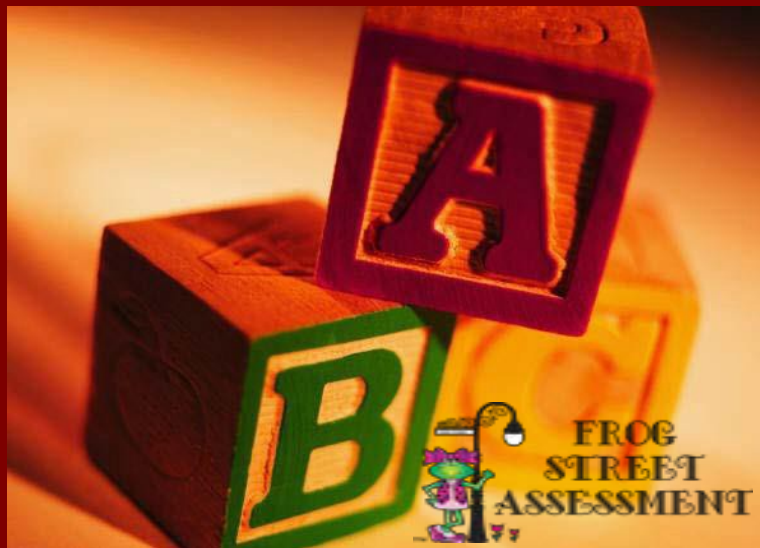


ANALYSIS OF FROG STREET ASSESSMENT (ENGLISH/SPANISH) FOR PRE-KINDERGARTEN



Prepared by
Center for Research & Development in Dual
Language and Literacy Acquisition (CRDLLA)
Department of Educational Psychology and
Educational Leadership Research Center (ELRC)
College of Education and Human Development
Texas A&M University
College Station, TX

2015

RESEARCH TEAM

- Dr. Rafael Lara-Alecio, Professor and Director of Bilingual Education Programs and the Center for Research and Development in Dual Language and Literacy Acquisition (CRDLLA), Department of Educational Psychology, Texas A&M University, College Station, TX
- Dr. Fuhui Tong, Associate Professor and Associate Director of the Center for Research and Development in Dual Language and Literacy Acquisition and Coordinator of the Bilingual Education Programs Data Center, Department of Educational Psychology, Texas A&M University, College Station, TX
- Dr. Beverly J. Irby, Professor, Director of Educational Leadership Research Center (ELRC), and Associate Department Head, Department of Educational Administration and Human Resource, and Co-Director of CRDLLA, College of Education, Texas A&M University, College Station, TX



TABLE OF CONTENTS

Research Team	2
----------------------	----------

Introduction	4
---------------------	----------

Review of Sub-Assessments	8
----------------------------------	----------

Pre-Kindergarten: English and Spanish

Conclusions and Recommendations	26
--	-----------



Introduction

Frog Street Assessment



The Frog Street Assessment (FSA) is a criterion-referenced assessment which has as its purpose to assess children individually at the pre-kindergarten (Pre-K) level. FSA, created in 2010, is based on the curriculum from Frog Street Pre-K, a comprehensive, research-based program that integrates instruction across developmental domains and early learning disciplines. Though the FSA also is considered a curriculum-based measure, it can be used independently of the Frog Street Pre-K. It is fully automated with a web-based application. FSA includes 30 Pre-K Skill Domains with a Beginning- of-Year (**BOY**), Middle-of-Year (**MOY**), and End-of-Year (**EOY**) assessment for most skills in English and in Spanish.

Reliability Analysis

The Research Team analyzed PreK 500 English and 500 Spanish student records per language of test. These 1000 student records were randomly selected from the FSA database. Each sub-assessment domain question by the time of administration (i.e., BOY, MOY, and EOY) was analyzed for reliability using Cronbach's alpha. For each of the domain's subtests, answers to each item is a binary coding, with 1 indicating correct and 0 indicating incorrect. Because items in most subtests are slightly different among BOY, MOY, and EOY, Cronbach's alpha is computed based on the data that were collected at each time point to examine the level of internal consistency. An average Cronbach's alpha was also calculated across the time of administration for each subtest. Additionally, t-tests, along with effect sizes, were conducted on subtests that were the same in the initial assessment and the end of year assessment to determine growth.

Benefits of the Reliability Analysis of the FSA

Reliability is an important characteristic of educational measurements. Reliability refers to the idea that test scores are free from non-systematic, or random, measurement errors, and is a critical technical characteristic of any measurement instrument (Gall, Gall, & Borg, 2007; Messick, 1995), because unreliable instruments cannot be interpreted in a valid way. Evidence of score reliability is normally supported by evidence of internal consistency, a measure of the consistency with which students respond to the items within a test. Cronbach's alpha is the most commonly used statistical index for internal consistency. Reliability estimates

for FSA are mainly obtained through analyses of internal consistency.

What is considered an acceptable level of reliability depends on the research context and how assessment results are used. In the early stages of test development, Nunnally (1978) recommended measures to have a reliability coefficient of .70 or above. For high-stakes tests, it is recommended to have a reliability of above .80 or .90. Considering the exploratory nature of this study, the research team determines that reliability coefficients from 0.70 to 0.79 are acceptable, 0.80 to 0.89 are adequate, and above 0.90 are excellent.

.)



Analysis of Sub-Assessments



Section 1. Reliability Analysis: English and Spanish

Language and Literacy Domain: 11 Pre-K Skill Domains

Domain 1

Following Directions: English

The subtest of Following Directions in English given to Head Start children has an average Cronbach's alpha of .807, with all three coefficients at BOY, MOY, and EOY suggesting adequate internal consistency among these items.

Sequir Instruccionnes: Spanish

The subtest of Sequir Instruccionnes in Spanish has an average Cronbach's alpha of .75, with all three coefficients at BOY, MOY and EOY suggesting an acceptable level of internal consistency among these items.

Domain 2

Vocabulary: English

For this subtest of Vocabulary (expressive), the average Cronbach's alpha is .884. The BOY vocabulary words are one-syllable, familiar words; the MOY vocabulary words are one- and



two-syllable words, and the EOY vocabulary words are two- and three-syllable, more complex words. As a result, the reliability coefficients slightly decrease from BOY (.891) to EOY (.857). However, these values are still considered to be adequate regarding internal consistency among the pictures.

Vocabulario: Spanish

The subtest of Vocabulario in Spanish has an average Cronbach's alpha of .91, with all three coefficients obtained at BOY, MOY and EOY at or above .885, suggesting an overall excellent level of internal consistency among these items/pictures.

Domain 3

Sentence Structure: English

The subtest of Sentence Structure in English has an average Cronbach's alpha of .768, with all three coefficients obtained at BOY, MOY and EOY at or above .759, suggesting an acceptable level of item consistency.

Estructura de un Enunciado: Spanish

Similarly to its English version, the subtest of Estructura in Spanish has an average Cronbach's alpha of .723, suggesting an acceptable level of item consistency.

Domain 4

Uppercase Letters: English

The subtest of Uppercase Letters in English has an average Cronbach's alpha of .972, with all three coefficients obtained at BOY, MOY, and EOY at or above .969, suggesting an excellent level of item consistency.



Letras Mayúsculas: Spanish

The subtest of Letras Mayúsculas in Spanish has an average Cronbach's alpha of .965, with all three coefficients obtained at BOY, MOY, and EOY at or above .964, suggesting an excellent level of item consistency.

Domain 5**Lowercase Letters: English**

The reliability coefficients in forms of Cronbach's alpha in the subtest of Lowercase Letters in English range between .968 and .970 for BOY, MOY, and EOY, indicating an excellent level of item consistency.

Letras Minúsculas: Spanish

The reliability coefficients in forms of Cronbach's alpha in the subtest of Letras Minúsculas in Spanish range between .962 and .966 for BOY, MOY, and EOY, indicating an excellent level of item consistency.

Domain 6**Syllables and Sounds: English**

In this subtest, reliability coefficients in forms of Cronbach's alpha somewhat vary among the three administrations. The value is highest in the BOY with .946, followed by .870 in the MOY, and the lowest in the EOY that falls below .700. The coefficients range between excellent to acceptable.

Sylabas y Sonidos: Spanish

Similar to the English version, the subtest of Sylabas and Sonidos in Spanish has reliability

coefficients ranging from .552 to .928, with the value being the highest in the BOY administration, and lowest in the EOY administration. Overall, this would still be an acceptable level of internal consistency.

Domain 7

Rhyming: English

The subtest of Rhyming in English has an average reliability coefficient of .794. This is an adequate level of internal consistency.

Rima: Spanish

Similar to the English version, the subtest of Rima in Spanish has an average reliability coefficient of .807, with the coefficient at EOY slightly below .80 (.782). This is an adequate level of internal consistency.

Domain 8

Alliteration: English

In this subtest, reliability coefficients range between .75 and .854, suggesting an overall coefficient of .802, suggesting an adequate level of item consistency.

Aliteración: Spanish

In this subtest, reliability coefficients range between .788 and .823, suggesting an adequate level of item consistency.

Domain 9

Letter Sounds: English

The reliability coefficients for Letter Sounds in English have an average of .939, suggesting an excellent level of item consistency.

Sonidos de las letras: Spanish

For the test, Sonidos de las letras, in Spanish, the reliability coefficients are in an excellent range with an average of .938.

Domain 10

Name Writing: English

Because this subtest (teacher rating) is an observation of the student's performance in class or from a portfolio of work samples that documents the students' ability to write his/her name, the child does not have to be present for the teacher to rate. As a result, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Escritura del nombre: Spanish

For the same reason stated above in the English test of Name Writing, no values are reported in the Spanish test either. We recommend the curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Domain 11**Writing: English**

No values with regards to internal consistency are reported here because of the same reason stated above—inter-rater/observer reliability may be more appropriate. We recommend the curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Escritura: Spanish

No values with regards to internal consistency are reported here because of the same reason stated above—inter-rater/observer reliability may be more appropriate. We recommend the curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Mathematics Domain: 12 Pre-K Skill Domains

Domain 1

Rote Counting: English

In this subtest, reliability coefficients range between .735 and .833, suggesting an overall adequate level of item consistency.

Contando en voz alta: Spanish

In this subtest, reliability coefficients in terms of Cronbach's alpha are .826 (BOY), .831 (MOY), and .824 (EOY), suggesting an adequate level of item consistency.

Domain 2

Numeral Recognition: English

In this subtest, reliability coefficients range between .920 and .922, suggesting an excellent level of item consistency.

Reconocimiento de números: Spanish

In this subtest, reliability coefficients range between .898 and .920, suggesting an excellent level of item consistency.

Domain 3

Subitize: English

In the subtest of Subitize in English, reliability coefficients range between .676 and .792, suggesting an overall acceptable level of item consistency.

Calculo mental: Spanish

In this subtest, reliability coefficients range between .627 and .764, with average of .700, suggesting an acceptable level of item consistency.

Domain 4**1-to-1 Counting: English**

In this subtest, reliability coefficients range between .894 and .915, suggesting an excellent level of item consistency.

Enumerando: Spanish

In this subtest, reliability coefficients range between .868 and .894, suggesting an adequate level of item consistency.

Domain 5**Cardinality: English**

In this subtest, reliability coefficients range between .923 and .94, suggesting an excellent level of item consistency.

Números cardinal: Spanish

In this subtest, reliability coefficients range between .907 and .929, suggesting an excellent level of item consistency.

Domain 6**Positional Words: English**

In this subtest of Positional Words, reliability coefficients range between .750 and .785, suggesting an acceptable level of item consistency.



Palabras de posición: Spanish

In this subtest, reliability coefficients range between .809 and .834, suggesting an adequate level of item consistency.

Domain 7**Sorting: English**

In this subset of sorting in English, reliability coefficients range between .745 and .878, suggesting an overall adequate level of item consistency.

Clasificar: Spanish

Similar to the English counterpart, the subtest of Clasificar in Spanish has an average reliability coefficient of .770, suggesting an overall acceptable level of item consistency.

Domain 8**Shapes: English**

In this subset of Shapes in English, reliability coefficients range between .750 and .815, with the average of .777, suggesting an overall adequate level of item consistency.

Figuras: Spanish

In this subset of Figuras in Spanish, reliability coefficients range between .753 and .812, with the average of .776, suggesting an overall adequate level of item consistency.

Domain 9**Pattern: English**

In this subset of Pattern in English, reliability coefficients have an overall average of .911, suggesting an excellent level of item consistency.



Patrones: Spanish

In this subset of Patrones in Spanish, reliability coefficients range between .851 and .919, with the average of .877, suggesting an overall excellent level of item consistency.

Domain 10**Ordinal Words: English**

The average reliability coefficients for Ordinal Words in English is .738, suggesting an acceptable level of item consistency.

Palabras ordinales: Spanish

The average reliability coefficients for Palabras ordinales in Spanish is .715, suggesting an overall acceptable level of item consistency.

Domain 11**Measurement: English**

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Medidas: Spanish

For the same reason stated above in the English subtest, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the

same student to complete this section.

Domain 12

Operations: English

For the subtest of Operations in English, the average reliability coefficient in the form of Cronbach's alpha is .838, suggesting an adequate level of item consistency.

Operaciones: Spanish

Similarly, for the subtest of Operaciones in Spanish for Head Start children, the average reliability coefficient in the form of Cronbach's alpha is .823, suggesting an adequate level of item consistency.

Generic Domain: 8 Pre-K Skill Domains

Domain 1

Social-Emotional: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Emocional social: Spanish

For the same reason stated above, no values are reported.

Domain 2

Gross Motor: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Motrices Gruesas: Spanish

For the same reason stated in the English subtest, no values are reported in the Spanish subtest of Motrices Gruesas.

Domain 3

Fine Motor: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Motrices Finas: Spanish

For the same reason stated in the English subtest, no values are reported in the Spanish subtest of Motrices Finas.

Domain 4

Science: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Ciencias Naturales: Spanish

For the same reason stated in the English subtest, no values are reported in the Spanish subtest of Ciencias Naturales.

Domain 5

Social Studies: English

In this subset of Social Studies in English, reliability coefficients range between .809 and .864, with the average of .829, suggesting an adequate level of item consistency.

Estudios Sociales: Spanish

In this subset of Estudios Sociales in Spanish, reliability coefficients average is .828, suggesting an adequate level of item consistency.

Domain 6

Creative Arts: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer



reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Artes Creativas: Spanish

For the same reason stated in the English subtest, no values are reported in the Spanish subtest of Artes Creativas.

Domain 7

Health Knowledge: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.

Domain 8

Approaches to Learning: English

Because this subtest is answered by a teacher based on their observation/judgment of the students, the test reliability may be appropriately assessed through inter-rater/observer reliability, instead of item consistency. Hence, no values are reported. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section.



Formas de Aprendizaje: Spanish

For the same reason stated in the English subtest, no values are reported in the Spanish subtest of Formas de Aprendizaje.

Section 2. Performance Analyses: Specific Subtests

Lowercase Letters: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.23. These results indicate that students significantly improved their performance during the academic year.

Letras Minúsculas: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 2.1. These results indicate that students significantly improved their performance during the academic year.

Uppercase Letters: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.1. These results indicate that students significantly improved their performance during the academic year.

Letras Mayúsculas: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.9. These results indicate that students significantly improved their performance during



the academic year.

Alliteration: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 0.6. These results indicate that students significantly improved their performance during the academic year.

Aliteración: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 2.6. These results indicate that students significantly improved their performance during the academic year.

Numeral Recognition: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 0.94. These results indicate that students significantly improved their performance during the academic year.

Reconocimiento de números: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.6. These results indicate that students significantly improved their performance during the academic year.



Rhyming: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 0.37. These results indicate that students significantly improved their performance during the academic year.

Rima: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 2.3. These results indicate that students significantly improved their performance during the academic year.

Syllables and Sounds: English

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.11. These results indicate that students significantly improved their performance during the academic year.

Sylabas y Sonidos: Spanish

Based on a dependent samples t-test from 500 student records, there was a statistically significant difference ($p < .001$) between BOY and EOY assessments with an effect size (Cohen's d) of 1.1. These results indicate that students significantly improved their performance during the academic year.



Conclusions and Recommendations



The Texas A&M University Research Team from the Center for Research and Development in Dual Language and Literacy Acquisition (CRLDLLA) and the Educational Leadership Research Center (ELRC) concludes that the Frog Street Assessment provides reliable information regarding the assessment in both English and Spanish measures on all domains from which data were collected and reported from a sample of 500 students attending Pre-K programs. These students have demonstrated a significant improvement from BOY to EOY on the six measures (Lowercase letters, Uppercase Letters, Alliteration, Rhyming, Numeral Recognition, and Syllables and Sounds) that were examined in both English and Spanish test. This is particularly evident with the Spanish measures with large effect sizes. Therefore, the TAMU CRDLLA and ELRC Team concludes that the FSA could be used to aid in benchmarking students as they move through the curriculum and could serve as a positive instrument for teachers as they reflect on their work in order to improve their instruction based on data. The research team also concluded through the use of the Frog Street Pre-K curriculum with fidelity, children are likely to make gains in skills are, according to the literature, related to success in kindergarten. In many of the subtests (e.g., Social-Emotional, Gross/Fine Motor, Science, etc.) in both English and Spanish in the Generic Domain, no values in terms of internal consistency

are reported because these subtests require teacher rating based on their observation/judgment of the students; therefore, the test reliability may be more appropriately assessed through inter-rater/

observer reliability, instead of item consistency. We recommend the FSA curriculum and assessment team consider multiple teacher/rater on the same student to complete this section so as to gather data to calculate inter-rater reliability for these subtests. We also recommend further appropriate analyses for a variety of types of reliability and validity.

References

- Gall, M., Gall, J., & Borg, W. (2007). *Educational research: An introduction*. Boston, MA: Pearson.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.
- Nunnally, J. C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.