% very usable

Table 4. Qualitative Teacher Survey Results: Phonological Awareness				
Question	Phonological Awareness tasks			
	Rhyming	First Sounds/ Primero Sonidos	What word is it?/Que palabra queda	Blending/ Mezclar
To what extent would this assessment measure phonological awareness?	0% not at all 0% slightly 67% somewhat 33% very much	0% not at all 0% slightly 0% somewhat 100% very much	0% not at all 0% slightly 0% somewhat 100% very much	0% not at all 0% slightly 33% somewhat 67% very much
What portion of Spanish speaking children in your class would be able to answer at least one item on this task?	0% none to a few 33% some 33% most 33% all	0% none to a few 0% some 33% most 67% all	7% none to a few 0% some 33% most 67% all	21% none to a few 29% some 33% most 67% all
How accurately would this assessment measure your students' Spanish early literacy skills?	0% not at all accurately 0% slightly accurately 33% somewhat accurately 0% accurately 67% very accurately	0% not at all accurately 0% slightly accurately 0% somewhat accurately 0% accurately 100% very accurately	0% not at all accurately 0% slightly accurately 33% somewhat accurately 0% accurately 67% very accurately	0% not at all accurately 0% slightly accurately 0% somewhat accurately 33% accurately 67% very accurately
How easy, if at all, is this assessment to use?	0% not at all easy 0% slightly easy 67% somewhat easy 0% easy 33% very easy	0% not at all easy 0% slightly easy 33% somewhat easy 0% easy 67% very easy	0% not at all easy 0% slightly easy 33% somewhat easy 0% easy 67% very easy	0% not at all easy 0% slightly easy 33% somewhat easy 0% easy 67% very easy
How easy, if at all is it to evaluate correct and incorrect responses on this assessment?	0% not at all easy 0% slightly easy 0% somewhat easy 0% easy 100% very easy	0% not at all easy 0% slightly easy, 0% somewhat easy 0% easy 100% very easy	0% not at all easy 0% slightly easy 0% somewhat easy 33% easy 67% very easy	0% not at all easy 0% slightly easy 0% somewhat easy 0% easy 100% very easy
To what extent, if at all, would this assessment measure the skills you teach in your classroom?	0% not at all 0% slightly 33% somewhat 67% very much	0% not at all 0% slightly 0% somewhat 100% very much	0% not at all 0% slightly 33% somewhat 46% very much	0% not at all 0% slightly 33% somewhat 67% very much
How usable, if at all, would this task's score be in planning for instruction?	0% not very usable 0% slightly usable 67% somewhat usable 33% very usable	0% not very usable 0% slightly usable 0% somewhat usable 100% very usable	0% not very usable 0% slightly usable 33% somewhat usable 67% very usable	0% not very usable 0% slightly usable 33% somewhat usable 67% very usable

Table 5. Qualitative Teacher Survey Results: Oral Language				
Question	Oral Language tasks			
	Which One Doesn't Belong & Categories ($\bar{X}$)	Functions	Verbs (expressive and receptive) ($\bar{X}$)	Picture Naming
To what extent would this assessment measure oral language?	0% not at all 0% slightly 60% somewhat 40% very much	0% not at all 0% slightly 57% somewhat 43% very much	0% not at all 7% slightly 36% somewhat 57% very much	0% not at all 0% slightly 43% somewhat 57% very much
What portion of Spanish speaking children in your class would be able to answer at least one item on this task?	8% none to a few 31% some 45% most 16% all	0% none to a few 29% some 43% most 29% all	7% none to a few 7% some 57% most 36% all	0% none to a few 0% some 71% most 29% all
How accurately would this assessment measure your students' Spanish early literacy skills?	0% not at all accurately 14% slightly accurately 63% somewhat accurately 0% accurately 23% very accurately	0% not at all accurately 0% slightly accurately 86% somewhat accurately 0% accurately 14% very accurately	0% not at all accurately 7% slightly accurately 50% somewhat accurately 0% accurately 43% very accurately	0% not at all accurately 0% slightly accurately 43% somewhat accurately 0% accurately 57% very accurately
How easy, if at all, is this assessment to use?	0% not at all easy 0% slightly easy 39% somewhat easy 0% easy 61% very easy	0% not at all easy 0% slightly easy 57% somewhat easy 0% easy 43% very easy	0% not at all easy 7% slightly easy 28% somewhat easy 0% easy 65% very easy	0% not at all easy 0% slightly easy 29% somewhat easy 0% easy 71% very easy
How easy, if at all is it to evaluate correct and incorrect responses on this assessment?	0% not at all easy 0% slightly easy 14% somewhat easy 34% easy 52% very easy	0% not at all easy 14% slightly easy 0% somewhat easy 43% easy 43% very easy	0% not at all easy 7% slightly easy 0% somewhat easy 50% easy 43% very easy	0% not at all easy 0% slightly easy 0% somewhat easy 50% easy 50% very easy
To what extent, if at all, would this assessment measure the skills you teach in your classroom?	0% not at all 14% slightly 71% somewhat 15% very much	0% not at all 14% slightly 57% somewhat 29% very much	0% not at all 7% slightly 64% somewhat 29% very much	0% not at all 0% slightly 43% somewhat 57% very much
How usable, if at all, would this task's score be in planning for instruction?	0% not very usable 22% slightly usable 54% somewhat usable 24% very usable	0% not very usable 14% slightly usable 43% somewhat usable 43% very usable	0% not very usable 0% slightly usable 43% somewhat usable 57% very usable	0% not very usable 0% slightly usable 43% somewhat usable 57% very usable

Rubric evaluations. Results for the qualitative rubric are provided in Table 6. Results were discussed in a cross-site team meeting where raters described the reasoning for each score. Results indicated a total of seven tasks were removed from the candidacy pool for field testing based on low qualitative scores. Results for tasks that were removed from the pool of measures for field-testing are described in the summary decisions section.

Table 6. Qualitative Rubric Results

Task	Qualitative Criteria							
	Active Engagement (1.a.)	Valid Response Patterns (1.b.)	Easy to Use (1.c.)	Timely to Deliver (1.d.)	Grand Total Score (sum of 1.a. through 1.d.)	Rank (based on **Grand Total Score**)	MN Rank (based qualitative and quantitative)	Utah Rank
PA:								
Rhyming	3	2	2	3	10	2	1	2
First Sounds	3	2	2	3	10	2	2	1
~~Detection~~	~~3~~	~~3~~	~~2~~	~~3~~	~~11~~	~~1~~	~~3~~	~~3~~
Blending	3	2	3	3	11	1	2	2
Elision	3	2	2	2	9	2	1	1
OL:								
Picture Naming	3	3	3	2	11	1	1	1
~~WODB~~	~~2~~	~~1~~	~~2~~	~~2~~	~~7~~	~~6~~	~~6~~	~~5~~
~~Categories~~	~~2~~	~~1~~	~~2~~	~~2~~	~~7~~	~~6~~	~~7~~	~~7~~
Functions	3	3	2	2	10	4	4	4
~~Receptives~~	~~3~~	~~2~~	~~3~~	~~3~~	~~11~~	~~1~~	~~2~~	~~2~~
Receptive Verbs	3	3	2	3	11	1	3	3
Expressive Verbs	3	2	2	2	9	5	5	6
AK:								
~~Letter Detection~~	~~3~~	~~2~~	~~2~~	~~3~~	~~10~~	~~2~~	~~3~~	~~1~~
~~Exp Letter Naming~~	~~2~~	~~1~~	~~3~~	~~3~~	~~9~~	~~4~~	~~4~~	~~3~~
Rec Letter Naming	3	2	3	3	11	1	1	2
Sound ID	3	2	2	3	10	2	2	4
Contextualized OL:								
Storybook	2	3	2	2	9	1	1	1
~~Vamos a Hablar~~	~~1~~	~~1~~	~~2~~	~~2~~	~~6~~	~~2~~	~~2~~	~~2~~

Quantitative Results

Empirical statistics for each task are presented in Table 7. Results indicated that few tasks obtained zero scores, which may generally suggest that the tasks are

developmentally appropriate for 4 to 5 year old students because floor affects were not achieved for most measures. Similarly, the majority of tasks produced valid response patterns with the sample group (n/N).

Table 7. Quantitative Results

Task			Empirical Criteria			
	n/N	Mean	Median	SD	Min	Max
Rhyming	9/10	6.89	6	3.06	3	10
First Sounds	8/10	6.25	6	1.91	4	9
~~Detection~~	~~9/10~~	~~9.89~~	~~10~~	~~0.33~~	~~9~~	~~10~~
Picture Naming	10/10	6.10	6.5	2.38	2	10
~~WODB~~	~~8/10~~	~~4.12~~	~~4~~	~~1.25~~	~~2~~	~~6~~
~~Categories~~	~~2/10~~	~~4.13~~	~~4~~	~~1.25~~	~~0~~	~~7~~
Functions	7/10	4.4	5	2.54	1	8
~~Receptives~~	~~10/10~~	~~9.5~~	~~10~~	~~0.71~~	~~8~~	~~10~~
Receptive Verbs	10/10	9.5	10	0.85	8	10
Expressive Verbs	6/10	5	5.5	1.26	3	6
~~Letter Detection~~	~~8/10~~	~~8.5~~	~~9~~	~~1.6~~	~~5~~	~~10~~
~~Exp Letter Naming~~	~~8/10~~	~~1.88~~	~~1.5~~	~~1.89~~	~~0~~	~~4~~
Rec Letter Naming	9/10	6.22	6	2.05	3	9
Sound ID	8/10	6.12	6	2.64	2	10
Blending	10/10	10	10	0	10	10
Elision	7/10	5.86	7.5	4.14	0	8
Storybook	10/10	17.3	20.5	8.19	5	25
~~Vamos a hablar~~	~~12/12~~	~~3.44~~	~~0~~	~~1.316~~	~~1~~	~~59~~

*n characterizes valid responses patterns, while N is the total number of students who interacted with the task.

Summary Decisions

Taken together, quantitative results indicated a pool of nine measures were appropriate candidates for pilot testing. Results for measures that were discarded are presented here.

Detection. While Detection scored well across the qualitative rubric, quantitative results indicated the measure was too easy, with nearly all students scoring a perfect score. Because of the limited variance in student performance and frequent ceiling effects, detection was eliminated from the pool of candidate measures.

Which One Doesn't Belong (WODB). WODB yielded low scores on the qualitative rubric such that the measure was difficult for children to engage in, potentially because the task represented both the domain of oral language, but also taps the complementary construct of comprehension in that the child must know and understand categories. Further, WODB, demonstrated the second lowest average score (4.12). As a result, WODB was not included in the pool of measures for further field-testing.

Categories. The Categories task obtained a low score on the qualitative rubric (7) and was ranked as the least promising of the Oral Language measures. This task was difficult for children to engage in and only two of the ten children tested were able to engage in the testing session. In addition, Categories was presented on the teacher survey and results indicated the teachers did not find it useful. As a result, Categories was not included in the pool of measures for further field-testing.

Receptives. The receptives task offered students an opportunity to use basic vocabulary skills to identify common objects. This task demonstrated ease of use and high qualitative scores. However, because a significant proportion of students reached the ceiling (answered every item correctly), this task was removed from the pool. In addition, *Receptives* captures a skill most often demonstrated in curriculum-based interactions and as a result yielded redundant information.

Letter Detection. Letter Detection was easy for students to engage in, however similar to other discontinued measures, many students obtained a perfect score and the measure would potentially be too easy for students. As a result we removed Letter Detection from the candidate pool for field-testing.

Expressive Letter Naming. The Expressive Letter Naming task was created to align with teacher and professional efforts to test alphabet knowledge. Because the task is very popular in preschool classrooms it was determined the task would yield redundant information. In addition, students' scores approximated a potential floor effect, with few children scoring above a 1.0. As a result, Expressive Letter Naming was removed from the candidate pool for field-testing.

Vamos a Hablar! *Vamos a hablar* was initially developed to elicit conversation with preschool age students about common interactions or suggested scenes. However, due to

poor accuracy, a significant time commitment, and limited utility in scoring, this measure was eliminated from the candidate pool for field-testing.

Conclusions

Taken together, the results from this pilot study yielded 11 S-IGDI measures for field-testing. Each measure was carefully vetted across qualitative and quantitative criteria to best ensure that the measure demonstrates appropriate construct validity and supports practical constraints of classroom use. Further, by carefully examining Spanish early literacy development, we were better able to remove measures that indicated limited utility within the context of Spanish development (see Tech. Report 1 for more information on how the nature of Spanish early literacy development).

These 11 measures will be included on a 2013 field test with over 120 students across two states. Each measure will be administered via a sampling scheme that allows for at least 100 responses per item, so the last building block of Wilson's model, the measurement model, can be applied. This study is further described in Technical Reports 3 and 4.

Developing measures for SEB students that accurately represent student skills from the perspective of how Spanish develops is a novel approach to assessment. For this reason, this process has been iterative in nature, with this report representing the second step in a sequence of revisions that will produce measures that demonstrate sound validity and reliability for use with SEB preschoolers.

References

Anthony, J. L., Williams, J. M., Duran, L. K., Gillam, S. L., Liang, L., Aghara, R., Swank, P. R., Assel, M. A., & Landry S. H. (2011). Spanish phonological awareness: Dimensionality and sequence of development during the preschool and kindergarten years. *Journal of Educational Psychology, 103*, 857-876.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M., & Francis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*(1), 170–181.

Cárdenas-Hagan, E., Carlson, C.D., & Pollard-Durodola, S.D. (2007). The cross-linguistic transfer of early literacy skills: The role of initial L1 and L2 skills and language of instruction. *Language, Speech, and Hearing in the Schools*, *38*(3), 249-259.

Cisero, C. A. & Royer, J. M. (1995). The development and cross-language transfer of phonological awareness. *Contemporary Educational Psychology, 20*, 275-292.

Davison, M. D., & Brea-Spahn, M. (2012). Literacy. In B. Goldstein (Ed.). *Bilingual language development & disorders in Spanish-English speakers*. Baltimore, MD: Brookes.

Dunst, C.J., Trivette, C.M., Masiello, T., Roper, N., & Robyak, A. (2006). Framework for developing evidence-based early literacy learning practices. *CELL Papers, 1*, 1-12.

Durgunoğlu, A.Y., Nagy, W.E., & Hancin-Bhatt, B.J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85,* 453-465.

Gorman, B. K. & Gillam, R. B. (2003). Phonological awareness in Spanish: A tutorial for speech-language pathologists. *Communication Disorders Quarterly, 25*(1), 13-22.

Hammer, C.S., & Rodriguez, B.L. (2012). Bilingual language acquisition and the child

    socialization process. In B. A. Goldstein (ed.), *Bilingual language development and*

    disorders in Spanish-English speakers. Baltimore, MD: Brookes Publishing Company

    (pp. 21-50).

Kennedy, C. A., Brown, N. J., Draney, K., & Wilson, M. (2005). Using progress variables and

    embedded assessment to improve teaching and learning. *American Education*

    Research Association, San Francisco, California.

Kuo, L. & Anderson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the

    impact of early bilingualism on phonological awareness. *Scientific Studies of Reading,*

    14, 365-385.

McBride-Chang, C. (1999). The ABCs of the ABCs: The development of letter-name and

    letter-sound knowledge. *Merrill-Palmer Quarterly, 45*, 285-308.

Morgan, P.L., & Meier, C.R. (2008). Dialogic reading's potential to improve children's

    emergent literacy skills and behavior. *Preventing School Failure, 52*, 11-16.

Peña, E.D. & Halle, T.G. (2011). Assessing preschool dual language learners: Traveling a

    multiforked road. *Child Development Perspectives, 5*(1), 28-32.

Wackerle-Hollman, A., Durán, L., Rodriguez, M. C., Brunner, S., Palma, J. & McConnell, S.

    (under review). Advances in preschool assessment for Spanish-English bilinguals:

    Pioneering the use of a measurement model in the design and evaluation of General

    Outcome Measures.

Wackerle-Hollman, A., Brunner, S., Durán, L., McConnell, S., Palma, J., Kohlmeier, T., Callard,

    C., & Rodriguez, M. (2012). *The development of early literacy skills in bilingual and*

    Spanish-speaking preschool-age children: A literature review (Technical Report #1).

Retrieved from http://www.cehd.umn.edu/ceed/projects/spanishigdis/

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah,

NJ: Lawrence Erlbaum.

Appendix A

Classroom Language Survey (Year 1)

Thank you for participating in this survey.
This survey should be completed by the lead teacher. If there is more than one lead teacher, all lead teachers should complete a copy of the survey.

We are collecting information about you and your classroom to help us with the development of the new Spanish Individual Growth and Development Indicators (S-IGDIs). Your participation is important and we appreciate your time in completing this survey.

1. How many years have you been teaching?
   _____

2. What is your highest level of education?
   ☐ High school graduate, GED or equivalent
   ☐ Child Development Associate's (CDA) Degree
      ☐ With CDA bilingual specialization
   ☐ Associate's degree (Please indicate major) _____
   ☐ Bachelor's degree (Please indicate major) _____
   ☐ Master's degree (Please indicate major) _____
   ☐ Other _____

3. What is your native language(s)?
   _____

4. How well do you **speak** English and Spanish?
Please circle your level.

| **English**: | Not at all | Poorly | Well | Excellently/Fluently |
| **Spanish**: | Not at all | Poorly | Well | Excellently/Fluently |

5. How well do you **read** English and Spanish?
Please circle your level.

| **English**: | Not at all | Poorly | Well | Excellently/Fluently |
| **Spanish**: | Not at all | Poorly | Well | Excellently/Fluently |

6. How well do you **write** English and Spanish?
Please circle your level.

| **English**: | Not at all | Poorly | Well | Excellently/Fluently |
| **Spanish**: | Not at all | Poorly | Well | Excellently/Fluently |

7. Please circle your level of knowledge of bilingual development.

Highly	Very	Somewhat	Not
Knowledgeable	Knowledgeable	Knowledgeable	Knowledgeable

8. In what type of program do you work (check all that apply)?
- ☐ Head Start
- ☐ Private Preschool
- ☐ ECFE
- ☐ School Readiness
- ☐ Migrant Head Start
- ☐ State-funded Preschool
- ☐ Other _____

9. How many students do you teach?

      Session 1: _____   Session 2: _____   Full day: _____

10. Of those students, how many speak Spanish?

      Session 1: _____   Session 2: _____   Full day: _____

11. What language or languages are **spoken** in your classroom?
- ☐ Only Spanish
- ☐ More Spanish than English
- ☐ Both, equally
- ☐ More English than Spanish
- ☐ Only English

12. In what language or languages do you **provide instruction**?
- ☐ Only Spanish
- ☐ More Spanish than English
- ☐ Both, equally
- ☐ More English than Spanish
- ☐ Only English

13. Who speaks Spanish in your classroom? (Select all that apply)
- ☐ No one
- ☐ Lead Teacher
- ☐ Teacher Assistant/Paraprofessional
- ☐ Other support staff
- ☐ Specialists
- ☐ Parent volunteer
- ☐ Children
- ☐ Other _____

14. What is the goal of your preschool program?
- ☐ To develop both English and Spanish skills
- ☐ To primarily develop English skills
- ☐ To primarily develop Spanish skills
- ☐ Other _____

Qualitative Criteria for selecting measures that should move forward to the Phase 2 testing:

1. Qualitative Criteria & GOM standards
   a. **The measure should demonstrate that child interaction with the content will produce meaningful data through active engagement of the child.**
      i. Did the children attend to the task?
      ii. Did the children seem to like the task?
      iii. Did the child respond to you when you asked a question?
   b. **The measure should demonstrate that child interaction with the content will produce meaningful data through valid response patterns.**
      i. Did the child appear to respond thoughtfully?(e.g. responding NOT thoughtfully would be always choosing an item on the right, left, middle, etc. Clearly guessing, being silly about responses etc.)
      ii. Did the child appear to understand what you were asking (directives)?
   c. **The measure should be easy to use by end users with minimal academic training (para-professional level).**
      i. Did the procedures allow for successful administration of the task?
      ii. Was scoring easy to complete?
   d. **The measures should be timely to deliver and score.**
      i. Did it take more than a minute or two to score the measures?
      ii. Could an assessor arrive at an appropriate score without additional analysis?
         1. By considering exemplars or other response patterns?

Qualitative Scoring Rubric (to be used per task)				
TASK TITLE:				
Criteria	Did not achieve the desired outcome (unsatisfactory and unresolvable)	Did not achieve the desired outcome, but approached it. (unsatisfactory, but resolvable)	Achieved desired outcome generally, but with some reservations (satisfactory, considered with some reservations, likely resolvable)	Achieved the desired outcome to the highest standard (satisfactory and considered without reservations)
Points	0	1	2	3
1.a. (OVERALL PERCEPTION)				
1.a.i.				
1.a.ii				
1.a.iii				
1.b.(OVERALL PERCEPTION)				
1.b.i				
1.b.ii				
1.c.(OVERALL PERCEPTION)				
1.c.i				
1.c.ii				
1.d.(OVERALL PERCEPTION)				
1.d.i				
1.d.ii				
SUM SCORES				
GRAND TOTAL SCORE for TASK:				