All New Spanish-IGDIs: Bilingual Measurement Considerations & S-IGDI Development

A product of the University of Minnesota and Utah State University

Lillian Durán
Alisha Wackerle-Hollman
Scott McConnell
Our Agenda

- What are the Spanish-IGDIs?
- Bilingual Measurement Considerations
- A review of the literature to support measurement targets in Spanish
- New Measure Examples (Hot off the presses!)
- The research and measurement process currently in progress for developing the new S-IGDIs
- Next Steps
Our Team

- University of Minnesota - Measurement content and IGDIs
  - Alisha Wackerle-Hollman
  - Scott McConnell
  - Michael Rodriguez
  - GRAS: Jose Palma & Stephanie Brunner

- Utah State University - SEB expertise
  - Lillian Durán
  - Tim Slocum
  - GRAs: Terry Kohlmeier, Chase Callard
  - Translation support: Lara Linares
What are the Spanish-IGDIs?

- The University of Minnesota and Utah State University were funded in July by the Institute of Educational Sciences under a goal 5 grant to develop a Spanish version of the IGDIs for screening language and literacy abilities of 3-5 year-olds.

- In year 4 of the grant we will compare performance of Spanish-English bilingual children on the English IGDIs 2.0 and the new Spanish-IGDIs to be able to provide preschool teachers with a way to consider performance in both languages when screening young bilinguals.
A Process of Discovery

- It is important to recognize that this will be a process of discovery for our team as we try new ideas, new measurement targets, and even new ways of measuring Oral Language, Phonological Awareness, and Alphabet Knowledge in Spanish.
Bilingual Measurement Considerations

- Spanish has unique linguistic features that need to be considered. Simply translating the test to Spanish is not psychometrically sound.

- Young Spanish speakers in the US have a wide range of variability in the amount of English and Spanish that they have been exposed to. It is important to consider language background when developing cut-off scores in English and Spanish.
Language and culture are inextricably connected. Young Spanish speakers often have experienced different language socialization patterns that might affect performance on assessment tasks given the type of adult-child interaction that is required. (Hammer & Rodriguez, 2012)

In addition children’s vocabulary knowledge will be influenced by their cultural context i.e. chile may be a more familiar item than apple.
Young simultaneous/relatively balanced bilinguals will have both vocabulary and syntactical skills distributed across both of their languages. It is critical to measure a child in both of his/her languages for an accurate diagnosis of delay.
Peña’s Four Tenets

To address these types of issues in bilingual measurement our team has adopted Peña’s four tenets (Peña, 2007) as our guiding principles:

1. Functional equivalence
2. Cultural equivalence
3. Metric equivalence
4. Linguistic equivalence
Functional equivalence

- Functional equivalence addresses the question, “Do items measure the same construct in each language?”

- Adequately addressing this question involves considering the unique functional and pragmatic features of the target language and the manner in which a skill might be elicited within the structure of that language.

- Assessments that are functionally equivalent in two languages may have different types of items or instructions to access the same construct.

- One example is the Bilingual English Spanish Assessment (BESA; Peña, Gutiérrez-Clellan, Iglesias, Goldstein, & Bedore, n.d.). For example on the BESA Spanish semantics section there are items requiring the child to provide a verb versus on the English version where the focus is primarily on naming nouns given the difference in frequency in early language production of nouns and verbs across English and Spanish.
Cultural equivalence directly addresses the differences that may be associated with test items and procedures based on cultural aspects. Item presentation or elicitation techniques may have different levels of importance, meaning, or even motivation based on the cultural background of participants. For instance, to ensure cultural equivalence the research team must consider not only the selection of words represented in vocabulary tests, but also the images that represent those words.
Metric Equivalence

- Metric equivalence drives the technical adequacy of any measure.
- In Peña’s model, metric equivalence is the extent to which two measures demonstrate similar relations to criterion measures and socially meaningful outcomes.
- To address metric equivalence the unique developmental progression of the target language must be considered. The level of difficulty of items may differ with respect to word and grammatical frequency.
Linguistic Equivalence

- Linguistic equivalence is the extent to which measures relate to essential linguistic features of a particular target language.

- English instruments are often translated into other languages using an expert translator and “back translated” to ensure accuracy.

- This process does not address linguistic features that may be unique to the other language. For example, there may be differences in frequency of occurrence, developmental or chronological sequencing, or familiarity with referenced words, phrases, or concepts in English and the other target language.
Framework for Design

Figure 4. Wilson’s Measurement Model

Construct Map \[ \xrightarrow{Causality} \] Item Response

Measurement Model \[ \xrightarrow{Inference} \] Outcome Space

*Figure 4.* The four building blocks of an item response model approach to measurement construction (Wilson, 2005).
New Measures

Spanish-Individual Growth and Development Indicators
Research Process

- Extensive Literature Reviews
- Robust analyses of component skills for each early literacy area: Alphabet Knowledge, Oral Language, and Phonological Awareness
- Research Design, New Measures and Pilot implementation
- Next Steps
Literature Review

- We conducted a thorough literature review to identify targets for measurement in Oral Language, Phonological Awareness, and Alphabet Knowledge in Spanish.

- Our goal is to choose targets in Spanish that have been found to be correlated with both later English and Spanish reading ability.
Oral Language

- We have used the research on the semantic development of young Spanish speakers to identify targets such as identifying functions of items, verb knowledge, and category recognition, in addition to picture naming as measurement targets (Peña, Bedore, and Rappazzo, 2003; Peña, Kester, & Sheng, 2012)

- We have also included both receptive and expressive items
We also used the literature on language socialization of young Spanish speakers in the US to consider items that include more naturalistic communication and we are in the process of developing a play-based assessment similar to the Early Communication IGDI and a new narrative assessment (Hammer & Rodríguez, 2012)
## Construct Definitions

<table>
<thead>
<tr>
<th>Spanish IGDIs Construct Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological Awareness</strong></td>
</tr>
<tr>
<td><strong>Oral Language</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Alphabet Knowledge</strong></td>
</tr>
</tbody>
</table>
Example Measures

Oral Language
Which one doesn’t belong?
Vamos a ver algunas figuras y decidir cuál de ellas no pertenece a las demás.
Primero, me toca a mí: Pelota, pantalones, camisa. *(Señale cada figura a medida que las va enumerando.)*
Ahora voy a decidir cuál figura no pertenece a las demás.
La pelota no pertenece. *(Señale a la pelota.)*
Los pantalones y la camisa son ropa, pero la pelota no pertenece porque es un juguete.
Functions
Vamos a ver algunas figuras y a continuación tú me dices para que sirven cada una de éstas.

Primero, me toca a mí: Cama. (Señale cada imagen a medida que las va enumerando.)

Ahora voy a decidir para qué se usa.

El uso de la cama es para dormir.

dormir
Verbs
Vamos a ver algunas figuras de personas haciendo algo y a continuación tú escoges la figura que mejor represente la acción que te digo.

Primero, me toca a mí. Voy a escoger la persona que come.

Esta niña come. *(Señale a la niña que come el helado.)*

La otra figura representa un niño que recorta un papel con tijeras.

comer  recortar
Picture Naming
Voy a mirar estas figuras y decir lo que representa cada figura.

león
Receptive Vocabulary
Vamos a jugar un juego en el que tienes que encontrar la figura de lo que te digo.

Primero me toca a mí. Voy a escoger el oso.

Este es el oso. *(Señale al oso.)*

correo  oso  abeja
Categories
Vamos a ver algunas imágenes y decidir a cuál categoría pertenecen.
Primero, me toca a mí: Mesa, plato, cuchara. *(Señale a cada imagen cuando lo denomina.)*
Ahora voy a decidir a cuál categoría pertenecen.
Todos los imágenes son de cosas que relatan a la cocina.

cuchara    plato    mesa

Respuestas aceptables:
1. La cocina
2. Comer
Example Measures

Phonological Awareness
Phonological Awareness

- There is significant evidence supporting the notion that there is a cross-linguistic transfer of PA skills between Spanish and English (Cárdenas-Hagan, Carlson, & Pollard-Durodola, 2007).
- Targets such as rhyming, alliteration (first sounds), blending, and elision are all related to later reading in English and Spanish.
- The syllable as the unit of manipulation may be more salient than the phoneme in Spanish given word structure (i.e. onset-rime does not work as well).
Phonological Awareness

- Varying between multiple choice and free response may produce more range in difficulty and better discrimination than traditional word level manipulation in English in blending and elision tasks (i.e. compound word, syllable, phoneme) (Anthony et al., 2011)
Rhyming
Vamos a jugar un juego para encontrar las figuras que rimen.
Primero me toca a mí. *(Señale cada figura a medida que las va enumerando.)*
Gato, manzana, pato.

Ahora voy a decidir cuales son las palabras que riman: ¿Es gato, manzana *(pausa)* o gato, pato?
First sounds
Vamos a jugar un juego para encontrar las palabras que empiezan con este sonido que te digo.
Primero, me toca a mí. (*Señale cada figura a medida que las va enumerando.*)
Ahora voy a encontrar la palabra que empieza con /r/.
Rosa (*señale a la rosa*), rosa empieza con /r/.
Escucha: /r/, rosa.

**bota**

**rosa**

/r/
Detection
Vamos a jugar un juego donde tú tienes que decirme cual de las palabras que yo te digo es la correcta y representa esta figura. Escucha con atención lo que te digo.

Es “fres” o “fresa”? 

fresa
Measures Examples

Alphabet knowledge
Alphabetic Knowledge

- **Alphabetic principle** is knowledge about the names and sounds of letters (McBride-Chang, 1999).
- There are 27 letters in Spanish including the ñ: ll and ch are no longer recognized as letters.
- We suspect that children’s knowledge of Spanish letter names and sounds may be dependent on their exposure to Spanish preschool instructional environments. This may not be commonly taught in home environments.
Letter detection

\[\downarrow \quad \mathbf{M} \quad \mathbf{+}\]
Vamos a jugar un juego para decidir lo que es una letra del alfabeto.
Primero me toca a mí.
La M es una letra. *(Señale la M.)*
Los otros dos objetos son símbolos que no son parte del alfabeto.
Ahora vamos a hacer más ejemplos.

Diga: Enséñame la letra del alfabeto.
Letter naming

B
Vamos a jugar otro juego donde diremos algunas letras del alfabeto.
Primero, me toca a mí.
Esta letra es B.
Ahora vas a escoger la letra que digo. ¿Cuál letra es L?
Sound identification
Vamos a ver estas tarjetas y decidir cuál letra hace el sonido que te digo.
Primero me toca a mí. Voy a mostrarte cuál letra hace el sonido /f/.
Esta letra hace el sonido /f/ (señale).
Research/measurement design

Rasch Modeling
The Rasch Model

- A specific item response theory (IRT) model, which describes the location of cards [items] on the measurement scale in relation to the trait or construct that underlies the measure.
  - Characterizes a construct on a linear scale
  - Locates items on the scale and in turn, locates people on this same scale.
  - Person’s ability is independent of items the person is administered.
  - Difficulty of items is independent of the sample of people who received the items.
The Rasch Model

- Rasch provides the probability of correct response for each item, modeled as a logistic function (Rasch, 1960).
  - Examinee ability level
  - Difficulty level of the item, or the ability at which an examinee has a predicted probability .5 of answering the item correctly.
  - Mathematical model of the relation between the probability of success, and the difference between an individual's ability and an item's difficulty.

- These scales typically range from about -4 to +4 with a standard deviation around 1, and are centered around the mean item location for the measure (zero represents the average item location).
<table>
<thead>
<tr>
<th>PERSON</th>
<th>MAP</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>.#</td>
<td>30160</td>
</tr>
<tr>
<td>5</td>
<td>.</td>
<td>30098 30151</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>30209</td>
</tr>
<tr>
<td>4</td>
<td>.#</td>
<td>30162 T 30124</td>
</tr>
<tr>
<td>3</td>
<td>#</td>
<td>30116 30207</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30155 30218 30224</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30048 30166</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30180 30190 30247</td>
</tr>
<tr>
<td>2</td>
<td>####</td>
<td>30148 30156 30161 30171 30211 30243 30249</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30078 30081 30163 30174 30178 30189 30214 30215</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30147 30168 30228 30229</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30050 30135 30172 30235 30241</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30122 30138 30222</td>
</tr>
<tr>
<td>1</td>
<td>####</td>
<td>30038 30146 30227 30233</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30107 30115 30144 30175 30208 30210</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30065 30093 30109 30181 30205 30240 30242 30251</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30005 30085 30136 30200 30212 30230 30248 30250 30253</td>
</tr>
<tr>
<td>0</td>
<td>####</td>
<td>30037 30104 30123 30130 30134 30154 30167 30206</td>
</tr>
<tr>
<td></td>
<td>####</td>
<td>30220 30231</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30083 30088 30182 30213 30238 30239</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30071 30084 30092 30095 30108 30117 30177 30191</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30196 30198 30201</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30041 30133 30232 30234 30244</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30086 30113 30137 30145 30165</td>
</tr>
<tr>
<td></td>
<td>#</td>
<td>30075 30076 30096 30097 30118 30119 30193 30245 30246</td>
</tr>
<tr>
<td>-1</td>
<td>.</td>
<td>30003 30100 30114 30158 30164 30184</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>30015 30080 30094 30127 30142 30145 30184 30204 30216</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>30068 30099 30103 30120 30179 30188 30204 30216</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>30101 30110 30143 30169 30197 30211 30226 30252</td>
</tr>
<tr>
<td>-2</td>
<td>.</td>
<td>30112 S 30170</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30111 30126 30128 30195 30219</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30102 30121</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30091 30217</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30011 30157</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30150 30199 30203</td>
</tr>
<tr>
<td>-3</td>
<td>.</td>
<td>30132 T 30153</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30149 30152 30223</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>30202</td>
</tr>
<tr>
<td>-4</td>
<td>.</td>
<td>T 30014 30159</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>30090 30105</td>
</tr>
<tr>
<td>-5</td>
<td>.###</td>
<td>&lt;less&gt;</td>
</tr>
</tbody>
</table>

Each "#" is 19. Each "." is 1 to 18
Tier Level Candidacy

IGDI Benchmarks for identification of students in need of Tier 1, Tier 2/3 services within a decision making framework
IGDIIs are one part of the information necessary to determine if a student needs additional services, but are NOT the entire picture.

- Other sources of data (criterion tests, mastery monitoring information, permanent product reviews)
- Interviews
- File history
- Observations
Benchmarks

- Benchmarks have utility within an identification model to help professionals select which students need additional services. They can tell us:
  - If a student performs below or above a reference point.
  - They can relate the reference point to RTI tier levels

They can’t tell us:
- If the student is making progress
- How close or how far away the student is from the reference point
The PLDs will ask teachers and parents to use their knowledge of students to rank performance based on operational definitions of each early literacy domain for the Fall, Winter and Spring seasons, with definitions changing respectively over the course of the year.

- Teachers and Parents placed students in Tier 1, Tier 2 or Tier 3
- Teachers and Parents indicated to what extent they were confident about the placement
Setting the benchmarks

- A combination of Rasch output, ROC analysis, Regression analysis and contrasting groups design methods will be used to produce Rasch values related to the reference point between Tier 1 and Tier 2/3 performance based on IGDI scores AND teacher + parent evaluation of student performance using Tier Level Descriptors.
Example descriptive of Contrasting group design

![Histograms showing frequency distribution of Picture Naming Measure Wave 1 across different tiers.](chart.png)
Example descriptive of Contrasting group design

![Histograms showing frequency of Sound Identification Measure Wave 1 for Tier 1 and Tier 2/3, with separate bars for Oral Language and Tier 1.](image-url)
Benchmarks

Performance of students who have been classified as Tier-1 by their teacher and parent

Performance of students who have been classified as Tier 2/3 by their teacher and parent

Underlying early literacy trait performance for 4-year olds
TABLE 12.2 Concurrent Calibration for Picture N ZOU496WS.TXT Aug 9 2:51 2012
INPUT: 3164 PERSON 182 ITEM REPORTED: 3161 PERSON 172 ITEM 2 CATS WINSTEPS
3.72.2

<table>
<thead>
<tr>
<th>PERSON - MAP - ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;more&gt;</td>
</tr>
</tbody>
</table>

| 6 | 30160 |
| 5 | 30098  30151 |
| 4 | 30209  30162 |
| 3 | 30116  30207  30048  30180  30190  30247  30148  30156  30161  30171  30221  30243  30249  30176  30187  30078  30081  30163  30174  30178  30189  30214  30215  30147  30168  30228  30229  30122  30138  30222  30038  30142  30227  30065  30093  30109  30181  30205  30240  30242  30251  30107  30115  30144  30175  30208  30210  30095  30096  30137  30139  30145  30165  30118  30119  30193  30245  30226  30248  30250 |
| 2 | 30020  30231 |
| 1 | 30083  30088  30120  30123  30213  30238  30239  30071  30084  30092  30095  30108  30117  30177  30191  30196  30198  30201  30041  30133  30232  30234  30244  30086  30113  30137  30145  30165  30075  30076  30096  30097  30118  30119  30193  30245  30246 |
| 0 | 30003  30100  30114  30158  30164  30184  30015  30080  30094  30127  30142  30080  30101  30110  30143  30169  30197  30211  30226  30252 |
| -1 | 30112  30170  30111  30126  30128  30195  30219 |
| -2 | 30102  30121  30091  30217  30011  30157  30150  30199  30203 |
| -3 | 30132  30153  30149  30152  30223  30202 |
| -4 | 30014  30159 |
| -5 | 30090  30105 |

\*TABLE 12.12 Concurrent Calibration for Picture N ZOU496WS.TXT Aug 9 2:51 2012
INPUT: 3164 PERSON 182 ITEM REPORTED: 3161 PERSON 172 ITEM 2 CATS WINSTEPS
3.72.2

EACH "#" IS 19. EACH "." IS 1 TO 18
Setting the Benchmarks

- Student abilities (as a function of IGDI responses) will then be converted to a number correct card-count score (number correct expected given the study ability related to tier placement).
Challenges in Benchmarking

- The brevity of the measures makes them less precise.
  - As a result, we aren’t yet able to differentiate Tier 2 from Tier 3
- When making decisions considering the transition from Tier 2 to Tier 3 we are still defining the relevant educational features:
  - Consider behavioral or presenting issues
  - Consider responsiveness to Tier 2 services.
Next Steps

- **Measurement R&D – “Tools”**
  - Defining “constructs” as a way of defining “samples”
  - Elaborating General Outcome Measurement within contemporary measurement models

- **Research – “Applications”**
  - Unpacking within- and cross-linguistic effects
  - Identifying factors that promote growth

- **Policy**
  - Providing information that informs what is possible, and what’s needed to make it likely
Conclusions

- Where are we in developing General Outcome Measures for language and early literacy for young Spanish speakers?
  - Improved psychometrics given adherence to our guiding principles
  - A focus on innovation in measurement given cross-linguistic and cross-cultural considerations

- How are we getting there?
  - Renewed research and development, based on many researchers’ and practitioners’ experience
  - Adaptation and addition of new methods, esp. IRT

- Coming attractions
  - New measures
  - Decision criteria designed specifically for young Spanish speakers for tiered intervention
Questions and Comments?