



# Istation

Supporting Educators. Empowering Kids.  
Changing Lives.

## Technical Report

### Istation's Indicators of Progress (ISIP) Español

Istation  
2000 Campbell Centre II  
8150 North Central Expressway  
Dallas, Texas 75206  
Toll-free: 1-866-883-READ (7323)  
[www.istation.com](http://www.istation.com)

Copyright © 2012 Istation-All rights reserved

## Chapter 1

### Introduction

*ISIP, istation's Indicators of Progress, Español* has also been introduced in Spanish as *istation seguimiento individualizado de progreso de la lectura en español*, or *ISIP Español*. This assessment is designed to comply with the requirements of Texas Statute TEC 28.006 and support the implementation of Spanish Language Arts and Reading (SLAR) in Special Language Programs under Texas Education Code, Chapter 29, Subchapter B, including bilingual education and dual language models.

*ISIP Español* is a computer-delivered screener and continuous progress monitoring system of Spanish early reading skills for students in kindergarten, first, second and third grade receiving language arts reading instruction in Spanish. It provides growth information in the following critical domains of early reading as it applies to Spanish language development: phonological awareness, grapheme-phoneme correspondence (symbol/sound: letters and syllables combinations), communication skills (listening comprehension and vocabulary), spelling, reading comprehension, and fluency. It is designed to (a) identify children at risk of not meeting SLAR standards for end-of-year expectations in kindergarten, first, second, and third grade, (b) offer automatic continuous progress monitoring of the above critical skills, and (c) provide immediate and automatic linkage of assessment data to students' learning needs, thus promoting differentiated instruction.

The development and implementation of this assessment system has been conducted with guidance and direction from a Spanish advisory board. This board is composed of state and nationally known researchers in the field, whose advice has been instrumental in ensuring that the assessment data are reliable, efficient, easy to access and use, and capable of wide-scale dissemination on both PC and Mac platforms. *istation* is currently delivering *ISIP Español* assessments to 59,249 users in six states, serving more than 100 districts, primarily in the state of Texas.

### Significance

#### **The Need to Improve Testing Practices in Bilingual Classrooms**

Districts implementing special language programs (including bilingual education, dual language models, and ESL) under Texas Education Code, Chapter 29, Subchapter B, are required to use a Language Proficiency Assessment Committee (LPAC) in order to design students' academic plans based on individualized needs. Consequently, school districts offering these programs are currently addressing academic issues in both Spanish and English. Tests results in both languages

are critical in order to inform instruction. However, current practices have proven to require much of teachers' instructional time as well as excessive cost to the districts.

In addition, current national discussions regarding academic services for students whose first language is Spanish have found that academic programs suffer from a dearth of assessments that prove to be non-biased and appropriate (Roseberry-McKibbin, Brice, O'Hanlon, 2005; Escamilla, 2006 Baker, Shanahan, Linan-Thompson, Collins, Scarcella, 2007).

Bilingual education programs, including dual language models, are in great need of improving current testing practices. Improved practices should (a) allow teachers to redirect time to instructional purposes, (b) target funding to other academic needs, and (c) adopt testing tools that are culturally, linguistically, and cognitively appropriate for programs that follow Spanish Language Arts and Reading standards.

Monitoring students' literacy ability and academic growth in each language is necessary in order to improve bilingual students' academic performance. Both *ISIP Early Reading* and *ISIP Español* can provide tools for monitoring literacy development in two languages. These tests were built individually, using different items, and based on separate field-tests data as recommended by TEA's, reading assessment guidelines, 2005.

Obtaining data results that are relevant, reliable, and valid improves assessment practices. To be relevant, data must be available on a timely basis and target important skills that are influenced by instruction. To be reliable, there must be a reasonable degree of confidence in the student scores. To be valid, the skills assessed must provide information that is related to student performance expectations. Hispanic students who rely on their native language to excel academically have been found to be impacted negatively by results of commonly used invalid, biased, or inadequate assessment practices. As a result, there has been an over-identification and/or under-identification of Spanish speaking students in special education programs (Espinosa & López, 2007). Identifying students' progress toward literacy has the potential to impact intervention results as well as special education services.

Roseberry, McKibbin and O'Hanlon (2005) reviewed surveys completed by public school speech-language pathologists on service delivery for non-native English speakers from 1990 to 2001 and found that there was a dearth of assessments that proved to be both non-biased and appropriate. Test items were generally outside of the students' cultural knowledge and therefore unfamiliar to speakers of other languages, resulting in students' inability to demonstrate the skill being tested. Additionally, test norms based on native speakers of English should not be used with individuals whose first language is not English; and those individuals' test results should be interpreted as reflecting, in part, their current level of proficiency rather than their ability, potential, aptitude, etc. (In order to assess Spanish-speaking students, tests must be adequately tested for cultural relevance and proper Spanish terminology that avoids regionalisms and

colloquial terms. At the same time, items must demonstrate internal consistency when tested and scored with the population that they intend to evaluate - in this case, the Hispanic student population is enrolled in bilingual education classrooms in the US public education system.

There are many reasons why a student's score at a single point in time under one set of conditions may be inaccurate: confusion, shyness, illness, mood or temperament, communication or language barriers between student and examiner, scoring errors, and inconsistencies in examiner scoring. However, by gathering assessments across multiple time points, student performance is more likely to reflect actual ability. By using the computer, inaccuracies related to human administration and recording errors are also reduced. In addition, computer-administered assessments increase opportunities to retest efficiently.

### **The Need to Link Early Spanish Reading Assessment to Instructional Planning**

Instructional time is utilized more effectively when assessment is linked to instruction. Early reading assessments of Spanish literacy development need to (a) identify students at risk for reading difficulties, (b) monitor student progress for skill growth on a frequent and ongoing basis in order to identify students that are falling behind, (c) provide information about students that will be helpful in planning instruction to meet their needs, and (d) assess students' movement toward grade level end-of-year expectations.

Teachers need to be able to identify students' risk of reading failure and/or struggle to meet end-of-year grade level expectations. Individualized student data supports and prescribes differentiated instruction, helping teachers to work effectively to meet the specific needs of each student.

Linking teacher instruction to the results of assessment is promoted by using formative assessments. Following progress through formative assessments needs to occur often enough that teachers may discover when instruction has not been effective in order to make modifications in a timely manner. According to current research, the best examples that follow a formative assessment structure are called "Online Formative Assessment" (Nicol, D.J. & Macfarlane-Dick, D., 2006). It is also envisioned that computer-based formative assessments will play an increasingly important role in learning. This is because of the use of banks of question items which fosters the construction and delivery of dynamic, on-demand assessments.

Research suggests that children with different levels of language proficiency who are also developing literacy skills (whether in one language or two) respond successfully to frequent formative assessments. These assessment results pinpoint skills as they are emerging and provide the best information as to which readers require additional support in specific reading skills (Gersten et al., 2007).

The purpose of formative assessment can be defined as assessment “for learning, rather than of learning” (Stiggins & Chappuis, 2006, p. 10). Equal educational opportunities for emergent readers should offer the use of formative assessments as a necessity, regardless of language of instruction. Formative assessments provide detailed pictures of the abilities that are measured, making critical instruction modifications possible. The primary goals of formative assessment are to guide curriculum strategies and guide instructional strategies. Districts, teachers, and curriculum developers use data to differentiate classroom instruction while monitoring academic progress. It is important to engage in an ongoing process rather than a single test when using formative assessment. Involving students in the measurement of their own progress enables opportunities for both teachers and students to work together toward common goals<sup>1</sup>. Assessment tools that support self-monitoring contribute to engaging students in self-driven progress practices. A systematic and collaborative process that involves self-monitoring and feedback benefits both teachers and students, because it promotes engagement in meta-cognitive processing that informs learning and increases student achievement (Stiggins & Chappuis, 2006). This type of assessment is most useful when (a) it is conducted periodically, (b) it provides information immediately, (c) it is easy and systematic in administration, and (d) it helps gather a more complete picture of each student’s ability (Gersten et al., 2007). Computer-delivered evaluations support all four strengths of formative assessment and allow students to self-monitor their progress.

### **The Need for Continuous Progress Monitoring (CPM) Tools**

Teachers who monitor their students’ progress and use this data to inform instructional planning and decision-making have higher student outcomes than those who do not (Conte & Hintze, 2000; Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Mathes, Fuchs, Roberts, 1998). These teachers also have a more realistic conception of the capabilities of their students than teachers who do not regularly use student data to inform their decisions (Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Stecker, 1991; Mathes et al., 1998).

The collection of sufficient, reliable assessment data on a continuous basis is a daunting task for schools and teachers. Screening and inventory tools for Spanish literacy such as the *Tejas LEE* (Brookes) and *Indicadores Dinámicos del Éxito en la Lectura* (Good & Kaminski, 2002) use a benchmark or screen schema, in which testers administer assessments three times a year. However, more frequent continuous progress monitoring is recommended for all low-performing students, but administration is a difficult process for already overburdened schools and teachers.

Districts currently use curriculum-based measurement (CBM) models to index student progress over time, which in turn can facilitate teachers' formative evaluation of their teaching

<sup>1</sup> “Students and teachers become partners in the classroom assessment process, relying on student-involved assessment, record keeping, and communication to help students understand what success looks like, see where they are now, and learn to close the gap between the two” (Stiggins & Chappuis, 2006, p.11).

effectiveness. Research indicates that CBM can accurately, meaningfully, and sensitively describe such progress (Marston, 1989). This is accomplished through the frequent administration of short, equivalent tests sampling all the skills in the curriculum. A student's past, present, and probable future growth is tracked. When students are not making adequate progress, teachers modify their instructional programs. The educational value of CBM would greatly benefit the outcomes of bilingual education programs, and research demonstrates that instructional programs designed with CBM can result in greater student achievement, enhanced teacher decision making, and improved student awareness of learning (e.g., Fuchs, Fuchs, Hamlett, & Stecker, 1991). Thus, CBM represents a logical model for helping bilingual teachers to identify those students for whom the standard curriculum in place is not having the desired effects. Once identified, teachers can intervene before failure occurs.

Although proven to be a great tool for classroom teachers, CBM has not been as widely embraced as would be hoped and has hardly been utilized in bilingual education classrooms. These assessments, even in their handheld versions, require a significant amount of work to be administered individually to each student. The examiners who implement these assessments must also receive extensive training in both the administration and scoring procedures to uphold the reliability of the assessments and avoid scoring errors. These assessments are labor intensive and expensive for school districts to implement. Bilingual education programs, already pressed for time to evaluate students' academic and proficiency needs in two languages, are unable to easily accommodate the requirements of CBM implementation. Therefore, it is difficult for bilingual teachers to be able to use CBM models for continuous progress monitoring and validation of test results.

The introduction of handheld technology has allowed for graphing of student results. Assessments like *Tejas LEE* can be recorded using palm-pilot devices, but information in this format is often not available on a timely basis for whole school or whole district results. Additionally, the time needed for one-on-one administration and the need for additional staff to support classroom teachers during testing periods make it difficult to implement with fidelity and consistency.

Computer applications have been found to be a reliable means by which to deliver CBM models. Computer administration is able to utilize similar equivalent test sampling that can be used over time and programs that can collect and store the data.

Computerized CBM applications are a logical step in increasing the likelihood that continuous progress monitoring occurs more frequently with monthly or even weekly assessments in both the general education and bilingual education classrooms. Computerized CBM applications have been developed and used successfully in upper grades in mathematics and spelling (Fuchs, Hamlett, Fuchs, 1995). Computerized applications save time and money. They eliminate burdensome test administrations and scoring errors by calculating, compiling, and reporting scores automatically. They provide immediate access to student results that can be used to affect

instruction. Information is organized in formats that automatically group children according to risk or tier level and recommend instructional interventions. Student results are instantly plotted on progress charts with trend lines projecting year-end outcomes based upon growth patterns, eliminating the need for teachers to manually create documentation of results.

### ***ISIP Español, also known as***

#### ***istation Seguimiento Individualizado de Progreso de la Lectura en Español***

*ISIP Español* has been designed to provide continuous progress monitoring of kindergarten through third grade student progress toward SLAR end of year expectations in critical areas of Spanish early reading development. Most importantly, there is no other continuous progress monitoring assessment tool that measures all domains in the same instrument. Many districts currently use different assessment tools in order to create a mix that combines all areas needed for assessment of students learning to read in bilingual education classrooms.

*ISIP Español* is delivered through the Internet in an interactive, animated, game-like format that allows students to accurately demonstrate their true abilities through a high-interest and risk-free format. The complete evaluation of students' reading ability in the domains assessed for each grade level can be administered individually or in a whole-group computer lab setting in approximately 30 minutes. The intention of this assessment is to enable the teacher to quickly evaluate each student's progress in Spanish literacy, thus providing more time for instructional purposes. Teachers can obtain valuable information through real-time data reports, allowing them to focus on quality instructional time. The assessment is designed as a continuous progress monitoring tool that can be delivered to students monthly (or as often as needed) throughout the year.

#### **Description of *ISIP Español***

The growing enrollment of Spanish-speaking students in the Texas public education system calls for the development of linguistically and culturally appropriate Spanish assessments that align to language arts standards. *ISIP Español* allows for more equitable educational opportunities for Spanish speaking students, the largest growing number of English learners in the state of Texas. Across the nation, this student population has the option to enroll in special language programs geared toward improved academic achievement for at-risk students. (Howard, Lindhojm-Leary, 2007; The National Task Force on Early Childhood Education for Hispanics/La Comisión Nacional para la Educación de la Niñez Hispana, 2007; García, Miller, 2008).

The specific domains and the order in which the domains and skills are presented in *ISIP Español* are based on an analysis of the findings and recommendations of the United States National Reading Panel, European and Latin American research, including the latest publications from *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*.

[Consejo de Europa, 2005]. In addition, the following research findings were considered when developing the assessment blueprint for *ISIP Español*:

*Es evidente que las prácticas educativas orientadas a exponer al niño a experiencias de comunicación, de intercambio comunicativo, de partir de sus experiencias previas, de tener sentido aquello que se trata de descodificar, etc., es algo que está plenamente justificado y que no importa para ello el contexto idiomático. Sin embargo, los hallazgos más recientes, desde una perspectiva psicolingüística, ponen de manifiesto que todo ello no sería suficiente ya que el proceso cognitivo de asociación grafía-fonema es un elemento imprescindible cuando se aprende a leer en un sistema alfabético (Jiménez & O'Shanahan, 2008).*

English Translation: It is evident that education practices that are oriented towards exposing a child to communicative experiences, communicative exchange, taking into consideration their prior knowledge and experiences, would have meaning and encourage decoding, etc, is something that is fully justified, and no idiomatic context would be important. Furthermore, the most recent findings, from a psycholinguistic perspective, conclude that it would not be sufficient given that the cognitive process of letter sound association is an essential element when one learns to read an alphabetic system.

Findings from recent research conducted at Renée Descartes University, France and Universidad de La Laguna, Spain compared English-, German-, French-, and Spanish-speaking children learning to read. Conclusions from these studies demonstrated that the performance of reading words for students learning to read in different linguistic contexts (such as English, French, and Portuguese) is systematically higher in Spanish than in other languages. In addition, it was also found that knowledge of the complex rules of grapheme/sound correspondence occurs earlier in Spanish than in English, French, or Portuguese. Similarly, Spanish-speaking students reach higher levels of word reading earlier, when compared to students who speak other languages. Such a finding indicates that the appropriate use of the phonological process occurs earlier in Spanish than in English, French, or Portuguese. (Jiménez , O'Shanahan;, 2008;

*ISIP Español* domains are paralleled to International Early Grade Reading Assessment (EGRA) shown in Table 1. The EGRA evaluates critical areas for Spanish early literacy and has been given to students in Spanish speaking countries in Latin American like Nicaragua and Guatemala.

Table 1. *Spanish literacy critical areas assessed using EGRA*

CONCIENCIA FONOLÓGICA (CF) <sup>2</sup>
---

<sup>2</sup> Phonological awareness allows students to segment oral language in phonemes without using the letter names

<i>La instrucción en CF consiste en enseñar a los niños a segmentar el lenguaje oral en fonemas sin el apoyo de las letras del alfabeto.</i>
<b>CONVERSIÓN GRAFEMA-FONEMA (CGF)</b> <sup>3</sup> <i>La instrucción de reglas de CGF es una forma de enseñar a leer que enfatiza la adquisición de las correspondencias símbolo-sonido.</i>
<b>LECTURA CON FLUIDEZ</b> <sup>4</sup> <i>La fluidez en la lectura es necesaria para la comprensión. Leer con velocidad y precisión, respetando los signos de puntuación, facilita la comprensión del texto.</i>
<b>VOCABULARIO</b> <sup>5</sup> <i>Hay dos tipos de vocabulario: el oral y el escrito. Cuando un lector encuentra una palabra en el texto puede descodificarla, es decir, convertirla en habla.</i>
<b>COMPRENSIÓN</b> <sup>6</sup> <i>Los datos sugieren que la comprensión mejora cuando los alumnos son capaces de relacionar las ideas que están representadas en el texto con su propio conocimiento y experiencias, así como representaciones mentales construidas en la memoria.</i>

Taking into consideration the studies conducted on the transparency of languages, there are common elements for assessment such as grapheme-phoneme correspondence, word reading, level of vocabulary, reading comprehension of both narrative and expository texts, listening comprehension, and fluency, but there are also elements that are critical to the support of reading development in each particular language (Jiménez, O’Shanahan, 2008). In the case of the Spanish language, writing development and orthography are closely tied to reading. This is the reason why reading in Spanish is referred to as *lectoescritura* (Ferreiro, 2002; Bazán, Acuña, Vega, 2008). This term posits an intrinsic relationship between writing and reading.

There are differences, as well as similarities, in emergent reading and writing behaviors of Spanish speaking children (). English writing rubrics cannot help to guide instruction in Spanish. Differences in writing development can impact outcomes in grade level and state-standards-based assessments. Issues that emerged from this research highlight Spanish primary students’ development in which vowels emerge before consonants. Primary students move from strings of letters to invented spelling in Spanish earlier than English speakers do. A writing component is critical for the assessment of emergent literacy skills in Spanish speaking children. *ISIP Español* domains include a writing component that aims to address the specific needs of children developing Spanish literacy skills based on the principles of Spanish *lectoescritura*.

### ***ISIP Español Domains***

<sup>3</sup> Grapheme-phoneme conversion comprises the reading rules to acquire symbol-sound correspondence

<sup>4</sup> Fluency is necessary to develop correct pace, observing punctuation and thus enhancing reading comprehension

<sup>5</sup> Vocabulary objectives can be divided in two categories: oral and written. Decoding enables conversion of text into a verbal outcome

<sup>6</sup> Comprehension is improved when the students are able to relate ideas from the text to their own experiences, knowledge, and memory

The domains selected for the assessment measures of *ISIP Español* were established as explained in the section above. Table 2 shows the domains of *ISIP Español*.

Table 2. *Domains and subtests of ISIP Español*

DOMAIN <sup>7</sup>	SUBTEST
<i>DESTREZA FONOLÓGICA Y FONÉTICA</i> Phonological Awareness and Phonetics	<i>Conversión grafema-sonido</i> <sup>8</sup> <i>Conciencia fonética y silábica</i> <sup>9</sup> <i>Unión de sonidos, sonido inicial y rima</i> <sup>10</sup>
<i>DESTREZA COMUNICATIVA</i> Listening Skills	<i>Comprensión auditiva</i> <sup>11</sup>
<i>VOCABULARIO</i> Vocabulary	<i>Vocabulario de lenguaje oral</i> <sup>12</sup> <i>Vocabulario para lectura y escritura</i> <sup>13</sup>
<i>COMPRESIÓN DE LECTURA</i> Reading Comprehension	<i>Entendimiento de lo leído en textos de ficción y no-ficción</i> <sup>14</sup>
<i>LECTURA CON FLUIDEZ</i> Fluency	<i>Registro de velocidad de la lectura con comprensión</i> <sup>15</sup>
<i>COMUNICACIÓN ESCRITA</i> Writing Conventions	<i>Ortografía y acentuación de palabras</i> <sup>16</sup> <i>Dictado</i> <sup>17</sup>

## Literature Review for *ISIP Español* Domains

### *Destreza fonológica y fonética* (Phonological Awareness and Phonetics)

The subtests in this domain intend to evaluate the early literacy skills associated with mechanics of reading (or pre-reading skills) which relate to the recent findings in neuropsychology studies. (Agrás G, Molina H, Bareche Monclús, R., 2007). These subtests are: grapheme-phoneme correspondence and phonemic/ syllabic awareness, including beginning, ending, and blending sounds.

- Grapheme-phoneme correspondence is the ability to pair sounds (phonemes) with the letters (graphemes) that represent them. The term phonics is also widely used to describe methods used for teaching children to read and decode words (Abadzi, et.al.2005). According to US-based studies, children begin learning to read using phonics, usually around the age of five or six (NRP, 2000). In the case of some alphabetic languages such as Spanish, the orthographic representations

<sup>7</sup> The labels used for this table are not necessarily a direct translation, but rather an equivalent task measured in English reading assessments

<sup>8</sup> Grapheme-phoneme correspondence

<sup>9</sup> Phonemic/ Syllabic Awareness

<sup>10</sup> Blending Sounds, Beginning and Ending Sound

<sup>11</sup> Listening Comprehension

<sup>12</sup> Vocabulary for Oral Language

<sup>13</sup> Vocabulary for Reading and Writing

<sup>14</sup> Reading Comprehension

<sup>15</sup> Silent Reading Fluency

<sup>16</sup> Spelling

<sup>17</sup> Dictation

of sounds are even simpler because there is nearly a one-to-one correspondence between letter patterns and the sounds that represent them. Even though studies conducted with Spanish-speaking children have not been completed in large quantities, as studies with English-speaking students have, the transparency of the language has been widely researched in the field of linguistics (Wimmer, Mayringer, Landerl, 2001; Ziegler, Perry, Ma-Wyatt, Ladner, Shulte-Körne, 2003), placing languages such as Spanish and French on the transparent side of the languages scale and English and German on the opaque side (Seymour, Aro, Erskine, 2003).

- Phonemic and syllabic awareness is “the ability to notice, think about, and work with the individual sounds in spoken words” (Armbruster, Lehr, & Osborn, 2006, p. 2). A broader term for this concept is phonological awareness. The concepts that describe phonological awareness suggest that before children learn to read print, they must understand that words are made up of speech sounds. The United States’ National Reading Panel (NRP, 2000) found that children's ability to read words, to comprehend what they read, and to spell is improved with phonemic awareness. Studies of phonemic awareness conducted with Spanish-speaking children have been used recently to confirm the different levels of phonological awareness that are relevant to the Spanish language (Serrano, Dfior, Jiménez, 2009). These levels comprise identification of phonemes in isolation, beginning and ending sounds and syllables, and intra-syllabic sounds, which impact the development of reading skills in unequal levels of relationship (Serrano et al, 2009).

*Los modos de representación prealfabética se suceden en cierto orden: primero varios modos de representación ajena a toda búsqueda de correspondencia entre la pauta sonora de una emisión y la escritura. Luego modos de representación silábicos (con o sin valor sonoro convencional) y modos de representación silábico-alfabético que preceden la aparición de la escritura. Estos niveles están caracterizados por formas de conceptualización que actúan en un sistema asimilador, absorbiendo la información dada (Ferreiro, 2002).*

Translation: Prior to any alphabetic representation, there are identifiable audible emissions of written text that find correspondence, beginning with a single sound, followed by syllabic representations (with or without conventional value), and ending with alphabetical syllabic representations that precede emerging writing. These levels of representation are assimilated in conceptual systems that absorb the information given. (Ferreiro, 2002).

*En el caso del castellano, diversos trabajos han mostrado la relevancia de la sílaba, señalando que la conciencia silábica se puede usar como un buen indicador de las habilidades lectoras importantes en una ortografía transparente como el castellano, debido a la correspondencia directa entre grafemas y fonemas (Carrillo, 1994; Jiménez & Ortiz, 2000).*

Translation: Studies conducted in Spanish language (terminology using “castellano” refers to Spanish) have demonstrated that syllabic awareness is a good predictor of reading skills, due to the direct influence of a transparent orthography over the grapheme-phoneme correspondence (Carrillo, 1994; Jiménez & Ortiz, 2000).

### Destreza comunicativa (Listening Skills)

The subtest in this domain intends to evaluate students’ listening comprehension proficiency levels as indicators of foundational early literacy skills.

*En todo proceso auditivo, para poder ser asimilada; la información debe ser integrada a un sistema previamente construido (o un sistema en proceso de construcción). No es la información de por sí, la que crea conocimiento, el conocimiento es el producto de la construcción de un sujeto cognoscente (Ferreiro, 2002).*

Translation: Listening proficiency assimilates new information based on existing constructed systems that integrate new information. The information alone does not constitute knowledge. Knowledge is a product constructed through a cognizant subject. Alfabetización: teoría y práctica Emilia Ferreiro, 2002).

- Listening Comprehension refers to the ability to effectively receive auditory input (receptive skills) in order to understand the information that was said. In bilingual education settings, listening skills developed in the native language benefit second language acquisition. In fact, listening often develops before the productive skill of speaking, so students may depend on listening skills while they are silent for an extended period during second language acquisition (Crawford & Krashen, 2007).

Students developing early reading skills continue to strengthen listening comprehension abilities while they are able to read silently. Reading difficulties have been found to reach the same levels that the listening comprehension disorders do. Even so, these two may not be evident simultaneously or may not be mutually exclusive (Junqué, Bruna, Mataró, 2004). The latest publications from the United States National Early Reading Panel (NEPL) and the National Institute for Literacy identify listening comprehension as one of the key foundational skills for later reading achievement.

### Vocabulario (Vocabulary)

Current scientific research overwhelmingly supports the idea that a dearth of vocabulary impedes reading comprehension and a broad vocabulary increases comprehension and facilitates further learning (Hirsch Jr., 2003). Adequate reading comprehension has been correlated to the number of

words in a text that a reader already knows. Experts consider that a good reader knows between 90 and 95 percent of the words in a text.

- Oral language vocabulary refers in general to “the words we must know to communicate effectively” (Armbruster, Lehr, & Osborn, 2003). On the other hand, reading vocabulary refers to words that a student needs to know in order to be able to understand what is read. The development of oral language proficiency—both productive (speaking) and receptive (listening)—is key to literacy growth. Furthermore, there is a rich vein of literature that suggests that vocabulary is an important precursor to literacy (Scarborough, 1998).

- Reading vocabulary demands knowledge of words and their relationships, as well as the ability to extract meaning from words in context. The percentage of words that a reader understands when reading a text either causes the reader to miss the gist of the reading or allows the reader to get a good idea of what is being said. Therefore readers need to be able to make correct inferences in order to determine the meaning of any unfamiliar words (Hirsch Jr., 2003). Bilingual education settings need to take advantage of the vocabulary that students acquire in their native language. First language vocabulary ability has been shown to correlate with later academic language development. “Academic language refers to the decontextualized, cognitively challenging language used not only in school, but also in business, politics, science, and journalism, and so forth. In the classroom, it means the ability to understand story problems, write book reports, and read complex texts” (Crawford & Krashen, 2007).

### Comprensión y reflexión de lectura (Reading Comprehension)

This domain measures reading comprehension skills similarly to state’s criterion reference tests. The comprehension questions in this battery examine students’ ability to identify elements such as main idea, summarization, conclusions, and predictions associated to fiction/ non-fiction and expository texts.

- Reading Comprehension is defined as the process through which meaning is extracted from the written language. Comprehension measures can be classified in two types: (a) literal comprehension, which focuses on the recognition or retrieval of primary details, main ideas, sequences, or cause-effect patterns from the information that is explicit in the text, and (b) inferential comprehension, which requires establishing logical connections and relationships among facts in texts, thus allowing readers to deduce events, make generalizations, interpret facts, and relate previous knowledge or personal experiences to information implicit in the text ().

### Lectura con fluidez (Fluency)

This domain allows students to read silently, indicating the number of words that they are able to read per minute while, at the same time, demonstrating accuracy.

- Silent Reading Fluency is the ability to read text correctly and with appropriate pace. Reading with fluency requires accuracy and speed. Therefore, a fluent reader is able to read aloud effortlessly using a natural expression, as if speaking; whereas a reader who has not yet developed

fluency reads slowly, word by word, with inconsistency and constant interruption. Accuracy refers to the percentage of words read correctly, and speed is the number of words read per minute. In order to measure fluency, a calculation of the number of words read correctly during the time allotted yields a score. Recent studies of fluency outcomes conclude that attention to connections among ideas and between these ideas as they relate to background knowledge are more characteristic of fluent readers than non-fluent readers (Armbruster, Lehr, & Osborn, 2003; NRP, 2000).

Table 3 exemplifies fluency rates (WPM- words per minute) for children learning to read in Spanish in kindergarten through grade 5. The various levels of fluency for each grade level take into consideration a variable rate of fluency acquisition, which depends primarily on practice. The fluency rates proposed in this table were established by *Educando Juntos*, Society for Primary Instruction in Chile, South America, and they take into account the socio-economic status of students learning to read in grades kindergarten through grade 5.

Table 3.

<i>Lectura</i> Reading fluency	<i>1°básico</i> Kindergarten	<i>2°básico</i> First Grade	<i>3°básico</i> Second Grade	<i>4°básico</i> Third Grade	<i>5°básico</i> Fourth Grade	<i>6°básico</i> Fifth Grade
<i>Muy rápida</i> Very fast	56	84	112	140	168	196
<i>Rápida</i> Fast	47-55	74-83	100-111	125-139	150-167	178-195
<i>Media alta</i> Medium-high	38-46	64-73	88-99	111-124	136-149	161-177
<i>Media baja</i> Medium-low	29-37	54-63	76-87	97-110	120-135	143-160
<i>Lenta</i> Low	22-28	43-53	64-75	85-96	104-119	125-142
<i>Muy lenta</i> Very low	21	42	63	84	103	124

### Comunicación escrita (Writing Conventions)

The subtests in this domain measure orthography development and dictation. Orthography measures comprise spelling and use of accent marks, while dictation measures a student's ability to follow grammatically correct sentence structures and emergent syntactic skills.

- Spelling refers to the ability to determine the fully-specified orthographic representations of words in the language. Knowing the spelling of a word makes the representation of it sturdy and accessible for fluent reading (Ehri, 2000; Snow et al., 2004).
- Dictation refers to receptive and productive syntactic skills that have been found to be related to reading ability (Scarborough, 1990). These studies found that there are evident discrepancies between the sentences produced by preschoolers who became poor readers and sentences by those who did not.

According to research by Snow, Burns, and Griffin (1998) with young learners, there are 3 components of first language ability that have been shown to correlate with later reading development. These components include story recall, lexical skills, and syntactic skills. If a student is able to find the relationship of words inside a sentence after hearing it, he or she should also be able to demonstrate it productively.

#### **Development of *ISIP Español* Items**

The items for each subtest were written according to sound standards for educational testing (Stiggins, Chappuis, 2006). These were used to design the framework for *ISIP*'s processes, including underlying constructs, item rules, and selection of distractors. As such, the process and experts involved as described below provides a strong source of validity evidence related to the content of the assessment (see Chapter 4 discussion on validity). The design of the test was guided by the recommendations for Early Grade Reading Assessment (EGRA), presented during the 2008 IRA World Symposium (). In addition, a Texas-based editorial firm, TRI-LIN, devoted to assessment development and special education services focusing on the Spanish and bilingual educational community, was contracted by Istation to write over 5,000 items. A review conducted by national experts, included language, content, and cognitive objectives. The language review was conducted to eliminate any Spanish language issues such as colloquial terms or regionalisms. In addition, a content review was conducted to eliminate any bias and ambiguity. The cognitive level review was done to align all forms to grade level expectation measures from pre-kindergarten to third grade. This review was conducted by members of the advisory board, listed below.

Dr. Iliana Alanis  
Assistant Professor  
University of Texas at San Antonio  
Department of Interdisciplinary Learning and Teaching

Dr. Igone Arteagoitia  
Research Associate  
Center for Applied Linguistics (CAL)  
Washington DC

Dr. Kathy Escamilla  
Professor of Education  
School of Education  
University of Colorado at Boulder

Dr. Gilda Evans  
Retired Assistant Superintendent over the Multi-language Department of Dallas ISD  
Current Vice-president of Bilingual Education Association of the Metroplex (BEAM)

Dr. Eugene E. García  
Professor and Vice President, Education Partnerships  
Arizona State University

Dr. William Pulte  
Associate Professor, Director of Bilingual Education Programs  
Simmons School of Education  
Southern Methodist University

Dr. Luis Rosado  
Professor – College of Education  
Director – Center for Bilingual Education  
University of Texas at Arlington

Dr. Annette Torres-Elias  
Assistant Professor of Education  
School of Education  
Texas Wesleyan University

## Item Development and Test Forms

Spanish Language Arts Reading (SLAR) and English as a Second Language (ESL) Elementary Standards are available under Texas Education Code 128.10. These standards were used by TRI-LIN, a Texas-based editorial firm, contracted by Istation to develop a “fusion” of standards combining the Texas SLAR with Spanish reading standards from selected states and countries (i.e., California, Puerto Rico, World-class Instructional Design and Assessment [WIDA] consortium, Colombia, Mexico and Spain). The merged SLAR standards were revised by national experts who collaborated with Istation as members of the ISIP Español Advisory Board. A team of item writers from TRI-LIN used the merged standards to determine two cornerstones for item development: 1) the skills (or domains) to be assessed, and 2) end of year expectations in grades pre-kindergarten through grade 3. Advisory board-established domains of *ISIP Español* per grade level are shown in Table 4.

Table 4. ISIP Español *domains by grade level*

	Listening Comprehension	Phonological Awareness and Phonics	Reading Comprehension	Vocabulary	Writing Conventions	Reading Fluency
K	✓	✓	✓	✓		
1		✓	✓	✓	✓	
2			✓	✓	✓	✓
3			✓	✓	✓	✓

More than 5,000 items were written using multiple choice answer options. The answer choice design follows specific elimination rules (these rules are available upon request). All items were originally written for use in this assessment; no items were translated or derived from any assessment delivered in English. In addition, all items underwent comprehensive analysis to ensure that no items contained linguistic or cultural bias and that all were age and grade level appropriate. Thirteen forms for each grade level were developed. All thirteen forms of the test per grade level are similar in difficulty and were designed to assess end-of-year grade level expectations. Each form contains the same number of items per administration as shown on Table 5.

Table 5. ISIP Español *number of items per domain and grade level*

Domain*	Grade	Total Items Developed	Total Items per Administration
Phonological Awareness	K	351	27
	1	298	23
Listening Comprehension	K	104	8
Vocabulary	K	155	12
	1	194	15
	2	364	28
	3	310	24
Reading Comprehension	K	104	8
	1	130	10
	2	195	15
	3	260	20
Writing Conventions	1	102	6
	2	102	6
	3	149	9

\*The total number of fluency items depend on each reader's ability and therefore are not included in this table.

### Teacher Friendly

*ISIP Español* is teacher-friendly. The assessment is computer-delivered with no manual scoring needed. It requires very little teacher/examiner or administration effort during testing. Teachers monitor student performance during assessment periods to ensure result reliability. In particular, teachers are alerted to observe specific students identified by *ISIP Español*, based on a three-tier normative data system - Tier 1, Tier 2, and Tier 3. Normative grouping is based in indices associated with the 20<sup>th</sup> and 40<sup>th</sup> percentiles. Students with a score above the 40<sup>th</sup> percentile for their grade are placed into Tier 1. Students with a score between the 21<sup>st</sup> and 39<sup>th</sup> percentile are placed into Tier 2. Students with a score below at the 20<sup>th</sup> percentile and below are placed into Tier 3. Teachers subsequently review student results to validate outcomes. For students whose skills may be a concern based upon performance level, teachers may easily validate student

results by re-administering the entire *ISIP Español* battery or individual skill assessments on an on-demand basis.

### **Student Friendly**

*ISIP Español* is also student-friendly. Each assessment session simulates a computer game. The ISIP assessment is introduced to the students as a challenge, called *A ver cuánto sabes* (Show what you know). In the beginning of the session, an animated owl named Don Buhiermo appears on the screen and plays the role of a game show announcer who invites students to participate by saying, “¡Bienvenidos! ¡En este juego vas a demostrar que sí puedes!” (Welcome! It’s time to show what you can do!) After the owl helps the students to understand the game rules, a model and a practice period follow before the assessment begins. At the end of each assessment, children see an animated graph of their progress. Each activity proceeds in a similar fashion.

### **Ongoing Revisions**

Istation is currently undertaking the IRT work necessary to develop a computerized adaptive testing (CAT) assessment. The use of CAT algorithms will further enhance test administration. The adaptive item algorithm will allow the assessment system to adjust item difficulty while a student is taking the test, quickly zeroing in on ability level. As mastery level is determined based on the algorithms, the subtests are dropped from the battery. Therefore, when the CAT version for *ISIP Español* is completed, the amount of time necessary to accurately determine student ability will be further reduced. Reliability and validity research will be updated once this phase of development is complete. See Chapter 5 for ongoing development details.

### **Rationale for Subtests Administration**

Children acquire the skills that they need to become proficient readers during the first years of school. These skills may be introduced and monitored separately, but teachers need to practice integrating these skills during daily reading routines as often and quickly as possible. Critical early reading skills are emphasized according to the grade level and the developmental stage of the child; however, daily practice of identified critical domains such as phonemic awareness, vocabulary, comprehension, and fluency is highly desirable (Vaughn & Linan-Thompson, 2004).

Based on research findings from the National Reading Panel (NRP, 2000), instruction in phonemic awareness is emphasized in kindergarten and is recommended for about 15 minutes a day (Vaughn & Linan-Thompson, 2004). The teaching of phonemic awareness has expanded through many countries (particularly Spanish-speaking countries) by introducing instructional methodologies that allow students to manipulate phonemic and syllabic sounds of spoken words. Such methods involve teaching and practicing blending, segmentation of sounds, and identification of sounds that represent sounds in speech with or without the use of letters (NRP, 2000).

Early literacy instruction that incorporates decoding and word study provides a strong foundation for emergent literacy. Therefore, related skills such as grapheme-phoneme correspondence and print awareness can be introduced as early as kindergarten. Based on the transparency of language studies that place Spanish as a language with shallow orthography (Wimmer, Mayringer, Landerl, 2000; Ziegler, Perry, Ma-Wyatt, Ladner, Shulte-Körne, 2003;), children learn these skills rather quickly, and it is important to integrate these emergent reading skills with reading comprehension questioning. Assessment items in the reading comprehension measures require connections to students' own experiences and prior knowledge. This questioning model was adapted and implemented in the current version of *ISIP Español* in preparation for future predictive validity studies using the State of Texas Assessment of Academic Readiness (STAAR). In 2002 Taylor et al. found that students in first grade grew more in comprehension and fluency when their teachers asked more high-level questions.

Once students acquire a solid foundation in word recognition and decoding, fluency instruction should be emphasized. English literacy studies have shown that this usually begins during the second semester of first grade (Vaughn & Linan-Thompson, 2004). For fluency instruction in Spanish, the transparency of the language must be taken into consideration. Fluency development may begin as early as kindergarten, based on the fact that Spanish phoneme combinations can be represented in only 45 variations (). Spanish grapho-phonemic conversions also make writing development equally accessible and rapidly attained. Tasks that require knowledge of grapheme-phoneme and symbol-sound correspondence can be evaluated through students' spelling and dictation performance. Studies which examine the productive and receptive syntactic skills of kindergarteners also show correlations with success in reading (Ballantyne, Sanderman, McLaughlin, 2008). Research has also shown that learning to spell and learning to read rely on much of the same underlying knowledge, such as the relationships between letters, letter units, and sounds. Knowing the spelling of a word makes the representation of it sturdy and accessible for fluent reading (Ehri, 2000 ; Snow et al., 2004).

Teaching vocabulary supports reading comprehension and increases speed, thus improving fluency. First and second grade vocabulary teaching impacts reading development, and the instructional methodology must incorporate familiar language in order to take advantage of the word superiority effect (). Students need to be able to identify with and connect to the reading material in order to be interested in it. Linguistic considerations are important, but cultural relevance may be a determining factor in assessing students' reading success (). When stories are interesting and written in simple language, they are very likely to encourage struggling students to persevere (Abadzi, 2006).

Successful reading development is also associated with small-group instruction. Hierarchical Linear Modeling (HLM) has been used to analyze classroom variables and compare outcomes. These observations confirm that teachers who engage often in small-group instruction have

students who demonstrate more gains in fluency, comprehension, and vocabulary.

### **Measures for Each Subtest**

Grapheme-phoneme Correspondence is a measure of alphabetic principle that assesses how many letter sounds a student can correctly identify. Item selections for this portion of the assessment represent a combination of both upper and lower case letters, including vowels and consonants. Symbol-Sound is a measure of symbol conversion based on auditory input that combines letter units (syllables), as opposed to single phonemes (letters). Item selections for this portion of the assessment include the following syllable types: opened (CV), closed (CVC), consonant combination (CCV), and vowel combination (VV). Items are presented according to kindergarten and first grade level expectations.

Beginning/Ending Sound is a measure of phonemic awareness that assesses a student's ability to recognize the initial sound in an orally presented word. Blending is a measure of phonemic awareness that assesses a student's ability to blend syllables and phonemes that make up spoken words. These skills are tested in kindergarten and first grade, as aligned to SLAR Standards. The subtests Grapheme-phoneme Correspondence and Phonemic /Syllabic Awareness are reported as a single score for *Destreza fonológica y fonética*. (Phonological Awareness and Phonetic Skills).

Listening Comprehension is a measure of a student's ability to listen and retain enough information in his or her working memory to be able to recall simple facts. This skill is tested in kindergarten and first grade, as aligned to SLAR standards. This subtest requires students to listen to two passages, one narrative and one expository. After listening, students answer questions about the passage. The questions follow the same pattern used for the *ISIP Español* reading comprehension battery (See page 17). This skill is tested in kindergarten, and first grade, in alignment to SLAR expectations. The resulting score is reported as *Comprensión auditiva* (Listening Comprehension)

Vocabulary is a measure of a student's knowledge of two types of vocabulary words: (1) oral vocabulary, or "common" words, which are primarily used in daily social interactions according to each developmental age (grade level appropriate) and (2) academic vocabulary, or "meaning" words, which are frequently encountered in text but not typically used in daily conversation (Beck, McKeown, & Kucan, 2002 ). In particular, this second evaluation target contains items that were developed to assess students' knowledge of specific Spanish language elements that support understanding of meaning, such as word association, word derivatives, word roots (prefixes/suffixes), and synonyms. The subtests in this domain require that students not only demonstrate knowledge of words, but also demonstrate knowledge of the relationships among words. For the first subtest, four pictures appear on the screen. The narrator asks the student to identify the picture that best illustrates the word spoken orally. The academic vocabulary subtest

requires the students to use a number of word strategies that are assessed using both pictures and words that appear in sets of four on the screen. After instructions are given, the student is asked to identify each word accordingly. These two subtests are scored together and reported as *Vocabulario* (Vocabulary).

Reading comprehension is a measure of a student's ability to read and understand grade level appropriate narrative and expository passages. According to the NRP research (NRP, 2000), *text comprehension* is a complex cognitive process that incorporates all foundational reading areas, including vocabulary and fluency. In order to assess a student's ability to comprehend the passages, this type of evaluation requires an intentional and thoughtful interaction between the student and the computer screen where the passages are presented. Students in kindergarten are able to listen to the answer choices associated with a picture before they select their answer. Students in first through third grade are able to apply reading comprehension strategies to enhance understanding. The passage appears on the screen, and the student prompts the computer to begin the questions. Once the questions begin, the passage moves to the left side of the screen, and each question changes after 50 seconds to avoid inactivity. The questioning design is similar to the multiple choice patterns used in state's criterion referenced tests that combine explicit and implicit answers as they apply to grade level requirements aligned to SLAR standards. The resulting score is reported as *Comprensión de lectura* (Reading Comprehension).

Fluency is a measure of a student's ability to read silently with comprehension. This subtest is constructed in a very different manner than others, using grade-leveled, culturally appropriate passages. Each of these passages was carefully written to conform to specific word level features, follow linear story grammar structure, and have readability according to a commonly accepted readability formula for end of grade level expectations as it applies to Spanish fluency (see fluency rates used in Table 3). In order to assess text reading on the computer, a maze task is utilized in which every seventh word is left blank, with a three word menu of choices to complete the sentence. The score obtained from this task incorporates the number and accuracy of Maze items completed in the allocated time, and it accounts for the number of words read between Mazes. This task has been shown to be highly correlated to measures of both fluency and comprehension, and it has high reliability and concurrent validity (Espin, Deno, Maruyama, & Cohen, 1989; Fuchs & Fuchs, 1990; Jenkins, Pious, & Jewell 1990; Shinn, Good, Knurson, Tilly, Collins, 1992). As opposed to fluency measures for first grade, in which the teacher relies on oral/observable measures, students in second and third grade would be more accurately assessed using tools that register receptive reading skills. Silent reading fluency is tested in second and third grade, and the resulting score is reported as *Lectura con fluidez* (Fluency).

Spelling and Dictation are measures designed to determine if students are developing fully-specified orthographic representations of words. Items were designed following different rules, depending on the grade level assessed. For first grade students, a word is given and an array of syllables appears on the screen, with which the student spells the word. Second and third grade

students use individual letters to spell the words. Items for the dictation subtest follow a similar functionality. Students choose, from a word bank, the necessary word to complete a sentence. These items have been carefully constructed to move from easier to harder, using a sequence of difficulty aligned to SLAR standards. Additionally, students in first and third grade are required to select correctly spelled words that exemplify commonly used accented patterns for word categories such as *palabras llanas*, *agudas*, and *esdrújulas*. The scores for each subtest are factored together into a single score under the domain *Comunicación escrita* (Writing Conventions).

### **The *ISIP Español* Link to Instructional Planning**

*ISIP Español* provides continuous assessment results that can be used in recursive assessment-instructional decision loops. Initially *ISIP Español* identifies students in need of support. If validation of student results is needed, re-administering the assessments can increase the reliability of the scores. The technology underlying *ISIP Español* delivers real-time reports on student progress immediately upon assessment completion. This data facilitates the evaluation of curriculum and instructional plans. Assessment reports automatically group students according to level of support needed based on each skill evaluated. Data are provided in both graphical and detailed numerical formats on every measure and at every level of a district's reporting hierarchy. Reports provide summary and skill information for the current and prior assessment periods that can be used to evaluate curriculum, plan instructional support, and manage resources.

At each assessment period, *ISIP Español* automatically alerts teachers to children in need of instructional support through a summary of students' results called the Priority Report. In this report, the students are grouped according to instructional level and skill need. Links are provided to teacher-directed lessons for each instructional level and skill category. There are downloadable lessons and materials appropriate for each level or tier of instruction.

In addition to the recommended activities, reading coaches and teachers have access to an entire library of teacher-directed lessons and support materials at [www.istation.com](http://www.istation.com). These downloadable, printable lessons support small group instruction through scripted lessons. These teacher-directed lessons are based on student individualized needs per the Priority Report. As the lessons are taught, teachers document intervention delivery on the Priority Report. This provides a visual reference of teacher intervention and its effectiveness. The ease of identification of skill needs and readily available lessons facilitates intervention and puts instructional time back in the classroom. All student information is automatically available by demographic classification as well as specially designated subgroups of students that need to be monitored by tier level.

A year-to-year history of *ISIP Español* results is available. Administrators, principals, and teachers may use their reports to evaluate and modify curriculum. Interventions, Adequate Yearly Progress (AYP), effectiveness of professional development, and personnel performance may also

be correlated to the growth model depicted from the reports.

## **Scoring and Reporting**

A raw score, based on the number of items answered correctly out of the total number of items, and a percent score are calculated. Each test item equates to one point. Scores are calculated and reported for each domain assessed. As soon as a student (or group of students) completes each testing session, several reports are immediately available with information about performance for each student, group of students (class), school, and school district. Teachers are able to immediately determine how each student performed. The real-time data reports are designed to show student performance over time. Teachers and administrators can use the results on a large scale to determine how a given group of students is progressing toward grade level expectations for each domain assessed. In addition, the *ISIP Español Informe de Prioridad*, Priority Report, automatically groups students who fail to answer a sufficient number of items within a skill set or domain. The structure and functionality of the Priority Report is based on a set of student performance level descriptors or tiers, grouped by grade level and domain. The results reported in these Priority Reports link to more than 150 lesson plans for Spanish early literacy development aimed to help students performing in three levels of risk, thus promoting small group instruction and intervention practices for struggling readers. A complete history of Priority Report notifications, including those from the current year and all prior years, is maintained for each student. On the report, teachers may acknowledge that suggested interventions have been provided. A record of these interventions is maintained with the student history as an Intervention Audit Trail. This history can be used for special education individualized education program (IEP) and in Response to Intervention (RTI), or other models of instruction that require modifications of a student's instructional plan.

## Chapter 4

### *ISIP Español – Reliability, Item Parameters, and Validity*

This report assesses the *ISIP*<sup>TM</sup>, *istation's Indicators of Progress, Español* examinations. It uses data collected from kindergarten to third grade students in six states<sup>1</sup> during the 2010-11 school year (Table 1). However, most of the testing (95.6%) was in Texas elementary schools. The testing was conducted in 37 school districts, covering 228 schools. Among these, 30 school districts and 217 schools were in Texas.

Table 1. *Student participants in the ISIP Español examinations\* (as of April 2011)*

		<i>ISIP Español</i> Subjects				Total
		Kindergarten	Grade First	Grade Second	Grade Third	
Gender	Male	424	502	481	411	1818
	Female	421	450	429	429	1729
	Total	845	952	910	840	3547
	Missing/Unidentified	123	76	77	72	348
<hr/>						
Special Education						
	Yes	41	56	79	65	241
	No	788	862	806	754	3210
	Total	829	918	885	819	3451
	Missing/Unidentified	139	110	102	93	444
<hr/>						
Economically Disadvantaged						
	Yes	494	817	791	739	2841
	No	336	101	94	79	610
	Total	830	918	885	818	3451
	Missing/Unidentified	138	110	102	94	444
<hr/>						
English Proficiency						
	Non-English Speaker	480	344	294	303	1421
	Fluent English Speaker	1	0	0	1	2
	Limited English Speaker	354	575	591	514	2034
	Total	355	575	591	515	3457
	Missing/Unidentified	133	109	102	94	438

\* Students used for assessing the parameters of CAT items

Collected data were used to determine the item parameters. Originally, 2,586 test questions were tried out (Table 2). Among these, 2,419 questions were within the desired ability range measured by the test ( $-3.50 \leq \theta \leq 3.50$ ). However, a total of 331 questions fell below the desired minimum discrimination range (0.50). These questions were referred to subject matter experts (SMEs) for

<sup>1</sup> Texas, Illinois, Indiana, Maryland, New Mexico, and Pennsylvania

review and possible re-inclusion in the tests at a later date if their respective parameters fulfill the desired minimum criteria.

Table 2. Test question review

<b>ISIP Español Examinations Item Exclusion</b>							
<b>Examination</b>	<b>Grade</b>	<b>Total No. Items</b>	<b>No. Items below Threshold <math>\theta &lt; -3.50</math></b>	<b>No. Items above Threshold <math>\theta &gt; 3.50</math></b>	<b>Total Excluded</b>	<b>No. Items Needing Review*</b>	<b>No. of Usable Items</b>
<b>ESC</b>	<b>1</b>	104	0	1	1	2	101
	<b>2</b>	104	0	2	2	7	95
	<b>3</b>	169	4	6	10	21	138
	<b>Total</b>	367	4	9	13	30	324
<b>FON</b>	<b>K</b>	419	96	1	97	19	303
	<b>1</b>	471	23	2	25	13	433
	<b>Total</b>	890	119	3	135	32	723
<b>LC</b>	<b>K</b>	104	0	1	1	10	93
<b>RC</b>	<b>K</b>	155	5	15	20	25	130
	<b>1</b>	135	8	27	35	32	68
	<b>2</b>	136	6	9	15	21	100
	<b>3</b>	201	3	15	18	42	141
	<b>Total</b>	472	17	51	68	120	284
<b>VOC</b>	<b>K</b>	164	7	2	9	16	139
	<b>1</b>	170	7	13	20	43	107
	<b>2</b>	295	6	10	16	34	245
	<b>3</b>	228	7	2	9	46	173
	<b>Total</b>	857	27	27	54	139	664

\* Discrimination below the desired minimum of 0.50

### Reliability

Reliability is an important quality of measurement data. The adage that a test cannot be valid without being reliable applies to item response theory (IRT), although the methodology is not exactly the same as in classical test theory. Reliability can be thought of as the consistency, either over items within a testing instance or over scores from multiple testing instances, and the precision of the test.

The *marginal reliability* in IRT is used as an overall index of precision in computerized adaptive testing (CAT) for comparison with the classical internal-consistency reliability estimated for paper-and-pencil forms (Dimitrov, 2003; Green et al., 1984; Thissen, 1990). In effect, this statistic estimates the reliability of an entire set of questions (item bank) for a CAT.

The IRT analogue to classical internal consistency is marginal reliability (Bock & Mislevy, 1982), which is thus applied to *ISIP Español*. Marginal reliability is a method of combining the variability in estimating abilities at different points on the ability scale into a single index (Table 3). Like Cronbach's alpha ( $\alpha$ ), marginal reliability is a unit-less measure bounded by 0 and 1, and it can be used with Cronbach's alpha to directly compare the internal consistencies of classical test data to IRT-based test data. The marginal reliability coefficient operates on the variance of the ability scores ( $\sigma_{\theta}^2$ ) and the average of the expected error variance ( $\bar{\sigma}_E^2$ ; Sireci, Thissen, &

Wainer, 1991):  $\rho^2 = \frac{\sigma_{\theta}^2 - \bar{\sigma}_E^2}{\sigma_E^2}$ . *ISIP Español* has a stopping criterion based on minimizing the

standard error of the ability estimate. As such, the lower limit of the marginal reliability of the data in any testing instance of *ISIP Español* will always be approximately 0.90.

As shown in Table 3, the majority of the tests depict very high reliability coefficients: from Vocabulary in kindergarten (.99) to Phonological Awareness and Phonetic Skills in first grade (.996). The largest variability occurs with the Spelling examinations (ESC) in second and third grades. While both tests have lower reliability coefficients, the standard deviations for these tests are about the same as for all of the other tests. In addition, as shown below in the test information function (TIF) graphs (Figures 6 and 7), they cover about the same range of ability in the same distribution as many of the other tests. They also cover the full range of abilities,  $-3.5 \leq \theta \leq 3.5$ . The variance in reliability as compared to the other tests is attributed to the fact that the tests contain a maximum of six questions. It is well-known that the number of questions on a test has a direct impact on the reliability. This will be carefully monitored as the tests progress. It is interesting to note that if second and third grade Spelling (ESC) tests are removed, the lowest reliability is .821, for Spelling (ESC) in first grade.

Table 3. Reliability and standard error of measurement (SEM)

<b>ISIP Español Reliability Coefficients</b>						
Test	Grade	Reliability	Standard Deviation*	No. of Items	No. of Subjects (N)	SEM**
Spelling (ESC)	1	0.821	0.67	104	4203	0.283
Spelling (ESC)	2	0.603	1.22	104	2927	0.769
Spelling (ESC)	3	0.65	1.27	159	1670	0.751
Vocabulary (VOC)	K	0.99	1.11	164	5131	0.111
Vocabulary (VOC)	1	0.969	1.56	170	5110	0.275
Vocabulary (VOC)	2	0.989	1.04	295	3974	0.109
Vocabulary (VOC)	3	0.989	1.32	228	2607	0.138
Reading Comprehension (RC)	K	0.888	1.44	119	4929	0.482
Reading Comprehension (RC)	1	0.918	1.14	135	4963	0.326
Reading Comprehension (RC)	2	0.949	0.93	136	3820	0.210
Reading Comprehension (RC)	3	0.96	1.08	201	2480	0.216
Listening Comprehension (LC)	K	0.9631	0.94	104	5198	0.181
Phonetics (FON)	K	0.986	1.02	471	5319	0.121
Phonetics (FON)	1	0.996	0.73	419	5226	0.046
Reading Fluency (FLU)	2	0.97	1.09	689	3706	0.189
Reading Fluency (FLU)	3	0.97	1.05	681	2408	0.182

\* in  $\theta$  units (Test ranges  $-3.5 \leq \theta \leq 3.5$ )

\*\* Standard deviation  $\times$  square root( $1 - r$ ),  $s\sqrt{1 - r}$

### Rasch Model Reliabilities of the Item Pools

Another index of reliability comes from a preliminary Rasch (or 1PL) analysis of each measure conducted using the scores of one single district. To accomplish the Rasch analyses, all items across the 13 forms within each measure were combined to improve estimation of item functioning (Rodriguez, 2011). This model considers the 13 forms to be samples of a larger pool of items, which is consistent with the intent to create CAT. This is also true for the 2PL model analysis, discussed above, that was conducted for parameter estimation. These estimates of reliability are based on an estimate of true variance (model-based score variance) and measurement error variance, given the theoretical definition of reliability as the ratio between true-score variance and observed-score variance (the proportion of observed variance that is true). Because one single district's samples are too small to provide stable estimates from a Rasch analysis, the results for this particular analysis should be interpreted as preliminary (Table 4). It is reported here as a concurrent study with the validity analysis in the next chapter.

*Person reliability*, which is similar to traditional test-score reliability, indicates the capacity of the sample to generate a stable ordering of person abilities based on their test scores. Low person reliabilities among a pool of items suggest a high degree of randomness in responses (guessing).

*Item reliability*, which has no traditional equivalent, indicates the capacity of the sample to generate a stable ordering of item difficulties. Low item reliabilities suggest that the sample is too small to provide stable estimates of item difficulty or that the range of item difficulties is narrow.

Table 4. *Preliminary Rasch Analysis Reliability*

Skill Area	Grade	<i>n</i>	No. Items	Person Reliability	Item Reliability
Listening Comprehension	K	52	104	.90	.80
Reading Comprehension	K	52	104	.48	.88
	1	56	130	.83	.74
	2	52	195	.88	.66
	3	59	260	.89	.77
Phonological Awareness and Phonetics	K	52	351	.96	.83
	1	55	298	.96	.78
Vocabulary	K	52	155	.88	.85
	1	55	194	.92	.91
	2	52	364	.96	.89
	3	59	310	.92	.90

These reliabilities provide an index of the measurement quality of the pool of items in each area, based on this specific sample. They might also be interpreted as potential score reliabilities based on the current pools of items.

### **Item Difficulty and Discrimination**

The criterion utilized by Istation for the range of ability covered in the *ISIP Español* examinations is  $-3.50 \theta$  to  $+3.50 \theta$ . This is to say that the desired coverage of each test area is from the lowest .02% (2 out of 10,000) to 99.98%. This range includes 99.96% of the student population. The reported parameter estimates are based on the year-long testing of kindergarten to third grade students, described previously. The items were tested in kindergarten through third grade, with 13 parallel forms for each grade.<sup>2</sup> A total of 2,690 questions were tried out. Questions with difficulty values outside of the desired range of interest for these tests ( $-3.50 \leq \theta \leq 3.50$ , which covers 99.78% of all students in their respective grades) were eliminated. This left 2,419 questions. Among these questions, 331 had discrimination values below the desired minimum of .50. These questions are in review to see if they can be improved for future usage. This leaves a total of 2,088 questions that are immediately usable without further attention. See

<sup>2</sup> The statistics for the forms are not included here, as they contribute nothing toward the CAT version of *ISIP Español*. This information is available upon request from Istation.

Figure 1 for a representation of the distribution of the item difficulty parameter for all of the tests across all of the grades for the *ISIP Español* examination (Hoelzle, 2011).

Figure 1. *Item difficulty across all tests and all grades in this analysis*

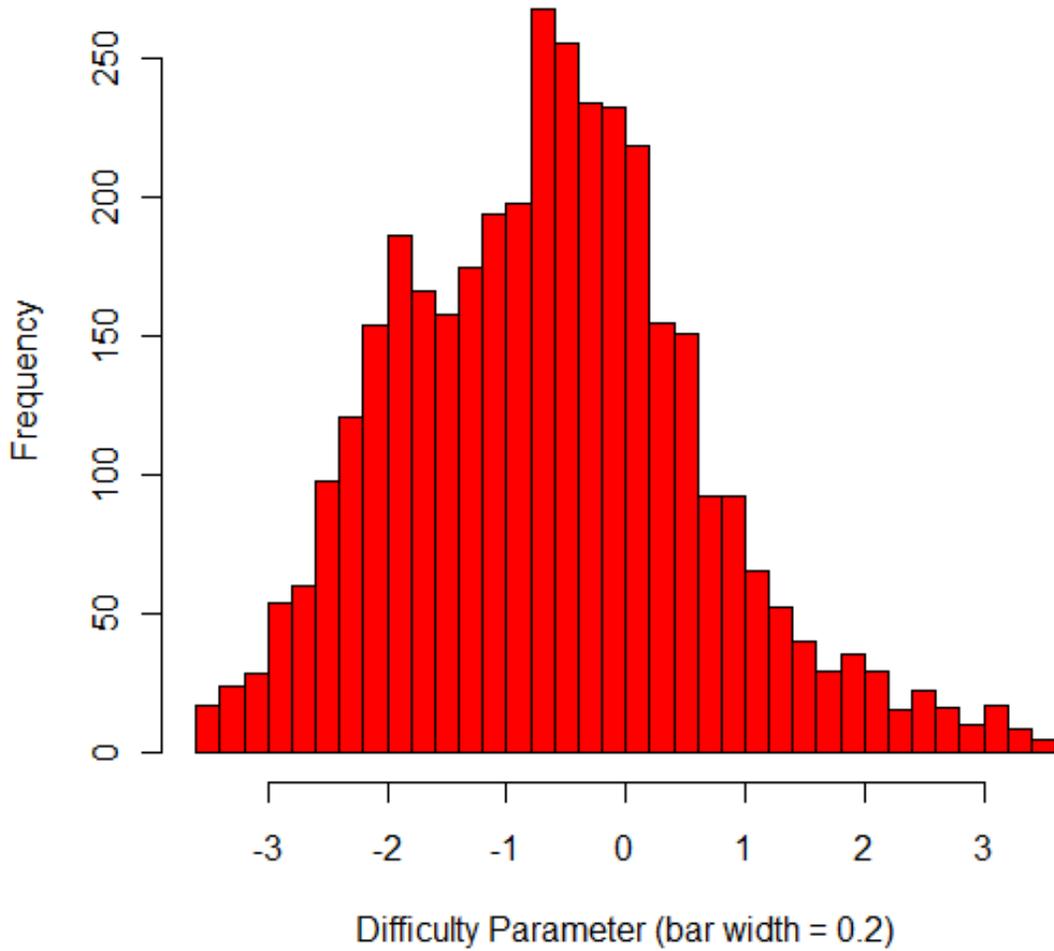


Table 4. *Item difficulty*

<b><i>ISIP Español Item Difficulty</i></b>					
Test	Grade	Mean*	Minimum*	Maximum*	Number of Items (i)**
Spelling (ESC)	1	-0.57556	-1.7717	1.438	102
Spelling (ESC)	2	-0.38182	-2.0374	3.0583	102
Spelling (ESC)	3	-0.11196	-3.4947	3.4349	149
Vocabulary (VOC)	K	-1.43996	-3.0867	3.0822	155
Vocabulary (VOC)	1	-0.63944	-3.4906	3.4278	174
Vocabulary (VOC)	2	-1.09738	-3.2644	2.6903	356
Vocabulary (VOC)	3	-0.98729	-3.4972	3.4338	296
Reading Comprehension (RC)	K	0.586979	-1.9911	3.2182	99
Reading Comprehension (RC)	1	0.462407	-3.4171	3.1819	100
Reading Comprehension (RC)	2	0.569773	-0.3165	3.3091	181
Reading Comprehension (RC)	3	0.58491	-1.7681	3.4026	245
Listening Comprehension (LC)	K	-0.58396	-2.5702	2.1699	103
Phonetics (FON)	K	-0.73663	-3.4643	3.4542	446
Phonetics (FON)	1	-1.98126	-3.4775	1.7558	322
Reading Fluency (FLU)	2	-0.34344	-3.4217	3.4283	604
Reading Fluency (FLU)	3	-0.8044	-3.4808	3.25455	654

\* in  $\theta$  values. Includes items being reviewed. Table used discrimination of  $\leq .2$  as cut-off

\*\* in item bank

Density plots (Figures 2 and 3, Hoelzle, 2011) are provided to graphically depict the item difficulty information of the tests. Information about each test's overall coverage is provided below (See Total Test Information).

Figure 2. *Density plots for difficulty parameter*

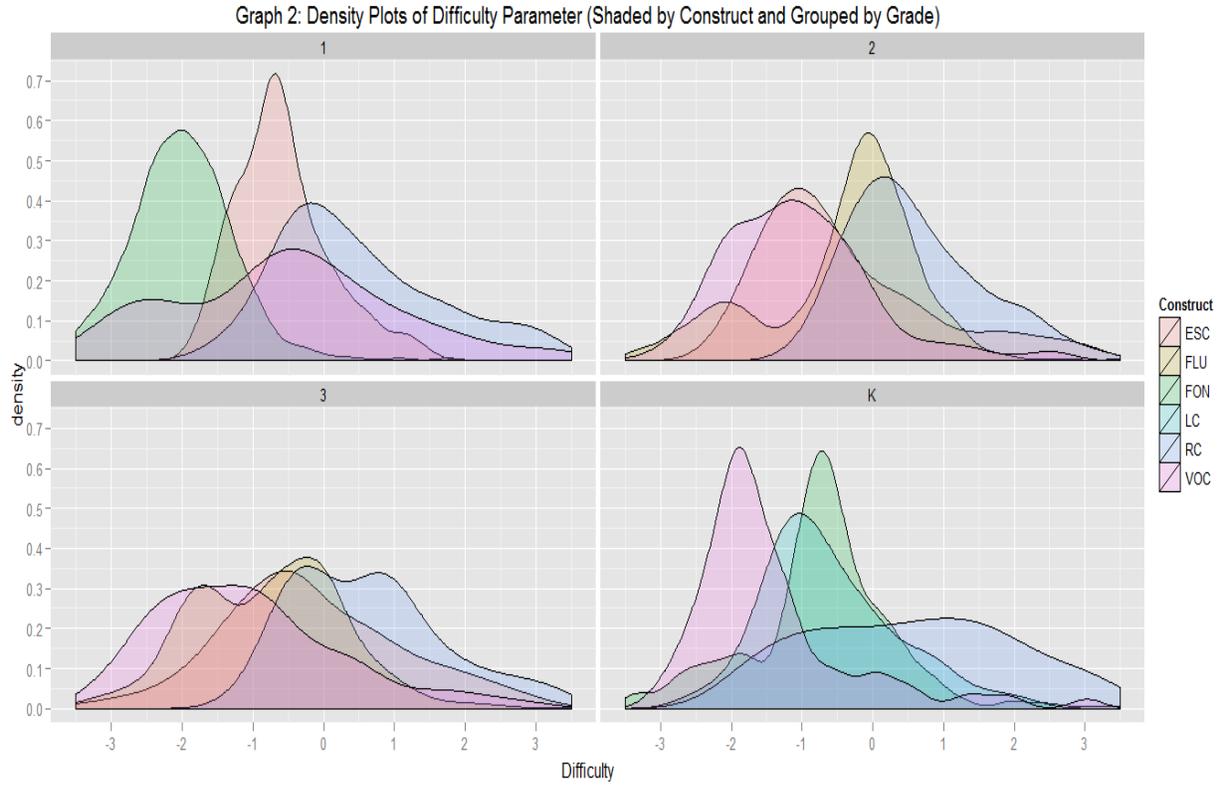


Figure 3. *Density plots for difficulty parameter*



Table 5. *Item discrimination*

<b>ISIP Español Item Discrimination</b>					
Test	Grade	Item Discrimination	Minimum*	Maximum*	Number of Items (i)**
Spelling (ESC)	1	1.36	0.46	2.38	102
Spelling (ESC)	2	1.12	0.28	2.88	102
Spelling (ESC)	3	0.92	0.26	2.02	149
Vocabulary (VOC)	K	1.03	0.25	1.91	155
Vocabulary (VOC)	1	0.86	0.21	1.95	174
Vocabulary (VOC)	2	1.04	0.27	2.45	356
Vocabulary (VOC)	3	0.91	0.23	1.82	296
Reading Comprehension (RC)	K	0.78	0.20	1.95	99
Reading Comprehension (RC)	1	0.80	0.25	2.13	100
Reading Comprehension (RC)	2	0.94	0.28	1.97	181
Reading Comprehension (RC)	3	0.88	0.21	2.64	245
Listening Comprehension (LC)	K	0.94	0.29	1.99	103
Phonological Awareness and Phonetics (FON)	K	1.10	0.29	2.28	446
Phonological Awareness and Phonetics (FON)	1	1.05	0.26	1.82	322
Reading Fluency (FLU)	2	1.08	0.21	3.84	604
Reading Fluency (FLU)	3	1.37	0.21	10.31	654

\* in  $\theta$  values. Includes items being reviewed.

\*\* in item bank

### Total Test Information

A TIF shows what a total test covers, at least in terms of the range of ability best measured by the examination. It is essentially the summation of the item information functions (IIFs) for an entire test. The IIF for the two parameter model is:

$$I(\theta) = a_i^2 p_i(\theta) q_i(\theta)$$

IIFs tend to have a bell-shaped distribution. This is allowed because of the fundamental IRT assumption of local independence. For this reason, IIFs are additive. The shape of their

summation (TIF) depends on the difficulty ( $p_i q_i$ ) and the discrimination ( $a_i$ ) of each item. This property, in a large item bank, allows TIFs to precisely control measurement error; the demonstration of how precisely something is being measured is the purpose of reliability (Baker, 1992, p. 81; Weiss & Yoes, 1991, p. 79).

The following 13 figures demonstrate the maximum amount of information that is available for each test. They also demonstrate the ranges of abilities covered by the tests, where further questions may be needed, and where the tests have their strongest (most precise) measures. Most of the TIFs tend toward normal distributions, as is expected given the range of the subjects. As expected, they also tend to be centered below the zero point of the abscissa.

Figure 4. *Kindergarten Listening Comprehension (LC)*

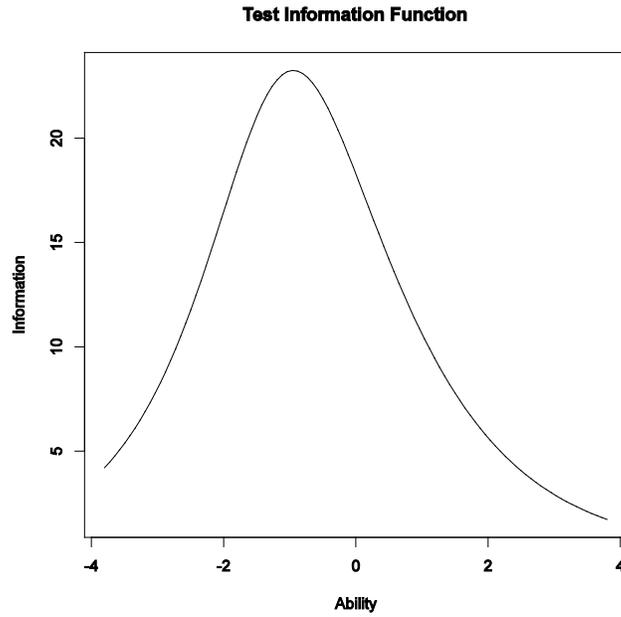


Figure 5. *First grade Spelling (ESC)*

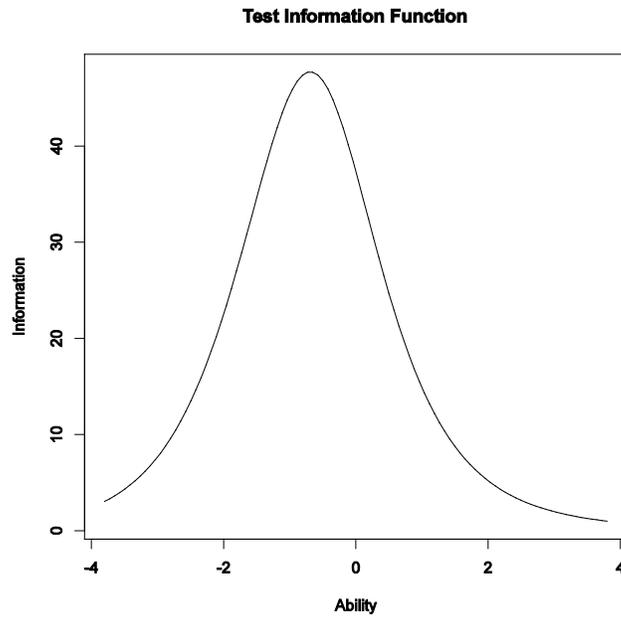


Figure 6. *Second grade Spelling (ESC)*

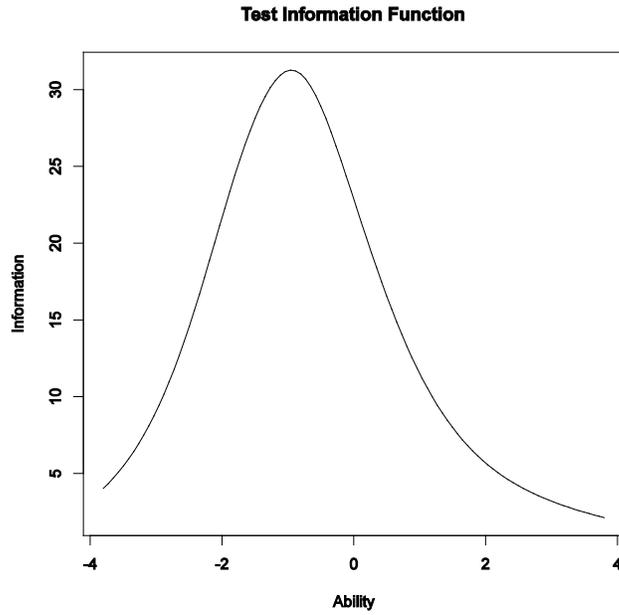


Figure 7. *Third grade Spelling (ESC)*

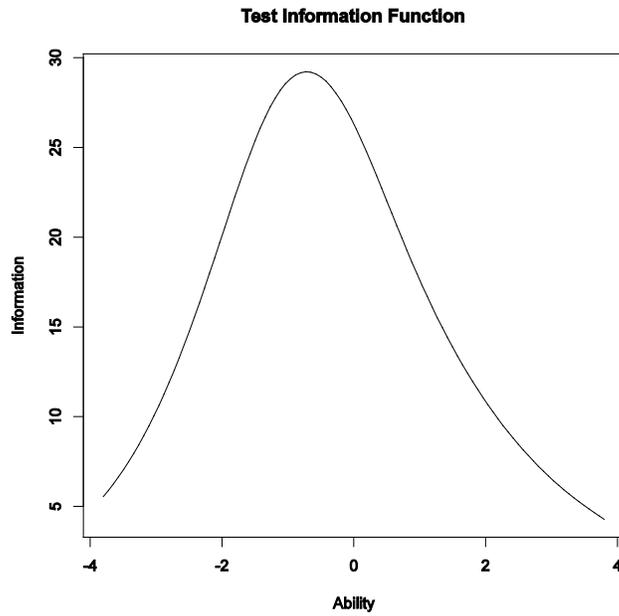


Figure 8. *Kindergarten Vocabulary (VOC)*

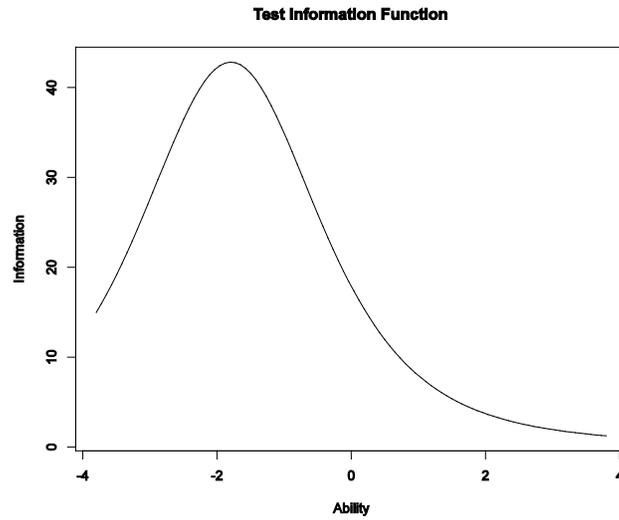


Figure 9. *First grade Vocabulary (VOC)*

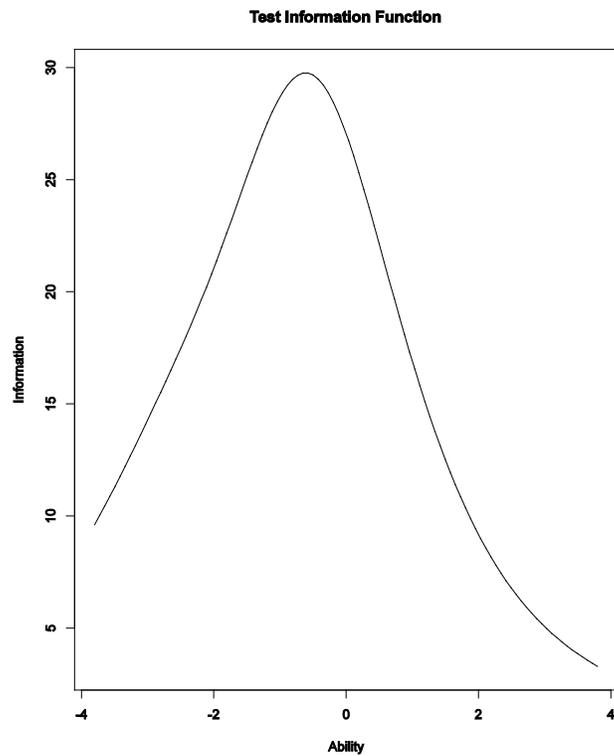


Figure 10. *Second grade Vocabulary (VOC)*

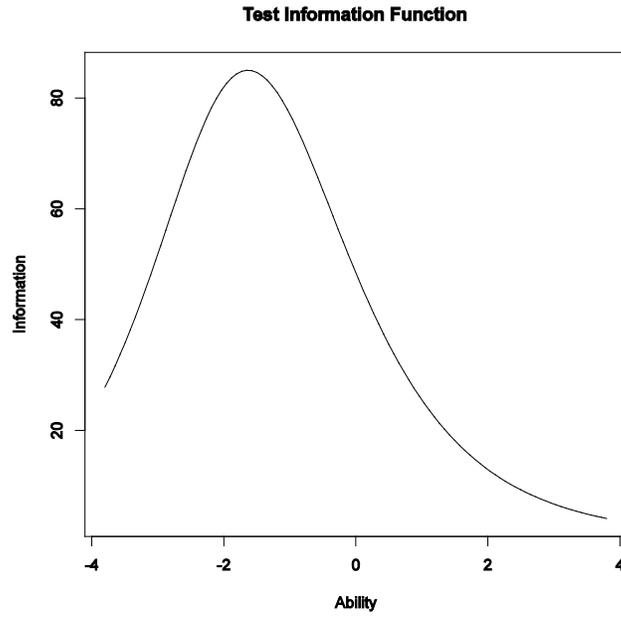


Figure 11. *Third grade Vocabulary (VOC)*

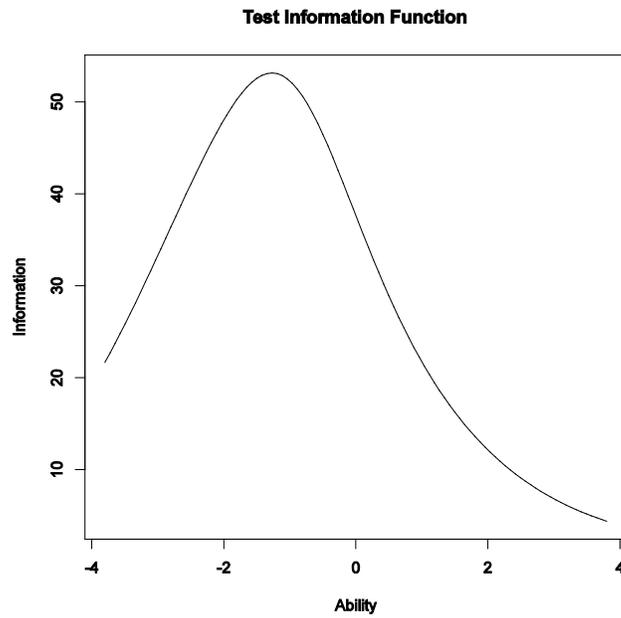


Figure 12. *Kindergarten Reading Comprehension (RC)*

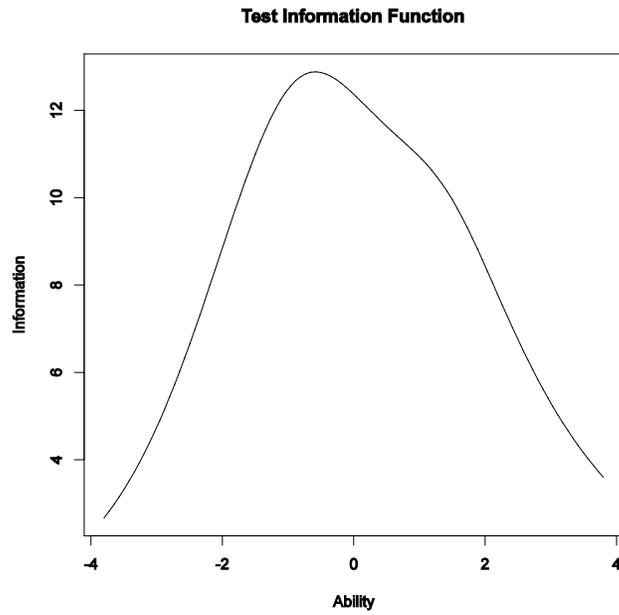


Figure 13. *First grade Reading Comprehension (RC)*

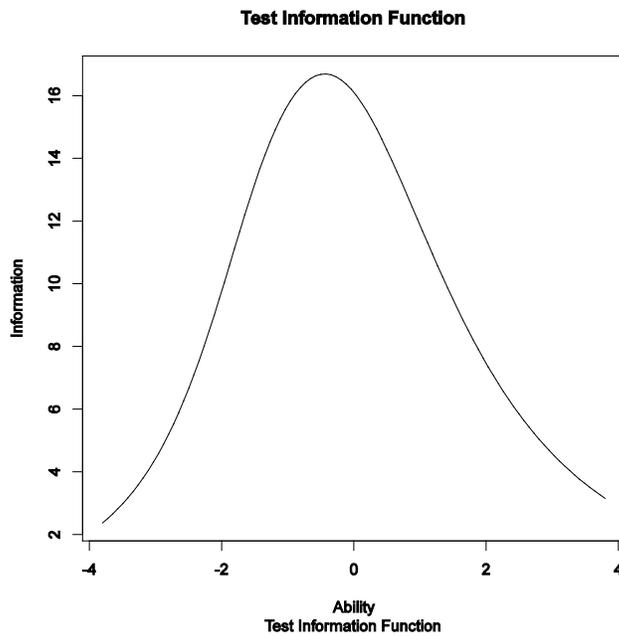


Figure 14. *Second grade Reading Comprehension (RC)*

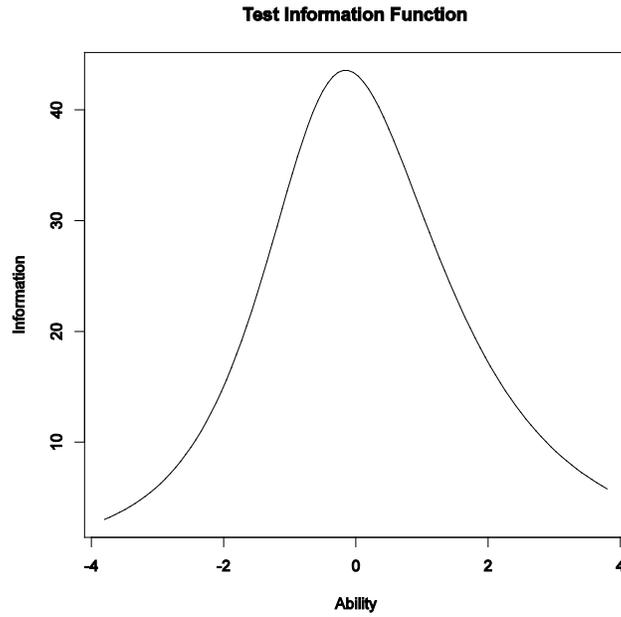


Figure 15. *Third grade Reading Comprehension (RC)*

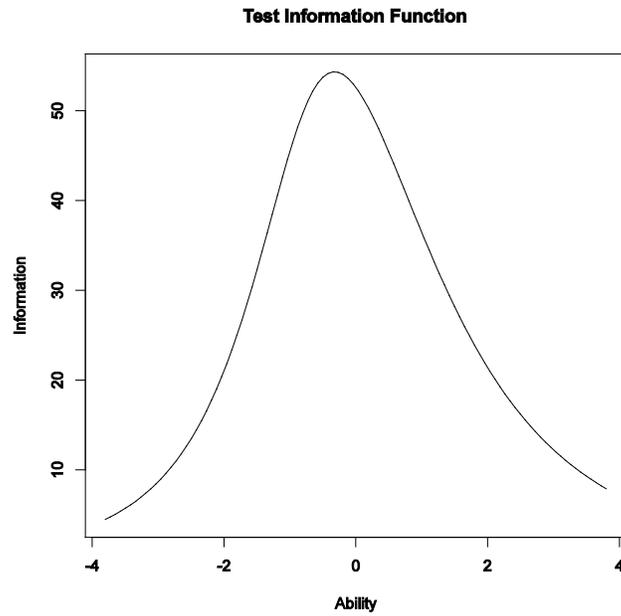


Figure 16. *Kindergarten Phonological Awareness and Phonetic Skills (FON)*

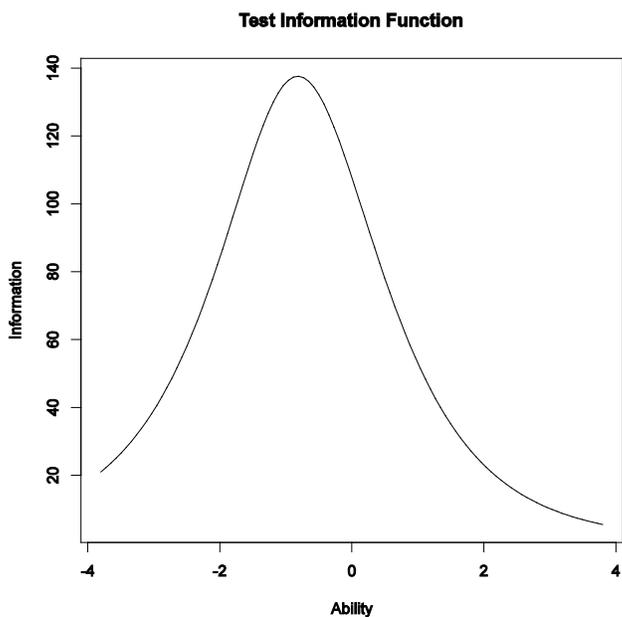
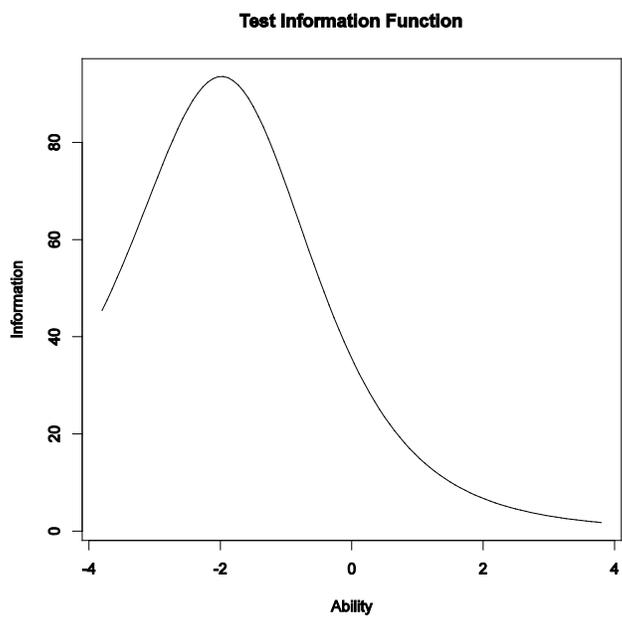


Figure 17. *First grade Phonological Awareness and Phonetic Skills (FON)*



## Validity

For a test to be usable, it needs to possess at least two things: *reliability* and *validity*. This is true regardless of how the test is delivered, whether by computer, paper-and-pencil, essay, orally, etc. While CAT has special concerns about delivery and can greatly enhance the reliability of a test, the validity of the test needs to be established in the same manner as in those using classical measurement theory. The focus of validity is that a test measures what it is supposed to measure. There are three principal types of validity evidence: *construct*, *content*, and *criterion-related*.

*Construct validity* is the degree to which a test measures a theoretical trait or construct it is designed to measure. *Intelligence* is a much-used example of this type of construct or trait. The *ISIP Español* examinations are not intended as measures of specific traits. Therefore, this type of validity evidence is not directly applicable to *ISIP Español* examinations.

*Content validity* is usually established through an analysis of the content of the test questions by SMEs. As presented in Chapter 1 of this report, the advisory committee systematically and logically reviewed and approved all of the test questions that were used in the original test item bank. No question was utilized in this study of the *ISIP Español* examinations that had not been previously approved by these experts.

*Criterion-related validity* is the relationship of test scores with an external measure or measures (criterion/criteria) that is established as being a measure of the characteristic(s) or behavior(s) of interest. Two types of criterion-related validity evidence of particular interest for this study are *predictive validity* and *concurrent validity*.

*Predictive validity*, is established based on future performance. It deals with the degree to which future performance can be predicted from a test score. It is established by giving a test to a representative sample, waiting an established period of time, getting criterion scores, and then seeing the association of the tests with the criterion. For this report, estimating the parameters for the items to be used in the CAT *ISIP Español* examinations was the objective and therefore, a predictive validity analysis was postponed. This will be done in the future, when the CAT version is implemented (see Chapter 5).

*Concurrent validity* is based on how well the test score(s) and the criterion or criteria score(s) are related when they are measured at the same time. In addition to the previously mentioned content validity, for the purpose of the *ISIP Español* examinations, the primary focus of the validity analysis is to see how well these tests relate to established examinations (standards) in the same areas. The Spanish language test areas include Vocabulary, Listening Comprehension, Reading Comprehension, Text Fluency, and Phonological Awareness and Phonemic Skills.

In the case of *ISIP Español* concurrent validity assumes that there is an established standard or test to measure the area of interest and that the test(s) of that area of interest can improve upon the established test. “Generally, concurrent validity is useful in that often there are poor tests of the same variable which the new test attempts to improve on. In cases such as these, concurrent validity studies would expect significant but modest correlations” (Kline, 1986, p. 4).

The comparative points, or validity focuses, of this study are two-fold: *El Inventario de Lectura en Español de Tejas (Tejas LEE)* and Texas Assessment of Knowledge and Skills (TAKS). Both

of these exams are specific to Texas. For that reason primarily, they were selected as the criteria for validity of the *ISIP Español* examinations.

The *Tejas LEE* subtests include the following: *Identificación de letras* (Letter Identification), *Conocimiento de los sonidos* (Knowledge of Letter Sounds), *Unión y segmentación de sílabas y de sonidos* (Syllable Blending and Segmentation), *Identificación de sonido inicial* (Beginning Sound Identification), *Identificación del sonido final* (End Sound Identification), *Unión de los sonidos* (Blending Sounds), *Omisión de sonido inicial y final* (Omitting Beginning and End Sound), *Reconocimiento de palabras* (Word Recognition), *Comprensión auditiva* (Listening Comprehension), *Dictado* (Dictation), and *Exactitud, fluidez y comprensión* (Accuracy, Speed, and Comprehension). *Tejas LEE* uses some combinations of subtest scores to provide total scores. The subtest combination for Phonological Awareness (PA TOTAL) is Phoneme/Syllabic Awareness + Letter Sound Fluency + Beginning/End Sound + Blending. Kindergarten subtests include rhyming as part of this combination, but that subtest was not used for the study. The subtests used for Grapheme-Phoneme Correspondence (GPC TOTAL) are Letter Identification + Knowledge of Letter Sounds + Word Recognition. Second grade subtests do not include Letter Identification or Knowledge of Letter Sounds; this total includes the score of the Word List subtest only. The subtests used for Reading Comprehension Total are Accuracy + Fluency + Comprehension.

TAKS is used to measure how well students are progressing in selected content areas. For this study, only the Reading test was used. The TAKS assesses student progress in conjunction with the state-mandated curriculum, the Texas Essential Knowledge and Skills (TEKS). It should be noted that the Texas Education Agency (TEA) is introducing a new state assessment, State of Texas Assessment of Academic Readiness (STAAR). Because the new STAAR was not available at the time of this study, a concurrent validity study using TAKS was conducted here, even with the reduced numbers of the TAKS Spanish version. When the third grade STAAR results can be obtained, a new validation study will be conducted for the third grade.

This study was conducted in ten schools of a suburban/urban school district located in the El Paso area (Table 7). These schools implemented a bilingual education program through a 90-10, 80-20, 30-70, and 50-50 model, in which they increase the instruction in the targeted language by grade level. Students in this study were receiving Spanish Language Arts Reading (SLAR) instruction through third grade, as recommended by the Language Proficiency Assessment Committee (LPAC). Exit criteria were used to stop instruction in the native language as needed. This implementation design lead to a notable reduction in the number of students who were assessed in the upper grade levels. *Tejas LEE* is given in kindergarten, first grade, and second grade. TAKS is given in third grade. 740 students were included in the comparison of the *Tejas LEE* scores with the *ISIP Español* scores. However, comparisons of the appropriate *ISIP Español* scores (Vocabulary, Reading Comprehension, and Fluency) with TAKS scores (Reading – Raw Score [RDRAW] and Reading – Scaled Score [RDSCAL]) were limited to 42, 40, and 39 students, respectively. Nonetheless, these samples and their results are sufficient for the preliminary validity study. A further study, with a larger sample size and more external instruments, will be conducted in the future when the *ISIP Español* tests are more commonly used. (Please see future discussion in Chapter 5.)

Table 7. District's student participants in ISIP Español validity study

	ISIP Español Subjects (N = 591*)				
	Grade (n, %)				
Gender	Kindergarten	First Grade	Second Grade	Third Grade	Total
Male	108(47)	86(49.7)	56(56)	33(45.2)	284(49.2)
Female	122(53)	87(50.3)	44(44)	40(54.8)	293(50.8)
Total	230(39.9)	173(30)	100(17.3)	73(12.7)	577
Missing/Unidentified	7	2	3	2	14
Special Education					
Yes	8(3.5)	6(3.5)	5(5)	2(2.8)	21(3.7)
No	222(96.5)	166(96.5)	95(95)	70(97.2)	554(96.3)
Total	230	172	100	72	570
Missing/Unidentified	7	3	3	3	16
Economically Disadvantaged					
Yes	144(62.6)	159(92.4)	89(89)	67(93.1)	459(80)
No	86(37.4)	13(7.6)	11(11)	5(6.9)	115(20)
Total	230	172	100	72	574
Missing/Unidentified	7	3	3	3	16
English Proficiency					
Non-English Speaker	89(38.7)	12(7)	17(17)	7(9.7)	125(21.8)
Fluent English Speaker	0	0	0	0	0
Limited English Speaker	141(61.3)	160(93)	83(83)	65(90.3)	449(78.2)
Total	230	172	100	72	574
Missing/Unidentified	7	3	3	3	16

\* Out of a total of 740 participating students, demographic information was acquired for 591 students (79.9%).

In many cases, a comparison of the beginning of year (BOY) with the end of year (EOY) shows an approximate reduction of 35–46% of the number of students tested. That is, the assessment results were not available as these students stopped receiving Spanish instruction, students moved to other schools, or the data could not be collected due to the testing policies. These testing policies limit the number of assessments that the students can take. After students with missing data are removed from the sample, only about one-half of the students who took the examinations at the beginning of the school year also took them at the end of the school year. A comparison of grades also shows a reduction in numbers, by about one-third, of BOY administrations from kindergarten to third grade. While the within-grade and between-grades samples are different from one another, they both reflect diminishment of the number of students over a school year and movement of students from native language instruction to English language instruction.

Educational curriculum design focuses on learning growth, which should be reflected in its measurements of the curriculum. This is clearly the case in the assessment of the BOY-with-EOY correlation coefficients in *Tejas LEE* and *ISIP Español* tests. This learning growth is accounted for by the higher correlations of the tests at the end of the year. As mentioned above, the significant, but modest, correlations shown here are expected for the validation of the *ISIP Español* examinations.

Because of the changes among the students during the school year, the last administration of *ISIP Español* for this study concluded with the following distribution of students: kindergarten 208 (60 eliminated), first grade 154 (141 eliminated), and second grade 75 (92 eliminated), for a total of 437 students at the end of the school year. In addition, *Tejas LEE* administration at the end of the school year was limited to students who had not been classified as “developed.” This made the number of students insufficient (IN) for some comparisons. *Phonological Awareness and Phonetics* subtests administration requires teachers to drop tasks in which the students have already demonstrated mastery. This significantly reduces the number of students who need to be tested at the end of the school year. Nonetheless, the results were adequate to show the highly significant, but modest, correlation between the *Tejas LEE* and *ISIP Español*.

Table 8. *Correlations between Tejas LEE and ISIP Español kindergarten*

	<i>ISIP Español</i>					
	Phonological Awareness & Phonetics		Reading Comprehension		Listening Comprehension	
	BOY	EOY	BOY	EOY	BOY	EOY
<i>Tejas LEE</i>	<i>r</i> ( <i>N</i> )	<i>r</i> ( <i>N</i> )	<i>r</i> ( <i>N</i> )	<i>r</i> ( <i>N</i> )	<i>r</i> ( <i>N</i> )	<i>r</i> ( <i>N</i> )
GPC Total	.457** (303)	.539** (165)				
PA Total	.413** (290)	IN (23)				
Word List	.351** (290)	.588** (136)				
Comprehension Total	.170* (310)	.417** (196)	.108 (303)	.315** (188)	.360** (303)	.306** (194)

\*  $p < .01$

\*\*  $p < .001$

IN(*n*): Insufficient number of takers to present meaningful result

Table 9. Correlations between Tejas LEE and ISIP Español first grade

	ISIP Español					
	Phonological Awareness & Phonetics		Vocabulary		Reading Comprehension	
	BOY	EOY	BOY	EOY	BOY	EOY
<i>Tejas LEE</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>
Comprehension Total	.362** (230)	.497** (154)	.403** (298)	.392** (149)	.230** (297)	.293** (137)
PA Total	.457** (297)	IN (3)	.233** (229)	IN (3)		
Word List	.576** (297)	.771** (45)	.414** (298)	IN (42)		

\*  $p < .01$

\*\*  $p < .001$

IN(n): Insufficient number of takers to present meaningful result

Table 10. Correlations Tejas LEE and ISIP Español second grade

	ISIP Español					
	Fluency		Vocabulary		Reading Comprehension	
	BOY	EOY	BOY	EOY	BOY	EOY
<i>Tejas LEE</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>	<i>r (N)</i>
Comprehension Total	.205 (127)	.350* (66)				
Reading Fluency	.420** (93)	.687** (64)			.339** (97)	.390** (64)
Word List	.203 (127)	.398* (56)	.538** (134)	.519** (57)		

\*  $p < .01$

\*\*  $p < .001$

Table 11. Correlations TAKS and ISIP Español third grade

TAKS	ISIP Español		
	Vocabulary	Reading Comprehension	Fluency
RD RAW	.357* (42)	.476 ** (40)	.416** (39)
RDSCAL	.413 ** (42)	.497** (40)	.447** (39)

\*  $p < .05$

\*\*  $p < .01$