

**Informing the Performance-Based  
Contract Between First 5 LA and  
LAUP: Assessing Child Progress**

Spring Report

December 10, 2010

Yange Xue  
Sally Atkins-Burnett  
Pia Caronongan  
Emily Moiduddin



**MATHEMATICA**  
Policy Research, Inc.

**This page has been left blank for double-sided copying.**

Contract Number:  
07110

Mathematica Reference Number:  
06791.302

Submitted to:  
First 5 LA  
Suite 300, 750 N. Alameda St.  
Los Angeles, CA 90012  
Project Officer: Sharon Murphy and  
Artineh Samkian

Submitted by:  
Mathematica Policy Research  
600 Maryland Avenue, SW  
Suite 550  
Washington, DC 20024-2512  
Telephone: (202) 484-9220  
Facsimile: (202) 863-1763  
Project Director: Emily Moiduddin

**Informing the Performance-Based  
Contract Between First 5 LA and  
LAUP: Assessing Child Progress**

Spring Report

December 10, 2010

Yange Xue  
Sally Atkins-Burnett  
Pia Caronongan  
Emily Moiduddin

**MATHEMATICA**  
Policy Research, Inc.

**This page has been left blank for double-sided copying.**

## **ACKNOWLEDGMENTS**

The authors of this report have many people to thank for making this study possible. Our previous project officer, Katie Fallin at First 5 LA, worked closely with us at every step along the way. Her colleague Christine Ong was equally involved, insightful, and supportive throughout the study. Staff at the Los Angeles Universal Preschool (LAUP) program kept us focused on what matters for children, families, and their programs and contributed tremendously to the substance of the project, so we are indebted to Kimberly Hall, Julia Love, Schellee Rocher, Delila Vasquez, and the LAUP program coaching staff. We would also like to thank LAUP's chief executive officer (CEO), Celia Ayala. As a special research consultant to First 5 LA, Michael López of the National Center for Latino Child & Family Research provided thoughtful input and constructive critiques.

As project director of UPCOS-3, John Love provided the guidance and wisdom necessary for completing a study of such magnitude. Susan Sprachman, as survey director, coordinated all aspects of the project related to data collection, working with Elisha Smith as deputy survey director. They were assisted valiantly in their efforts by Anne Self, Richard Godwin, and David Eden. All their efforts were facilitated by the LAUP program directors who provided tremendous support. The LAUP teachers gave generously of their class time, which we know was not always convenient, allowing us to assess the children. Especially central to the success of the study was the participation of the LAUP children. We are deeply indebted to them for their time and cooperation. The children worked hard to complete the assessments.

Enrolling programs into the study, arranging the schedules for data collection, and working with the programs to establish the eligible sample and obtain parental consent are critical tasks on a study like UPCOS; we are grateful to the Mathematica Policy Research, Juárez & Associates, and AIR staff who coordinated this process.

Lynne Beres and Dorothy Bellows formatted the numerous forms and training materials. Susan Golden oversaw all data entry and data quality control activities. Our programming team was led by Scott Reid and included Neil DeLeon and Susan Shillaber. The analysis team consisted of programmers Judy Cannon, Lauren Bernstein, and Kathleen Kohl.

Finally, we give special thanks to the editing team led by Amanda Bernhardt and to Alfreda Holmes for the tremendous effort that went into producing the report.

**This page has been left blank for double-sided copying.**

## CONTENTS

	EXECUTIVE SUMMARY .....	ix
<b>I</b>	OVERVIEW AND METHODS.....	1
	The Child Outcomes Proxy Measure: the Provider Administered Assessment (PAA) .....	1
	Standardized Assessment Battery: The Full UPCOS Battery .....	3
	Sample .....	3
	Procedures .....	6
	Scoring .....	7
	Approach to Analysis .....	8
<b>II</b>	RESULTS OF STANDARDIZED ASSESSMENTS .....	11
	Description of Measures and Scoring Methods .....	11
	Independently Administered Assessments .....	11
	Teacher-Administered Assessments .....	13
	Scoring Procedures .....	13
	Children’s Progress from Fall 2009 to Spring 2010 .....	13
	Overall Sample .....	13
	Children’s Progress by Subgroups .....	15
	Language Subgroups .....	15
	Program Subgroups: Center-Based and Family Child Care Settings .....	18
	Program Subgroups: High and Low Concentrations of ELLs .....	18
	Summary .....	19
<b>III</b>	CHILDREN’S PROGRESS AS MEASURED BY THE PRE-KINDERGARTEN OBSERVATION FORM.....	21
	Children’s Progress from Fall 2009 to Spring 2010 for the Overall Sample ...	22
	Children’s Progress by Subgroup .....	23
	Language Subgroups .....	23
	Program Subgroups: Center-Based and Family Child Care Settings .....	23
	Program Subgroups: High and Low Concentrations of ELLs .....	24

III (continued)

Inter-Factor Correlations for Spring Data ..... 24

    Inter-Factor Correlations in the Overall Sample ..... 24

    Inter-Factor Correlations by Subgroup..... 25

Evidence of Validity ..... 25

    Concurrent Validity for the Overall Sample..... 25

    Relationships Between Change Measured by the P-KOF and by Standardized Measures..... 26

    Predictive Validity for the Overall Sample ..... 26

    Concurrent Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC) ..... 27

    Relationships Between Change Measured by the P-KOF and by Standardized Measures for Subgroups Based on Program Type (Center-Based and FCC) ..... 27

    Predictive Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC) ..... 28

    Concurrent Validity Estimates for Subgroups Based on Concentration of ELLs ..... 28

    Relationships Between Change Measured by the P-KOF and by Standardized Measures for Subgroups Based on Concentration of ELLs28

    Predictive Validity Estimates for Subgroups Based on Concentration of ELLs ..... 29

Summary ..... 29

IV CHILDREN’S PROGRESS AS MEASURED BY THE PROVIDER-ADMINISTERED ASSESSMENT ..... 31

Children’s Progress from Fall 2009 to Spring 2010 for the Overall Sample ... 32

    ASQ-3 ..... 32

    Rapid Letter Naming..... 33

    ASQ-3 Screening Scores and Cut-Off Scores..... 33

Children’s Progress by Subgroups ..... 34

    Language Subgroups ..... 34

    Program Subgroups: Center-Based and Family Child Care Settings ..... 34

    Program Subgroups: High and Low Concentrations of ELLs..... 35

Inter-Factor Correlations for Spring Data ..... 35

    Inter-Factor Correlations in the Overall Sample ..... 35

    Inter-Factor Correlations by Subgroup..... 36



IV (continued)

Evidence of Validity .....	36
Concurrent Validity for the Overall Sample.....	36
Relationships Between Change Measured by the PAA and by the Standardized Measures.....	37
Predictive Validity for the Overall Sample .....	37
Concurrent Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC) .....	38
Relationships Between Change Measured by the PAA and by Standardized Measures for Subgroups Based on Program Type (Center-Based and FCC) .....	38
Predictive Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC) .....	39
Concurrent Validity Estimates for Subgroups Based on Concentration of ELLs .....	39
Relationships Between Change measured by the PAA and by Standardized Measures for Subgroups Based on Concentration of ELLs.....	39
Predictive Validity Estimates for Subgroups Based on Concentration of ELLs .....	39
Summary .....	40
<b>V</b> SUMMARY AND CONCLUSIONS.....	41
Findings from Fall 2009 Data.....	41
Fall to Spring Change.....	41
Ceiling Effects.....	42
Validity .....	42
Evidence of Validity for the P-KOF .....	42
Evidence of Validity for the PAA.....	43
Measurement Concerns .....	44
Limitation .....	44
Conclusion .....	44
REFERENCES .....	47
APPENDIX A: ANALYSIS OF RATER EFFECTS BASED ON FALL 2009 ASSESSMENTS .....	A.1
APPENDIX B: TABLES OF RESULTS PRESENTED IN CHAPTER II: STANDARDIZED MEASURES.....	B.1

APPENDIX C: TABLES OF RESULTS PRESENTED IN CHAPTER III: PRE-KINDERGARTEN OBSERVATION FORM ..... C.1

APPENDIX D: TABLES OF RESULTS PRESENTED IN CHAPTER IV: PROVIDER ADMINISTERED ASSESSMENT ..... D.1

**TABLES**

**I.1** Standardized Measures of the Full UPCOS Battery, Their Purpose, and Scales of the Provider Administered Assessment and P-KOF Expected to Correlate with Each ..... 4

**I.2** Percentage of Children with Various Characteristics in the UPCOS-3 Sample, Fall 2009 and Spring 2010 ..... 5

**This page has been left blank for double-sided copying.**

## EXECUTIVE SUMMARY

In 2009-2010, as part of Phase 3 of the Universal Preschool Child Outcomes Study (UPCOS-3), Mathematica Policy Research administered a full battery of direct child assessments to inform the performance-based contract between First 5 LA and Los Angeles Universal Preschool (LAUP). In addition, First 5 LA asked Mathematica to explore the feasibility of using a brief assessment that both center-based teachers and family child care (FCC) providers could easily administer and that could serve as a proxy for the full battery. The goal was to learn whether providers could administer a brief set of measures combining ratings and direct assessments that would yield data that approximate the scores obtained through more intensive direct assessments. This study described children's growth from fall to spring as captured by the different measures and examined the reliability and validity of the teacher-administered measures.

The full battery of standardized measures that Mathematica administered included the English and Spanish PreLAS (Duncan and DeAvila 2002); Expressive One-Word Picture Vocabulary Test (EOWPVT; English version); Expressive One-Word Picture Vocabulary Test–Spanish Bilingual Edition (EOWPVT-SBE; Brownell 2000); the Woodcock-Johnson III (WJ-III; Mather and Woodcock 2001) and Woodcock-Muñoz Bateria III (WM-III; Woodcock and Muñoz-Sandoval 2005) Letter-Word Identification, Applied Problems, and Spelling tests; Preschool Kindergarten Behavior Scales–2 (PKBS-2; Merrell 2003); Pencil Tapping Task (Smith-Donald et al. 2007); and the Leiter Examiner Ratings (Roid 1997).

### Teacher-Administered Measures

The child outcomes proxy measures (which we refer to as the Provider-Administered Assessment, or PAA) are a set of relatively straightforward and brief measures that LAUP teachers can administer in future years in connection with LAUP's performance-based contract. The set of proxy measures is to be used for the purpose of monitoring children's learning from fall to spring of their LAUP year across school readiness dimensions, specifically social-emotional development, language and literacy, English-language development, mathematics, approaches to learning, and physical and motor development. During the 2009-2010 program year, we worked in close collaboration with First 5 LA and LAUP to select and evaluate a set of measures for this purpose. Here we describe the three measures that make up the PAA, and an additional measure that LAUP administered:

**Ages and Stages Questionnaires-3rd Edition** (ASQ-3; Squires et al. 2009) was originally designed as a parent-report monitoring system, but for UPCOS-3 we adapted it for teachers to complete based on their direct observation. The final UPCOS-3 version includes items from three age-specific questionnaires (48-month, 54-month, and 60-month forms) in order to capture a greater range of skills and enable us to assess change across the preschool year. In consultation with the authors of the ASQ-3, Mathematica standardized the prompts for use in this study. The ASQ-3 scales include Communication (English and Spanish), Problem Solving (including cognition and mathematics), Gross Motor, and Fine Motor development.<sup>1</sup>

---

<sup>1</sup> Because the Personal-Social scale on the ASQ-3 is more a measure of adaptive behavior than of social-emotional development, we excluded it from the PAA. Instead, we used the Preschool and Kindergarten Behavior Scales–2 (PKBS-2) to assess social-emotional development.

**Rapid Letter Naming (RLN)** was developed for use in UPCOS-2 and addresses the literacy domain. It is a conceptually scored measure of the child's ability to easily name uppercase and lowercase alphabet letters. The child receives credit for correct answers in both English and other languages (in particular, Spanish).<sup>2</sup>

**Preschool and Kindergarten Behavior Scales-2 (PKBS-2; Merrell 2003)** is a standardized assessment of social and emotional development that examines both positive and negative behaviors. Scales include Social Cooperation, Social Independence, Positive Social Interactions, Externalizing<sup>3</sup> and Internalizing<sup>4</sup> Problem Behaviors, Total Positive Social Skills, and Total Problem Behaviors. Teachers rate the frequency with which children exhibit behaviors using a four-point frequency scale ranging from never to often.

In addition to the three measures that constitute the PAA, we analyzed data from the Preschool-Kindergarten Observation Form (P-KOF), another teacher-administered measure, which LAUP had independently selected to use in 2009-2010. Its features are described here:

**Preschool-Kindergarten Observation Form (P-KOF; Applied Survey Research 2009)** is an observational measure that LAUP selected for use in all programs beginning in the 2009-2010 program year. The P-KOF consists of 24 items assessing children's school readiness skills. Teachers complete ratings on the 24 items based on their observations of each sample child during daily classroom activities. The item ratings have differing levels of specificity—some require reporting on the number of colors a child can name correctly while others require the teacher to rate whether a child has expressive abilities. Item ratings are combined to yield an overall score as well as scores on the following subscales: Self-Care & Motor Skills, Self-Regulation, Social Expression, and General Knowledge.

Once the final set of measures to be tested was selected, teachers in a sample of LAUP center-based and FCC programs administered the assessments for a subset of the children in their classes in both the fall and the spring. At each time point, Mathematica's independent assessors also conducted a full battery of direct, standardized child assessments for the purpose of testing the validity of the teacher-administered measures (that is, characteristics of the independently administered standardized assessments were to serve as a point of comparison for the teacher-administered measures).

## Sample

The UPCOS-3 study included a sample of programs in LAUP that were selected randomly from four groups: (1) center-based programs with a high concentration of English language learners (ELLs) (greater than 50 percent), (2) center-based programs with a low proportion of ELLs, (3) large FCC programs (with capacity of nine or more), and (4) small FCC programs. Within each center or FCC, we recruited all children. The study included the first 10 children in the selected groups from centers, and the first 4 from FCCs, to return parental permission slips (N=875; 656 in centers and 219 in FCCs). Thus, although the teachers were representative of LAUP classrooms in center-based

---

<sup>2</sup> To give a child credit for a response in the RLN task in a language other than English, teachers must be familiar with the letter names being used. Not all other languages use the English alphabet.

<sup>3</sup> These include acting out behaviors such as fighting, arguing, or acting impulsively.

<sup>4</sup> These are behaviors that suggest that the child is sad, lonely, or anxious.

and FCC programs, the children were not necessarily representative of the children in LAUP. It is likely that those who were the first to return their permission slips differ from children who returned the slips later.

## Children's Progress from Fall 2009 to Spring 2010

Overall, children made progress in all areas of development on the standardized measures in the full UPCOS-3 battery that Mathematica-trained assessors administered. Greater progress was noted on assessments administered in English than those administered in Spanish. The magnitude of change was greater for raw scores and continuous growth scores (such as the W-scores) than for standard scores that measured growth relative to same-age peers. Effect sizes (ES)<sup>5</sup> greater than .50 were noted for scores capturing absolute change on most standardized assessments in English. When looking at change relative to same age peers, the effect sizes were much smaller ( $ES \leq .21$ ). Effect sizes were also smaller for the sample of children who took the Spanish bilingual assessments at both time points (ES for absolute change ranged from .28 to .55). In general, the progress on the standardized measures did not differ across subgroups defined by child language, program type, or concentration of ELL children in the program, with the exception of language measures, where Spanish only and primarily groups made greater progress than English only and English primarily groups.

Children also made significant progress according to the teacher-administered PAA and P-KOF. Effects sizes greater than .60 were noted for the English assessment on the PAA and the P-KOF. Other than the Problem-Solving scale of the ASQ-3, the magnitude of change on the PAA was comparable to the progress captured by the IRT scores on the independently administered standardized measures.<sup>6</sup> We observed some subgroup differences in the progress on the PAA that were not evident on the independently administered standardized assessments: children in centers gained more than did children in FCCs on English Communication; Spanish only children and children in high ELL concentration programs made greater progress on English Communication, Problem Solving, and Rapid Letter Naming.

The changes observed on the P-KOF scales were larger than those observed on the standardized measures. Part of the reason for this difference is the narrower distribution of fall P-KOF scores. In addition, children who had lower scores in the fall, that is, children in the Spanish only language group, in center-based programs, and in programs with high concentrations of ELLs, had larger gains on the P-KOF. However, ceiling effects on the P-KOF scales may have led to an underestimation of the progress made by children who were already scoring near the maximum in the fall. As a result, the P-KOF may also overestimate between-group differences in change scores.

---

<sup>5</sup> We calculated effect sizes for growth in the standardized measures and teacher-administered measures as the difference between the spring and fall scores divided by the fall standard deviation. Thus, the magnitude of the effect is directly related to the variation in scores (how spread out children's scores were) in the fall.

<sup>6</sup> Several items were added to the ASQ-3 Problem Solving scale in the spring due to concerns about the low ceiling. Some children may have been able to answer the additional items in the fall but were not given the opportunity so the change scores did not accurately reflect the fall to spring difference on that scale.

## Evidence of Validity

### Validity of the P-KOF

The P-KOF had strong internal consistency and weak to moderate evidence of concurrent validity in the fall and spring, with similar correlations in the fall and spring between the independently administered standardized assessments and the P-KOF. The General Knowledge scale was the section most strongly associated with standardized measures of language and cognition in the spring ( $r = .43$  to  $.53$  with measures in English and the EOWPVT-SBE). The P-KOF General Knowledge scores in the fall were also most strongly related to spring scores on the WJ-III measures ( $r = .42$  to  $.54$ ). Most of the other correlations were low. Further, change on the P-KOF was not strongly associated with change on any of the standardized measures.

The analyses identified several weaknesses of the P-KOF:

- In the fall we found indications of rater effects: much higher intraclass correlations (ICCs)<sup>7</sup> for the P-KOF than for the standardized assessments. The ICCs for the P-KOF subscales were as high as  $.39$  to  $.56$ , while the independently administered assessments had ICCs of  $.01$  to  $.30$ ,<sup>8</sup> suggesting that as much as half of the variance in P-KOF scores may be attributable to the way in which the teacher uses the rating scales rather than to the children's skills and abilities.
- The P-KOF showed subgroup differences in fall to spring change that were not seen for the independently administered assessments, adding to evidence of potential rater effects.
- Unexpectedly, we found moderate correlations between the P-KOF General Knowledge scale and the standardized PKBS-2 (which taps different constructs). These correlations were only slightly lower than correlations with standardized measures in language, literacy, and math. This may be due to shared method variance—both measures were based on teacher ratings by the same teacher.
- Although there was some indication of ceiling effects in the fall, by the spring strong ceiling effects (that is, a large proportion of children with the maximum possible score) were found, with 23 percent (General Knowledge) to 78 percent (Self-Care & Motor) of the children receiving the maximum scores.<sup>9</sup> Ceiling effects limit the ability to detect change because children who already scored at the maximum have no room for change.

---

<sup>7</sup> The ICC is the proportion of variance that is attributable to teachers/classrooms. This value gives an indication of the degree to which scores within the same group of children as designated by their teacher or classroom are similar to each other. High ICCs can occur because of how children are grouped within classrooms (design effects) or to differences in how teachers administer the measures (rater effects). The ICCs for the standardized assessments (where independent assessors do not know the children) should identify the between-classroom differences in children's development in the various domains that are attributable to how children are grouped in classrooms (design effects).

<sup>8</sup> The ICCs at the high end of this range for the independently administered assessments were on the language related measures (English PreLAS) and are likely a function of the design. We sampled some programs with high proportions of ELLs and some with low proportions of ELLs.

<sup>9</sup> On the other scales, 38 percent (Self-Regulation) and 53 percent (Social Expression) of children were at the maximum score in the spring.



Ceiling effects can decrease the variance in the scores and that can limit the ability to find relationships with other variables.

## Validity of the PAA

The PAA had adequate to strong reliability, and the evidence of concurrent validity for the cognitive and language portions was moderate to strong in both the fall and the spring ( $r = .45$  to  $.76$  with independently administered standardized assessments in similar domains of development). The strongest correlations were between the most similar measures. For the WJ-III Letter-Word Identification and Rapid Letter Naming, the fall correlation was  $.76$ . Surprisingly, the magnitude of the relationships for similar measures administered in English was stronger in the fall than in the spring. The evidence of predictive validity was also moderate to strong ( $r = .50$  to  $.66$ ) for the cognitive and language domains. The change on Spanish Communication was moderately correlated with change on the EOWPVT and WM-III Letter-Word Identification. The change on Rapid Letter Naming was moderately correlated with change on WJ-III Letter-Word Identification. However, the correlations were in the low range for change scores on the ASQ-3 English Communication and Problem Solving with change on the standardized measures in the same domains.

We found a ceiling effect only for the ASQ-3 Gross Motor subscale, on which 50 percent of children scored at the maximum in the spring. The other subscales had skewed distributions but minimal ceiling problems, with only 3 to 11 percent of children scoring at the ceiling.

The measurement of Spanish speakers is problematic. Overall, we found more-limited relationships between the teacher-administered measures (both the P-KOF and the PAA) and the standardized Spanish measures. It is possible that this is due to the limited Spanish language proficiency of some teachers. However, the required administration procedures on some of the standardized measures may also have contributed to lower correlations. The independently administered standardized measures of cognition required either Spanish or English administration.<sup>10</sup> This was particularly problematic for the WJ-III/WM-III Letter-Word Identification test. In many classrooms, children were learning letter names in English, but the test required them to provide the letter names in Spanish if the children were being tested in Spanish. A much lower correlation was noted between the conceptually scored RLN and the WM-III Letter-Word ( $r = .28$  to  $.29$ ) than with the WJ-III Letter-Word ( $r = .59$  to  $.62$ ). Thus, it is not clear whether some of the weaker relationships are due more to the inflexibility in the language of assessment on the standardized measures or to differences in teachers' Spanish fluency.

## Conclusion

Validity is not inherent in a measure, but in its use. The purpose of this study was to investigate the evidence of validity for measures completed by preschool teachers for use in an accountability framework (within a performance-based contract). There was particular interest on the part of First 5 LA and LAUP in setting targets for change in children's skills and abilities across the program year in linguistically diverse preschool programs, and so our analysis focused on evaluating measures for that potential use.

Our analyses indicate that the P-KOF is not a good candidate for use in setting benchmarks or targets related to children's fall-spring progress: The results indicate weak evidence of convergent

---

<sup>10</sup> Bilingual assessors administered the standardized assessments to children who spoke Spanish.

validity in the spring, ceiling problems, limited variance in the spring, potential rater effects (in the fall), and subgroup differences in change scores. We did not find these problems on the independently administered standardized assessments. In addition, the weak correlations of the P-KOF change scores with change scores on the standardized measures suggest problems with concurrent validity.

The PAA has evidence of stronger reliability and validity in both fall and spring than the P-KOF. Change scores were moderately correlated for two of the scales (Spanish Communication and RLN), but we found lower correlations of English Communication and Problem Solving with fall-spring change scores on their related standardized measures. Available evidence of predictive validity was relatively strong for all scales except Spanish Communication. Ceiling effects on the PAA were minimal, although concerns about rater effects persist. With additional training provided to teachers, it is possible that measurement using the PAA would be stronger. Nevertheless, the issues of teachers' language fluency and the language used for assessment remain a concern. In addition, the rater effects of a teacher-administered assessment are likely to increase when assessments are used for accountability purposes.

## I. OVERVIEW AND METHODS

### The Child Outcomes Proxy Measure: the Provider Administered Assessment (PAA)

During the 2009-2010 program year, we worked in close collaboration with First 5 LA and LAUP to select a set of measures to be considered for the purpose of monitoring children's growth across the program year. We began by conducting an extensive review of teacher-administered measures. We also explored how other states and districts are monitoring school readiness. To the extent possible, we looked for assessments with the following characteristics:

- Collectively cover as many domains of children's early development and learning as possible
- Have some evidence of reliability and validity
- Pose minimal burden (less than 15 minutes per child to complete the assessment)
- Can be administered at relatively low cost
- Have been used with Spanish-speaking children
- Are sensitive to change in development over a six-month period
- Show minimal problems with floor and ceiling effects

In addition to these general characteristics, we considered other factors specific to both observational measures and direct assessment methods. For observational measures, the issue of inter-rater reliability is paramount. Observational measures that offer clear behavioral descriptions and provide adequate training materials for teachers are more likely to have stronger inter-rater reliability than rating scales, which can be influenced by differences in how teachers interpret the items. Thus, with observational measures it is more likely that two trained adults observing the same child will give that child the same rating.

For the direct assessments, three additional issues emerged. First, in light of the diversity in LAUP programs, a Spanish version of the assessment is needed. Second, the tasks cannot be curriculum-specific; they must be familiar to diverse groups of children whose teachers or providers take different approaches. Third, for any measure that does not include the full universe of items (for example, all the letters of the alphabet in a letter-naming task), parallel forms should be available so teachers are not tempted to teach only particular items. If a teacher teaches only the sample of items on the assessment, the test result is no longer representative of the child's skills.

Once we completed the review, we worked with First 5 LA and LAUP to select those measures that address key domains of kindergarten readiness and are straightforward for teachers to administer. Based on this collaboration, we included the following measures for studying as part of the PAA.

**Ages and Stages Questionnaires-3rd Edition (ASQ-3;** Squires et al. 2009) (5 minutes per subscale). The ASQ-3 was originally designed as a parent-report monitoring system, but for UPCOS-3 we adapted it to be completed by teachers based on direct observation. The final UPCOS-3 version includes items from three age-specific questionnaires (48-month, 54-month, and

60-month forms) in order to capture a greater range of skills and enable us to assess change across the preschool year. In consultation with the authors of the ASQ-3, Mathematica standardized the prompts for use in this study. The ASQ-3 scales include Communication (English and Spanish), Problem Solving (including cognition and mathematics), Gross Motor, and Fine Motor development.<sup>11</sup>

**Rapid Letter Naming (RLN)** (1 to 2 minutes). This brief instrument was developed for use in UPCOS-2 and addresses the literacy domain. It is a conceptually scored measure of the child's ability to easily name uppercase and lowercase alphabet letters. The child receives credit for correct answers in both English and other languages (in particular, Spanish<sup>12</sup>).

**Preschool and Kindergarten Behavior Scales-2 (PKBS-2; Merrell 2003)** (8-12 minutes). This standardized assessment of social and emotional development examines both positive and negative behaviors. Scales include Social Cooperation, Social Independence, Positive Social Interactions, Externalizing<sup>13</sup> and Internalizing<sup>14</sup> Problem Behaviors, Total Positive Social Skills and Total Problem Behaviors. Teachers rate the frequency with which children exhibit behaviors using a four-point frequency scale.

In addition, because LAUP had independently selected another teacher-administered measure to use in 2009-2010, we analyzed its data as well.

**Preschool-Kindergarten Observation Form (P-KOF; Applied Survey Research 2009)** (8-10 minutes). The P-KOF is an observational measure that LAUP selected for use in all programs beginning in the 2009-2010 program year. The P-KOF consists of the 24 items from an observational rating scale known as the Kindergarten Observation Form (KOF). The KOF was designed as a school readiness assessment to be used by kindergarten teachers after the first five weeks of school. For the P-KOF, teachers make ratings on 24 items based on their observations of the child during daily classroom activities. The item ratings have differing levels of specificity—some depend on the number of colors a child can name correctly while others require the teacher to rate whether a child has expressive abilities. Item ratings are combined to yield an overall score as well as scores on the following subscales: Self-Care & Motor Skills, Self-Regulation, Social Expression, and General Knowledge.

As noted, we refer to the ASQ-3, Rapid Letter Naming, and PKBS-2 as the “Provider Administered Assessment” or PAA. We refer to the combination of the PAA and P-KOF as teacher-administered assessments or measures. Note that in this study, the PKBS-2 served two purposes: it provided a measure of social and emotional development for the PAA and, as a standardized measure, also served as a tool in our validity analysis (discussed later).

---

<sup>11</sup> The Personal-Social scale on the ASQ-3 is more a measure of adaptive behavior than of social-emotional development. Thus, we excluded it from the PAA. Social-emotional development was addressed by the Preschool and Kindergarten Behavior Scales–2 (PKBS-2).

<sup>12</sup> To give a child credit for a response in the RLN task in a language other than English, teachers must be familiar with the letter names being used. Not all other languages use the English alphabet.

<sup>13</sup> These include acting out behaviors such as fighting, arguing, or acting impulsively.

<sup>14</sup> These are behaviors that suggest that the child is sad, lonely, or anxious.

## Standardized Assessment Battery: The Full UPCOS Battery

For the purpose of testing the functioning of the teacher-administered measures, independent assessors from Mathematica administered standardized assessments of language, literacy, mathematics, executive functioning, and social-emotional development (Table I.1). Additional detail regarding each measure is provided in Chapter II. This group of independently administered assessments is referred to as the Full UPCOS Battery. As discussed below, bivariate correlations between the independently administered standardized assessments and the teacher-administered assessments serve as an indicator of the validity of the teacher-administered measures.

### Sample

The UPCOS-3 study includes a sample of programs in LAUP that were selected randomly from four groups: (1) center-based programs with a high concentration of English language learners (ELLs) (greater than 50 percent), (2) center-based programs with a low proportion of ELLs, (3) large FCC programs (with capacity of nine or more), and (4) small FCC programs. Of the 75 center-based programs sampled, 53 agreed to participate in child assessments. Of the 86 FCC programs sampled, 54 agreed to participate. In some centers, more than one group of children was selected as long as each group included a unique lead teacher. Within each center or FCC, we recruited all children. The first 10 children in the selected groups from centers, and the first 4 from FCCs, to return parental permission slips were included in the study. Thus, although the teachers were representative of LAUP classrooms in center-based and FCC programs, the children were not necessarily representative of the children in LAUP. It is likely that those who were the first to return their permission slips differ from children who returned the slips later.

A total of 875 parents gave permission for their children to participate in the study; 829 completed at least some part of the direct child assessment in the fall; 740 did so in the spring. Table I.2 presents basic characteristics of the children in the sample at each time. Across the children who completed at least some portion of the Full UPCOS Battery, about three-quarters were in centers and one-quarter in FCCs at both times, reflecting the larger number of children sampled from each center-based classroom (10) than from each FCC program (4). A slightly lower percentage of children were in programs with a low concentration of ELL children in the spring (56 percent) than in the fall (59 percent).<sup>15</sup> The sample is split roughly evenly between boys (51 percent) and girls (49 percent) in the fall and spring. In terms of language, 40 percent reportedly spoke English only, 23 percent spoke English primarily, 24 percent spoke Spanish primarily, 10 percent spoke Spanish only, and 3 percent spoke a language other than Spanish or English only or primarily. On average, children in the sample were 53.2 months of age in the fall and 58.5 months in the spring. There were no differences in children's age across language groups, between centers and FCCs, or between programs with a low or high concentration of ELL children.

In detailed tables in Chapters III (P-KOF) and IV (PAA), we present the distribution of responses for the PAA and P-KOF measures. Overall, teachers completed at least some portion of the PAA for 725 children in the fall and 605 in the spring. As described below, teachers were asked to complete only a subset of the scales in the PAA to reduce burden. For the P-KOF, teachers completed forms for 757 children in the fall and 606 in the spring.

---

<sup>15</sup> LAUP data regarding ELL concentration were not available for all programs for this year. We imputed missing data based on the data from the previous year when available. We did not have data on ELL concentration from either year for some programs. Thus, ELL concentration was missing for 49 children.

**Table I.1. Standardized Measures of the Full UPCOS Battery, Their Purpose, and Scales of the Provider Administered Assessment and P-KOF Expected to Correlate with Each**

Scales of the Full UPCOS Battery Standardized Measures	Purpose (Domain Addressed)	Scales of the Provider-Administered Assessment and P-KOF
Spanish Pre-LAS: Tío Simon Dice (Duncan and DeAvila 2002)	Language Screener: Spanish receptive proficiency	PAA Communication
Spanish PreLAS: Exposición de Arte	Language Screener: Spanish expressive proficiency	PAA Communication P-KOF Social Expression
English PreLAS: Simon Says (Duncan and DeAvila 2002)	Language Screener: English receptive proficiency	PAA Communication
English PreLAS: Art Show	Language Screener: English expressive proficiency	PAA Communication P-KOF Social Expression
Expressive One-Word Picture Vocabulary Test-Spanish Bilingual Edition (EOWPVT-SBE; Brownell 2000)	Language Development: Vocabulary – Conceptually Scored	PAA Communication P-KOF Social Expression <sup>16</sup>
Woodcock-Johnson III (WJ-III; Mather and Woodcock 2001) and Woodcock-Muñoz Bateria III (WM-III; Woodcock and Muñoz-Sandoval 2005) Letter-Word Identification	Literacy development	PAA Rapid Letter Naming PAA Problem Solving P-KOF General Knowledge
Woodcock-Johnson III (WJ-III) and Woodcock-Muñoz Bateria III (WM-III) Applied Problems	Cognitive Development: Math concepts	PAA Problem Solving P-KOF General Knowledge
Woodcock-Johnson III (WJ-III)—Test 7 Spelling and Woodcock-Muñoz III (WM-III)—Test 7 Ortografía	Literacy Development and Fine Motor Skills: Writing and spelling/Ability to take dictation/Fine motor control	PAA Problem Solving P-KOF General Knowledge
Preschool Kindergarten Behavior Scales-2 (PKBS-2)	Rating social-emotional skills	P-KOF Self-Regulation P-KOF Social Expression
Executive Functioning—Pencil Tapping Task (Smith-Donald et al. 2007)	Executive functioning/self-regulation	P-KOF Self-Regulation PKBS-2 Problem Behaviors
Leiter Examiner Ratings (Roid 1997)	Executive functioning	P-KOF PKBS-2 (All Scales)
N/A	Motor	PAA Fine Motor PAA Gross Motor P-KOF Self-Care & Motor

<sup>16</sup> Although the P-KOF social expression is not designed to assess language, it is the only area of the P-KOF that includes language items other than naming letters, numbers, and shapes (these items are included in the P-KOF general knowledge scale).

**Table I.2. Percentage of Children with Various Characteristics in the UPCOS-3 Sample, Fall 2009 and Spring 2010**

	Fall 2009	Spring 2010
Program Type		
Center	76	77
FCC	24	23
Concentration of ELL Children		
Low	59	56
High	41	44
Gender		
Male	51	51
Female	49	49
Child Language Group		
English Only	40	40
English Primarily	23	23
Spanish Primarily	24	25
Spanish Only	10	10
Another Language Only or Primarily	3	3

## Procedures

For all portions of the PAA measure, in the fall, teachers received a self-study ( $n = 37$ ) or in-person training ( $n = 103$ ) about how to administer the items. For the P-KOF, Lisa Colvig-Amir, director of evaluation services at Applied Survey Research, trained Mathematica's lead trainers to train LAUP providers on the modified version of the P-KOF that ASR created for LAUP. Some teachers ( $n = 21$ ) were trained to use the P-KOF by LAUP, with the remainder of the teachers trained by Mathematica lead trainers. In the spring, new teachers received a self-study. Returning teachers were offered online refresher training.

LAUP required that all teachers complete the P-KOF in the 2009-2010 program year. However, for the PAA, in order to reduce burden on teachers we randomly assigned different combinations of subscales or measures to the teachers. For the approximately 850 children whose parents gave consent for them to participate in the study in the fall, we asked teachers to complete the Problem Solving and Communication subscales of the ASQ-3, RLN, and the PKBS-2 for 64 percent of them and the Fine Motor and Gross Motor subscales of the ASQ-3 for 21 percent of them.

To ensure consistency in the administration of the various child assessments, in the fall, we had teachers who were part of the UPCOS-3 sample complete the P-KOF in the week before Mathematica's independent assessors came to assess children. Teachers then completed the assigned ASQ-3 or PKBS-2 scales the week after the direct assessments. Administration of the P-KOF preceded the ASQ-3 so that teachers would be administering the P-KOF under conditions similar to those they would be experiencing in a typical year, when only the P-KOF might be used. Nineteen of the 24 items on the P-KOF are observational items and, as described by ASR, responses are to be based on teachers' prior knowledge of the child.<sup>17</sup> If teachers were to complete the ASQ-3—which primarily includes direct assessment items—prior to the P-KOF, their observations for the P-KOF would draw on knowledge they would not have in a typical year (that is, in a year in which only one assessment measure was used). In the spring, we once again requested that teachers complete the PAA in the week following the direct assessments. However, LAUP requested that all P-KOF forms be completed within a two-week window in May. Thus, the sequencing of measures differed across teachers.

At the child level, our language routing procedures (developed in UPCOS-2) determined whether children completed the language-specific measures in the Full UPCOS Battery in English or Spanish (WJ-III or WM-III Letter-Word Identification and Spelling/Ortografía; for these two measures, responses are only accepted in one language). They also determined whether conceptually scored measures—for which answers in either language could be accepted—were administered in English or Spanish (WJ-III or WM-III Applied Problems, EOWPVT).

Language routing procedures began with parent responses to three questions about home language use. Children whose parents indicated a primary home language other than English completed the English PreLAS as a screener and (if they passed the screener) proceeded to complete the assessment in English. Children whose parents indicated a primary language other than English or Spanish completed the English PreLAS as well. If these children passed the PreLAS, they completed the entire assessment in English. If they failed the screener, they were routed out of the assessment. The remaining children—those who spoke English primarily but whose home language

---

<sup>17</sup> <http://www.laup.net/images/stories/providers/forms/PKOF/instructions.pdf>



was Spanish, and those who spoke Spanish primarily or only—received the assessment in English or Spanish depending on their score on the English PreLAS screener. All children from homes where Spanish was spoken completed the Spanish PreLAS screener.

## Scoring

We followed the publishers' guidelines for all measures in the Full UPCOS Battery. We provided both the standard scores and the *W* scores for all the Woodcock-Johnson and Woodcock-Muñoz tests. The *W* score is not affected by differences in age, and provides the best estimate of growth across time. Standard scores compare the performance of children to a nationally representative sample of same age peers. The PKBS-2 and Leiter Examiner Ratings analyses used standard scores, and the PreLAS and Pencil Tapping Task analyses used raw scores.

We estimated standard scores with two sets of norms for one of the standardized measures<sup>18</sup>, the EOWPVT, because we administered two versions of it with overlapping items. The English version of the EOWPVT was administered to children who speak English only or primarily in their homes and to all other children that passed the English language screening test, regardless of home language. The EOWPVT-Spanish Bilingual Edition (SBE) was administered to Spanish-speaking children who did not pass the English language screening test; this version of the test is conceptually scored (that is, correct responses are accepted regardless of the language of the response). Some of the items in the English EOWPVT are not included in the EOWPVT-SBE version. Thus, we used the norms for the English version to create standard scores for children who passed the English-language screening test and used the norms for the SBE version to create standard scores for all children.<sup>19</sup> To place all the children on a single scale, we also used an item response theory (IRT) model<sup>20</sup> to estimate scores on the EOWPVT. IRT places the children's abilities and the difficulty of the items on the same scale using all of the responses of all of the children. IRT allows some missing data so the skipped items on the EOWPVT-SBE were not a problem in creating scores. Similar to *W* scores, the EOWPVT IRT score is not dependent upon the different standardization samples and provides an examination of absolute change, rather than change relative to a sample of children of the same age.

The scoring methods for the teacher-administered measures (the P-KOF, the ASQ-3, and the RLN) are described in detail in the chapters discussing the results of each measure. For the P-KOF, we followed the publisher's scoring guidelines. For the ASQ-3, we used a sum score on each scale. The RLN score is based on the sum of correct responses.

---

<sup>18</sup> The standard scores based on the English sample compare children to a nationally representative sample of children in the United States that includes a broad range of socioeconomic advantage. The standard scores for the SBE version compare children to a sample that is representative of Hispanic children in the United States and more than 50 percent of the standardization sample came from homes where mothers had less than a high school education.

<sup>19</sup> The bilingual/Spanish norms are based on a nationally representative sample of Spanish-speaking children in the United States. Children from California were overrepresented in the sample. More than 50 percent of the children in the standardization sample for the bilingual norms live in families with a mother whose education is less than high school. The English norms are based on a standardization sample that represents a greater range of maternal education and cultural backgrounds.

<sup>20</sup> We used a Rasch model to estimate scores on the EOWPVT. This is the same measurement model used in creating *W* scores on the WJ-III and the WM-III.

## Approach to Analysis

The two goals of our analysis were to (1) assess the reliability and validity of the teacher-administered assessments and (2) to describe how children's growth from fall to spring is captured on these different measures. To that end, we undertook a number of analyses.

For all measures (PAA, P-KOF, Full UPCOS Battery), we first calculated descriptive statistics (means, variance) and reliability (internal consistency) for the overall sample and for key subgroups defined by child language group<sup>21</sup> (children speaking Spanish primarily, Spanish only, English only, English primarily, and another language only or primarily), program type (center-based versus FCC setting), and concentration of ELL children in the program (high/50 percent or greater ELL versus low/less than 50 percent). Reliability is a necessary, though not sufficient, condition for validity. Typically, a statistic of .70 is considered an indication of adequate or acceptable internal consistency reliability for research (Bacon 2004; Kisker et al 2004; Nunnally 1978)<sup>22</sup> as measured by Cronbach's alpha,<sup>23</sup> and .80 is considered strong. Lower internal consistency suggests that other factors are influencing the measurement and can limit the magnitude of the measure's relationships with other factors.

To look at how children enrolled in LAUP programs in fall 2009 developed during the preschool year, we examined children's scores on the direct assessments as well as on the PAA and P-KOF for those who were assessed in both fall and spring. We examined means and standard deviations of spring scores in our sample relative to the fall baseline and tested the statistical significance of the difference between fall and spring scores (for standardized measures, we used both scale scores and standard scores when available). We also calculated the means and standard deviations of the change scores from fall to spring. We conducted these analyses for the full sample of children and for the three key subgroups defined by children's language, program type, and ELL concentration in program. We tested whether progress between the fall and spring was significant for the overall sample and key subgroups using t-tests. In addition, we conducted analysis of variance (ANOVA, *F*-tests) to determine whether progress differed across subgroups. We report *p* levels at the .10, .05, and .01 levels. We consider  $p < .05$  to indicate statistical significance; results of  $p < .10$ , indicating a trend, are noted in tables but are not identified as significant in our discussion. Note that significance tests were only conducted for this component of the analysis.

For the sake of comparability across measures in the Full UPCOS Battery, PAA, and P-KOF, we calculated effect sizes for growth as the difference between the spring and fall scores divided by the fall standard deviation. Thus, the magnitude of the effect is directly related to the variation in scores (how spread out children's scores were) in the fall. Certain scales within both the PAA and P-KOF were found to have ceiling effects in the fall; that is, a large proportion of children scored at or near the maximum possible score, limiting the variation. When the variation is limited in this way,

---

<sup>21</sup> Based on parent reports regarding language use in the home and with peers at the beginning of the school year, children were placed in one of six language groups: (1) English only (scores of 14-15), (2) English primarily (scores of 10-13), (3) Spanish only, (4) Spanish primarily, (5) other language only, or (6) other language primarily. For all analyses by language group, we collapsed other language only and primarily into one group. .

<sup>22</sup> When making decisions about individuals, minimum reliability estimates should be greater than .80 (Nunnally and Bernstein 1994; Salvia and Ysseldyke 2004). However, in UPCOS-3, we focused on group data.

<sup>23</sup> The size of the sample and the amount of variance can affect the coefficient alpha.

the effect size is inflated. This should be taken into account when interpreting certain effect sizes for both the PAA and P-KOF relative to measures in the Full UPCOS Battery.

Note that we compare progress on the P-KOF and PAA only to *W* scores for the WJ-III and WM-III and IRT scores for the EOWPVT, because these scores all measure children's absolute progress along a continuum of skill. Standard scores, on the other hand, are adjusted for the developmental progress children are expected to make compared to their same-age peers. We expect effect sizes based on standard scores to be smaller than those based on absolute (unadjusted) change.

Next, for the PAA and P-KOF we calculated inter-factor correlations (that is, bivariate correlations among the subscales within each measure) to examine both the distinctiveness of factors within each measure and evidence of halo effects (wherein teachers' ratings are driven by their overall positive or negative perceptions of a child's skills). Note that we did not examine inter-factor correlations for the PKBS-2, because we would expect high correlations among the PKBS-2 subscales, as all components of the measure address social-emotional development. We also looked for evidence of ceiling effects by examining what proportion of children reached the maximum possible score. If a large proportion of children reach the maximum score in the spring, this would indicate that the measure likely underestimates children's growth across the year. We examined inter-factor correlations and evidence of ceiling effects for the full sample of children and key subgroups defined by program type and concentration of ELL children in the program.

We then compared the teacher-administered measures and the Full UPCOS Battery to explore evidence of validity.<sup>24</sup> First, we looked for evidence of convergent and divergent (discriminant) validity by examining the bivariate correlations between the measures of the Full UPCOS Battery and the teacher-administered assessments of the PAA and P-KOF.<sup>25</sup> We also created change scores for each measure and examined the bivariate correlations between change on the UPCOS battery and change on the PAA and P-KOF. Correlations between the spring scores on the various measures of the same developmental area provide evidence of concurrent validity. Correlations between the fall scores on the PAA and P-KOF and spring scores on the Full UPCOS Battery for each developmental area provide evidence of predictive validity. Correlations between the change scores on the PAA or P-KOF and the Full UPCOS indicate whether the extent of the change is attributed to the same children on both measures; this is another indicator of validity. We examined bivariate correlations for the full sample of children and also for key subgroups defined by program type and concentration of ELL children.

---

<sup>24</sup> In the fall, we also compared the teacher-administered measures with measures in the Full UPCOS Battery to explore evidence of rater effects. Rater effects indicate that characteristics of the teacher may be influencing the scores he or she is giving the children on the assessment; thus, scores may not be an accurate reflection of children's development. The baseline information—that is, the analysis of rater effects completed in the fall—is sufficient for establishing the degree to which teacher-administered measures may be problematic for assessing growth (the higher the rater effects, the less accurate a picture of growth we are likely to have). Thus, those calculations were not repeated in the spring. An overview of the fall analysis of rater effects is included in Appendix A.

<sup>25</sup> In the fall, we examined evidence of convergent validity by using hierarchical linear models (HLM) to look at the amount of variance explained in the standardized measures by each of the related teacher-administered measures after controlling for children's age, sex, and language group. Evidence regarding convergent validity from the bivariate correlations and the hierarchical models converged. Thus, it was sufficient to repeat only the bivariate correlations, particularly since we examined them by subgroups as well.

In Table I.1, the final column shows which PAA or P-KOF scales were expected to correlate with each scale from the Full UPCOS Battery shown in the first column. Because the teacher-administered PKBS-2 is a standardized measure with existing evidence of reliability and validity, we used it to examine validity for the social-emotional subscales of the P-KOF. For the same reason (existing evidence of validity) we did not conduct a validity analysis for the PKBS-2 in this sample. When considering convergent validity, we expect the correlations to be greater than .40 for subscales measuring the same construct; the more similar the dimensions are, the higher we expect the correlation to be. For example, we expect WJ-III Letter-Word Identification scores to be more strongly associated with Rapid Letter Naming (RLN) scores than with ASQ-3 Problem Solving or the P-KOF General Knowledge scores. For divergent validity, we expect lower correlations between measures of different constructs. For example, the ASQ-3 Communication score should be more strongly related to the EOWPVT than to the Pencil Tapping Task.

In the remaining chapters of this report, we present findings regarding the performance of the teacher-administered measures in the spring of the program year. In addition, we describe children's growth from fall to spring. In Chapter II we summarize children's progress as measured by the independently administered standardized assessments as well as by one standardized assessment administered by teachers. In Chapters III and IV we present findings regarding the functioning of the teacher-administered measures, as well as children's progress as assessed by those measures. Tables supporting findings presented in Chapters II, III, and IV can be found in the appendices. Finally, in Chapter V we present conclusions regarding the functioning of the teacher-administered measures, drawing on the findings presented in both this report and the interim report prepared following the fall data collection (Moiduddin et al. 2010).

## II. RESULTS OF STANDARDIZED ASSESSMENTS

### Description of Measures and Scoring Methods

In Chapter I we provided an overview of the standardized measures included in this study, and in Table I.1 we outlined the domains of child development addressed by each. As a reminder, Mathematica's independent assessors administered measures of language (PreLAS English/PreLAS Spanish, EOWPVT English version and EOWPVT-SBE), literacy (WJ-III and WM-III Letter-Word Identification and Spelling/Ortografía), mathematics (WJ-III and WM-III Applied Problems), executive functioning (Pencil Tapping Task) and social-emotional development (Leiter Examiner Ratings). Teachers completed an additional standardized assessment of social-emotional development (the PKBS-2). We now provide additional detail regarding each measure in the battery.

#### Independently Administered Assessments

**Preschool Language Assessment Survey 2000 (PreLAS 2000).** Simon Says and Art Show are two subtests from the Oral Language Development Scale (OLDS) of the PreLAS 2000 (Duncan and DeAvila 2002) available in both English and Spanish. Simon Says assesses a child's listening comprehension of basic instructions. Art Show is a picture vocabulary test that measures a child's oral language skills. Both were included in the Head Start National Reporting System assessment battery, the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B) preschool assessment, and the kindergarten and first grade assessments of the ECLS-K study. They are also used in the Head Start Family and Child Experiences Survey (FACES). In UPCOS, we used Simon Says and Art Show in combination with parent and teacher reports of the child's primary home language to determine whether children should receive the direct child assessments in English or Spanish (or not at all, in cases where the child's primary language was other than English or Spanish). Children who scored fewer than 6 items incorrect on the English PreLAS received the assessment in English. The Spanish versions of each subtest were used as a warm-up activity for Spanish-speaking children; it is also a brief indicator of Spanish language development.

**Expressive One-Word Picture Vocabulary Test: Spanish-Bilingual Edition (EOWPVT-SBE).** The EOWPVT-SBE (Brownell 2000) is a measure of children's expressive vocabulary. Given that children who are ELLs often distribute their vocabulary across two languages, the EOWPVT-SBE has been proposed as a more meaningful measure of ELL children's conceptual knowledge. The EOWPVT-SBE has been specifically designed and normed based on an approach that allows for conceptual scoring—that is, it provides prompts in both English and Spanish and accepts responses in either language (including various Spanish dialects) so as not to underestimate the language skills of ELL children. The EOWPVT-SBE has been used with diverse populations, has strong evidence of reliability and validity, and has demonstrated sensitivity to the effects of early childhood interventions (Wasik, Bond, and Hindman 2006).

The standardization samples for the EOWPVT and the EOWPVT-SBE are very different. The English version of the EOWPVT has a nationally representative standardization sample of English-speaking children in the U.S. and as such represents a full range of income and maternal education. The SBE version of the EOWPVT-SBE has a nationally representative standardization sample of Spanish bilingual children in the U.S. and as such represents a less advantaged group of children than the EOWPVT. More than half of the children in the EOWPVT-SBE standardization sample live in homes with low levels of maternal education (less than a high school diploma). These

differences are important to keep in mind in interpreting the standard scores on the EOWPVT and the EOWPVT-SBE.

**Woodcock-Johnson III Battery (WJ-III) and Woodcock-Muñoz Bateria III (WM-III) (Mather and Woodcock 2001; Muñoz-Sandoval et al. 2005).** The WJ-III and WM-III measures have been used with children from diverse backgrounds in national studies such as FACES and the Oklahoma Study, and have established reliability and validity. We used three subtests from each measure in the present study to address literacy and mathematics development.

We used the Letter-Word Identification subtest of the WJ-III for English-speaking children and the corresponding Identificación de Letras y Palabras subtest from the WM-III for assessments in Spanish. This subtest yields a measure of letter knowledge and recognition of simple words; as testing progresses, items increase in difficulty to yield a measure of early reading identification skills.

The Spelling subtest from the WJ-III and the corresponding Ortografía subtest from WM-III provide a measure of children's early writing skills. The first six items tap fine motor coordination and pre-writing skills, such as drawing a line and copying letters. The remaining items measure a child's ability to provide written responses when asked to write specific uppercase or lowercase letters of the alphabet, words, and phrases.

The Applied Problems subtest of the WJ-III and the corresponding Problemas Aplicados subtest of the WM-III assess a child's ability to analyze and solve practical problems in mathematics. To solve the problems, the child must perform simple counting, addition, and subtraction computations. Although children receive instructions in the language identified during the screening process, responses are accepted in either English or Spanish on both the WJ-III and WM-III versions of the subtest.

**The Pencil Tapping Task.** The Pencil Tapping Task (Smith-Donald et al. 2007) is an adaptation of a peg-tapping task (Blair 2002; Diamond and Taylor 1996) that involves having a child do the opposite of what the assessor does (tap once when the assessor taps twice and tap twice when the assessor taps once) across 16 trials. It yields a measure of children's inhibitory control, which has been shown to relate to young children's development across a number of domains, including vocabulary, literacy, and mathematics (Blair and Razza 2007; Espy et al. 2004; McClelland et al. 2007). In addition, the task involves children's attention and working memory. Pencil Tapping has been used with diverse samples of children (including children in UPCOS-2; Love et al. 2009), has some evidence of reliability and validity in small-scale studies (Blair and Razza 2007), and has demonstrated developmental sensitivity. English and Spanish versions are available.

**Leiter Examiner Rating Scales.** We used subtests from the Leiter Examiner Rating Scales to obtain a picture of each child's Attention, Activity Level, and Sociability. The scales have been used successfully in three large-scale studies that include children with diverse racial, ethnic, and language backgrounds (FACES 2006, Early Head Start Transition to Prekindergarten [ACF 2006]), and Home Visiting 2000 [Olds et al. 2004]) and demonstrate good reliability and predictive validity. The assessor completed the ratings based on observations made during the administration of the assessment. Higher scores on the subscales indicate higher levels of positive approaches to learning and social skills.

## Teacher-Administered Assessments

**Preschool Kindergarten Behavior Scales-2 (PKBS-2).** The PKBS-2 (Merrell 2002) is a comprehensive, teacher-report measure of children's social-emotional competence that examines 34 positive child behaviors and 42 negative behaviors along 5 scales: Social Cooperation, Social Independence, Social Interaction, Externalizing Problem Behaviors, and Internalizing Problem Behaviors. These subscales can be further aggregated into Total Positive Social Skills and Total Problem Behaviors (the discussion below focuses primarily on these aggregated scales). The measure was standardized with a diverse sample of preschool and kindergarten children, has evidence of reliability and validity, and is available in English and Spanish.

## Scoring Procedures

Scoring procedures are described in detail in Chapter I. For all standardized measures, scoring procedures followed publisher guidelines. In addition, we used an IRT model to develop alternative scores for the EOWPVT that would enable us to assess children's absolute progress (rather than progress relative to a standardization sample). For the EOWPVT-SBE and English version, WJ-III, WM-III, and PKBS-2 mean standard scores are 100 points, with a standard deviation of 15 points. The national mean for the Leiter examiner ratings scale scores is 10, with a standard deviation of 2. For the Leiter we also calculated the percent of children that were in the clinical range and in the possible clinical range (at-risk) for all subscales. The PreLAS and Pencil Tapping Task analyses used only raw scores

## Children's Progress from Fall 2009 to Spring 2010

Table II.2 in Appendix B outlines the average time between fall and spring assessments for the overall sample and key subgroups based on child language and program type. Overall, 5.09 months passed between the two time points and averages did not differ markedly for subgroups defined by child language or program characteristics.

In the remainder of this chapter, we describe children's progress from fall to spring. All tables are included in Appendix B. As a reminder, we conducted tests to determine (1) whether children's progress from fall to spring was statistically significant and (2) whether progress differed between subgroups. Statistical significance of progress from fall to spring is indicated in tables by asterisks on fall scores. Statistical significance of differences in progress between subgroups is indicated in tables by asterisks on the change score for the first subgroup shown in each table. All findings noted in the text regarding progress from fall to spring or differences in progress across groups are statistically significant at the  $p < .05$  level unless otherwise noted. In a few instances we describe notable patterns that are not statistically significant. We did *not* conduct statistical tests comparing children's fall scores or spring scores between groups; comparisons of these point-in-time scores identify patterns rather than statistical significance.

## Overall Sample

Table II.3 presents progress in language, literacy, and mathematics for the full sample of children. As the table indicates, children made progress in all areas of development. Beginning with language, children's scores on the portions of the English PreLAS measuring both receptive and expressive proficiency increased from fall to spring, from 6.93 to 8.23 for Simon Says ( $ES = .44$ ) and from 7.41 to 8.33 for Art Show ( $ES = .30$ ). (As a reminder, we used the standard deviations of the fall scores to calculate the effect sizes shown throughout the report.) Children also made



progress in Spanish receptive language skills (Tío Simón Dice), with scores progressing from 5.63 in fall 2009 to 6.70 in spring 2010 (ES = .38). (As a reminder, the English PreLAS was used as a screener to determine whether children should complete the language-specific measures in the Full UPCOS Battery in English or Spanish.)

Language also progressed according to the EOWPVT, but only for the SBE version. For the English version, children did not make significant progress compared to a sample of English-speaking peers (that is, the standard score did not significantly increase), and the mean spring score remained below those of children of the same age nationally (90.55). Children's EOWPVT-SBE scores, on the other hand, progressed across the five-month period between assessments, from 105.94 to 108.49 (ES = .11). This indicates that children's expressive vocabulary (English plus Spanish) increased. Thus, as in the fall, children's spring skills were above their Latino peers nationally. As described previously, IRT scores provide a measure of children's absolute change across the year—how much progress they made without adjusting for the change in their age. IRT scores also indicated statistically significant growth, from 44.58 in the fall to 49.31 in the spring (ES = .44). Note that IRT scores, like the SBE scores, address progress of overall expressive vocabulary skills (English plus Spanish). Standard scores however, are adjusted for the age of the child and compare the child's performance to other children of the same age. When significant, positive changes in standard scores indicate that children made greater progress than would be expected due to developmental change over time.

On average, the children in the sample who took the assessments of literacy (WJ-III Letter-Word Identification, Spelling) and mathematics (WJ-III Applied Problems) in English made progress across the year, but only in absolute terms for mathematics. Letter-Word standard scores increased from 100.65 to 103.70 (ES = .21) and W scores from 332.28 to 349.46 (ES = .60). Spelling standard scores increased from 103.75 to 106.08 (ES = .16) and W scores from 380.17 to 398.08 (ES = .63). Applied Problems standard scores increased from 100.98 to 102.38 (ES = .11, but the change was not significant) and W scores from 397.47 to 409.91 (ES = .53, a significant change). Thus, as indicated by the standard scores, LAUP children continued to perform as well or better than children of the same age nationally.

Children who took the assessments in Spanish made progress only in absolute terms. Standard scores on the assessments of literacy (WM-III Letter-Word Identification, Ortografía) and mathematics (WM-III Applied Problems) did not change between fall and spring, but W scores did change in all three cases (a range of 8.35 to 17.43 points, ES = .36 to .53). Thus, the change that children made was similar to the developmental change expected for children of this age. In the spring, children continued to score below those of the same age in the normative sample. The mean standard score was 91.71 for WM-III Letter-Word, 89.26 for WM-III Ortografía, and 84.25 for WM-III Applied Problems.

In the spring, about 90 children who were tested in Spanish in the fall switched to English tests. We compared these children's fall Spanish test scores to the fall scores of children who received the Spanish tests in both the fall and spring. The two groups were not different in WM-III Letter-Word Identification or Spelling; however, children who switched to English tests in the spring had significantly higher WM-III Applied Problems scores and higher EOWPVT-SBE scores in the fall.

In Table II.4 we present children's progress in the area of social-emotional development. The Pencil Tapping Task is a measure of executive functioning that addresses children's inhibitory control. The mean for the pencil tapping trials was 52 percent (that is, children responded correctly about half of the time) in the fall and 71 percent in the spring. By way of comparison, the



representative sample of children in UPCOS-2 responded correctly 43 percent of the time in the fall; by spring they responded correctly 67 percent of the time.

In terms of social-emotional development, both independent assessors and teachers rated the children. The mean ratings of social skills—Leiter Examiner Ratings of Attention, Activity, and Sociability—from the independent assessors were in the low-average range in the fall (means = 8.36 to 8.51)<sup>26</sup> and spring (means = 8.49 to 9.05). From fall to spring, ratings significantly increased for both Attention (ES = .29) and Sociability (ES = .25). Both the fall and spring Leiter Examiner Ratings indicated that less than 10 percent of the children were in the clinical range (3 to 7 percent in the fall and 1 to 3 percent by the spring), and the percentage in the clinical range fell for all three social skills. Similarly, the percentage of children in the *possible* clinical range (that is, those at-risk) declined (from 14 to 12 percent in the fall to 7 to 11 percent in the spring). Teachers also reported that children’s social skills improved from fall to spring (mean standard scores for Total Positive Social Skills increased from 98.62 to 107.56, ES = .59) and Total Problem Behaviors declined (from 92.92 to 88.18, ES = -.35). Although teachers rated children as having better social skills than other preschoolers in the nation (and fewer problem behaviors), these ratings are not as positive as those found in the representative sample in UPCOS-2, where mean standard scores on social skills ranged from 115 to 119 in the fall.

**Reliability.** In the spring, the reliability estimates of the standardized assessments of language, literacy, and mathematics were all strong but, as in the fall, they were higher for the English (or bilingual) measures ( $\alpha = .91$  to  $.99$ ) than for the Spanish measures ( $\alpha = .79$  to  $.91$ ). Reliability estimates for the Pencil Tapping Task and for the teacher ratings of social-emotional development on the PKBS-2 ranged from  $.85$  to  $.97$ . The reliability of the Leiter Examiner Ratings ranged from the lowest internal consistency for the four-item Activity scale ( $\alpha = .76$ ) to the highest internal consistency for the 10-item Attention scale ( $\alpha = .97$ ).

## Children’s Progress by Subgroups

### Language Subgroups

For this analysis, we grouped children in five language groups: (1) English only, (2) English primarily, (3) Spanish only, (4) Spanish primarily, and (5) other language only or primarily. We consider the latter three groups to be ELLs. (As a reminder, these groups are based on parent responses to a series of questions about the child’s home language use.) Children identified as ELLs completed the assessment in English or Spanish depending on their performance on the language screener (English PreLAS). Note that for the Spanish only and other language only or primarily groups, sample sizes for certain assessments fell below 10 children; to avoid less stable/precise estimation as well as to protect confidentiality, these children’s scores have been deleted from the tables and are not discussed in the text.

We found differences in children’s performance and progress between fall and spring on language, literacy, and math standardized measures across the language groups (fall and spring scores are in Table II.5 and change scores appear in Table II.6). Children in the Spanish only group had the lowest average scores on the total English PreLAS and the two subscales in the fall and spring, followed by children in the Spanish primarily group. However, children in all groups made progress

---

<sup>26</sup> In UPCOS-2, the fall ratings ranged from 7.8 to 8.2.

between the two assessments in both receptive and expressive proficiency. For the total score, effect sizes for progress ranged from .21 for the English only group to .93 for the Spanish only group. Statistical tests indicate that progress differed between groups. Children in the Spanish only and Spanish primarily groups both progressed in their Spanish receptive skills (Tío Simón Dice scores went from 5.15 in the fall to 6.96 in the spring for the Spanish only group,  $ES = .64$ , and from 5.82 to 6.60 for the Spanish primarily group,  $ES = .28$ ); progress was greater for the Spanish only group.

Children who speak English only or primarily performed higher on the EOWPVT English version than ELL children in fall and spring, but all groups scored lower than the norming sample of English-speaking peers. By the spring, children in the English only group had a mean standard score of 92.55 compared to 90.17 for the English primarily group, 81.96 for the Spanish primarily group, and 88.59 for the other language group (the Spanish only group is excluded from the analysis due to sample size fewer than 10). None of the groups made significant progress in English expressive skills. Similarly, children in the English only group outperformed children in the other four groups on the EOWPVT-SBE, and none of the groups made significant progress relative to peers. Spring standard scores were 118.42 for the English only group, 114.47 for the other language group, 111.53 for the English primarily group, 96.86 for the Spanish primarily group, and 90.09 for the Spanish only group. The IRT scores, on the other hand, indicate that all groups made absolute progress in expressive language; effect sizes ranged from .38 to .56.

Language groups differ in their progress in the literacy domain, although the differences in change scores are not significant except in one case noted below. Beginning with WJ-III Letter-Word identification, children in the English only and primarily groups and the Spanish primarily group made progress relative to a sample of same-age peers between fall and spring. For the English only group, standard scores went from 101.75 to 104.17 ( $ES = 0.17$ ), for the English primarily group from 100.26 to 104.29 ( $ES = .28$ ) and for the Spanish primarily group from 94.15 to 99.38 ( $ES = .36$ ). Children in the other language group had a mean standard score of 109.76 in the spring but had not made significant progress relative to peers; however, W scores, which capture absolute change, did increase ( $ES = .62$ ). (The Spanish only group is excluded from the analysis due to sample size fewer than 10.) The Spanish primarily group made the greatest progress compared to the English only, English primarily, or other language group. For spelling, only the English only or primarily groups showed significant progress relative to a sample of same-age peers (that is, based on standard scores); standard scores went from 103.49 in the fall to 106.29 in the spring ( $ES = .20$ ; for W scores,  $ES = .66$ ) for the English only group and from 103.75 in the fall to 106.03 in the spring ( $ES = .16$ ; for W scores,  $ES = .62$ ) for the English primarily group. Children in the Spanish primarily group performed at a similar level to the English group (mean standard score = 102.55) but only made progress in an absolute sense; W scores increased significantly ( $ES = .57$ ). Although children in the other language group outperformed children in the other groups as well as a sample of same-age peers in the spring (mean standard score = 112.24) neither standard scores nor W scores changed. (The Spanish only group is excluded from the analysis due to sample size fewer than 10.)

Shifting to the Spanish-language measures of literacy<sup>27</sup>, standard scores for the WM-III Letter-Word Identification subtest in the spring ranged from 84.36 for the English primarily group to 95.26 for the Spanish only group. We found no significant changes in standard scores between fall and spring. However, children in the Spanish only group made absolute progress according to W scores ( $ES = .62$ ). For the WM-III Ortografía, standard scores for all language groups were around 90

---

<sup>27</sup> None of the English only children took the Spanish measures.

points in the spring. As with Letter-Word Identification, children in the Spanish only group made progress based on their W scores ( $ES = .60$ ). Across the Spanish and English measures of literacy, change scores differed significantly between language subgroups only for WM-III Letter-Word Identification standard scores.

Finally, in the mathematics domain children made progress in an absolute sense but, relative to a sample of same-age peers, performed similarly in the fall and spring. Beginning with WJ-III Applied Problems, children in all four groups scored at or above the mean standard score, but scores did not progress beyond what was expected developmentally between fall and spring. (The Spanish only group is excluded from the analysis due to sample size fewer than 10.) Nonetheless, children made absolute progress in the English only ( $ES = .53$ ), English primarily ( $ES = .54$ ), and Spanish primarily ( $ES = .49$ ) groups. Scores were unchanged for the other language group. Finally, for the WM-III Applied Problems, children's performance relative to same-age peers was unchanged for the English primarily and Spanish groups and remained below that of their peers (standard scores ranged from 76.36 in the spring for the English primarily group to 87.83 for the Spanish only group). Nonetheless, children made statistically significant absolute progress in both the Spanish only ( $ES = .62$ ) and Spanish primarily ( $ES = .46$ ) groups (progress was similar in magnitude for the English primarily group but not statistically significant). Note that change scores did not differ significantly across groups.

There were also differences across the language groups in the level of social-emotional development in the spring, although progress between fall and spring rarely differed across groups (Tables II.8 and II.9). Beginning with independently administered assessments, as in the fall, children who speak Spanish only had the lowest scores on the Pencil Tapping Task. However, scores increased for the Spanish only group (39 to 60 percent) as well as for the English only group (58 to 78 percent), English primarily group (48 to 70 percent) and the Spanish primarily group (50 to 64 percent). Change did not differ significantly across groups.

Children in the Spanish only group also had the lowest scores on the Leiter Examiner Ratings in the spring (8.18 to 8.40), although differences with the Spanish primarily group appear small. Further, children in all groups scored in the low-to-average range in the spring (scores ranged from 8.27 to 9.57 depending on the subscale and language group). Assessors rated children in the Spanish only group as making no progress in social skills, but children in the English only, English primarily and Spanish primarily groups made progress in both Attention ( $ES = .25, .33, \text{ and } .33$ ) and Sociability ( $ES = .28, .22, \text{ and } .27$ ). Children in the other language only or primarily group reportedly had decreases in Activity between fall and spring ( $ES = -.50$ ). For Activity only, changes differed across groups, with the Spanish only and Spanish primarily groups showing more change. The English and Spanish groups all experienced declines in the percentage of children in the clinical range for Attention; although we found no significant differences across groups in the amount of change, the Spanish only group experienced a particularly large decline, from 10 percent in the fall to 4 percent in the spring. Note that for the percentage of children in the clinical range for Sociability, where none of the declines were significant, both the Spanish only and other language only or primarily groups experienced a particularly large decline compared to the English only, English primarily, and Spanish primarily groups (10 percent versus 2 to 4 percent).

Finally, teachers reported that Total Positive Social Skills were above average for the English and Spanish groups in the spring as measured by the PKBS-2, although the mean was lowest for the Spanish only group. (The other language only or primarily group is excluded from the analysis due to sample size below 10.) Total Positive Social Skills scores also increased in all three groups ( $ES = .56 \text{ to } .67$ ). In terms of Total Problem Behaviors, teachers rated children in the English only, English

primarily, and Spanish only groups similarly in the spring (approximately 88-90 standard score points for each group); ratings were even lower for children in the Spanish primarily group (83.62). For the English only and Spanish primarily groups, Total Behavior Problems declined significantly between assessments (ES = -.42 and -.44, respectively). Neither changes in Total Positive Social Skills nor in Total Behavior Problems differed across groups.

**Reliability.** The reliability estimates (internal consistency; Tables II.7 and II.10) were generally similar (and strong) across the language groups on language, literacy, and mathematics measures. Exceptions included lower reliability estimates on the PreLAS Art Show for the English only, English primarily, and other language only or primarily groups (.55, .63, and .42, respectively) and on the WM-III Letter-Word Identification for the English primarily group (.56). Reliability estimates were also similar across language groups for the social-emotional measures, with just one exception: Cronbach's alpha was .75 for the Pencil Tapping Task in the other language only or primarily group compared to .92 or higher for the three remaining language groups.

### **Program Subgroups: Center-Based and Family Child Care Settings**

In Tables II.11 and II.12 we present children's performance on standardized measures of language, literacy, and mathematics, and in tables II.14 and II.15 children's performance on the social-emotional measures. Children in center-based and FCC programs performed in a similar manner across most of the measures. Exceptions include mean spring scores for Spanish-language measures of spelling (WM-III Ortografía) and mathematics (WM-III Applied Problems), which appear slightly higher in FCC settings. Patterns of progress were also similar (generally mirroring the full sample results). In some instances the fall-spring change was significant for one group but not the other (for example, WM-III Applied Problems scores changed significantly in centers but not in FCCs). Nonetheless, the magnitude of progress did not significantly differ for any of the language, literacy, and mathematics measures and for only one of the social-emotional measures (the decline in the percentage of children in the clinical range for Sociability according to the Leiter Examiner Ratings was larger in centers [-5.73 percent] than FCCs [-.63 percent]).

Consistent with the full sample results, the reliability estimates of the standardized assessments were generally stronger for the English (or bilingual) measures than for the Spanish measures in both settings, although alphas were generally strong for both types of measures (Tables II.13 and II.16). For the English measures, the alpha was greater than .85 (and usually greater than .90) for all measures with the exception of the Simon Says subtest of the English PreLAS ( $\alpha = .82$  to  $.84$ ), the Activity subscale of the Leiter Examiner Ratings ( $\alpha = .74$  to  $.82$ ), and the Social Independence subscale of the PKBS-2 in centers ( $\alpha = .84$ ). The estimates for measures in Spanish were also typically higher than .85 with the exception of the Spanish subtests of the PreLAS ( $\alpha = .79$  to  $.82$  in centers,  $.83$  to  $.86$  in FCC programs) and for the WM-III Letter-Word Identification scores ( $\alpha = .77$  to  $.83$ ). Typically, the reliability was similar in both types of settings.

### **Program Subgroups: High and Low Concentrations of ELLs**

In Tables II.17 and II.18 we present children's performance on standardized assessments of language, literacy, and mathematics, and in tables II.20 and II.21 we present children's performance on the social-emotional measures by high and low concentration of ELLs in the program. Beginning with language, literacy, and mathematics, as might be expected, children in programs with a high concentration of ELLs had lower mean scores on the assessments administered in English in the spring and had higher mean scores on the assessments administered in Spanish, although in some instances differences appear quite small. Progress differed by program type for only two measures.

First, children in programs with a high concentration of ELLs made more progress on the English PreLAS (ES = .49 for the total score and .40 for Art Show) than children in programs with a low concentration of ELLs (ES = .34 for the total score and .23 for Art Show). Second, in terms of Spanish-language literacy, children in programs with a low concentration of ELLs experienced larger declines in WM-III Letter-Word Identification standard scores. In high concentration programs, mean standard scores shifted from 93.17 in the fall to 92.70 in the spring (ES = -.04) and in low concentration programs from 95.96 to 88.44 (ES = -.63); the decline was significant only for the latter subgroup. Note that in low concentration programs, although WM-III Letter-Word Identification scores declined relative to a national sample of same-age peers, W scores did not change (that is, children's own skills remained steady between the two assessments, but their national peers increased in skills). Note, there are additional instances where fall-spring change was significant for one group but not the other, even though the magnitude of change for the groups did not statistically differ.

Ratings of social-emotional development by both independent assessors (Leiter Examiner Ratings) and teachers (PKBS-2) were generally positive for children in both types of programs in the spring. In addition, children in both types of programs made significant progress over the assessment period (improvements in social skills, declines in behavior problems). Statistical tests indicate that progress was greater for children in programs with a high concentration of ELLs than in low concentration programs for Leiter Examiner Ratings of Attention (ES = .39 and .22, respectively) and for teacher assessments of Total Problem Behaviors (ES = -.23 and -.50, respectively). Note that there were additional differences for the changes in the percentage of children in the clinical range for Leiter Examiner Ratings of both Attention and Sociability. There was a greater decrease in the percentage of children in the clinical range on Attention and Sociability subscales in programs with a concentration of ELLs (6 and 7 percent, respectively) than in programs with a low concentration of ELLs (1 and 2 percent, respectively).

Consistent with the full sample results, the reliability estimates of the standardized assessments were slightly higher for the English measures than for the Spanish measures in programs with different concentrations of ELLs. For the English measures, the alpha was greater than .85 for all measures except the Simon Says subscale of the PreLAS ( $\alpha = .82$  to  $.84$ ), the Art Show subscale in low concentration programs ( $\alpha = .80$ ), and the Leiter Examiner Ratings of Activity ( $\alpha = .75$  to  $.77$ ). For measures administered in Spanish, reliability estimates fell below .85 for the Spanish PreLAS ( $\alpha = .79$  to  $.83$ ), WM-III Letter-Word Identification ( $\alpha = .73$  to  $.80$ ), and WM-III Ortografía in low concentration settings ( $\alpha = .84$ ). The reliability estimates were similar for the classrooms from programs with high and low concentrations of ELLs except in the case of WM-III Letter-Word Identification ( $\alpha = .73$  for low concentration and  $.80$  for high concentration).

## Summary

Overall, children made progress in all areas of development (language, literacy, mathematics, social-emotional development) in both the full sample and in subgroups defined by child language, program type, or concentration of ELL children in the program in the five-month period captured by the assessments. For the Spanish-language measures in particular, progress was more likely to occur in absolute terms than relative to a national sample of peers; in other words, children made progress over the assessment period, but no more change than would be expected developmentally for children of this age. In some instances patterns of change did differ slightly across groups, but those differences were not necessarily statistically significant.

By the spring, on average, children who took assessments of language, literacy, and mathematics in English or bilingually continued to perform as well or better than children of the same age nationally with the exception of English expressive vocabulary for all children (as measured by the EOWPVT English version) and total expressive vocabulary (English plus Spanish) for the Spanish only and primarily subgroups (as measured by the EOWVPT-SBE). Children who took assessments in Spanish in both the fall and spring continued to perform below a national sample of peers on average. For all children across all subgroups, teachers rated social skills as better than other preschoolers in the nation by the spring; they also rated children as having fewer behavior problems.

### III. CHILDREN'S PROGRESS AS MEASURED BY THE PRE-KINDERGARTEN OBSERVATION FORM

The Pre-Kindergarten Observation Form (P-KOF) was developed by Applied Survey Research (ASR 2008, 2009) for teachers to use in assessing children's school readiness skills. For each item on the P-KOF, teachers are asked to indicate whether, and to what extent, a child is demonstrating a particular skill. Ratings for items 1 to 20 are "not yet" (1), "beginning" (2), "in progress" (3), "proficient" (4), and "don't know/not observed."<sup>28</sup> For items 21-24, ratings denote the number of letters, shapes, colors, or objects that the child names or counts. For example, the ratings for item 24 indicate the number of primary shapes that a child names: "none" (1), "1 object" (2), "2 objects" (3), or "3 objects or shapes" (4), and item 20 specifies how many objects a child can count with ratings of "none" (1), "1-5 objects" (2), "6-9 objects" (3), or "all 10 objects" (4). Teachers may assign a score of (5) "don't know/not observed" for items 21-24.

Scores on individual items are averaged to yield an overall score (24 items), and 4 subscale scores: Social Expression (6 items), General Knowledge (7 items), Self-Regulation (8 items), and Self-Care & Motor Skills (3 items). Ratings of "don't know/not observed" are recoded to missing. ASR recommends calculating the overall score when the teacher rates at least eight items and subscale scores when they rate at least two items.

The test developers provide additional guidelines for administering and scoring the P-KOF for ELL children. Teachers are asked to skip 10 items (items 7, 12, 13, 15, 16, 20, 21, 22, 23, 24) if they are unable to "communicate well enough with the child to make an accurate assessment" (ASR 2009). If a child is identified as an ELL, the 10 items are to be recoded to missing even if a teacher provides a response. In previous analyses, we examined how the psychometric properties of the P-KOF scales varied depending on the items included. We found that excluding the 10 items resulted in a decrease in the validity estimates for the scores (Moiduddin et al. 2010). For this reason, our discussion of the P-KOF in this report focuses on results using all available data provided by teachers.

In this chapter, we present the findings from our analysis of both fall and spring P-KOF data. We first describe children's progress from fall 2009 to spring 2010. Next, we report children's progress by subgroups: language, program type (centers and FCCs), and concentration of ELLs. We then report evidence of the concurrent and predictive validity of the P-KOF in comparison to the standardized measures (the Full UPCOS Battery—see Chapter II) for the overall sample and by subgroups. We conclude the chapter with a summary of the major findings for the P-KOF. We obtained P-KOF data for 605 children in spring 2010. We provide the distribution of children by program type and language group in Table III.1 (Appendix C contains all Chapter III tables and figures.)

---

<sup>28</sup> Prior to spring data collection, ASR provided additional clarification about how to rate children's skills and abilities on the P-KOF. The information was available on the LAUP website for a short period of time, however, Mathematica and LAUP decided not to promote the use of the new guidelines with teachers so that the fall and spring data would be measuring children on the same criteria so that we could look at growth.



## Children's Progress from Fall 2009 to Spring 2010 for the Overall Sample

In Table III.2 we present means and standard deviations of P-KOF scores in the fall and spring. Mean scores for all subscales were higher in the spring compared to the fall. Mean spring scores ranged from 3.53 to 3.87, indicating that most children were close to proficient on the skills assessed. Consistent with the fall data, the highest mean score in the spring was on Self-Care & Motor Skills while the lowest mean score was on General Knowledge.

Children showed considerable gains on the P-KOF scales from fall to spring (Table III.3). Effect sizes indicate that average increases on the P-KOF were moderate to large. Although the lowest mean score in the spring was on the General Knowledge subscale, it was also in this area that children showed the largest increase from the fall. Average General Knowledge scores increased from 2.81 in the fall to 3.53 in the spring ( $ES = 1.01$ ). This increase is larger than the change we observed on any of the WJ-III or WM-III W scores. The smallest change we observed on the P-KOF was on the Self-Care & Motor Skills scale, where the average score increased from 3.46 in the fall to 3.87 in the spring ( $ES = .68$ ). Increases on Social Expression and Self-Regulation were both large in terms of effect size. In contrast, we found only small to medium increases on the Pencil Tapping Task, PKBS-2, and Leiter Examiner Ratings. We also did not see as large a change on the PreLAS or EOWPVT, both of which assess skills related to the Social Expression scale. The larger magnitude of change on the P-KOF scores may be related to the range of skills that the P-KOF assesses compared to the standardized assessments, which tend to be broader in scope although they measure skills in the same domain. The P-KOF, for example, has just one item assessing letter recognition whereas the WJ-III and WM-III devote an entire subtest to naming letters and words. The P-KOF combines early literacy and math skills in one scale, while these domains are assessed separately in the standardized assessments.

Figure III.1 shows the distribution of fall and spring scores on each of the P-KOF scales, as well as the overall scores. The spring distributions all show considerable skew, with a large number of children receiving the maximum score of 4. This ceiling effect is most pronounced on the Self-Care & Motor Skills scale, with 78 percent of children receiving the maximum score. However, the percentage of children receiving the maximum score is still quite high for the other scales: 53 percent on Social Expression, 38 percent on Self-Regulation, and 23 percent on General Knowledge.

The shape of the spring distributions may result in underestimation of progress from fall to spring, particularly for children whose scores were already above 3 in the fall. The magnitude of change on each of the subscales may be affected, at least in part, by the skewed distribution of scores in the fall and spring. As previously reported, most children already had very high scores (close to the maximum of 4) on the P-KOF scales in the fall, particularly on Self-Care & Motor Skills (Moiduddin et al. 2010). On this subscale, it was not possible for mean scores to increase very much. Conversely, because children tended to score lower on General Knowledge compared to other P-KOF scales in the fall, it is not surprising to see the most growth in that area between fall and spring.

Our estimates of Cronbach's alpha using the spring scores closely mirror those from the fall analyses (Moiduddin et al. 2010). Most of the scales exhibit strong internal consistency. The only exception is the Self-Care & Motor Skills scale, which shows a lower reliability in the spring ( $\alpha = .67$ ). This may be due to the limited variance in scores on that scale.



## Children's Progress by Subgroup

### Language Subgroups

The mean P-KOF scores of all language subgroups increased from fall to spring (Table III.3). Similar to what we found in the fall, children who speak Spanish only had the lowest average scores across all of the P-KOF scales in the spring. However, all of the language subgroups had mean scores above 3 in the spring.

Table III.4 displays mean change scores for each of the language groups. The Spanish only subgroup exhibited the largest mean increase in scores on all P-KOF scales, followed by the Spanish primarily group, English primarily group, English only group, and the other language only or primarily group.

In general, differences in fall to spring change score effect sizes between language groups were not as pronounced on the standardized assessments as they were on the P-KOF scales; for all P-KOF scales, change scores significantly differed between child language groups. The larger between-group differences in change scores on the P-KOF may be due in part to ceiling effects. The Spanish only group had the lowest mean scores in the fall, with all but one of the mean scores below 3. The other language groups had already received mean scores of 3 and above in the fall. The standardized measures allow for a wider range of possible scores and more between-child variation. As such, we see smaller differences in the progress made by the various language groups.

We did not find substantial differences in internal consistency of the P-KOF scales between language groups (Table III.5). With the exception of Self-Care & Motor Skills, most of the scales had strong reliability for all groups. The Self-Care & Motor Skills scale had stronger internal consistency for the Spanish only and other language only or primarily groups than for the other subgroups.

### Program Subgroups: Center-Based and Family Child Care Settings

Children in both center-based and FCC programs had spring mean scores above 3 on all P-KOF scales (Table III.6). As in the fall, the highest mean scores for both groups were on Self-Care & Motor Skills, followed by Self-Regulation, Social Expression, and General Knowledge. Differences between mean scores for the two groups were smaller in the spring compared to the fall, with mean scores for all subscales except General Knowledge within .03 across center-based and FCC programs.

The fall scores showed differences between center-based and FCC programs; we also found differences between centers and FCCs in the change scores. Fall to spring change scores were larger for children in center-based programs on all subscales except General Knowledge, where the effect size for both groups was large ( $ES \sim 1.00$ ; Table III.7). Children in center-based programs also demonstrated large increases from fall to spring on the Social Expression and Self-Regulation scales ( $ES = .83$  to  $.84$ ), while children in family child care programs showed medium-sized changes ( $ES = .55$  to  $.60$ ) on these scales. The magnitude of change we observed on the P-KOF for both groups was larger than we found on the standardized measures, where effect sizes for both groups were in the small to medium range. Progress did not differ by program type for the overall score.

Internal consistency of the P-KOF scales in centers and FCCs mirrored estimates from the full sample of children, with Cronbach's alpha ranging from .63 to .94 for FCCs and from .68 to .92 in

centers (Table III.8). In both types of program, only Self-Care & Motor Skills showed lower internal consistency. As previously mentioned, this is likely due to the limited variance in scores on that scale.

### **Program Subgroups: High and Low Concentrations of ELLs**

Children in programs with both high and low concentrations of ELL children had high average scores on all P-KOF scales in the spring (Table III.9). As in the fall, both groups received the highest average scores on Self-Care & Motor Skills. The lowest score for children in high concentration programs was in General Knowledge, whereas the lowest score for children in low concentration programs was in Self-Regulation. Though not as close in the spring as the center-based and FCC estimates, the mean scores for the high and low concentration programs also were much closer than they had been in the fall.

Although all subscale scores significantly increased for both groups, children in high concentration programs showed larger mean increases from fall to spring on all the P-KOF scales than did children in low concentration settings; the size of the differences varied by subscale and was not significant in the case of Social Expression (Table III.10). Both program types demonstrated their largest gains on General Knowledge. For the high concentration group, the average score increased from 2.66 in the fall to 3.43 in the spring ( $ES = 1.08$ ). Children in low concentration programs had their average scores increase from 2.95 in the fall to 3.62 in the spring ( $ES = .94$ ). These increases are larger than the change we observed on the WJ-III and WM-III measures. On Self-Regulation, the average score for children in high concentration programs increased from 3.01 to 3.66 ( $ES = .89$ ), whereas the average score for children in low concentration programs increased from 3.07 in the fall to 3.56 in the spring ( $ES = .69$ ). These changes are only slightly larger in magnitude than the increases we observed on standardized measures of social-emotional development.

Estimates of Cronbach's alpha indicate strong internal consistency reliability for most of the P-KOF scales across programs with both high and low concentrations of ELL children (Table III.11). Similar to our findings for the overall sample and by program type, only the Self-Care & Motor Skills scale yielded lower reliability estimates for both groups.

## **Inter-Factor Correlations for Spring Data**

### **Inter-Factor Correlations in the Overall Sample**

Table III.12 presents correlations among the subscale scores on the P-KOF. The inter-factor correlation coefficients are all positive, ranging from .36 to .65, with the strength of the relationships consistent with the theoretical constructs. That is, stronger relationships were found between the most similar areas, and weaker relationships between areas of development that tap very different domains. The strongest correlation was between Self-Regulation and Social Expression ( $r = .65$ ) and the weakest relationship was between the Self-Care & Motor Skills and General Knowledge scales ( $r = .36$ ).

The inter-factor correlations in the spring were lower than the correlations we obtained with the fall data ( $r = .51$  to  $.82$ ; Moiduddin et al. 2010). This may suggest that teachers' spring ratings on individual scales were less affected by their overall perception of a child's skills. However, it is likely that the smaller sample sizes and reduced variation on all the P-KOF subscales in the spring compared to the fall (as shown in Figure III.1) contributed to the lower correlations.

## Inter-Factor Correlations by Subgroup

**Center-Based and FCC Programs.** Inter-factor correlations between P-KOF scales were higher in FCCs ( $r = .42$  to  $.81$ ) compared to center-based programs ( $r = .34$  to  $.59$ ). This was particularly true for the Self-Regulation and Social Expression scales, where we estimated a correlation of  $.59$  in centers and  $.81$  in FCCs. Consistent with the overall sample, the weakest correlation for both program types was between General Knowledge and Self-Care & Motor Skills.

**High and Low ELL Concentrations.** The patterns of inter-factor correlations by ELL concentration were very similar to those obtained from the overall sample and by program type. We found minimal differences between high and low concentration programs. The largest difference was in the correlation between Self-Regulation and Self-Care & Motor Skills, which was higher in high concentration programs ( $r = .64$ ) than in low concentration programs ( $r = .39$ ).

## Evidence of Validity

In this section, we present evidence of concurrent and predictive validity for the P-KOF. To examine concurrent validity, we calculated bivariate correlations between P-KOF subscale scores and scores on direct child assessments<sup>29</sup> in the spring. We examined predictive validity by calculating bivariate correlations between children's P-KOF scores in the fall and scores on direct child assessments in the spring. Throughout our discussion, we compare associations between P-KOF subscales and direct assessments measuring skills in the same domain (convergent validity) and relationships between subscales and direct assessments in different domains (divergent validity).

### Concurrent Validity for the Overall Sample

**P-KOF General Knowledge.** The P-KOF General Knowledge subscale includes items about print knowledge, rhyming, writing one's own first name, naming letters, shapes and colors, and counting 10 objects. General Knowledge was most strongly associated with WJ-III scores ( $r = .43$  to  $.53$ ). We also found strong associations with the PreLAS English and the EOWPVT-SBE. Relationships between General Knowledge and other Spanish measures, including PreLAS Spanish and WM-III, were weaker. In fact, correlations between General Knowledge and Applied Problems and Ortografía on the WM-III were close to zero.

Correlations between spring General Knowledge scores and standardized social-emotional measures were only slightly lower than correlations with standardized measures in language, literacy, and math. In particular, the correlation between General Knowledge and the PKBS-2 positive social skills scales ranged from  $.31$  to  $.42$ . This was unexpected given the types of skills assessed by the items on the General Knowledge scale, but consistent with estimates we obtained in previous analyses (Moiduddin et al. 2010). It is likely that these associations are due in part to shared method variance; that is, both the PKBS-2 and P-KOF scales are obtained through provider ratings

**P-KOF Social Expression.** The P-KOF Social Expression subscale assesses how a child handles frustration, relates with adults and peers, and verbally expresses himself or herself. Social Expression was moderately to highly correlated with the positive social skills scales on the PKBS-2 ( $r = .42$  to  $.66$ ). The relationship between these measures is not surprising because both assess social-emotional development. Although the correlation is likely strengthened due to shared method

---

<sup>29</sup> Chapter II presents the results for the direct child assessments.

variance, it is worth noting that of all the P-KOF scales, PKBS-2 positive social skills scores were most strongly related to Social Expression. Social Expression was weakly correlated with standardized measures of expressive and receptive language and with the WJ-III and WM-III subscales.

**P-KOF Self-Regulation.** The P-KOF Self-Regulation subscale measures a child's ability to stay focused, follow directions, and work and play cooperatively with peers. As expected, this scale was most strongly associated with the PKBS-2. Specifically, we found correlations ranging from .43 to .58 between Self-Regulation and PKBS-2 positive social skills scores. We also found moderate negative associations between Self-Regulation and Externalizing Problems on the PKBS-2 ( $r = -.53$ ). Correlations with the Pencil Tapping task ( $r = .22$ ) and Leiter Examiner Ratings ( $r = .18$  to  $.28$ ) were weak.

Self-regulation was weakly correlated with standardized measures of language, literacy, and math, including the PreLAS English and Spanish and EOWPVT, demonstrating good evidence of divergent validity. We also found low correlations with WJ-III and WM-III, though correlations with some of the WJ-III tests were higher than expected. WJ-III Applied Problems and Spelling were more strongly correlated with Self-Regulation than with the Pencil Tapping task, for example.

**P-KOF Self-Care & Motor Skills.** The P-KOF Self-Care & Motor Skills scale consists of three items measuring a child's use of small manipulatives, general coordination, and ability to perform self-care tasks. We did not expect a strong association between Self-Care & Motor Skills and any of the standardized measures because there was not much overlap in skills. As expected, we found correlations of less than .40 across all standardized measures.

### Relationships Between Change Measured by the P-KOF and by Standardized Measures

We examined the correlations between fall-spring change scores on the P-KOF and change scores on the standardized measures (Table III.15). General Knowledge change scores were weakly correlated with change on the standardized measures in the same domains, including the WJ-III, for which we found evidence of concurrent validity. In fact, change on General Knowledge was more strongly associated with change in the PKBS-2 Social Interaction scale ( $r = .22$ ) and the PreLAS English ( $r = .22$ ) than with change on the WJ-III ( $r = .05$  to  $.11$ ), although all of these correlations were weak in magnitude. Correlations between change on Social Expression and Self-Regulation and change on the PKBS-2 scales were also weak. Most correlations were lower than .20.

### Predictive Validity for the Overall Sample

**P-KOF General Knowledge.** The P-KOF General Knowledge scores in the fall were most strongly related to spring scores on the WJ-III ( $r = .42$  to  $.54$ ). Similar to our concurrent validity findings, we found moderate correlations with the spring PreLAS English and EOWPVT-SBE, and weaker correlations with other Spanish measures (Table III.16).

Fall General Knowledge scores were weakly associated with the spring PKBS-2 positive social skills scales, with correlations ranging from .23 to .30. These correlations are lower than those we found with the spring General Knowledge scores. Correlations with the spring Pencil Tapping Task ( $r = .39$ ) and Leiter Examiner Ratings ( $r = .28$  to  $.29$ ) were also weak, but larger in magnitude than relationships with P-KOF Self-Regulation.

**P-KOF Social Expression.** Fall scores on the P-KOF Social Expression subscale showed evidence of predictive validity with all but one of the PKBS-2 Positive Behavior scales ( $r = .40$  to  $.48$ ). These correlations were lower than those between the spring Social Expression and the same scales on the PKBS-2. Fall Social Expression scores were not strongly associated with any other spring standardized measures. Correlations with the PreLAS and EOWPVT were weak ( $r = .08$  to  $.27$ ) and, as expected, so were correlations with the WJ-III and WM-III scales ( $r = .01$  to  $.31$ ).

**P-KOF Self-Regulation.** Evidence of convergent validity for the Self-Regulation scale was limited. Fall P-KOF Self-Regulation scores were moderately associated with spring scores on the PKBS-2 positive social skills scales ( $r = .42$  to  $.51$ ), but associations with spring problem behavior scales were low ( $r = -.15$  to  $-.33$ ). We also found weak correlations between fall Self-Regulation scores and spring scores on Pencil Tapping ( $r = .27$ ) and Leiter Examiner Ratings ( $r = .20$  to  $.27$ ).

**P-KOF Self-Care & Motor Skills.** As previously stated, there is limited overlap between skills assessed on the P-KOF Self-Care & Motor Skills scale and any of the standardized measures. We did not find strong associations between fall scores on this scale and spring scores on the standardized measures.

### Concurrent Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC)

Correlations between spring P-KOF scores and spring scores on standardized measures were generally higher in FCCs than in center-based programs (Table III.17). The pattern of correlations in centers mirrored those for the overall sample. Specifically, we found moderate associations between General Knowledge and the WJ-III ( $r = .42$  to  $.55$ ), PreLAS English ( $r = .41$ ), and EOWPVT-SBE ( $r = .43$ ). We also found moderate to strong relationships between Social Expression and Self-Regulation and most of the PKBS-2 positive social skills scales ( $r = .40$  to  $.63$ ).

The relationships described above were also apparent in FCCs. In addition, spring PKBS-2 positive social skills scores were moderately to strongly associated with spring scores on all the P-KOF scales ( $r = .43$  to  $.86$ ), not just Social Expression and Self-Regulation, which assess skills in the same domain. We also found evidence of convergent validity between Self-Regulation and Pencil Tapping ( $r = .41$ ) in FCCs that we did not see in center-based programs. However, the correlation between General Knowledge and Pencil Tapping ( $r = .45$ ) was somewhat stronger, though still only moderately strong. These findings, coupled with higher inter-factor correlations in FCCs than centers, suggest that halo effects may be more of a concern in FCCs.

### Relationships Between Change Measured by the P-KOF and by Standardized Measures for Subgroups Based on Program Type (Center-Based and FCC)

Correlations between change scores on the P-KOF and on the standardized measures from fall to spring by program type are displayed in Table III.18. In centers, we did not find strong associations between change scores on the P-KOF scales and those on any of the standardized measures. All of the correlations were lower than  $.40$ , and many were close to zero. In FCCs, we found strong correlations between change on the WM-III Letter-Word Identification standard score and change on General Knowledge ( $r = .55$ ), Self-Regulation ( $r = .53$ ), and Self-Care & Motor Skills ( $r = .59$ ). The correlations with Self-Regulation and Self-Care & Motor Skills were unexpected given the limited overlap in skills assessed. However, it should be noted that these correlations are based on a small sample of children from FCCs with WM-III Letter Word Identification scores ( $n = 26$ ).

Also in FCCs, change on the PKBS-2 positive social skills scales was more strongly correlated with change in General Knowledge than with change in Social Expression or Self-Regulation.

### **Predictive Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC)**

Correlations between fall P-KOF scores and spring standardized measures for centers and FCCs are shown in Table III.19. In centers, we found moderate associations between General Knowledge and the WJ-III ( $r = .42$  to  $.53$ ), as well as the PreLAS English ( $r = .44$ ) and EOWPVT ( $r = .40$  to  $.48$ ). Fall Social Expression scores were moderately correlated with Social Interaction ( $r = .45$ ) and the PKBS-2 Total Positive Social Skills ( $r = .40$ ) in the spring. We found slightly larger correlations between fall Self-Regulation and three of the positive social skills scales on the PKBS-2 ( $r = .40$  to  $.48$ ). Self-Regulation was weakly correlated with the Pencil Tapping task ( $r = .28$ ) and, in fact, General Knowledge demonstrated a stronger association with the same measure ( $r = .40$ ).

In FCCs, we found similar evidence of relationships between General Knowledge and the WJ-III ( $r = .41$  to  $.57$ ), the PreLAS English ( $r = .45$ ), and EOWPVT-SBE ( $r = .45$ ) as in centers. Fall Social Expression scores were more strongly related to PKBS-2 Positive Behaviors ( $r = .50$  to  $.72$ ) in FCCs than in centers. We also found a moderate relationship between PKBS-2 Internalizing Behaviors and Social Expression ( $r = -.40$ ). Fall Self-Regulation scores were more highly correlated with the spring PKBS-2 positive social skills scales ( $r = .61$  to  $.68$ ) and problem behaviors scales ( $r = -.35$  to  $-.58$ ) in FCCs compared to centers. However, the PKBS-2 Internalizing Problems scale was more strongly correlated with General Knowledge than Self-Regulation. We also found moderate correlations ( $>.40$ ) between fall scores on the Self-Care & Motor Skills scale and several scales on the PKBS-2.

### **Concurrent Validity Estimates for Subgroups Based on Concentration of ELLs**

Correlations between spring P-KOF scores and spring scores on standardized measures in programs with high and low concentrations of ELLs are displayed in Table III.20. In general, the patterns of relationships between the P-KOF and standardized measures were the same in settings with varying concentrations. Spring General Knowledge scores were associated with WJ-III scores, with correlations in the low to moderate range ( $r = .32$  to  $.56$ ). Social Expression was most strongly associated with the PKBS-2 positive social skills scales, though correlations were stronger in high concentration programs ( $r = .48$  to  $.73$ ) than in low concentration programs ( $r = .48$  to  $.50$ ). Self-Regulation was also most strongly associated with PKBS-2 scores. In high concentration settings, correlations ranged from  $.44$  to  $.62$  with the positive social skills scales, and the correlation was  $.49$  with Externalizing Problems. Correlations were within a similar range in low concentration settings, although the correlation between Externalizing Problems and Self-Regulation was higher ( $r = -.59$ ). The Self-Care & Motor Skills scale score was not strongly associated with scores on any of the standardized measures.

### **Relationships Between Change Measured by the P-KOF and by Standardized Measures for Subgroups Based on Concentration of ELLs**

In Table III.21 we show the correlations between change scores on the P-KOF and the standardized measures from fall to spring by ELL concentration. In high concentration settings, correlations between P-KOF change scores and change on the standardized measures were all quite low. The strongest associations we found were between General Knowledge and the PreLAS English ( $r = .29$ ), Social Expression and the PKBS-2 Social Interaction ( $r = .33$ ). In low

concentration settings, correlations were similarly low, and some were in an unexpected direction. In particular, we found some negative associations between General Knowledge and W scores on the WM-III.

### **Predictive Validity Estimates for Subgroups Based on Concentration of ELLs**

Correlations between fall P-KOF scores and spring scores on standardized measures for programs with high and low concentrations of ELLs are displayed in Table III.21. As with our concurrent validity estimates, we did not find substantial differences in relationships between the P-KOF and standardized measures across levels of ELL concentration. Fall General Knowledge scores were most strongly associated with WJ-III scores in both high and low concentration programs ( $r = .40$  to  $.53$ ). Social Expression was most strongly related to positive social skills on the PKBS-2 in both high and low concentration settings, with correlations ranging from  $.30$  to  $.54$ . Self-regulation was more strongly related to positive social skills ( $r = .46$  to  $.55$ ) and Externalizing Problems ( $r = -.40$ ) on the PKBS-2 in high concentration settings compared to settings with low concentrations of ELLs, where correlations ranged from  $.36$  to  $.49$  on positive social skills scales, and the  $r$  was  $-.31$  for Externalizing Problems.

### **Summary**

Results show that across language groups, program types, and programs with varying concentrations of ELLs, children performed at a higher level on the P-KOF in the spring than in the fall. Children who had lower scores in the fall made larger gains; this was true for children in the Spanish only language group, in center-based programs, and in programs with high concentrations of ELLs. However, ceiling effects on the P-KOF scales may underestimate the progress made by children who were already scoring near the maximum in the fall. As a result, the P-KOF may also overestimate between-group differences in change scores.

Of all the subscales, the largest increases tended to be on the General Knowledge scale. Distributions of fall and spring scores for each of the scales demonstrate that scores on the General Knowledge scale are more spread out and less skewed than the other scales. This is likely due to the specificity and larger number of items on General Knowledge, as we discussed in a previous report (Moiduddin et al. 2010). In general, the changes we observed on the P-KOF scales were larger than those seen for the standardized measures. Part of the reason for this difference is the narrower distribution of fall P-KOF scores.

In the spring, the P-KOF scales continued to demonstrate strong internal consistency, except for Self-Care & Motor Skills for which we saw very limited variation in spring scores. Concurrent and predictive validity estimates demonstrated some expected relationships between fall and spring scores on the P-KOF scales and with spring scores on standardized measures. However, most of the correlations were low. Further, change on the P-KOF was not strongly associated with change on any of the standardized measures.

The analysis of rater effect completed in the fall indicates fairly substantial differences in the magnitude of the ICCs between the P-KOF and the standardized direct assessments (see Appendix A for an overview). These differences remained when we looked separately by ELL status. The different picture of growth captured across subgroups by the P-KOF compared to the growth across subgroups captured by the standardized assessments adds to the evidence of rater effects and calls into question the validity of the scores for this sample. In addition, using change scores from this measure for program evaluation is not recommended given the problems with the ceiling.

**This page has been left blank for double-sided copying.**



#### **IV. CHILDREN'S PROGRESS AS MEASURED BY THE PROVIDER-ADMINISTERED ASSESSMENT**

The Provider-Administered Assessment (PAA) piloted in UPCOS-3 drew from three sources to collectively assess the key domains of children's school readiness defined by the National Education Goals Panel (Kagan et al. 1995). Sources were the Ages and Stages Questionnaires-3rd Edition (ASQ-3; Squires et al. 2009); the Rapid Letter Naming Task (Atkins-Burnett, Sprachman, and Sonnenfeld 2007) from UPCOS-2; and the Preschool Kindergarten Behavior Scales-2 (PKBS-2; Merrell 2003). To minimize burden on the teachers, we randomly assigned each classroom to only two or three subscales or tests.

From the ASQ-3 we used the subscales for Communication (English and Spanish), Problem Solving, Fine Motor, and Gross Motor. The ASQ-3 was originally designed as a parent-report monitoring system, with age-specific forms ranging from 1 month to 5.5 years. For UPCOS-3 we adapted the measure to be completed by teachers based on direct observation. Each of the adapted subscales included items from the 48-, 54-, and 60-month-old forms, with a total of 9 items (Gross Motor and Problem Solving) or 10 items (English and Spanish Communication and Fine Motor) per subscale in the fall. In the fall analysis, we found some evidence of a potential ceiling effect for the Problem Solving subscale. Therefore, in the spring we added two more items for Problem Solving to reduce the ceiling problem.

The Communication section of the ASQ-3 does not include any measure of literacy, so we also included the Rapid Letter Naming Task from UPCOS-2. Letter naming is a quick, easy-to-administer measure of early literacy that has been a consistent predictor of literacy in the primary grades. The teachers who administered the Problem Solving assessment also administered the Rapid Letter Naming (RLN) Task (30 items). We used alternate forms in the fall and spring.

To collect information about social and emotional development, we asked teachers to complete the PKBS-2. It includes subscales for Social Cooperation (10 items), Social Independence (10 items), Positive Social Interactions (10 items), and both Externalizing (27 items) and Internalizing (15 items) Problem Behaviors, resulting in ratings on a total of 34 positive social skills and 42 problem behaviors.

In this chapter, we present the findings from our analysis of both fall and spring PAA data. We first describe children's progress from fall 2009 to spring 2010. Next, we report children's progress by subgroups: language, program type (centers and FCCs), and concentration of ELLs. We then report evidence of the concurrent and predictive validity of the PAA in comparison to the standardized measures (the Full UPCOS Battery—see Chapter II) for the overall sample and by subgroups. We conclude the chapter with a summary of the major findings for the PAA. One component of the PAA, the PKBS-2, served two purposes, as previously described: (1) It provided a measure of social-emotional development, but (2) as a standardized measure, it also served as a tool in our validity analysis. Thus, we presented children's progress on the PKBS-2 in Chapter II but refer to the PKBS-2 in this chapter when we report evidence of the validity of the PAA. In Table IV.1 in Appendix D, we present the sample sizes for the ASQ-3 subscales and Rapid Letter Naming task overall and by program type and language group. (Appendix D contains all Chapter IV tables

## Children's Progress from Fall 2009 to Spring 2010 for the Overall Sample

In this section we describe the results of our analysis of fall-spring change for the overall sample for each of the elements of the PAA (the ASQ-3 and RLN), show the difference when the ASQ-3 is analyzed for screening scores or cut-off scores, and compare progress based on the PAA with that identified by the standardized measures in the Full UPCOS Battery.

### ASQ-3

According to the teacher-administered ASQ-3 total subscale scores, children made significant progress over the course of the preschool year in Communication, Gross Motor, Fine Motor, and Problem Solving (Table IV.2). The progress in Communication was similar for children assessed in English and Spanish. The average scores for English Communication increased from 54.07 in the fall to 69.99 in the spring ( $ES = .62$ ), and the average ratings in Spanish Communication increased from 55.61 in the fall to 69.69 in the spring ( $ES = .55$ ). When administering the Communication scale, teachers were instructed to administer the English scale in addition to the Spanish scale to all children with enough proficiency to complete it, in their estimation, even if the child's home language was Spanish (or the child spoke primarily Spanish). Similarly, teachers administered the Spanish scale to all children whose home language was Spanish even if the child spoke English only or primarily. Therefore, many children had ratings on both versions. Because routing procedures may have resulted in children being assessed in a language in which they had limited skill, mean scores may be underestimating the communication abilities of children who speak English or Spanish only or primarily.

On average, children's scores increased significantly on Gross Motor and Fine Motor from fall to spring ( $ES = .64$  and  $0.59$ , respectively). We noted that the mean score in the spring was very high (82 out of a total of 90) and upon investigation found that 50 percent of children scored at the maximum, indicating a ceiling problem for Gross Motor (Figure IV.1). Therefore, children's growth on Gross Motor is underestimated with the current measure.<sup>42</sup> The other subscales had skewed distributions but minimal ceiling problems: 11 percent of children scored at the ceiling on Problem Solving, 6 percent on Spanish Communication, 5 percent on English Communication, and 3 percent on Fine Motor. For the RLN, 9 percent of the children scored at the maximum. (Figure IV.1)

Children showed the greatest gains from fall to spring on Problem Solving, with scores increasing from 61.38 in the fall to 86.86 in the spring ( $ES = 1.15$ ). This may be a slight overestimate, however, due to the fact that we added two more items to the Problem Solving subscale in the spring, as noted earlier. By spring, only 11 percent of children scored at the maximum on Problem Solving compared with 15 percent in the fall. For children who were at the ceiling in the fall, we may have underestimated their fall ability and overestimated their change—they may have been able to answer correctly the two new items in the fall as well as in the spring.

The internal consistency coefficients are adequate to strong for the spring ASQ-3 subscales (alpha coefficients range from  $.72$  for Problem Solving to  $.90$  for Spanish Communication).

---

<sup>42</sup> Although we found some evidence of ceiling problem for Gross Motor in the fall (with 27 percent of children scoring at the maximum), we did not add more difficult items because space limitations would make them challenging to administer in some of the programs, and programs do not have a strong focus on this area of development.

## Rapid Letter Naming

Teachers reported that children increased the number of letters that they named correctly from fall to spring (Table IV.2). The Rapid Letter Naming Task was conceptually scored; that is, children received credit for any letter they named whether they named it in English or Spanish. The task used two different forms with 30 letters that included all 26 letters in either upper or lower case, and 4 additional letters in the opposite case to aid in creating an overlap between forms. The score used for analysis is a raw sum score. Children scored a mean of 12.79 out of 30 letters in the fall; by spring, the number increased to 19.37 (ES = .61). The internal consistency is excellent for Rapid Letter Naming ( $\alpha = .96$ ).

## ASQ-3 Screening Scores and Cut-Off Scores

We also calculated two other scores for assessing children's ASQ-3 performance: screening and cut-off scores. The screening scores are based on the subset of items that are age appropriate for screening for developmental problems (6 items per area). The reliability estimates are adequate to strong for the ASQ-3 Communication and Problem Solving screening scores in the spring ( $\alpha = .70$  to  $.85$ ). However, with smaller samples of children per six items, the reliability estimates are lower for Gross Motor screening scores ( $\alpha = .63$  to  $.70$ ). The reliability estimates are poor for Fine Motor screening scores ( $\alpha = .32$  to  $.53$ ) (Table IV.3). Because of ceiling effects inherent in screening scores, we did not examine children's progress on these scores.

The cut-off scores for the screening instrument are based on a nationally representative sample of children across a range of socioeconomic status; the majority of the standardization sample has English as a first language. In our sample, all children were scored on the English Communication, though any child who came from a home where Spanish is spoken was assessed in Spanish as well (if the provider was fluent enough to assess the child in Spanish). This may mean that some children were assessed in a language in which they had limited skills. With these caveats in mind, the proportion of children who were identified as needing continued monitoring and followup decreased substantially from fall to spring (Table IV.4). In the fall, approximately 40 percent of the children scored in the at-risk range on Communication (2 or more standard deviations below the mean) in both English and Spanish. By spring, the at-risk rate dropped to 25 percent for English Communication. However, the proportion of children who scored in the at-risk range on Spanish Communication stayed about the same. The percentage of children who scored in the at-risk range on Problem Solving, Gross Motor, and Fine Motor decreased from 10-14 percent in the fall to 4-5 percent in the spring.

## Comparisons with Progress Identified by the Standardized Measures

Generally, the magnitude of children's progress measured by the PAA was similar to progress captured by the standardized measures (W or IRT score) in the same domain. The effect sizes of fall-to-spring change on the ASQ-3 English Communication, and Rapid Letter Naming were around  $.60$ , which was comparable to the effect sizes for the EOWPVT IRT score and WJ-III Letter-Word Identification and Spelling (ES =  $.44$ -. $60$ ). The effect size for the ASQ-3 Fine Motor was equivalent to that for WJ-III Spelling. Although slightly higher, the effect size for Spanish Communication (ES =  $.55$ ) was analogous to that for WM-III Letter-Word Identification and Spelling (around  $0.4$ ). The only exception is Problem Solving, for which the effect size was twice as large as that for WJ-III/WM-III Applied Problems. As discussed above, we added two more items for Problem Solving in the spring, which could be the reason for indicating greater progress in this area.

## Children's Progress by Subgroups

### Language Subgroups

As Table IV.5 shows, almost all language groups made significant progress from fall to spring on all areas of the ASQ-3 and Rapid Letter Naming. The only exception was the Gross Motor score for Spanish only children, which increased from 73.30 in the fall to 83.47 in the spring ( $ES = .52$ , Table IV.6), but the small sample size for this group ( $n = 11$ ) limited the ability to detect significance. The sample sizes for the other language group are small (mostly under 10), thus we excluded that group from the analysis to protect confidentiality and to provide more stable estimates.

When comparing gains on the ASQ-3 subscales across language groups, we found that, similar to findings on the PreLAS English and Spanish, the Spanish only group made the greatest progress on English and Spanish Communication when compared with the other groups. The Spanish primarily group made the greatest progress on Problem Solving among the language groups, followed by the Spanish only, English primarily, and English only groups. The Spanish only, Spanish primarily, and English primarily groups showed greater progress on Rapid Letter Naming than the English only group, with the greatest gains in the Spanish Primarily group—a pattern similar to the gains on the WJ-III Letter-Word Identification.<sup>43</sup> All language groups showed similar amounts of progress in the Gross Motor and Fine Motor areas (Table IV.6).

Overall, the patterns of differences in children's progress on language and literacy across language groups were similar to those captured by the standardized measures. The Spanish only and Spanish primarily groups made greater progress on PreLAS and WJ-III Letter-Word Identification than the other groups.

The reliability estimates were adequate to strong for the ASQ-3 scores across language groups ( $r = .66$  to  $.92$ ), with the exception of the Spanish only group on Gross Motor (Table IV.7), for which the alpha coefficient was very low ( $\alpha = .25$ ). It should be noted that only 14 children in this group had Gross Motor scores, so the estimate for this group was not as reliable. Rapid Letter Naming scores had similarly excellent internal consistency across language groups.

### Program Subgroups: Center-Based and Family Child Care Settings

Children in center-based programs made significant progress from fall to spring in all areas of the ASQ-3 and Rapid Letter Naming (Table IV.8). The gains for children in FCCs were also significant in ASQ-3 English and Spanish Communication, Problem Solving, and Rapid Letter Naming. FCC children's Fine Motor scores increased from 60.29 in the fall to 75.71 in the spring ( $ES = .63$ , Table IV.9), and Gross Motor scores increased from 63.22 in the fall to 74.33 in the spring ( $ES = .57$ , Table IV.9), but these did not reach statistical significance, probably due to the small sample for this group on Fine Motor ( $n = 14$ ) and Gross Motor ( $n = 13$ ).

Comparing the magnitudes of fall-spring progress of center-based programs and FCCs, children in centers made greater progress on English Communication than did children in FCCs ( $ES = .70$

---

<sup>43</sup> On the WJ-III and WM-III Letter-Word Identification, children must name the letter (or word) correctly in the same language as the administration. Responses naming the letter correctly in English when the WM-III (Spanish) is administered are scored incorrect. On the RLN, correct answers in either language are accepted as correct.

and .38 for centers and FCCs, respectively). Children's gains in other ASQ-3 areas and Rapid Letter Naming are similar in both types of programs (Table IV.9).

The difference noted between center-based and FCCs in children's progress on the ASQ-3 English Communication was not seen on the standardized measures. According to standardized measures, there were no differences in children's progress in any domains between centers and FCCs.

The reliability estimates were adequate to strong for center-based programs ( $r = .73$  to  $.89$ ). For FCCs, the reliability estimates were lower on Problem Solving and Fine Motor (with  $r$  around  $.65$ ), but still in the acceptable range. The reliability estimates for Rapid Letter Naming were excellent for both types of programs (Table IV.10).

### **Program Subgroups: High and Low Concentrations of ELLs**

Children in both high and low ELL concentration programs showed significant gains from fall to spring in all areas of ASQ-3 and Rapid Letter Naming (Table IV.11).

When comparing the magnitude of children's progress from fall to spring in the two groups, we found that children in high ELL concentration programs made greater progress than did children in low ELL concentration programs on English Communication (ES =  $.78$  and  $.47$  for high and low concentration programs, respectively), Problem Solving (ES =  $1.25$  and  $1.07$  for high and low concentration programs, respectively), and Rapid Letter Naming (ES =  $.70$  and  $.51$  for high and low concentration programs, respectively). Children in both types of programs made similar progress on Spanish Communication, Gross Motor, and Fine Motor (Table IV.12).

The patterns of differences in children's progress on the PAA for high and low ELL concentration programs did not consistently match the patterns of differences on standardized measures. For language development, children in high ELL concentration programs also gained more on the English PreLAS than did children in low concentration programs (ES =  $.49$  and  $.34$  for high and low ELL concentration programs, respectively; a  $.15$  difference in effect sizes), but the magnitude of the difference was smaller than that for the ASQ-3 English Communication (a  $.31$  difference in effect sizes). Moreover, standardized measures indicate that children in high and low ELL concentration programs made similar progress on the Spanish PreLAS, EOWPVT, WJ-III Letter-Word Identification, and Applied Problems. These differences may be related to differences in the language proficiencies of the teachers in the programs.<sup>44</sup>

The reliability estimates are similarly adequate to strong in the two groups for all ASQ-3 subscale scores and Rapid Letter Naming (Table IV.13).

## **Inter-Factor Correlations for Spring Data**

### **Inter-Factor Correlations in the Overall Sample**

We examined the inter-factor correlations for the ASQ-3 (Table IV.14). Some of the correlations between subscales could not be calculated because children did not have ratings on all

---

<sup>44</sup> Mathematica assessors were fluent in both languages.

subscales, as a result of the study design.<sup>45</sup> The inter-factor correlations among the subscales were in the moderate to high range ( $r = .34$  to  $.62$ ), with the two correlations for Problem Solving slightly higher ( $r = .62$  and  $.58$  with English Communication and Fine Motor, respectively).

### Inter-Factor Correlations by Subgroup

**Center-Based and FCC Programs.** In both centers and FCCs, English Communication was moderately correlated with Problem Solving ( $r$ s around  $.60$ ). The remaining inter-factor correlations were higher in centers than in FCCs. The correlations in centers were in the moderate to high range ( $r = .40$  to  $.61$ ), while the correlations in FCCs were in the low range ( $r = .03$  to  $.33$ ; Table IV.15).

**High and Low ELL Concentrations.** The correlations between Gross Motor and Fine Motor with Problem Solving were higher in low ELL concentration programs than in high ELL concentration programs ( $.39$  to  $.66$  in low ELL concentration programs and  $.31$  to  $.36$  in high ELL concentration programs). The rest of the inter-factor correlations were higher in high concentration programs than in low concentration programs (ranging from  $.62$  to  $.76$  in high concentration programs and from  $-.03$  to  $.52$  in low ELL concentration programs; Table IV.16). This again may be partly a function of the language proficiency of the teachers across programs. The lowest inter-factor correlations were with the Spanish Communication in low ELL classrooms.

## Evidence of Validity

In this section, we present the validity evidence by looking at concurrent and predictive validity, as well as convergent and discriminant validity. For evidence of concurrent validity, we calculated bivariate correlations between children's PAA scores and scores on the direct child assessments<sup>46</sup> in the spring. For evidence of predictive validity, we calculated bivariate correlations between children's PAA scores in the fall and their scores on the direct child assessments in the spring. We also examined evidence of convergent and discriminant validity by looking at bivariate correlations between children's PAA scores and scores on direct child assessments in the same or different domains.

### Concurrent Validity for the Overall Sample

**ASQ-3 Communication.** The ASQ-3 Communication subscales (English and Spanish) include assessment of both receptive and expressive language and look at morphology, syntax, and pragmatics. The ASQ-3 English Communication subscale had moderate correlations with the English PreLAS ( $r = .53$ ) and EOWPVT IRT score ( $r = .45$ ). The correlation with WJ-III Letter-Word Identification W scores was low ( $r = .18$ ) due to the fact that WJ-III Letter-Word Identification measures letter knowledge rather than the oral communication assessed by the ASQ-3<sup>47</sup> (Table IV.17).

As expected, the correlations between ASQ-3 English Communication and the standardized measures of other constructs were typically weaker. Correlations with the Pencil Tapping Task or

---

<sup>45</sup> The gross and fine motor scales were not administered to children who responded to the English and Spanish Communication scales.

<sup>46</sup> Chapter II presents the results for the direct child assessments.

<sup>47</sup> We specifically added RLN to measure literacy because the PAA Communication scales did not include literacy items.



the Leiter Examiner Rating scaled scores were around .3, which were lower than the correlation of ASQ-3 English Communication with English PreLAS scores ( $r = .53$ ) or with EOWPVT IRT scores ( $r = .45$ ). The correlations of the ASQ-3 English Communication with the PKBS-2 Positive Social Skills ( $r = .20$  to  $.30$ ) and Problem Behaviors (virtually not correlated) or with the WJ-III Applied Problems and Spelling scores (around  $.30$ ) were also lower than the correlations between ASQ-3 English Communication and English PreLAS scores or with EOWPVT IRT scores.

The Spanish Communication scale does not have good evidence of concurrent validity. The ASQ-3 Spanish Communication had low correlations with the standardized Spanish measures of language and expressive vocabulary ( $r = .19$  to  $.25$ ; Table III.17). However, the correlations of Spanish Communication with Spanish PreLAS or EOWPVT IRT scores were similar to its correlations with the Leiter Examiner Ratings, Pencil Tapping Task, or with the WM-III Applied Problems and Spelling (with correlations around  $.20$ ). The correlation between Spanish Communication and WM-III Letter-Word Identification was negative, although very low ( $r = -.10$ ). Interestingly, the correlations of Spanish Communication with the PKBS-2 positive social skills scales ( $r = .30$  to  $.50$ ) were the highest among all the correlations.

**ASQ-3 Problem Solving.** The ASQ-3 Problem Solving subscale also demonstrated moderate relationships with the standardized measures of Language, Expressive Vocabulary, Literacy, and Mathematics for children assessed in English (with correlations around  $.40$  to  $.50$ ; Table III.17). The ASQ-3 Problem Solving was not correlated with Total Problem Behaviors, although the correlations with standardized measures of social skills ( $r = .30$  to  $.50$ ) were close to its correlations with language, literacy, and mathematics, indicating weaker evidence of discriminant validity.

**Rapid Letter Naming.** Rapid Letter Naming had moderate to high correlations with English PreLAS, EOWPVT IRT scores, and WJ-III Letter-Word Identification and Spelling ( $r = .40$  to  $.60$ ). These correlations were higher than the correlations between Rapid Letter Naming and standardized social-emotional measures, which were in the low range ( $r = .10$  to  $.30$ ) for social skills and trivial for problem behaviors. The correlations of Rapid Letter Naming and WM-III Letter-Word Identification and Spelling were in the low range ( $r = .20$  to  $.30$ ).

### Relationships Between Change Measured by the PAA and by the Standardized Measures

We examined the correlations between change scores on the PAA and on the standardized measures from fall to spring (Table IV.18). The correlations were in the low range for the ASQ-3 English Communication and Problem Solving with standardized measures in the same domains. These correlations were lower than the correlations with change scores on the PKBS-2. The change on Spanish Communication was moderately correlated with change on the EOWPVT IRT score and WM-III Letter-Word Identification. The change on Rapid Letter Naming was moderately correlated with change on WJ-III Letter-Word Identification.

### Predictive Validity for the Overall Sample

**ASQ-3 Communication.** The fall ASQ-3 English Communication scores had moderate to high correlations with spring English PreLAS and EOWPVT IRT scores ( $r = .66$  and  $.56$  for PreLAS and EOWPVT, respectively). These correlations were higher than the correlations with WJ-III Applied Problems and Spelling and standardized social-emotional measures, indicating evidence of convergent and discriminant validity. However, the correlation with WJ-III Letter-Word Identification was low ( $r = .28$ ); it was even lower than the correlations with WJ-III Applied

Problems and Spelling ( $r = .48$  and  $.38$  for Applied Problems and Spelling, respectively; Table IV.19).

The ASQ-3 Spanish Communication lacks evidence of predictive validity. The fall ASQ-3 Spanish Communication scale had low correlations with the spring standardized Spanish measures of language and expressive vocabulary ( $r = .23$  to  $.30$ ; Table IV.19). These correlations were lower than its correlations with the spring PKBS-2 positive social skills ( $r = .34$  to  $.59$ ). A number of factors may contribute to these findings. Children in the fall were assessed in Spanish if any Spanish was spoken in their homes. In the spring, only those children with limited proficiency in English were assessed in Spanish; however, the PKBS-2 scores were available on children independent of their language proficiency. Children who were bilingual may have more opportunities to socially interact with peers, particularly in high ELL concentration programs.

**ASQ-3 Problem Solving.** The fall ASQ-3 Problem Solving scores were moderately to highly correlated with spring standardized measures of language, literacy, and mathematics in English ( $r = .50$  to  $.65$ ; Table IV.19). These correlations were higher than the correlations with standardized social-emotional measures in the spring. However, the correlations with Spanish standardized measures were low to zero.

**Rapid Letter Naming.** Fall Rapid Letter Naming scores were moderately to highly correlated with spring WJ-III Letter-Word Identification and Spelling ( $r = .52$  and  $.65$ , respectively). These correlations were higher than the correlations with standardized social-emotional measures in the spring. However, the correlations with spring WM-III Letter-Word identification and Spelling were low to zero (Table IV.19).

### **Concurrent Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC)**

When examined by program type, the correlation between spring ASQ-3 English Communication and English PreLAS was higher in FCCs ( $r = .65$ ) than in centers ( $r = .50$ ). The correlation of Spanish Communication with EOWPVT IRT scores was also higher in FCCs than in centers ( $r = .30$  versus  $.14$ ). The correlations between English Communication and EOWPVT IRT scores and between Spanish Communication and Spanish PreLAS were similar in centers and FCCs. Centers and FCCs had a similar range of correlations between spring ASQ-3 Problem Solving and English PreLAS, expressive vocabulary, literacy, and mathematics ( $r = .45$  to  $.54$ ). The correlation of spring Rapid Letter Naming and WJ-III Letter-Word Identification was moderately strong in both types of programs ( $r = .65$  and  $.58$  for centers and FCCs, respectively), while the correlation of Rapid Letter Naming and WM-III Letter-Word Identification was higher in FCCs than in centers ( $r = .52$  versus  $.20$ ; Table IV.20).

Spanish Communication had low to moderate correlations with social-emotional standardized measures in centers, but was highly correlated with PKBS-2 in FCCs ( $r = .81$  to  $.93$  for social skills and  $-.49$  to  $-.79$  for problem behaviors).

### **Relationships Between Change Measured by the PAA and by Standardized Measures for Subgroups Based on Program Type (Center-Based and FCC)**

The correlations between change measured by the PAA and by standardized measures were low in centers, but in FCCs there were moderate to high correlations between change on Spanish Communication and the WJ-III Letter-Word Identification, WM-III Letter-Word Identification, and



WM-III Spelling. We also found low to moderate correlations between Problem Solving and WM-III Letter-Word Identification, Applied Problems, and Spelling (Table IV.21). The correlations between change in Problem Solving and change in PKBS-2 Social Skills are also moderately high in FCCs.

### **Predictive Validity Estimates for Subgroups Based on Program Type (Center-Based and FCC)**

The correlation of fall ASQ-3 English Communication with English PreLAS was higher in FCCs than in centers ( $r = .80$  versus  $.62$ ), while the correlation of English Communication with EOWPVT was higher in centers than in FCCs ( $r = .58$  versus  $.47$ ). These correlations were higher than the correlations with WJ-III Applied Problems, and social-emotional measures in both types of programs. FCCs and centers had similar correlations between English Communication and WJ-III Letter-Word identification (with correlations around  $.30$ ). In both centers and FCCs, spring Spanish Communication had low to moderate correlations with standardized measures in the same domains (Table IV.22).

The evidence of predictive validity for ASQ-3 Problem Solving and Rapid Letter Naming was stronger in centers than in FCCs (Table IV.22). The correlations of fall Problem Solving with language, literacy, and mathematics standardized measures were higher in centers ( $r = .54$  to  $.68$ ) than in FCCs ( $r = .45$  to  $.62$ ). Rapid Letter Naming scores were more highly correlated with EOWPVT, WJ-III Letter-Word Identification, and Spelling in centers ( $r = .52$  to  $.69$ ) than in FCCs ( $r = .32$  to  $.50$ ). The correlations of PAA scores with standardized measures in the same domains were higher than the correlations with social-emotional standardized measures in both types of programs, indicating evidence of discriminant validity.

### **Concurrent Validity Estimates for Subgroups Based on Concentration of ELLs**

The correlations between spring ASQ subscale scores and scores on standardized measures were in the moderate range in both high ELL concentration programs ( $r = .36$  to  $.56$ ) and low ELL concentration programs ( $r = .38$  to  $.55$ ; Table IV.23). However, Problem Solving was also moderately correlated with social-emotional measures in both high and low ELL concentration programs ( $r = .36$  to  $.47$ ), indicating weaker evidence of discriminant validity.

### **Relationships Between Change measured by the PAA and by Standardized Measures for Subgroups Based on Concentration of ELLs**

The correlations between change in the ASQ-3 English Communication and change in the standardized measures were higher in high ELL concentration programs (although still in the low range) than in low ELL concentration programs. In low ELL concentration programs, the change in Spanish Communication was highly correlated with change on WJ-III measures ( $r = .73$  to  $.94$ ), but the calculations were based on only five cases (Table IV.24). The change in Problem Solving had moderate to high correlations with change in WM-III measures in low ELL concentration programs, but the calculations were also based on a small number of cases ( $n = 4$ ).

### **Predictive Validity Estimates for Subgroups Based on Concentration of ELLs**

The correlations of fall ASQ-3 English Communication with spring English PreLAS and EOWPVT IRT scores and the correlation of fall Problem Solving with spring EOWPVT IRT scores were higher in high ELL concentration programs ( $r = .55$  to  $.68$ ) than in low ELL concentration

programs ( $r = .47$  to  $.57$ ), but the correlations of fall Problem Solving scores with spring standardized measures of literacy and mathematics and the correlation between fall Rapid Letter Naming and spring WJ-III Letter-Word Identification were in the same range in high ELL concentration programs ( $r = .50$ -.68) and low ELL concentration programs ( $r = .50$  to  $.62$ ; Table IV.25). These correlations with language and cognitive measures were higher than the correlations with social-emotional measures, suggesting good evidence of discriminant validity.

## Summary

According to teacher-administered PAA, children made progress from fall to spring on the ASQ-3 subscales of Communication, Gross Motor, Fine Motor, and Rapid Letter Naming for the overall sample. The mean growth measured by the PAA was comparable to that measured by the standardized assessments we administered in the Full UPCOS Battery.

Spanish only children made the greatest progress compared to children in other language groups. Children in centers and FCCs had similar gains in Rapid Letter Naming and all ASQ-3 subscales except for English Communication, where children in centers gained more than did children in FCCs. Children in high ELL concentration programs made greater progress on English Communication, Problem Solving, and Rapid Letter Naming than did children in low ELL concentration programs.

The PAA (ASQ-3 and RLN) demonstrated adequate to strong internal consistency and moderate to strong evidence of concurrent and predictive validity for ASQ-3 English Communication, ASQ-3 Problem Solving, and Rapid Letter Naming. Measurement of Spanish language was very weak; while the Rapid Letter Naming provided a strong indication of literacy. However, the increased inter-factor correlations in the spring, along with the somewhat higher ICCs compared to the standardized direct assessments in the fall (see Appendix A) particularly for language specific measures, suggest that rater effects are also a concern for the PAA.

## V. SUMMARY AND CONCLUSIONS

The goal of this study was to learn whether a brief set of measures could be administered by center-based teachers and FCC providers and yield data that approximate the scores obtained on standardized direct assessments. We selected a representative random sample of LAUP *programs* for the study, but we had a convenience sample of *children* within the programs. Although we invited all children in the selected centers and FCC programs to participate, the sample consisted of the first 10 children in center-based programs and the first 4 children in FCC programs whose parents returned a consent form. Thus, it is possible that the children in the study sample lived in families that were more engaged with the program than those that were late in returning their consent forms. As a consequence, we might expect higher levels of growth across the year in this study than we would find in a truly representative sample of children.

### Findings from Fall 2009 Data

In our fall report (Moiduddin et al. 2010) we summarized child outcomes measured soon after the official start of the 2009-2010 school year<sup>36</sup>. In addition, we presented the initial psychometric performance of the teacher-administered measures as indicated by assessments of internal consistency reliability, validity, and rater effects. We briefly summarize those findings here.

Our fall analysis of the standardized assessments indicated that on average, the children in the sample for this study who took the assessments of literacy (WJ-III Letter-Word Identification, Spelling) and mathematics (WJ-III Applied Problems) in English did as well as other children of the same age nationally (Standard Score [SS] = 101 to 104). The children who took the assessments in Spanish (WM-III Letter-Word Identification, Spelling, and Applied Problems) did not do as well as other children of their age (SS = 83 to 94). The conceptual vocabulary of the children in this sample was below average (SS = 89) compared with a nationally representative sample of English-speaking children, but above average (SS = 106) when compared to a nationally representative Spanish-speaking sample. Social-emotional development as measured by the PKBS-2 fell within an average range (SS = 97), and the children in this sample exhibited fewer problem behaviors (SS = 93) than a nationally representative sample (Moiduddin et al. 2010).

Our analysis of the relationships in the fall assessments found adequate reliability for all measures, and significant relationships between the teacher-administered measures and the independently administered standardized assessments. However, we also found indications of potential rater effects, particularly for the P-KOF, and ceiling effects. The evidence of convergent validity with assessments administered in English was weak to moderate for the P-KOF and moderate to strong for the PAA. Validity evidence for assessments administered in Spanish was weak for the P-KOF and weak to moderate for the PAA (Moiduddin et al. 2010).

### Fall to Spring Change

The majority of the independently administered and teacher-administered measures indicated a significant increase in the skills of children in LAUP from fall to spring. Effect sizes (for the magnitude of the fall-spring change) were greater than .50 for scores capturing absolute change for

---

<sup>36</sup> Some programs started in the summer before the LAUP funding start time in September.

most standardized assessments in English<sup>37</sup> and greater than .60 for the teacher-administered assessments in English. For assessments administered in Spanish or bilingually, the effect sizes are smaller for both the standardized assessments (ES for scores measuring absolute change range from .28 to .53) and for Spanish communication on the PAA (ES = .55). When adjusting for the age of the children and expected change over time (that is, when using standard scores), the effect sizes were much smaller (ES = .09 to .21), and not significant for the EOWPVT using English norms or for the WM-III subtests (although these Spanish measures were administered to smaller samples).

## Ceiling Effects

The ceiling effect on the P-KOF was much more pronounced in the spring than in the fall, with 78 percent of children receiving the maximum score on Self-Care & Motor Skills, 53 percent on Social Expression, 38 percent on Self-Regulation, and 23 percent on General Knowledge.

On the PAA, only the ASQ Gross Motor scale had a substantial ceiling problem, with 50 percent of children scoring at the maximum. The other scales had skewed distributions but minimal ceiling problems: 11 percent of children scored at the ceiling on Problem Solving, 6 percent on Spanish Communication, 5 percent on English Communication, and 3 percent on Fine Motor. For the RLN, 9 percent of the children scored at the maximum.

## Validity

We examined the relationships (bivariate correlations) between the spring scores on the teacher-administered measures and spring scores on the independently administered assessments to examine concurrent validity. We also compared the magnitude of change noted in different subgroups on both the teacher-administered and independent assessments and examined the bivariate correlations between change scores to explore whether the independent assessments and the teacher assessments identified change in the same children.<sup>38</sup> In addition, we looked at evidence of predictive validity for the teacher-administered measures, examining correlations between fall scores on the teacher administered assessments and spring scores on the standardized assessments.

### Evidence of Validity for the P-KOF

Evidence of concurrent validity for the P-KOF was similar in the spring compared to the fall. General Knowledge continued to be the P-KOF scale most strongly associated with standardized language and cognitive measures ( $r = .43$  to  $.53$  with measures in English and the EOWPVT-SBE). It was weakly related or not related to the Spanish measures ( $r = -.06$  to  $.28$ ). Change scores for General Knowledge were weakly associated with change scores for the PKBS-2 Social Interaction scale and the PreLAS English scale ( $r = .22$ ), but not significantly related to the WJ-III subtests ( $r = .05$  to  $.11$ ). The P-KOF Social Expression and Self-Regulation scales were moderately to strongly related to the PKBS-2 positive social skills ( $r = .42$  to  $.66$  for concurrent validity). The P-KOF and the PKBS-2 were based on ratings by the same teacher and thus share method variance.

---

<sup>37</sup> The English PreLAS ES = .41.

<sup>38</sup> It is important to note that change scores typically have lower reliability than point-in-time scores. Thus, we expected weaker relationships between the change scores on the teacher-administered and independently administered assessments.

We found some evidence that the validity of the change scores may differ by subgroup. We found greater between-group differences in change scores by language group on the P-KOF when compared with the standardized assessments; that is, estimated change was more likely to differ by language group for the P-KOF than for the standardized assessments. This may be due in part to ceiling effects on the P-KOF scales; children in the Spanish only group had the lowest scores in the fall and thus had the most opportunity for growth. Similarly, fall-to-spring change scores were larger for children in center-based programs than for children in FCC programs on all subscales except General Knowledge (where there was less of a ceiling problem). This is related to the differences noted in the fall scores. The scores of children in center-based programs and children in FCC programs grew closer together in the spring.

Though associations between measures of skills in the same domain typically grow weaker with more time elapsed between assessments in early childhood, the evidence of predictive validity of the fall P-KOF scores to the spring standardized measures was similar to the convergent validity evidence. The ceiling problems and limited variance noted for the P-KOF in the spring<sup>39</sup> may have contributed to this finding and certainly contributed to the weak correlations between change scores. Among the P-KOF scales, General Knowledge again had the strongest relationship with the WJ-III measures ( $r = .42$  to  $.54$ ), the PreLAS English ( $r = .44$ ), and the EOWPVT-SBE ( $r = .39$  to  $.52$ ). The P-KOF Social Expression and Self-Regulation scales were moderately correlated with the spring PKBS-2 Positive Behaviors scales ( $r = .40$  to  $.51$ ).

### **Evidence of Validity for the PAA**

The evidence of concurrent validity for the cognitive and language portions of the PAA remained moderate to strong in the spring ( $r = .45$  to  $.62$  with standardized scales in the same domain) across the scales, although relationships were slightly weaker in the spring than in the fall.<sup>40</sup> The strongest relationship was again between the scales with the greatest similarity of content, the RLN and the WJ-III Letter-Word Identification.

Evidence from the change scores was more mixed, though positive overall. Change scores on RLN and the Spanish Communication scale were moderately correlated with change scores on the standardized measures in the same domain of development. We found low correlations between change scores on the English Communication and Problem Solving scales and the standardized measures in the same domain. On most subscales, the magnitude of change noted on the PAA was similar to the progress captured by scores on the standardized measures that captured absolute progress in the same domain. The exception is the ASQ-3 Problem Solving scale, which had an effect size much larger than the effect sizes for the WJ-III/WM-III Applied Problems. However, this is likely due to changes made to the ASQ-3 Problem Solving in the spring. The fall administration of the problem-solving measure indicated a potential ceiling problem so two items were added in the spring, potentially inflating the change for some of the children.<sup>41</sup>

---

<sup>39</sup> When variance is limited, correlations are generally lower. The variance among children who are at the ceiling on an assessment is not measured, thus limiting the overall variance on that measure

<sup>40</sup> This may be due to the skew in the PAA scores limiting the variance when compared with the adaptive standardized assessments. However, it is not as serious as a ceiling problem.

<sup>41</sup> Fifteen percent of the children had the maximum score in the fall. Eleven percent of the children were at maximum score in the spring. It is possible that at least some of these children could have correctly answered the two new items in the fall.

The only notable subgroup difference in the pattern of estimates of fall to spring change (when comparing the PAA to the independently administered assessments) was between center-based and FCC programs on the ASQ-3 English Communication scale. We found subgroup differences on English Communication that were not evident on the standardized measures. This suggests that children in centers may respond differently to their teachers on English Communication than do children in FCCs at one of the time points, or that the teachers in centers and FCCs have different proficiency in administering or rating children's English Communication.

Similar to the findings for concurrent validity, the evidence of predictive validity was moderate to strong for the ASQ-3 English Communication, ASQ-3 Problem Solving, and the RLN. We found weak to moderate correlations between the fall ASQ-3 Spanish Communication and the spring standardized measures of language and expressive vocabulary ( $r = .23$  to  $.30$ ). The ASQ-3 Spanish Communication scale was administered to all children who lived in homes where any Spanish was spoken, and many were not proficient in Spanish. This likely contributed to the lower correlations.

## Measurement Concerns

Although we identified many strengths in the measures administered in UPCOS-3, measurement of the cognitive and language skills of children who are not proficient in English remains problematic. Overall, we found more-limited relationships between the teacher-administered measures and the standardized Spanish measures. It is possible that this is due to the limited language proficiency of some teachers. However, the administration procedures on some of the standardized measures may also have contributed to lower correlations. The independently administered standardized measures of cognition required either Spanish or English administration.<sup>42</sup> This was particularly problematic for the WJ-III/WM-III Letter-Word Identification test. In many classrooms, children were learning letter names in English, but the test required them to provide the letter names in Spanish. A much lower correlation was noted between the conceptually scored RLN and the WM-III Letter-Word ( $r = .28$  to  $.29$ ) than with the WJ-III Letter-Word ( $r = .59$  to  $.62$ ). Thus, it is not clear whether some of the weaker relationships are due more to the inflexibility in the language of assessment on the standardized measures or to differences in teachers' Spanish fluency.

## Limitation

We conducted multiple significance tests and did not adjust significance for the number of tests conducted so it is possible that some of our findings were spurious. That is, we may have identified some significant differences or relationships that were due to chance.

## Conclusion

Validity is not inherent in a measure, but in its use. Overall, the results indicate that the P-KOF is not a good candidate for use in setting benchmarks related to children's fall-spring progress. The analyses found weaker evidence of convergent validity for the P-KOF compared to the PAA, ceiling problems particularly in the spring, limited variance in the spring, potential rater effects noted in the fall along with subgroup differences in change noted on the P-KOF but not on the independently administered assessments. The correlations of the change scores on the P-KOF and the standardized measures were weak.

---

<sup>42</sup> Bilingual assessors administered the standardized assessments to children who spoke Spanish.

The PAA has evidence of stronger reliability and validity in both fall and spring than the P-KOF. Change scores were moderately correlated for two of the scales, but the results indicate lower correlations of English Communication and Problem Solving with fall-spring change scores on their related standardized measures. Available evidence of predictive validity was relatively strong for all scales except Spanish Communication. Ceiling effects on the PAA were minimal, though concerns about rater effects persist. With additional training provided to teachers, it is possible that measurement using the PAA would be stronger. Nevertheless, the issues of teachers' language fluency and the language used for assessment remain a concern. In addition, the rater effects of a teacher-administered assessment are likely to increase when assessments are used for accountability purposes.

**This page has been left blank for double-sided copying.**



**REFERENCES**

- Applied Survey Research. *Pre-Kindergarten Observation Form 2009*. Santa Cruz, CA: Author, 2009.
- Applied Survey Research. *School Readiness in San Mateo County: Results of the 2008 Assessment*. Santa Cruz, CA: Author, October 2009.
- Bacon, Donald. "The Contributions of Reliability and Pretests to Effective Assessment." *Practical Assessment, Research & Evaluation*, vol. 9, no. 3, 2004.
- Brownell, R. "Expressive One-Word Picture Vocabulary Tests." San Antonio, TX: Harcourt Assessment, Inc., 2000.
- Duncan, Sharon, and Edward DeAvila. "Pre-Language Assessment Scales [PreLAS2000]." Monterey, CA: CTB-McGraw Hill, 2002.
- Mather, Nancy, and Richard W. Woodcock. "Woodcock-Johnson III Tests of Achievement Examiner's Manual." Itasca, IL: Riverside Publishing, 2001.
- Merrell, Kenneth W. "The Preschool and Kindergarten Behavior Scales – Second Edition." Austin, TX: PRO-ED, 2003.
- Moiduddin, Emily, Sally Atkins-Burnett, Yange Xue, Pia Caronongan, Elisha Smith, and Marta Induni. "Results of Activities Informing the Performance-Based Contract Between First 5 LA and LAUP." Report submitted to First 5 LA. Princeton, NJ: Mathematica Policy Research, June 2010.
- Kisker, E.E., Boller, K., Nagatoshi, C.T., Sciarrino, C., Jethwani, V., Zavitsky, T., Ford, M., & Love, J.M. "Resources for measuring services and outcomes in Head Start programs serving infants and toddlers." Princeton, NJ: Mathematica Policy Research, 2004.
- Nunnally, Jum C., and Ira Bernstein. *Psychometric Theory, 3rd Edition*. New York: McGraw-Hill, 1994.
- Roid, Gale H., and Lucy J. Miller. "Leiter-R Performance Scale—Revised." Wood Dale, IL: Stoelting Co., 1997.
- Salvia, Jlnh, and James E. Ysseldyke. *Assessment in Special Education and Inclusive Education – Ninth Edition*, Boston: Houghton Mifflin Company, 2004.
- Smith-Donald, Radiah, C. Cybele Raver, Tiffany Hayes, and Breeze Richardson. "Preliminary Construct and Concurrent Validity of the Preschool Self-Regulation Assessment (PSRA) for Field-Based Research." *Early Childhood Research Quarterly*, vol. 22, 2007, pp. 173-187.
- Squires, Jane, Elizabeth Twombly, Diane Bricker, and LaWanda Potter. "Ages and Stages Questionnaires-Third Edition (ASQ-3) User's Guide." Baltimore, MD: Paul H. Brookes Publishing Co. Inc., 2009.
- Woodcock, Richard W., and F. Munoz-Sandoval. "Woodcock-Muñoz Language Survey Revised." Itasca, IL: Riverside Publishing, 2005.

**This page has been left blank for double-sided copying.**

## UPCOS-3 COPYRIGHT PERMISSIONS

*Woodcock-Johnson® III (WJ III®), WJ III® Tests of Achievement.* Copyright © 2001, 2007, The Riverside Publishing Company. Reproduced with permission of the publisher. All rights reserved.

No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording or by any information storage or retrieval system without the proper written permission of The Riverside Publishing Company unless such copying is expressly permitted by federal copyright law. Address inquiries to Contracts and Permissions Department, The Riverside Publishing Company, 3800 Golf Road, Rolling Meadows, Illinois 60008-4015.

*Batería III Woodcock-Muñoz®.* Copyright © 2004, 2007, The Riverside Publishing Company. Reproduced with permission of the publisher. All rights reserved.

No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording or by any information storage or retrieval system without the proper written permission of The Riverside Publishing Company unless such copying is expressly permitted by federal copyright law. Address inquiries to Contracts and Permissions Department, The Riverside Publishing Company, 425 Spring Lake Drive, Itasca, Illinois 60143-2079.

*PreLAS 2000*, by Sharon E. Duncan, Ph.D., and Edward A. DeAvila, Ph.D. Copyright © 1998 CTB/McGraw-Hill LLC, a subsidiary of The McGraw-Hill Companies, Inc. Reproduced by permission of CTB/McGraw-Hill LLC.

*Leiter International Performance Scale-Revised Examiner Ratings.* Copyright © 1997, 2002 Stoelting Co., 620 Wheat Lane, Wood Dale, IL 60191. All rights reserved.

*Expressive One Word Picture Vocabulary Test (EOWPVT).* Copyright © 2000, Academic Therapy Publications., 20 Commercial Boulevard, Novato, CA, 94949-6191. All rights reserved. Reproduced by permission of Academic Therapy Publications.

*Preschool and Kindergarten Behavioral Scales – Second Edition (PKBS-2)*, by Kenneth W. Merrell. Copyright © 2002 PRO-ED Inc., Austin, Texas. Reproduced and used by permission of PRO-ED, Inc.

**This page has been left blank for double-sided copying.**

**APPENDIX A:**

**ANALYSIS OF RATER EFFECTS BASED ON FALL 2009 ASSESSMENTS**

**This page has been left blank for double-sided copying.**

Rater effects indicate that characteristics of the assessor or rater may be influencing scores on an assessment; as a consequence, scores may not be an accurate reflection of children's development. To examine rater effects in the fall 2009 data, we calculated the intraclass correlations (ICCs) for the independently administered and teacher-administered measures using hierarchical linear modeling (HLM). HLM partitions the variance in a particular measure into within- and between-classroom components. The ICC is the proportion of variance that is attributable to teachers/classrooms. This value gives an indication of the degree to which scores within the same unit—in this case, within the same group of children as designated by their teacher or classroom—are similar to each other. High ICCs can occur due to the ways in which children are grouped within classrooms (design effects) or to differences in how teachers administer the measures (rater effects). In order to separate the design effects from the rater effects, we examined the fall ICCs for the teacher-administered assessments in relation to the standardized assessments administered by Mathematica assessors who went to multiple classrooms.

The ICCs for the independently administered direct assessments should identify the between-classroom differences in children's development in the various domains that are attributable to how children are grouped in classrooms (design effects). We expect that there are at least some differences between classrooms, particularly given that we stratified our sample by ELL concentration to allow us to represent both high ELL and low ELL concentration classrooms/programs in the study. In other words, by design, children in some classrooms may resemble one another more than they resemble children in other classrooms, and this could result in increased ICCs. We examined rater effects in the teacher-administered measures by comparing the ICCs for the Mathematica-administered assessments with the ICCs for the teacher-administered assessments. There is no rule of thumb for evaluating the differences; however, we look for ICCs that are as similar as possible across modes of administration and our expectation was that the ICCs would be within .10 of one another in magnitude. We anticipated slightly stronger ICCs for teacher-administered assessments than for independently administered assessments, because children often respond differently to familiar versus unfamiliar assessors and some class sizes were smaller, thus offering children greater opportunity to develop a relationship with the teacher.

In Table A.1, we present the ICCs for the independently administered assessments; A.2 presents the ICCs for the teacher-administered assessments. The ICCs for the direct child assessments administered by Mathematica ranged from .01 to .30. The ICCs for the ASQ-3 subscale scores and the Rapid Letter Naming (RLN) scores ranged from .27 (for the RLN) to .47 (for Spanish Communication). The ICCs for the PKBS-2 ranged from .24 for Social Cooperation to .47 for Internalizing Problems. The ICCs for the P-KOF range from .39 (for General Knowledge) to .56 (for Self-Care & Motor Skills).

We also calculated the ICCs separately for the subgroups of ELL children and for children who speak English only or primarily (groups are based on the language used in the direct assessment). This was to help address the fact that some classrooms included children who spoke multiple languages, and teachers differed in their fluency. Teachers in some classrooms might administer the assessment in the strongest language for each child, while teachers in other classrooms might only be able to assess in English. The results, shown in Table A.3, indicate that the ICCs of the P-KOF and PAA were higher than the ICCs of the direct child assessments for both ELL and non-ELL children. For the non-ELL language group, the differences in the magnitude of the ICCs for independent assessors and teachers ranged from .04 to .18 for the PAA, and from .22 to .35 for the P-KOF. For the ELL language group, the differences ranged from .17 to .27 on the PAA, and from .14 to .38 on the P-KOF. The greater differences found with the ELL language group suggest that at least some of the rater effects are due to differences in the teacher's language proficiency.

Finally, we examined the distribution of children's age across classrooms, because an uneven distribution of ages could influence the ICCs if scores are related to child age. The results indicate that some classrooms had a higher proportion of younger children than did others; similarly, some of the classrooms had a higher proportion of older children than did others. The distribution of ages across classrooms may have contributed to the ICCs.

In summary, although the clustering of children in classrooms by ELL status and age contributes to higher ICCs, the fairly substantial differences in the magnitude of the ICCs between the teacher-administered and the standardized direct assessments ( $> .30$ ) clearly indicate the operation of rater effects in all scales completed by teachers. These differences remained when we looked separately by ELL status, particularly for the P-KOF. The pattern of differences by ELL status suggests that at least some of the rater effects are due to differences in the teachers' language proficiency.



**Table A.1. Intraclass Correlations (ICCs) for Direct Child Assessments Fall 2009**

Standardized Measures	ICC
PreLAS total language screener score English	0.30
PreLAS total language screener score Spanish	0.17
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score	0.16
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.24
Woodcock-Johnson III (WJ-III)	
Letter-Word Identification Standard Score	0.13
Letter-Word Identification W Score	0.12
Spelling Standard Score	0.12
Spelling W Score	0.11
Applied Problems Standard Score	0.13
Applied Problems W Score	0.13
Woodcock-Muñoz-III (WM-III)	
Letter-Word Identification Standard Score	0.17
Letter-Word Identification W Score	0.12
Spelling Standard Score	0.12
Spelling W Score	0.12
Applied Problems Standard Score	0.01
Applied Problems W Score	0.01
Executive Functioning	
Pencil Tapping	0.11
Leiter Examiner Ratings Scaled Score	
Attention	0.21
Activity	0.23
Sociability	0.23

Source: UPCOS-3 Fall 2009 Direct Child Assessment

Note: The ICC is the proportion of variance that is attributable to teachers/classrooms. The analyses include 875 children and 150 teachers.

**Table A.2. Intraclass Correlations (ICCs) for Teacher-Administered Measures Fall 2009**

Teacher-Administered Measures	ICC
ASQ-3 Subscale Total Score	
Communication, English	0.31
Communication, Spanish	0.47
Problem Solving	0.37
Gross Motor	0.37
Fine Motor	0.34
Rapid Letter Naming Raw Score	0.27
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score	
Social Cooperation	0.24
Social Interaction	0.46
Social Independence	0.27
Total Positive Social Skills	0.34
Externalizing Problems	0.25
Internalizing Problems	0.47
Problem Behaviors	0.43
P-KOF	
Social Expression	0.49
General Knowledge	0.39
Self-Regulation	0.41
Self-Care & Motor Skills	0.56
Overall score	0.47

Source: UPCOS-3 Fall 2009 teacher-administered measures.

Note: The ICC is the proportion of variance that is attributable to teachers/classrooms. The analyses include 875 children and 150 teachers.

**Table A.3. Comparison of Intraclass Correlations (ICCs) for Direct Child Assessment and Teacher-Administered Measures Fall 2009, by Language Group**

Measures	Language Group	
	Non-ELL	ELL
ASQ-3 Subscale Total Score vs. Direct Child Assessment		
ASQ-3 Communication, English	0.28	--
EOWPVT, English	0.11	--
WJ-III Letter Word Identification	0.24	--
ASQ-3 Communication, Spanish		
ASQ-3 Communication, Spanish	--	0.39
EOWPVT, SBE	--	0.12
WM-III Letter Word Identification	--	0.22
ASQ-3 Problem Solving		
ASQ-3 Problem Solving	0.31	0.23
WJ-III		
Letter Word Identification	0.08	--
Applied Problems	0.13	--
Spelling	0.17	--
WM-III		
Letter Word Identification	--	0.27
Applied Problems	--	0.05
Spelling	--	0.04
Rapid Letter Naming vs. Direct Child Assessment		
Rapid Letter Naming Raw Score	0.27	0.06
WJ-III Letter Word Identification	0.18	--
WM-III Letter Word Identification	--	0.26
P-KOF vs. Direct Child Assessment		
Social Expression	0.50	0.49
EOWPVT English	0.15	--
EOWPVT SBE	--	0.11
General Knowledge		
General Knowledge	0.36	0.32
WJ-III		
Letter Word Identification	0.14	--
Applied Problems	0.12	--
Spelling	0.12	--
WM-III		
Letter Word Identification	--	0.18
Applied Problems	--	0.02
Spelling	--	0.13

Source: UPCOS-3 Fall 2009 Direct Child Assessment and teacher-administered measures.

Note: The ICC is the proportion of variance that is attributable to teachers/classrooms. The ICCs were calculated for children who have both the specific PAA or P-KOF ratings and the corresponding direct assessment scores.

**This page has been left blank for double-sided copying.**

**APPENDIX B.**

**TABLES OF RESULTS PRESENTED IN CHAPTER II:  
STANDARDIZED MEASURES**

**This page has been left blank for double-sided copying.**

**Table II.1** Number of Children with Full UPCOS Battery Data Fall 2009 and Spring 2010, by Program Type and Language Group

<i>Language group</i>	Fall 2009			Spring 2010		
	Center	FCC	Total	Center	FCC	Total
English only	231	93	324	208	73	281
English primarily	155	41	196	143	36	179
Spanish only	64	18	82	55	17	72
Span primarily	155	42	197	149	36	185
Other language only or primarily	23	3	26	19	3	22
Missing	1	3	4	1	.	1
Total	629	200	829	575	165	740

**Table II.2 Average Time in Months Between Fall 2009 and Spring 2010 Standardized Assessments, by Program Type and Language Group**

<i>Language group</i>	Fall-Spring Change		
	Center	FCC	Total
English only	5.23	4.96	5.16
English primarily	5.02	5.13	5.05
Spanish only	5.18	4.95	5.13
Span primarily	4.96	5.25	5.01
Other language only or primarily	5.16	4.62	5.09
Missing	5.53	.	5.53
Total	5.10	5.06	5.09



**Table II.3. Means, Standard Deviations, and Cronbach's Alphas for UPCOS-3 Language, Literacy and Math Standardized Measures: Fall 2009 – Spring 2010**

Outcome	N	Fall 2009		Spring 2010		Cronbach's Alpha	Fall-Spring Change		
		Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation		Mean/ Percentage	Standard Deviation	Effect Size
Pre-LAS total language screener score English <sup>a</sup>	707	14.33***	5.45	16.54	4.06	0.89	2.21	3.45	0.41
Simon Says Score	707	6.93***	2.95	8.23	2.15	0.82	1.30	2.42	0.44
Art Show Score	705	7.41***	3.06	8.33	2.37	0.87	0.92	1.91	0.30
Pre-LAS total language screener score Spanish <sup>a</sup>	245	11.53***	4.08	12.69	4.20	0.83	1.16	3.90	0.28
Tío Simón Dice Score	245	5.63***	2.85	6.70	2.34	0.81	1.07	2.75	0.38
Exposición de Arte Score	245	5.90	2.52	5.99	2.80	0.81	0.09	2.50	0.04
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	467	89.16	15.03	90.55	15.14	0.99	1.39	10.09	0.09
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	699	105.94**	22.26	108.49	20.60	0.99	2.55	12.71	0.11
Expressive One Word Picture Vocabulary Test—IRT Score	700	44.58***	10.84	49.31	10.68	0.99	4.73	6.55	0.44
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	466	100.65***	14.45	103.70	13.90	0.93	3.06	11.52	0.21
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	466	332.28***	28.79	349.46	28.75	0.93	17.19	22.82	0.60
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>	103	93.85	11.99	91.71	10.81	0.79	-2.15	13.34	-0.18
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	121	313.57***	23.20	321.92	21.78	0.79	8.35	26.01	0.36
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	466	103.75**	14.19	106.08	13.61	0.91	2.33	12.43	0.16
Woodcock-Johnson III (WJ-III) Spelling W Score	466	380.17***	28.24	398.08	25.44	0.91	17.91	22.93	0.63
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>	121	90.60	16.51	89.26	21.01	0.85	-1.34	23.61	-0.08
Woodcock-Muñoz-III (WM-III) Ortografía W Score	121	352.40***	30.33	364.77	37.19	0.85	12.37	42.82	0.41
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	466	100.98*	12.45	102.38	11.46	0.93	1.40	9.47	0.11

Outcome	N	Fall 2009		Spring 2010		Cronbach's Alpha	Fall-Spring Change		
		Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation		Mean/ Percentage	Standard Deviation	Effect Size
Woodcock-Johnson III (WJ-III) Applied Problems W Score	466	397.47***	23.61	409.91	20.33	0.93	12.44	17.70	0.53
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	121	81.52	14.89	84.25	15.91	0.91	2.73	15.94	0.18
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	121	354.91***	33.14	372.34	33.18	0.91	17.43	34.44	0.53

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted t-tests to compare fall and spring scores.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.4. Means, Standard Deviations, and Cronbach's Alphas for Social-Emotional Standardized Measures: Fall 2009 – Spring 2010**

Outcome	N	Fall 2009		Spring 2010		Cronbach's Alpha	Fall-Spring Change		
		Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation		Mean/ Percentage	Standard Deviation	Effect Size
Executive Functioning (Percentage)									
Pencil Tapping <sup>a</sup>	697	52.20***		71.04		0.94	18.84		
Leiter Examiner Ratings Scaled Score <sup>b</sup>									
Attention	701	8.36***	2.17	8.99	1.77	0.97	0.63	2.30	0.29
Activity	701	8.43	1.81	8.49	1.53	0.76	0.06	2.10	0.03
Sociability	701	8.51***	2.11	9.05	1.69	0.92	0.53	2.37	0.25
Leiter Examiner Ratings Scaled Score in the Clinical Range <sup>c</sup> (Percentage)									
Attention	701	6.28***		3.42			-2.85		
Activity	701	2.71***		1.14			-1.57		
Sociability	701	7.13***		2.85			-4.28		
Leiter Examiner Ratings Scaled Score in Possible Clinical Range <sup>d</sup> (Percentage)									
Attention	701	12.27***		6.99			-5.28		
Activity	701	13.84***		10.41			-3.42		
Sociability	701	13.41		8.99			-4.42		
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score <sup>e</sup>									
Social Cooperation	273	102.01***	14.80	107.01	14.90	0.94	5.00	12.54	0.34
Social Interaction	272	94.94***	16.56	105.31	14.74	0.92	10.37	12.29	0.63
Social Independence	274	99.70***	15.36	107.24	12.36	0.85	7.53	13.11	0.49
Total Positive Social Skills	272	98.62***	15.26	107.56	14.20	0.96	8.94	12.62	0.59
Externalizing Problems	259	94.78***	13.21	91.05	14.10	0.97	-3.74	10.11	-0.28
Internalizing Problems	267	92.39***	14.30	89.03	12.90	0.90	-3.36	11.45	-0.23
Problem Behaviors	271	92.92***	13.42	88.18	13.84	0.97	-4.74	11.09	-0.35

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: We conducted t-tests to compare fall and spring scores.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

<sup>b</sup>The scaled scores range from 1 to 10 (truncated at 10), with 10 being "average", or "of no clinical concern".

<sup>c</sup>Scores of 4 or less.

<sup>d</sup>Scores of 5 and 6.

<sup>e</sup>National mean for standard scores is 100 with a standard deviation of 15.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.5. Means and Standard Deviations for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 and Spring 2010, by Language Group**

Outcome	English Only		English Primarily <sup>d</sup>		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010
	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)
Pre-LAS total language screener score English <sup>a</sup>	17.42*** (2.99)	18.58 (2.01)	15.50*** (4.34)	17.17 (3.07)	6.91*** (5.36)	11.97 (5.08)	11.36*** (5.31)	14.48 (4.60)	14.71*** (3.51)	17.62 (2.52)
Simon Says Score	8.27*** (1.94)	9.01 (1.52)	7.21*** (2.66)	8.26 (2.01)	3.85*** (3.07)	6.94 (2.49)	5.95*** (3.16)	7.52 (2.49)	6.19** (3.12)	8.38 (2.06)
Art Show Score	9.16*** (1.46)	9.58 (0.88)	8.34*** (2.06)	8.97 (1.31)	3.06*** (3.16)	5.03 (3.13)	5.42*** (3.02)	6.99 (2.64)	8.52* (1.29)	9.24 (1.00)
Pre-LAS total language screener score Spanish <sup>a</sup>					11.10*** (4.01)	13.25 (3.64)	11.69* (4.10)	12.48 (4.39)		
Tío Simón Dice Score					5.15*** (2.84)	6.96 (1.97)	5.82*** (2.84)	6.60 (2.47)		
Exposición de Arte Score					5.96 (2.62)	6.29 (2.87)	5.88 (2.48)	5.88 (2.77)		
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	92.25 (15.26)	92.55 (14.95)	87.05 (14.39)	90.17 (16.10)			80.91 (10.60)	81.96 (11.53)	84.53 (13.33)	88.59 (10.60)
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	118.03 (17.89)	118.42 (16.55)	107.85* (19.94)	111.53 (19.99)	85.87 (19.77)	90.09 (19.15)	93.07* (19.43)	96.86 (17.37)	108.94 (17.81)	114.47 (13.87)
Expressive One Word Picture Vocabulary Test—IRT Score	50.10*** (10.10)	54.21 (9.57)	45.15*** (9.58)	50.71 (10.42)	35.09*** (9.18)	40.30 (8.84)	39.14*** (8.67)	43.64 (8.51)	46.35** (7.57)	52.40 (7.20)
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	101.75** (13.42)	104.17 (13.14)	100.26** (16.03)	104.29 (15.77)			94.15** (13.09)	99.38 (10.73)	107.29 (13.04)	109.76 (8.33)
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	333.85*** (28.45)	349.75 (27.16)	331.45*** (32.13)	350.33 (32.75)			321.68*** (26.40)	343.47 (23.26)	347.41** (24.74)	365.12 (15.14)
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>			90.27 (11.37)	84.36 (7.89)	93.29 (12.05)	95.26 (9.84)	94.89* (12.12)	90.95 (11.12)		
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score			306.93 (22.41)	310.57 (14.46)	312.50*** (22.81)	326.98 (21.44)	315.60 (23.62)	321.27 (22.51)		

Outcome	English Only		English Primarily <sup>a</sup>		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010
	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	103.49** (14.20)	106.29 (14.42)	103.79** (15.29)	106.03 (13.35)			101.36 (11.45)	102.55 (10.43)	114.24 (10.27)	112.24 (9.13)
Woodcock-Johnson III (WJ-III) Spelling W Score	378.88*** (28.45)	397.64 (27.53)	380.05*** (29.93)	397.60 (23.80)			379.38*** (23.98)	395.51 (19.32)	401.76* (17.51)	412.59 (14.85)
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>			85.86 (20.40)	88.86 (13.25)	89.40 (14.21)	91.20 (16.22)	92.31 (16.90)	88.19 (24.68)		
Woodcock-Muñoz-III (WM-III) Ortografía W Score			343.57 (40.57)	363.36 (23.84)	349.88*** (27.96)	368.05 (31.70)	355.75 (29.24)	363.10 (42.42)		
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	101.27 (12.80)	102.73 (11.68)	100.45* (12.85)	102.12 (11.78)			99.17 (10.29)	100.26 (10.63)	106.94 (10.08)	106.18 (8.48)
Woodcock-Johnson III (WJ-III) Applied Problems W Score	397.28*** (24.30)	409.87 (20.75)	396.33*** (24.61)	409.04 (20.97)			397.43*** (19.50)	409.04 (18.00)	409.18* (17.36)	419.29 (17.17)
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>			73.21 (12.80)	76.36 (15.70)	83.48 (15.36)	87.83 (13.84)	82.09 (14.63)	83.76 (16.66)		
Woodcock-Muñoz-III (WM-III) Applied Problems W Score			336.14 (27.87)	354.50 (33.19)	359.13*** (35.03)	379.65 (30.41)	356.31*** (32.11)	371.70 (33.73)		

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted t-tests to compare fall and spring scores within each subgroup.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

<sup>d</sup> In the English primarily group, 129 children took the WJ-III tests, and only 14 children took the WM-III tests.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.6. Change Scores for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 – Spring 2010, by Language Group**

Outcome	English Only			English Primarily			Spanish Only			Spanish Primarily			Other Language Only or Primarily		
	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size
Pre-LAS total language screener score English <sup>a</sup>	266	1.16*** (2.59)	0.21	175	1.67 (3.10)	0.31	68	5.06 (4.36)	0.93	176	3.12 (3.71)	0.57	21	2.90 (3.03)	0.53
Simon Says Score	266	0.75*** (1.87)	0.25	175	1.05 (2.35)	0.36	68	3.09 (2.82)	1.05	176	1.57 (2.56)	0.53	21	2.19 (3.01)	0.74
Art Show Score	266	0.42*** (1.35)	0.14	174	0.62 (1.37)	0.20	68	1.97 (2.79)	0.64	175	1.58 (2.35)	0.52	21	0.71 (1.49)	0.23
Pre-LAS total language screener score Spanish <sup>a</sup>							68	2.15*** (3.60)	0.53	177	0.79 (3.95)	0.19			
Tío Simón Dice Score							68	1.81** (3.00)	0.64	177	0.79 (2.60)	0.28			
Exposición de Arte Score							68	0.34 (2.60)	0.13	177	0.00 (2.47)	0.00			
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	266	0.30*** (10.71)	0.02	130	3.12 (9.71)	0.21				47	1.04 (8.29)	0.07	17	4.06 (5.75)	0.27
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	266	0.40*** (12.46)	0.02	174	3.68 (11.58)	0.17	67	4.22 (12.99)	0.19	175	3.79 (14.00)	0.17	17	5.53 (8.06)	0.25
Expressive One Word Picture Vocabulary Test—IRT Score	266	4.12 (7.14)	0.38	174	5.57 (6.27)	0.51	68	5.20 (6.23)	0.48	175	4.50 (6.13)	0.42	17	6.05 (3.83)	0.56
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	266	2.42 (10.89)	0.17	129	4.04 (12.61)	0.28				47	5.23 (10.89)	0.36	17	2.47 (10.99)	0.17
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	266	15.89 (21.53)	0.55	129	18.88 (25.11)	0.66				47	21.79 (21.74)	0.76	17	17.71 (21.19)	0.62
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>				11	-5.91** (11.57)	-0.49	35	1.97 (12.41)	0.16	57	-3.95 (13.78)	-0.33			
Woodcock-Muñoz-III (WM-III) Letter-Word				14	3.64 (22.32)	0.16	40	14.48 (23.58)	0.62	67	5.67 (27.72)	0.24			

Outcome	English Only			English Primarily			Spanish Only			Spanish Primarily			Other Language Only or Primarily		
	N	Mean/Percentage (SD)	Effect Size	N	Mean/Percentage (SD)	Effect Size	N	Mean/Percentage (SD)	Effect Size	N	Mean/Percentage (SD)	Effect Size	N	Mean/Percentage (SD)	Effect Size
Identification W Score															
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>a</sup>	266	2.80 (11.84)	0.20	129	2.24 (14.43)	0.16				47	1.19 (10.40)	0.08	17	-2.00 (11.24)	-0.14
Woodcock-Johnson III (WJ-III) Spelling W Score	266	18.76 (21.85)	0.66	129	17.55 (26.71)	0.62				47	16.13 (19.22)	0.57	17	10.82 (20.19)	0.38
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>				14	3.00 (20.29)	0.18	40	1.80 (18.54)	0.11	67	-4.12 (26.66)	-0.25			
Woodcock-Muñoz-III (WM-III) Ortografía W Score				14	19.79 (36.36)	0.65	40	18.18 (33.85)	0.60	67	7.36 (48.37)	0.24			
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	266	1.46 (9.41)	0.12	129	1.67 (9.69)	0.13				47	1.09 (9.83)	0.09	17	-0.76 (9.68)	-0.06
Woodcock-Johnson III (WJ-III) Applied Problems W Score	266	12.59 (17.41)	0.53	129	12.71 (19.26)	0.54				47	11.62 (16.52)	0.49	17	10.12 (17.34)	0.43
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>				14	3.14 (17.29)	0.21	40	4.35 (13.28)	0.29	67	1.67 (17.20)	0.11			
Woodcock-Muñoz-III (WM-III) Applied Problems W Score				14	18.36 (39.39)	0.55	40	20.53 (28.54)	0.62	67	15.39 (36.86)	0.46			

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted ANOVA to compare change scores between language groups.

<sup>a</sup> These scores are raw counts of children’s correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.7. Internal Consistency Reliability of UPCOS-3 Language, Literacy and Math Standardized Measures Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha
Pre-LAS total language screener score English <sup>a</sup>	271	0.76	176	0.78	69	0.89	180	0.88	22	0.76
Simon Says Score	271	0.80	176	0.78	69	0.82	180	0.83	22	0.81
Art Show Score	271	0.55	175	0.63	69	0.86	178	0.82	22	0.42
Pre-LAS total language screener score Spanish <sup>a</sup>					69	0.78	181	0.84		
Tío Simón Dice Score					69	0.75	181	0.82		
Exposición de Arte Score					69	0.83	181	0.80		
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	271	0.99	156	0.99	27	0.96	97	0.97	20	0.99
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	271	0.99	175	0.99	69	0.99	178	0.99	20	0.99
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	271	0.91	156	0.93	27	0.91	97	0.92	20	0.94
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	271	0.91	156	0.93	27	0.91	97	0.92	20	0.94
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>			19	0.56	42	0.77	80	0.80		
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score			19	0.56	42	0.77	80	0.80		
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	271	0.90	156	0.89	27	0.91	97	0.91	20	0.91
Woodcock-Johnson III (WJ-III) Spelling W Score	271	0.90	156	0.89	27	0.91	97	0.91	20	0.91
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>			19	0.85	42	0.84	80	0.85		



Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha
Woodcock-Muñoz-III (WM-III) Ortografía W Score			19	0.84	42	0.84	80	0.85		
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	271	0.92	156	0.92	27	0.92	97	0.93	20	0.93
Woodcock-Johnson III (WJ-III) Applied Problems W Score	271	0.92	156	0.92	27	0.92	97	0.93	20	0.93
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>			19	0.88	42	0.90	80	0.90		
Woodcock-Muñoz-III (WM-III) Applied Problems W Score			19	0.88	42	0.90	80	0.90		

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

**Table II.8. Means and Standard Deviations for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 and Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010
	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)	Mean/ Percentage (SD)
Executive Functioning Pencil Tapping <sup>a</sup>	57.94 ***	77.86	47.90***	70.19	38.62***	59.70	50.18***	63.69	80.15	92.65
Leiter Examiner Ratings Scaled Score <sup>b</sup>										
Attention	8.66*** (2.07)	9.20 (1.69)	8.33*** (2.20)	9.05 (1.66)	7.64 (2.45)	8.27 (1.99)	8.10*** (2.15)	8.80 (1.90)	9.19 (1.57)	9.57 (1.08)
Activity	8.57 (1.72)	8.61 (1.51)	8.49 (1.80)	8.47 (1.54)	7.97 (1.97)	8.18 (1.77)	8.18 (1.90)	8.41 (1.48)	9.62*** (0.80)	8.71 (1.23)
Sociability	8.71*** (2.08)	9.30 (1.59)	8.60** (2.02)	9.07 (1.68)	7.85 (2.42)	8.40 (1.85)	8.34*** (2.08)	8.90 (1.74)	8.86 (2.15)	9.00 (1.52)
Leiter Examiner Ratings Scaled Score in the Clinical Range <sup>c</sup>										
Attention	5.26 ***	3.76	5.20	2.89	10.45***	4.48	7.47**	3.45	4.76	0.00
Activity	2.63***	1.13	2.31	1.16	4.48	1.49	2.87***	1.15	0.00	0.00
Sociability	6.39***	2.26	6.36***	2.89	13.43	2.99	6.32	4.02	9.52	0.00
Leiter Examiner Ratings Scaled Score in Possible Clinical Range <sup>d</sup>										
Attention	10.15	3.38	13.29*	7.51	17.91	10.45	13.79***	11.49	0.00	0.00
Activity	13.16	7.52	14.45**	10.98	16.42	17.91	14.94	12.07	0.00	4.76
Sociability	10.90**	6.39	10.98	5.78	20.90	20.90	17.24	10.92	9.52	14.29
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score <sup>e</sup>										
Social Cooperation	98.38** (15.49)	104.10 (15.75)	103.77** (13.39)	109.03 (13.49)	101.16 (13.28)	104.81 (14.58)	105.12* (14.89)	109.55 (14.90)		
Social Interaction	97.13*** (13.75)	106.18 (13.41)	95.19*** (16.45)	107.66 (13.06)	87.00*** (16.31)	98.50 (14.01)	94.32*** (19.22)	105.13 (17.28)		
Social Independence	98.67*** (15.66)	106.23 (13.25)	100.77*** (15.14)	109.34 (9.87)	94.58** (16.30)	103.87 (13.90)	101.38*** (14.81)	107.91 (12.63)		
Total Positive Social Skills	97.71*** (15.43)	106.37 (14.73)	99.63*** (14.98)	109.87 (12.63)	93.60*** (13.22)	103.13 (13.65)	100.13*** (15.91)	108.73 (15.05)		
Externalizing Problems	98.94 (15.14)	95.28 (16.02)	93.34 (10.87)	90.51 (13.39)	94.40 (10.64)	92.00 (13.23)	91.26*** (12.40)	86.09 (11.17)		
Internalizing Problems	96.37* (15.17)	92.31 (13.28)	90.29 (10.66)	88.78 (13.43)	93.03 (16.60)	88.55 (12.87)	89.66* (14.23)	85.66 (11.65)		
Problem Behaviors	97.47** (15.00)	91.82 (15.46)	90.90 (9.14)	88.03 (13.69)	92.81 (13.17)	89.00 (12.59)	89.50*** (13.23)	83.62 (11.45)		

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: We conducted t-tests to compare fall and spring scores within each subgroup.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

<sup>b</sup>The scaled scores range from 1 to 10 (truncated at 10), with 10 being "average", or "of no clinical concern".

<sup>c</sup>Scores of 4 or less.

<sup>d</sup>Scores of 5 and 6.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.9. Change Scores for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 – Spring 2010, by Language Group**

Outcome	English Only			English Primarily			Spanish Only			Spanish Primarily			Other Language Only or Primarily		
	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size	N	Mean/ Percentage (SD)	Effect Size
Executive Functioning Pencil Tapping <sup>a</sup>	266	19.92		173	22.29		67	21.08		174	13.51		17	12.50	
Leiter Examiner Ratings Scaled Score															
Attention	266	0.54 (2.23)	0.25	173	0.72 (2.34)	0.33	67	0.63 (2.56)	0.29	174	0.71 (2.32)	0.33	21	0.38 (1.94)	0.18
Activity	266	0.05*** (1.97)	0.03	173	-0.02 (2.18)	-0.01	67	0.21 (2.28)	0.12	174	0.23 (2.20)	0.13	21	-0.90 (1.18)	-0.50
Sociability	266	0.59 (2.27)	0.28	173	0.47 (2.19)	0.22	67	0.55 (2.64)	0.26	174	0.56 (2.59)	0.27	21	0.14 (2.22)	0.07
Leiter Examiner Ratings Scaled Score in the Clinical Range															
Attention	266	-1.50		173	-2.31		67	-5.97		174	-4.02		21	-4.76	
Activity	266	-1.50		173	-1.16		67	-2.99		174	-1.72		21	0.00	
Sociability	266	-4.14		173	-3.47		67	-10.45		174	-2.30		21	-9.52	
Leiter Examiner Ratings Scaled Score in Possible Clinical Range															
Attention	266	-6.77		173	-5.78		67	-7.46		174	-2.30		21	0.00	
Activity	266	-5.64		173	-3.47		67	1.49		174	-2.87		21	4.76	
Sociability	266	-4.51***		173	-5.20		67	0.00		174	-6.32		21	4.76	
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score															
Social Cooperation	94	5.71 (12.06)	0.39	62	5.26 (12.86)	0.36	31	3.65 (13.67)	0.25	78	4.44 (13.00)	0.30			
Social Interaction	94	9.05** (12.11)	0.55	62	12.47 (12.00)	0.75	30	11.50 (11.60)	0.69	78	10.81 (12.99)	0.65			
Social Independence	95	7.568 (14.02)	0.49	62	8.56 (12.92)	0.56	31	9.29 (13.75)	0.60	78	6.53 (11.92)	0.43			
Total Positive Social Skills	94	8.66 (12.88)	0.57	62	10.24 (13.39)	0.67	30	9.53 (13.15)	0.62	78	8.60 (11.89)	0.56			
Externalizing Problems	88	-3.66 (10.82)	-0.28	59	-2.83 (10.85)	-0.21	30	-2.40 (10.68)	-0.18	74	-5.16 (8.82)	-0.39			
Internalizing Problems	93	-4.05 (10.95)	-0.28	58	-1.52 (13.17)	-0.11	31	-4.48 (10.73)	-0.31	77	-4.00 (11.19)	-0.28			
Problem Behaviors	94	-5.65 (11.66)	-0.42	60	-2.87 (11.72)	-0.21	31	-3.81 (10.84)	-0.28	78	-5.88 (10.28)	-0.44			

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: Reported by teachers. We conducted ANOVA to compare change scores between language groups.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.10. Internal Consistency Reliability of UPCOS-3 Social-Emotional Standardized Measures Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha
Executive Functioning Pencil Tapping	271	0.94	175	0.94	69	0.92	178	0.94	20	0.75
Leiter Examiner Ratings Scaled Score										
Attention	271	0.96	175	0.9	69	0.97	178	0.97	22	0.95
Activity	271	0.75	175	0.81	69	0.81	178	0.74	22	0.78
Sociability	271	0.91	175	0.92	69	0.92	178	0.90	22	0.93
Leiter Examiner Ratings Scaled Score in the Clinical Range										
Attention	271	0.96	175	0.97	69	0.97	178	0.97	22	0.95
Activity	271	0.75	175	0.81	69	0.81	178	0.74	22	0.78
Sociability	271	0.91	175	0.92	69	0.92	178	0.90	22	0.93
Leiter Examiner Ratings Scaled Score in Possible Clinical Range										
Attention	271	0.96	175	0.97	69	0.97	178	0.97	22	0.95
Activity	271	0.75	175	0.81	69	0.81	178	0.74	22	0.78
Sociability	271	0.91	175	0.92	69	0.92	178	0.90	22	0.93
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score										
Social Cooperation	114	0.94	67	0.93	33	0.93	80	0.94		
Social Interaction	114	0.91	67	0.92	33	0.92	80	0.93		
Social Independence	115	0.86	67	0.86	33	0.86	80	0.86		
Total Positive Social Skills	114	0.96	67	0.95	33	0.95	80	0.96		
Externalizing Problems	108	0.97	62	0.97	32	0.97	76	0.96		
Internalizing Problems	113	0.89	62	0.91	33	0.91	79	0.91		
Problem Behaviors	114	0.97	64	0.96	33	0.96	80	0.97		

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

**Table II.11. Means and Standard Deviations for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 and Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Pre-LAS total language screener score English <sup>a</sup>	14.32***	5.35	16.56	3.95	14.34***	5.79	16.46	4.45
Simon Says Score	6.93***	2.86	8.25	2.11	6.93***	3.25	8.17	2.30
Art Show Score	7.40***	3.04	8.32	2.34	7.43***	3.15	8.34	2.49
Pre-LAS total language screener score Spanish <sup>a</sup>	11.80***	3.99	12.92	4.08	10.54	4.29	11.85	4.58
Tío Simón Dice Score	5.76***	2.75	6.77	2.28	5.15**	3.16	6.46	2.57
Exposición de Arte Score	6.04	2.38	6.16	2.66	5.38	2.94	5.38	3.22
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	89.36	15.35	90.36	15.52	88.55	14.04	91.14	13.97
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	105.40*	22.60	107.89	21.12	107.75	21.03	110.53	18.68
Expressive One Word Picture Vocabulary Test—IRT Score	44.46***	11.08	49.09	11.00	45.00***	10.00	50.04	9.51
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	100.44***	14.55	103.85	14.37	101.26	14.18	103.25	12.38
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	332.20***	29.12	350.17	29.96	332.52***	27.87	347.30	24.69
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>	93.99	11.75	92.17	9.70	93.46	12.91	90.35	13.70
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	313.71***	22.57	322.38	19.95	313.16	25.30	320.58	26.74
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	103.38***	14.01	106.21	13.02	104.87	14.75	105.66	15.32
Woodcock-Johnson III (WJ-III) Spelling W Score	379.89***	28.03	398.71	24.27	381.04***	28.96	396.15	28.73

Outcome	Center				Family Child Care			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>a</sup>	89.13	16.54	87.64	21.73	94.87	15.92	93.97	18.29
Woodcock-Muñoz-III (WM-III) Ortografía W Score	349.38**	30.74	361.47	38.69	361.16*	27.76	374.35	31.05
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	101.42	12.58	102.54	11.24	99.66	11.98	101.92	12.14
Woodcock-Johnson III (WJ-III) Applied Problems W Score	398.50***	23.60	410.48	19.78	394.32***	23.44	408.16	21.94
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	79.67	15.23	82.72	16.01	86.90	12.57	88.68	15.01
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	350.44***	33.00	368.80	32.92	367.87*	30.46	382.61	32.25

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted t-tests to compare fall and spring scores within subgroups.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.12. Change Scores for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 – Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Pre-LAS total language screener score English <sup>a</sup>	546	2.24	3.41	0.41	161	2.12	3.61	0.39
Simon Says Score	546	1.32	2.40	0.45	161	1.24	2.47	0.42
Art Show Score	545	0.92	1.92	0.30	160	0.91	1.89	0.30
Pre-LAS total language screener score Spanish <sup>a</sup>	193	1.12	3.94	0.27	52	1.31	3.77	0.32
Tío Simón Dice Score	193	1.01	2.66	0.35	52	1.31	3.06	0.46
Exposición de Arte Score	193	0.12	2.51	0.05	52	0.00	2.50	0.00
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	352	1.00	9.89	0.07	115	2.59	10.63	0.17
Expressive One Word Picture Vocabulary Test— Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	539	2.49	12.69	0.11	160	2.78	12.81	0.12
Expressive One Word Picture Vocabulary Test—IRT Score	540	4.63	6.38	0.43	160	5.04	7.11	0.46
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>o</sup>	351	3.40	11.24	0.24	115	1.99	12.32	0.14
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	351	17.97	22.23	0.62	115	14.78	24.47	0.51
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>o</sup>	77	-1.82	12.20	-0.15	26	-3.12	16.50	-0.26
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	90	8.67	23.81	0.37	31	7.42	31.98	0.32
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	351	2.83	12.29	0.20	115	0.79	12.76	0.06
Woodcock-Johnson III (WJ-III) Spelling W Score	351	18.82	22.62	0.67	115	15.10	23.75	0.53
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>	90	-1.49	25.29	-0.09	31	-0.90	18.21	-0.05
Woodcock-Muñoz-III (WM-III) Ortografía W Score	90	12.09	45.84	0.40	31	13.19	33.16	0.43



Outcome	Center				Family Child Care			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	351	1.12	9.37	0.09	115	2.26	9.75	0.18
Woodcock-Johnson III (WJ-III) Applied Problems W Score	351	11.98	17.38	0.51	115	13.83	18.66	0.59
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	90	3.06	16.42	0.21	31	1.77	14.66	0.12
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	90	18.36	35.63	0.55	31	14.74	31.11	0.44

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted ANOVA to compare change scores between centers and FCCs.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.13. Internal Consistency Reliability of UPCOS-3 Language, Literacy and Math Standardized Measures Spring 2010, by Program Type**

Outcome	Center		Family Child Care	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Pre-LAS total language screener score English <sup>a</sup>	557	0.88	162	0.91
Simon Says Score	557	0.82	162	0.84
Art Show Score	555	0.86	161	0.88
Pre-LAS total language screener score Spanish <sup>a</sup>	199	0.82	52	0.85
Tío Simón Dice Score	199	0.80	52	0.83
Exposición de Arte Score	199	0.79	52	0.86
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	442	0.99	129	0.99
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	553	0.99	161	0.99
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	442	0.93	129	0.92
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	442	0.93	129	0.92
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>	110	0.77	32	0.83
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	110	0.77	32	0.83
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	442	0.90	129	0.91
Woodcock-Johnson III (WJ-III) Spelling W Score	442	0.90	129	0.91
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>	110	0.85	32	0.87
Woodcock-Muñoz-III (WM-III) Ortografía W Score	110	0.85	32	0.87
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	442	0.93	129	0.93
Woodcock-Johnson III (WJ-III) Applied Problems W Score	442	0.93	129	0.93
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	110	0.90	32	0.92
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	110	0.90	32	0.92

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

**Table II.14. Means and Standard Deviations for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 and Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Executive Functioning Pencil Tapping <sup>a</sup>	51.92***		71.47		53.13***		69.57	
Leiter Examiner Ratings Scaled Score <sup>b</sup>								
Attention	8.30***	2.22	8.94	1.79	8.56***	1.98	9.13	1.71
Activity	8.31	1.88	8.40	1.47	8.81	1.48	8.80	1.69
Sociability	8.39***	2.22	8.98	1.68	8.93*	1.66	9.27	1.70
Leiter Examiner Ratings Scaled Score in the Clinical Range <sup>c</sup>								
Attention	6.84***		3.51		4.38*		3.13	
Activity	3.33**		0.92		0.63***		1.88	
Sociability	8.69***		2.96		1.88		2.50	
Leiter Examiner Ratings Scaled Score in Possible Clinical Range <sup>d</sup>								
Attention	12.94***		7.95		10.00*		3.75	
Activity	14.97***		11.65		10.00		6.25	
Sociability	14.23		9.43		10.63		7.50	
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score <sup>e</sup>								
Social Cooperation	101.82***	14.78	106.57	15.11	103.10**	15.05	109.58	13.48
Social Interaction	94.17***	16.54	104.98	14.72	99.56**	16.07	107.28	14.92
Social Independence	99.79***	14.94	107.46	11.88	99.23*	17.79	105.95	14.98
Total Positive Social Skills	98.23***	15.06	107.29	14.00	100.95**	16.43	109.15	15.40
Externalizing Problems	94.60***	12.91	91.16	14.05	95.89	15.02	90.35	14.52
Internalizing Problems	91.94**	13.85	89.08	12.64	95.11*	16.70	88.76	14.61
Problem Behaviors	92.51***	12.98	88.38	13.62	95.28**	15.65	87.05	15.21

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: We conducted t-tests to compare fall and spring scores within subgroups.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

<sup>b</sup>The scaled scores range from 1 to 10 (truncated at 10), with 10 being “average”, or “of no clinical concern”.

<sup>c</sup>Scores of 4 or less.

<sup>d</sup>Scores of 5 and 6.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.15. Change Scores for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 – Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Executive Functioning Pencil Tapping <sup>a</sup>	537	19.55			160	16.45		
Leiter Examiner Ratings Scaled Score								
Attention	541	0.65	2.31	0.30	160	0.56	2.28	0.26
Activity	541	0.09	2.13	0.05	160	-0.01	2.00	-0.01
Sociability	541	0.59	2.40	0.28	160	0.34	2.24	0.16
Leiter Examiner Ratings Scaled Score in the Clinical Range								
Attention	541	-3.33	27.01		160	-1.25	25.05	
Activity	541	-2.40*	19.57		160	1.25	11.15	
Sociability	541	-5.73**	30.80		160	0.63	20.97	
Leiter Examiner Ratings Scaled Score in Possible Clinical Range								
Attention	541	-4.99	39.83		160	-6.25	33.06	
Activity	541	-3.33	45.01		160	-3.75	38.67	
Sociability	541	-4.81	44.46		160	-3.13	41.09	
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score								
Social Cooperation	233	4.74	12.95	0.32	40	6.48	9.84	0.44
Social Interaction	233	10.81	12.19	0.65	39	7.72	12.74	0.47
Social Independence	234	7.67	13.51	0.50	40	6.73	10.53	0.44
Total Positive Social Skills	233	9.06	12.99	0.59	39	8.21	10.27	0.54
Externalizing Problems	222	-3.44	10.42	-0.26	37	-5.54	7.89	-0.42
Internalizing Problems	229	-2.86	10.81	-0.20	38	-6.34	14.54	-0.44
Problem Behaviors	231	-4.13*	10.72	-0.31	40	-8.23	12.60	-0.61

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: We conducted ANOVA to compare change scores between centers and FCCs.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.16. Internal Consistency Reliability of UPCOS-3 Social-Emotional Standardized Measures Spring 2010, by Program Type**

Outcome	Center		Family Child Care	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Executive Functioning Pencil Tapping	553	0.94	161	0.95
Leiter Examiner Ratings Scaled Score				
Attention	555	0.97	161	0.97
Activity	555	0.74	161	0.82
Sociability	555	0.91	161	0.95
Leiter Examiner Ratings Scaled Score in the Clinical Range				
Attention	555	0.97	161	0.97
Activity	555	0.74	161	0.82
Sociability	555	0.91	161	0.95
Leiter Examiner Ratings Scaled Score in Possible Clinical Range				
Attention	555	0.97	161	0.97
Activity	555	0.74	161	0.82
Sociability	555	0.91	161	0.95
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score				
Social Cooperation	258	0.94	44	0.93
Social Interaction	258	0.91	44	0.95
Social Independence	259	0.84	44	0.89
Total Positive Social Skills	258	0.95	44	0.97
Externalizing Problems	246	0.97	40	0.97
Internalizing Problems	253	0.90	42	0.92
Problem Behaviors	255	0.97	44	0.97

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

**Table II.17. Means and Standard Deviations for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 and Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Pre-LAS total language screener score English <sup>a</sup>	12.38***	5.90	15.05	4.54	15.86***	4.54	17.71	3.20
Simon Says Score	6.19***	3.16	7.66	2.37	7.55***	2.59	8.70	1.81
Art Show Score	6.19***	3.38	7.42	2.74	8.33***	2.43	9.03	1.76
Pre-LAS total language screener score Spanish <sup>a</sup>	11.73**	4.11	12.82	4.17	11.22**	3.79	12.53	4.14
Tío Simón Dice Score	5.85***	2.79	6.79	2.36	5.21***	2.90	6.56	2.25
Exposición de Arte Score	5.88	2.54	6.04	2.81	6.01	2.40	5.97	2.74
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	84.84	13.33	86.37	13.46	91.07	15.36	92.40	15.48
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	97.42*	22.14	100.66	20.83	112.42	20.10	114.49	18.33
Expressive One Word Picture Vocabulary Test—IRT Score	40.85***	10.31	45.53	10.16	47.43***	10.37	52.20	40.16
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	97.07**	14.26	100.85	12.67	102.29**	14.27	105.03	14.27
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	326.11***	29.19	344.69	26.71	335.10***	28.25	351.66	29.46
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>	93.17	11.68	92.70	10.52	95.96**	13.15	88.44	11.41
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	312.98***	22.49	324.55	21.49	315.87	25.85	315.06	21.53

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	102.71	13.04	103.60	12.55	104.17***	14.68	107.18	13.96
Woodcock-Johnson III (WJ-III) Spelling W Score	379.42***	26.82	394.95	24.26	380.39***	28.86	399.44	25.89
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>	90.00	16.34	91.06	20.09	91.42	17.35	83.74	23.46
Woodcock-Muñoz-III (WM-III) Ortografía W Score	351.52***	29.97	368.06	36.16	353.68	32.30	355.03	40.06
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	98.05	11.70	99.58	9.33	102.29	12.58	103.64	12.12
Woodcock-Johnson III (WJ-III) Applied Problems W Score	393.35***	23.49	406.77	17.32	399.29***	23.46	411.28	21.45
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	81.34*	14.99	85.73	15.14	82.65	15.02	80.26	17.86
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	354.77***	33.19	375.49	31.02	357.00	33.77	364.19	38.37

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: We conducted t-tests to compare fall and spring scores within subgroups.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.18. Change Scores for UPCOS-3 Language, Literacy and Math Standardized Measures Fall 2009 – Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Pre-LAS total language screener score English <sup>a</sup>	306	2.68***	3.81	0.49	397	1.85	3.12	0.34
Simon Says Score	306	1.47*	2.65	0.50	397	1.16	2.21	0.39
Art Show Score	305	1.22***	2.24	0.40	396	0.69	1.58	0.23
Pre-LAS total language screener score Spanish <sup>a</sup>	171	1.09	3.88	0.27	73	1.32	4.00	0.32
Tío Simón Dice Score	171	0.94	2.68	0.33	73	1.36	2.93	0.48
Exposición de Arte Score	171	0.15	2.59	0.06	73	-0.04	2.32	-0.02
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	146	1.53	9.30	0.10	320	1.33	10.46	0.09
Expressive One Word Picture Vocabulary Test— Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	303	3.24	13.63	0.15	392	2.07	11.99	0.09
Expressive One Word Picture Vocabulary Test—IRT Score	303	4.68	6.49	0.43	393	4.77	6.63	0.44
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>o</sup>	146	3.78	11.49	0.26	319	2.73	11.56	0.19
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	146	18.58	22.28	0.65	319	16.56	23.10	0.58
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>o</sup>	77	-0.47**	12.81	-0.04	25	-7.52	14.00	-0.63
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	88	11.57**	25.15	0.50	31	-0.81	27.38	-0.03
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	146	0.89*	12.02	0.06	319	3.02	12.58	0.21
Woodcock-Johnson III (WJ-III) Spelling W Score	146	15.53	22.15	0.55	319	19.05	23.25	0.67
Woodcock-Muñoz-III (WM-III) Ortografía Standard Score <sup>b</sup>	88	1.06	20.94	0.06	31	-7.68	29.93	-0.47
Woodcock-Muñoz-III (WM-III) Ortografía W Score	88	16.53	37.93	0.55	31	1.35	54.57	0.04



Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	146	1.53	9.59	0.12	319	1.34	9.44	0.11
Woodcock-Johnson III (WJ-III) Applied Problems W Score	146	13.42	19.15	0.57	319	11.98	17.04	0.51
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	88	4.39*	13.72	0.29	31	-2.39	20.48	-0.16
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	88	20.72	29.73	0.63	31	7.19	44.29	0.22

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

Note: Reported by teachers. We conducted ANOVA to compare change scores between high and low concentration ELL programs.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.19. Internal Consistency Reliability of UPCOS-3 Language, Literacy and Math Standardized Measures Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs		Low Concentration of ELLs	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Pre-LAS total language screener score English <sup>a</sup>	312	0.89	403	0.87
Simon Says Score	312	0.82	403	0.80
Art Show Score	310	0.86	402	0.84
Pre-LAS total language screener score Spanish <sup>a</sup>	176	0.83	74	0.83
Tío Simón Dice Score	176	0.81	74	0.79
Exposición de Arte Score	176	0.81	74	0.79
Expressive One Word Picture Vocabulary Test (EOWPVT)—English Edition Standard Score <sup>b</sup>	205	0.98	364	0.99
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score <sup>b,c</sup>	309	0.99	401	0.99
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard Score <sup>b</sup>	205	0.92	364	0.93
Woodcock-Johnson III (WJ-III) Letter-Word Identification W Score	205	0.92	364	0.93
Woodcock-Muñoz-III (WM-III) Letter-Word Identification Standard Score <sup>b</sup>	103	0.80	37	0.73
Woodcock-Muñoz-III (WM-III) Letter-Word Identification W Score	103	0.80	37	0.73
Woodcock-Johnson III (WJ-III) Spelling Standard Score <sup>b</sup>	205	0.91	364	0.90
Woodcock-Johnson III (WJ-III) Spelling W Score	205	0.91	364	0.90
Woodcock-Muñoz-III (WM-III) Spelling Standard Score <sup>b</sup>	103	0.86	37	0.84
Woodcock-Muñoz-III (WM-III) Spelling W Score	103	0.86	37	0.84
Woodcock-Johnson III (WJ-III) Applied Problems Standard Score <sup>b</sup>	205	0.92	364	0.93
Woodcock-Johnson III (WJ-III) Applied Problems W Score	205	0.92	364	0.93
Woodcock-Muñoz-III (WM-III) Applied Problems Standard Score <sup>b</sup>	103	0.90	37	0.91
Woodcock-Muñoz-III (WM-III) Applied Problems W Score	103	0.90	37	0.91

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessments.

<sup>a</sup> These scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

<sup>b</sup> National mean for standard scores is 100 with a standard deviation of 15.

<sup>c</sup> This measure was conceptually scored and the standard score was generated for all children in the sample.

**Table II.20. Means and Standard Deviations for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 and Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
Executive Functioning Pencil Tapping <sup>a</sup>	47.94***		67.72		55.64***		73.77	
Leiter Examiner Ratings Scaled Score <sup>b</sup>								
Attention	8.07***	2.37	8.90	1.83	8.59***	1.98	9.07	1.71
Activity	8.37	1.96	8.52	1.51	8.46	1.69	8.46	1.55
Sociability	8.24***	2.28	8.94	1.68	8.72***	1.96	9.15	1.68
Leiter Examiner Ratings Scaled Score in the Clinical Range <sup>c</sup>								
Attention	9.54***		3.62		3.82***		3.05	
Activity	3.62***		0.99		2.04***		1.27	
Sociability	10.20		2.96		4.83***		2.80	
Leiter Examiner Ratings Scaled Score in Possible Clinical Range <sup>d</sup>								
Attention	14.14		7.89		10.69***		6.36	
Activity	14.14		10.53		13.74***		10.43	
Sociability	15.46		10.20		11.96		7.63	
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score <sup>e</sup>								
Social Cooperation	103.71***	12.90	107.97	14.97	99.84***	16.72	105.78	14.78
Social Interaction	92.30***	18.22	103.83	16.56	98.28***	13.52	107.18	11.85
Social Independence	100.42***	15.20	107.34	12.85	98.80***	15.57	107.11	11.77
Total Positive Social Skills	98.53***	14.96	107.41	14.93	98.73***	15.70	107.75	13.27
Externalizing Problems	91.66**	10.28	88.80	12.53	98.75**	15.34	93.90	15.46
Internalizing Problems	88.18	11.56	86.37	11.53	97.71***	15.63	92.40	13.79
Problem Behaviors	88.80**	10.29	85.68	12.06	98.10***	15.04	91.33	15.28

Source: UPCOS-3 Fall 2009 and Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: Reported by teachers. We conducted t-tests to compare fall and spring scores within subgroups.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

<sup>b</sup>The scaled scores range from 1 to 10 (truncated at 10), with 10 being "average", or "of no clinical concern".

<sup>c</sup>Scores of 4 or less.

<sup>d</sup>Scores of 5 and 6.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.21. Change Scores for UPCOS-3 Social-Emotional Standardized Measures Fall 2009 – Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	N	Mean/ Percentage	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
Executive Functioning Pencil Tapping <sup>a</sup>	303	19.78			390	18.13		
Leiter Examiner Ratings Scaled Score								
Attention	304	0.84**	2.46	0.39	393	0.47	2.15	0.22
Activity	304	0.15	2.22	0.08	393	0.00	2.00	0.00
Sociability	304	0.69	2.57	0.33	393	0.42	2.18	0.20
Leiter Examiner Ratings Scaled Score in the Clinical Range								
Attention	304	-5.92**	29.81		393	-0.76	23.13	
Activity	304	-2.63	19.73		393	-0.76	16.73	
Sociability	304	-7.24**	33.70		393	-2.04	24.66	
Leiter Examiner Ratings Scaled Score in Possible Clinical Range								
Attention	304	-6.25	42.14		393	-4.33	35.09	
Activity	304	-3.62	46.17		393	-3.31	41.82	
Sociability	304	-5.26	46.37		393	-4.33	41.11	
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score								
Social Cooperation	153	4.26	12.94	0.29	120	5.93	11.99	0.40
Social Interaction	152	11.53*	12.32	0.70	120	8.90	12.15	0.54
Social Independence	153	6.92	13.46	0.45	121	8.31	12.67	0.54
Total Positive Social Skills	152	8.88	12.73	0.58	120	9.03	12.54	0.59
Externalizing Problems	145	-2.86	10.11	-0.22	114	-4.85	10.04	-0.37
Internalizing Problems	149	-1.81**	9.48	-0.13	118	-5.31	13.31	-0.37
Problem Behaviors	151	-3.13***	9.80	-0.23	120	-6.77	12.27	-0.50

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

Note: We conducted ANOVA to compare change scores between high and low concentration ELL programs.

<sup>a</sup>Pencil Tapping is percent of the time the child responded correctly.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table II.22. Internal Consistency Reliability of UPCOS-3 Social-Emotional Standardized Measures Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs		Low Concentration of ELLs	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Executive Functioning Pencil Tapping	309	0.94	401	0.94
Leiter Examiner Ratings Scaled Score				
Attention	310	0.97	402	0.97
Activity	310	0.77	402	0.75
Sociability	310	0.91	402	0.93
Leiter Examiner Ratings Scaled Score in the Clinical Range				
Attention	310	0.97	402	0.97
Activity	310	0.77	402	0.75
Sociability	310	0.91	402	0.93
Leiter Examiner Ratings Scaled Score in Possible Clinical Range				
Attention	310	0.97	402	0.97
Activity	310	0.77	402	0.75
Sociability	310	0.91	402	0.93
Preschool Kindergarten Behavior Scale—Second Edition (PKBS-2) Standard Score				
Social Cooperation	154	0.95	148	0.94
Social Interaction	154	0.93	148	0.90
Social Independence	154	0.86	149	0.84
Total Positive Social Skills	154	0.96	148	0.96
Externalizing Problems	145	0.96	141	0.97
Internalizing Problems	149	0.90	146	0.90
Problem Behaviors	151	0.96	148	0.97

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment (PAA).

**This page has been left blank for double-sided copying.**

**APPENDIX C**

**TABLES OF RESULTS PRESENTED IN CHAPTER III:  
PRE-KINDERGARTEN OBSERVATION FORM**

**This page has been left blank for double-sided copying.**



**Table III.1 Number of Children with Pre-Kindergarten Observation Form (P-KOF) Data Fall 2009 and Spring 2010, by Program Type and Language Group**

<i>Language group</i>	Fall 2009			Spring 2010		
	Center	FCC	Total	Center	FCC	Total
English only	199	78	277	161	73	234
English primarily	141	34	175	104	33	137
Spanish only	60	15	75	51	16	67
Span primarily	146	41	187	118	30	148
Other Language only or Primarily	19	3	22	15	3	18
Missing	9	8	17	1		1
Total	574	179	753	450	155	605

**Table III.2. Means, Standard Deviations, and Cronbach's Alphas for of Pre-Kindergarten Observation Form Fall 2009 – Spring 2010**

Outcome	N	Fall 2009		Spring 2010		Cronbach's Alpha	Fall-Spring Change		
		Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation		Mean/ Percentage	Standard Deviation	Effect Size
Social Expression	532	3.14***	0.74	3.72	0.42	0.88	0.58	0.64	0.78
General Knowledge	532	2.81***	0.71	3.53	0.46	0.81	0.72	0.53	1.01
Self-Regulation	545	3.04***	0.72	3.61	0.44	0.91	0.57	0.58	0.79
Self-care & Motor Skills	545	3.46***	0.59	3.87	0.28	0.67	0.40	0.55	0.68
Overall score	546	3.05***	0.61	3.65	0.34	0.92	0.59	0.48	0.97

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures

Note: Reported by teachers. We conducted t-tests to compare fall and spring scores.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.3. Means and Standard Deviations for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 and Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010	Fall 2009	Spring 2010
	Mean (SD) <sup>a</sup>	Mean (SD)	Mean (SD) <sup>a</sup>	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Social Expression	3.24*** (0.71)	3.74 (0.38)	3.16*** (0.70)	3.74 (0.40)	2.75*** (0.84)	3.58 (0.51)	3.12*** (0.75)	3.73 (0.46)	3.33* (0.64)	3.72 (0.44)
General Knowledge	3.00*** (0.69)	3.60 (0.43)	2.87*** (0.65)	3.62 (0.42)	2.32*** (0.63)	3.32 (0.49)	2.62*** (0.67)	3.39 (0.49)	3.48** (0.37)	3.78 (0.32)
Self-Regulation	3.07*** (0.71)	3.53 (0.46)	3.04*** (0.67)	3.66 (0.40)	2.80*** (0.72)	3.57 (0.52)	3.07*** (0.77)	3.69 (0.41)	3.41* (0.48)	3.70 (0.38)
Self-Care & Motor Skills	3.56*** (0.52)	3.86 (0.28)	3.40*** (0.60)	3.88 (0.27)	3.28*** (0.64)	3.81 (0.38)	3.43*** (0.64)	3.91 (0.22)	3.79 (0.34)	3.81 (0.40)
Overall score	3.16*** (0.60)	3.65 (0.32)	3.06*** (0.58)	3.70 (0.32)	2.71*** (0.60)	3.53 (0.41)	3.00*** (0.63)	3.64 (0.35)	3.46** (0.36)	3.74 (0.35)

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers. We conducted t-tests to compare fall and spring scores within each subgroup.

<sup>a</sup>SD = Standard deviation.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.4. Change Scores for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 – Spring 2010, by Language Group**

Outcome	English Only			English Primarily			Spanish Only			Spanish Primarily			Other Language Only or Primarily		
	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size
Social Expression	197	0.50** (.56)	0.68	122	0.59 (0.67)	0.80	59	0.83 (0.77)	1.12	138	0.61 (0.61)	0.82	16	0.39 (0.72)	0.53
General Knowledge	197	0.60*** (.49)	0.85	122	0.75 (0.52)	1.06	59	1.00 (0.54)	1.41	138	0.77 (0.54)	1.08	16	0.29 (0.22)	0.41
Self-regulation	205	0.46*** (.53)	0.64	124	0.63 (0.57)	0.88	60	0.77 (0.66)	1.07	139	0.62 (0.60)	0.86	16	0.29 (0.48)	0.40
Self-care and Motor Skills	205	0.30*** (.47)	0.51	124	0.48 (0.57)	0.81	60	0.53 (0.63)	0.90	139	0.48 (0.57)	0.81	16	0.02 (0.56)	0.03
Overall score	205	0.49*** (.43)	0.80	124	0.64 (0.49)	1.05	60	0.81 (0.52)	1.33	140	0.65 (0.49)	1.07	16	0.28 (0.40)	0.46

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

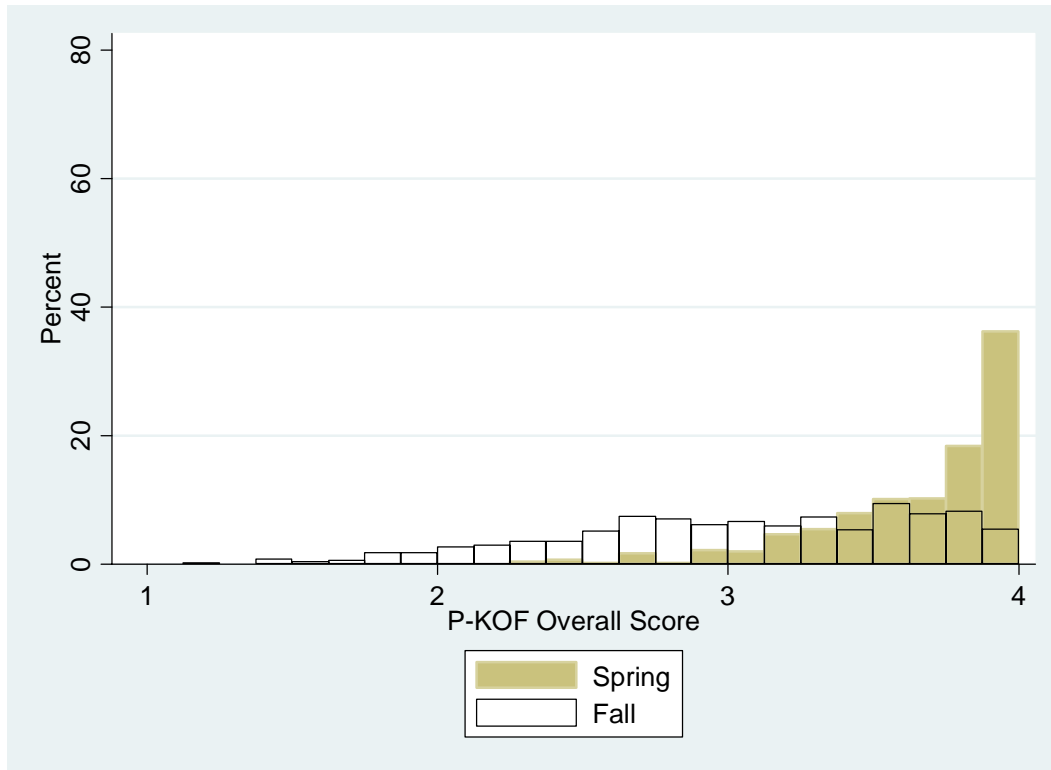
Note: Reported by teachers. We conducted ANOVA to compare change scores between language groups.

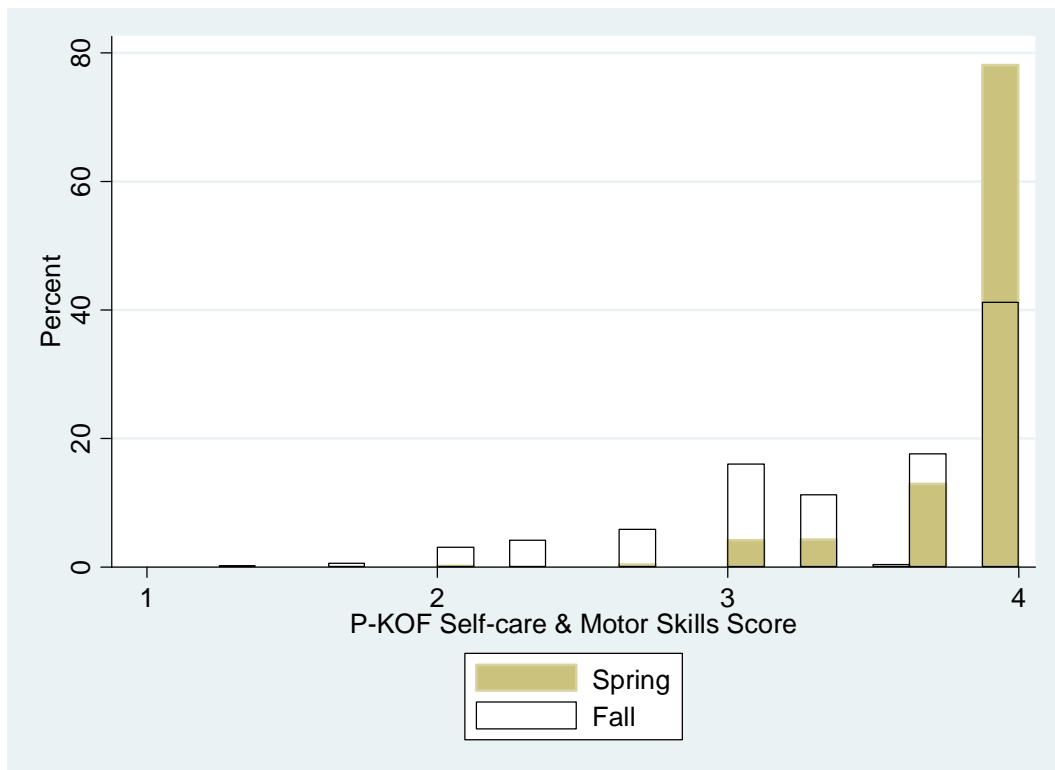
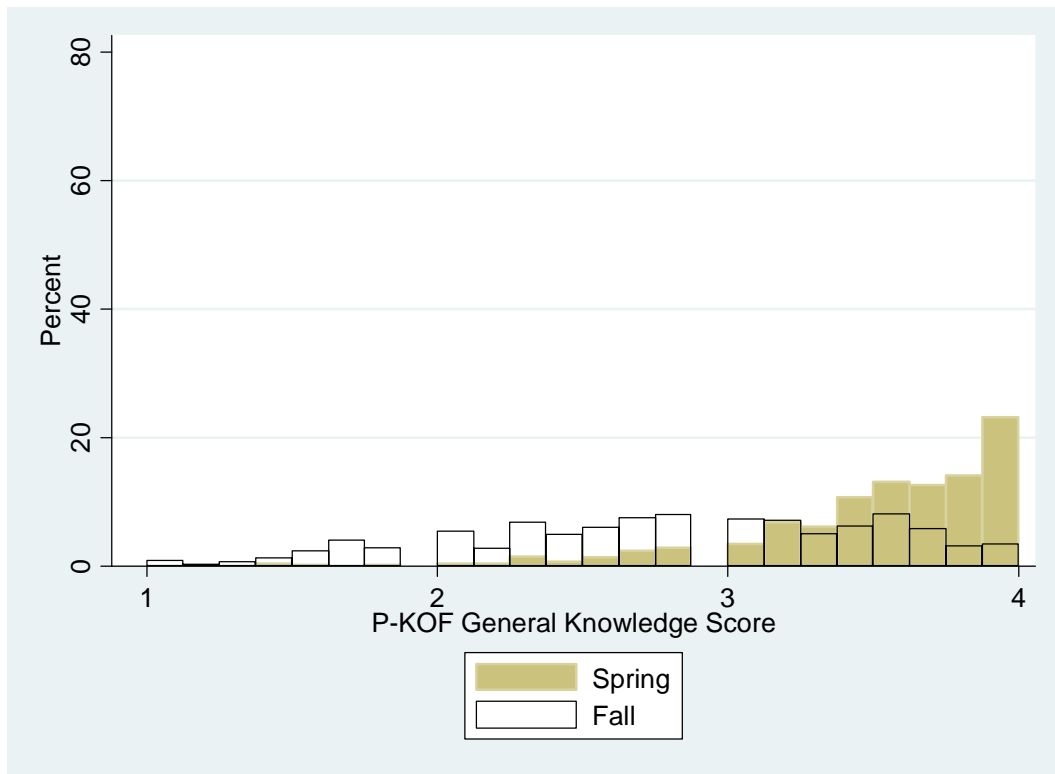
\*Significantly different from zero at the .10 level, two-tailed test.

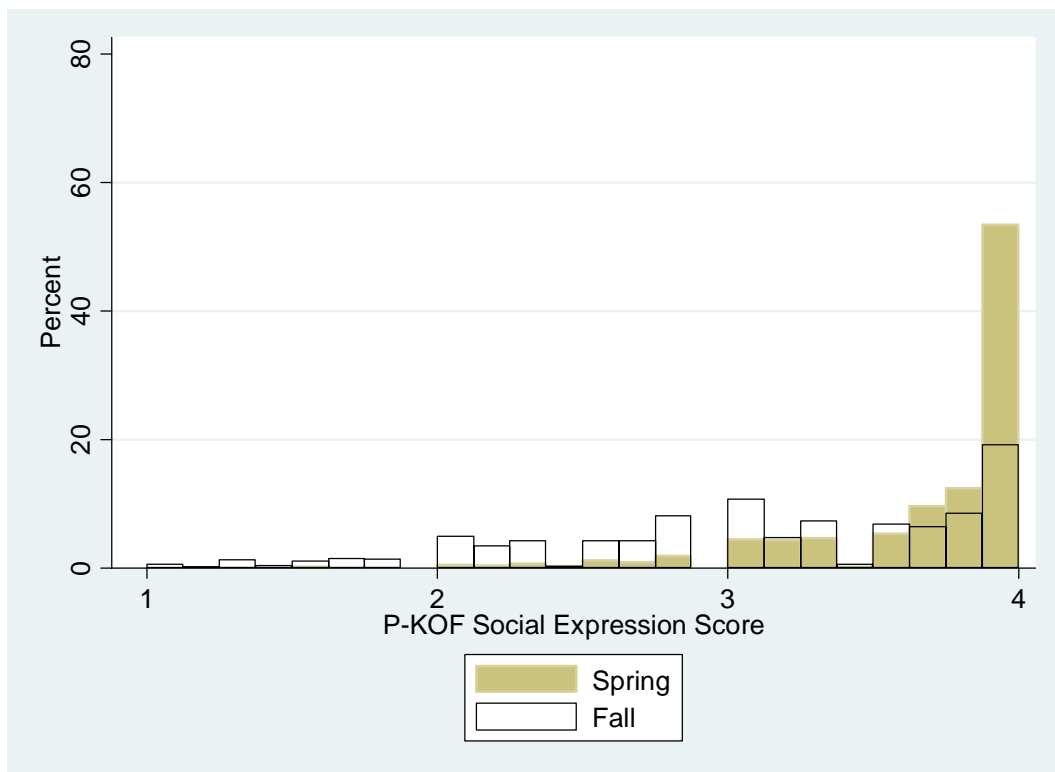
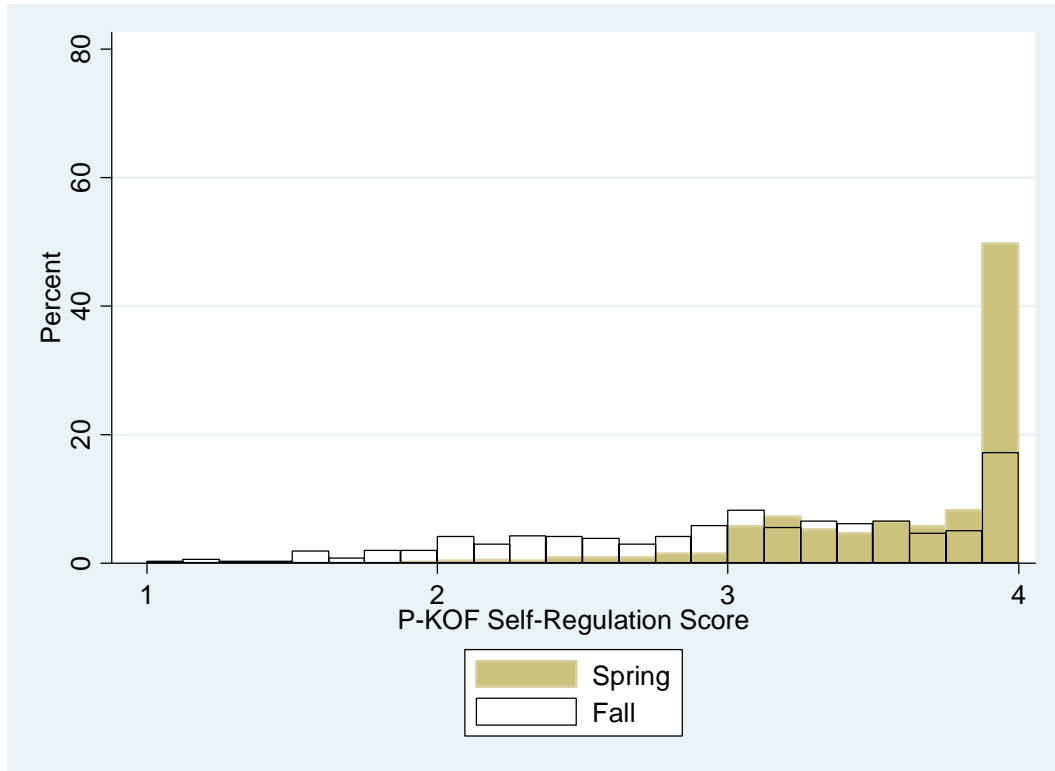
\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

Figure III.1 Distributions of Fall and Spring Pre- Kindergarten Observation Form (P- KOF) Scores







**Table III.5. Internal Consistency Reliability of Pre-Kindergarten Observation Form (P-KOF) Scales Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily		Other Language Only or Primarily	
	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha
Social Expression	233	0.85	137	0.86	67	0.93	148	0.89	18	0.97
General Knowledge	233	0.84	137	0.80	67	0.81	148	0.77	18	0.73
Self-Regulation	234	0.91	137	0.88	67	0.94	148	0.91	18	0.92
Self-Care & Motor Skills	234	0.57	137	0.62	67	0.85	148	0.58	18	1.00
Overall score	234	0.92	137	0.92	67	0.95	148	0.93	18	0.96

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.



**Table III.6. Means and Standard Deviations for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 and Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Social Expression	3.10***	0.74	3.71	0.41	3.27***	0.73	3.74	0.45
General Knowledge	2.80***	0.71	3.51	0.47	2.85***	0.71	3.59	0.44
Self-Regulation	3.01***	0.70	3.61	0.42	3.14***	0.75	3.62	0.50
Self-Care & Motor Skills	3.43***	0.60	3.86	0.29	3.56***	0.53	3.89	0.26
Overall score	3.02***	0.61	3.64	0.33	3.14***	0.60	3.68	0.37

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers. We conducted t-tests to compare fall and spring scores within each subgroup.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.7. Change Scores for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 – Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	N	Mean	Standard Deviation	Effect Size	N	Mean	Standard Deviation	Effect Size
Social Expression	404	0.62**	0.64	0.84	128	0.48	0.60	0.65
General Knowledge	404	0.71	0.51	1.00	128	0.74	0.58	1.04
Self-Regulation	413	0.60**	0.59	0.83	132	0.48	0.55	0.67
Self-Care & Motor Skills	413	0.43**	0.56	0.73	132	0.33	0.49	0.56
Overall score	414	0.61	0.48	1.00	132	0.54	0.48	0.89

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers. We conducted t-tests to compare change scores between centers and FCCs.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.8. Internal Consistency Reliability of Pre-Kindergarten Observation Form (P-KOF) Scales Spring 2010, by Program Type**

Outcome	Center		Family Child Care	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Social Expression	449	0.87	155	0.89
General Knowledge	449	0.81	155	0.83
Self-Regulation	450	0.90	155	0.93
Self-Care & Motor Skills	450	0.68	155	0.63
Overall score	450	0.92	155	0.94

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.

**Table III.9. Means and Standard Deviations for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 and Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Social Expression	3.10***	0.76	3.73	0.44	3.17***	0.72	3.71	0.40
General Knowledge	2.66***	0.68	3.43	0.50	2.95***	0.70	3.62	0.40
Self-Regulation	3.01***	0.73	3.66	0.44	3.07***	0.70	3.56	0.44
Self-Care & Motor Skills	3.41***	0.62	3.86	0.29	3.52***	0.54	3.87	0.27
Overall score	2.98***	0.62	3.63	0.36	3.12***	0.60	3.66	0.32

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers. We conducted t-tests to compare fall and spring scores within subgroups.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.10. Change Scores for Pre-Kindergarten Observation Form (P-KOF) Scales Fall 2009 – Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	N	Mean	Standard Deviation	Effect Size	N	Mean	Standard Deviation	Effect Size
Social Expression	260	0.62	0.69	0.84	272	0.54	0.58	0.73
General Knowledge	260	0.77**	0.54	1.08	272	0.67	0.51	0.94
Self-Regulation	269	0.64***	0.62	0.89	276	0.50	0.54	0.69
Self-Care & Motor Skills	269	0.46**	0.59	0.78	276	0.35	0.51	0.59
Overall score	269	0.65***	0.51	1.07	277	0.54	0.45	0.89

Source: Fall 2009 and Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers. We conducted t-tests to compare change scores between high and low ELL programs.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table III.11. Internal Consistency Reliability of Pre-Kindergarten Observation Form (P-KOF) Scales Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs		Low Concentration of ELLs	
	N	Cronbach's Alpha	N	Cronbach's Alpha
Social Expression	290	0.90	314	0.85
General Knowledge	290	0.80	314	0.82
Self-Regulation	290	0.91	315	0.91
Self-Care & Motor Skills	290	0.72	315	0.62
Overall score	290	0.93	315	0.92

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.

**Table III.12. Inter-factor Correlations for Pre-Kindergarten Observation Form (P-KOF) Subscales and Overall Score Spring 2010**

Outcome	Social Expression	General Knowledge	Self-Regulation	Self-care & Motor Skills
Social Expression				
General Knowledge	0.50			
Self-Regulation	0.65	0.45		
Self-Care & Motor Skills	0.52	0.36	0.51	
Overall score	0.84	0.78	0.86	0.62

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.

**Table III.13. Inter-factor Correlations for Pre-Kindergarten Observation Form (P-KOF) Subscales and Overall Score Spring 2010, by Program Type**

Outcome	Social Expression	General Knowledge	Self-Regulation	Self-care & Motor Skills
Center				
Social Expression				
General Knowledge	0.48			
Self-Regulation	0.59	0.40		
Self-Care & Motor Skills	0.49	0.34	0.47	
Overall score	0.82	0.77	0.84	0.60
Family Child Care				
Social Expression				
General Knowledge	0.56			
Self-Regulation	0.81	0.57		
Self-Care & Motor Skills	0.59	0.42	0.61	
Overall score	0.89	0.80	0.93	0.67

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.



**Table III.13. Inter-factor Correlations for Pre-Kindergarten Observation Form (P-KOF) Subscales and Overall Score Spring 2010, by Concentration of ELLs**

Outcome	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
High Concentration of ELLs				
Social Expression				
General Knowledge	0.55			
Self-Regulation	0.65	0.48		
Self-Care & Motor Skills	0.50	0.34	0.64	
Overall score	0.85	0.80	0.87	0.65
Low Concentration of ELLs				
Social Expression				
General Knowledge	0.45			
Self-regulation	0.67	0.48		
Self-care & Motor Skills	0.53	0.38	0.39	
Overall score	0.83	0.76	0.88	0.59

Source: Spring 2010 Teacher-Administered Measures.

Note: Reported by teachers.

**Table III.14. Concurrent Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010**

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Pre-LAS total language screener score English	0.27	0.43	0.11	0.09
Simon Says Score	0.24	0.32	0.14	0.11
Art Show Score	0.22	0.41	0.06	0.05
Pre-LAS total language screener score Spanish	0.19	0.20	0.08	0.08
Tío Simón Dice Score	0.25	0.22	0.18	0.12
Exposición de Arte Score	0.06	0.10	-0.02	0.02
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.12	0.32	0.07	0.03
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.22	0.44	0.10	0.02
Expressive One Word Picture Vocabulary Test—IRT Score	0.23	0.46	0.13	0.07
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.15	0.45	0.14	0.08
Letter-Word Identification W Score	0.19	0.51	0.19	0.13
Applied Problems Standard Score	0.21	0.43	0.25	0.15
Applied Problems W Score	0.26	0.50	0.31	0.22
Spelling Standard Score	0.20	0.47	0.27	0.21
Spelling W Score	0.25	0.53	0.32	0.27
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.01	0.22	0.11	-0.12
Letter-Word Identification W Score	0.06	0.28	0.15	-0.06
Applied Problems Standard Score	0.01	-0.06	0.09	-0.01
Applied Problems W Score	0.03	-0.03	0.11	0.02
Ortografía Standard Score	-0.11	-0.01	0.05	-0.04
Ortografía W Score	-0.07	0.03	0.08	-0.00

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.42	0.31	0.58	0.28
Social Interaction	0.66	0.42	0.43	0.26
Social Independence	0.61	0.33	0.50	0.37
Total Positive Social Skills	0.64	0.41	0.57	0.34
Externalizing Problems	-0.23	-0.09	-0.53	-0.15
Internalizing Problems	-0.29	-0.01	-0.32	-0.13
Problem Behaviors	-0.29	-0.07	-0.46	-0.16
Executive Functioning				
Pencil Tapping	0.17	0.32	0.22	0.10
Leiter Examiner Ratings				
Attention	0.25	0.36	0.28	0.13
Activity	0.17	0.28	0.22	0.05
Sociability	0.18	0.30	0.18	0.08

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

Table III.15. Correlations (Pairwise) Between Change Scores on Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Pre-LAS total language screener score English	0.09	0.22	0.16	0.10
Simon Says Score	0.09	0.13	0.15	0.07
Art Show Score	0.05	0.22	0.10	0.09
Pre-LAS total language screener score Spanish	0.03	0.08	0.12	0.04
Tío Simón Dice Score	-0.01	0.13	0.08	-0.03
Exposición de Arte Score	0.07	-0.01	0.09	0.10
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	-0.09	-0.02	-0.07	-0.06
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.00	0.01	0.03	-0.01
Expressive One Word Picture Vocabulary Test—IRT Score	-0.02	-0.03	0.00	-0.02
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.00	0.11	0.06	0.02
Letter-Word Identification W Score	-0.02	0.08	0.05	0.00
Applied Problems Standard Score	-0.08	0.05	-0.04	-0.07
Applied Problems W Score	-0.04	0.07	-0.01	-0.05
Spelling Standard Score	0.02	0.07	0.05	0.03
Spelling W Score	0.01	0.07	0.05	0.02
Woodcock-Muñoz Bateria III (WM-III)				
Letter-Word Identification Standard Score	-0.02	0.15	-0.10	-0.03
Letter-Word Identification W Score	-0.04	0.19	-0.10	0.04
Applied Problems Standard Score	-0.12	-0.07	0.01	-0.03
Applied Problems W Score	-0.11	-0.07	0.02	-0.02
Ortografía Standard Score	0.07	0.05	-0.02	0.06
Ortografía W Score	0.07	0.05	-0.01	0.06

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.14	0.05	0.06	-0.01
Social Interaction	0.28	0.22	0.19	0.18
Social Independence	0.16	0.10	0.07	0.02
Total Positive Social Skills	0.23	0.15	0.12	0.06
Externalizing Problems	-0.08	-0.06	-0.08	0.02
Internalizing Problems	-0.08	-0.04	-0.11	-0.03
Problem Behaviors	-0.05	-0.01	-0.06	0.01
Executive Functioning				
Pencil Tapping	-0.01	0.01	-0.05	-0.07
Leiter Examiner Ratings				
Attention	0.07	-0.04	0.06	0.03
Activity	0.05	0.00	0.03	-0.01
Sociability	0.02	-0.06	0.02	-0.02

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

**Table III.16. Predictive Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscale Scores from Fall 2009 and Standardized Measures from Spring 2010**

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Pre-LAS total language screener score English	0.27	0.44	0.19	0.18
Simon Says Score	0.25	0.38	0.18	0.16
Art Show Score	0.21	0.40	0.15	0.16
Pre-LAS total language screener score Spanish	0.17	0.22	0.11	0.06
Tío Simón Dice Score	0.21	0.22	0.15	0.11
Exposición de Arte Score	0.08	0.13	0.03	-0.00
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.21	0.39	0.19	0.14
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.25	0.48	0.19	0.17
Expressive One Word Picture Vocabulary Test—IRT Score	0.27	0.52	0.22	0.18
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.12	0.46	0.16	0.10
Letter-Word Identification W Score	0.18	0.54	0.20	0.14
Applied Problems Standard Score	0.25	0.44	0.31	0.20
Applied Problems W Score	0.31	0.52	0.36	0.24
Spelling Standard Score	0.15	0.42	0.23	0.18
Spelling W Score	0.21	0.51	0.29	0.23
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.10	0.15	0.11	-0.02
Letter-Word Identification W Score	0.14	0.20	0.14	-0.02
Applied Problems Standard Score	0.15	-0.03	0.14	0.17
Applied Problems W Score	0.18	-0.00	0.16	0.17
Ortografía Standard Score	0.01	-0.04	0.06	0.04
Ortografía W Score	0.04	-0.01	0.09	0.04

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.29	0.23	0.49	0.28
Social Interaction	0.48	0.30	0.43	0.31
Social Independence	0.40	0.28	0.42	0.31
Total Positive Social Skills	0.44	0.30	0.51	0.34
Externalizing Problems	-0.10	-0.02	-0.33	-0.07
Internalizing Problems	-0.10	0.03	-0.15	-0.06
Problem Behaviors	-0.12	-0.03	-0.27	-0.09
Executive Functioning				
Pencil Tapping	0.24	0.39	0.27	0.18
Leiter Examiner Ratings				
Attention	0.23	0.34	0.27	0.15
Activity	0.16	0.28	0.22	0.15
Sociability	0.20	0.29	0.20	0.14

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Teacher-Administered Measures.

Table III.17. Concurrent Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010, by Program Type

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
	Center			
Pre-LAS total language screener score English	0.24	0.41	0.07	0.09
Simon Says Score	0.21	0.28	0.11	0.11
Art Show Score	0.19	0.41	0.02	0.05
Pre-LAS total language screener score Spanish	0.14	0.16	0.03	0.06
Tío Simón Dice Score	0.14	0.17	0.08	0.07
Exposición de Arte Score	0.10	0.09	-0.01	0.03
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.14	0.31	0.06	0.04
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.21	0.43	0.07	0.03
Expressive One Word Picture Vocabulary Test—IRT Score	0.23	0.45	0.10	0.07
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.20	0.47	0.15	0.09
Letter-Word Identification W Score	0.24	0.53	0.19	0.14
Applied Problems Standard Score	0.19	0.42	0.24	0.13
Applied Problems W Score	0.24	0.51	0.29	0.20
Spelling Standard Score	0.19	0.47	0.23	0.15
Spelling W Score	0.25	0.55	0.28	0.23
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.01	0.15	0.09	-0.11
Letter-Word Identification W Score	0.05	0.21	0.11	-0.07
Applied Problems Standard Score	0.01	-0.09	0.14	-0.02
Applied Problems W Score	0.02	-0.06	0.15	0.00
Ortografía Standard Score	-0.11	-0.09	0.06	-0.02
Ortografía W Score	-0.08	-0.04	0.07	0.01



Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.40	0.29	0.56	0.22
Social Interaction	0.63	0.37	0.36	0.20
Social Independence	0.56	0.25	0.45	0.31
Total Positive Social Skills	0.61	0.36	0.53	0.28
Externalizing Problems	-0.23	-0.09	-0.55	-0.14
Internalizing Problems	-0.30	0.04	-0.33	-0.09
Problem Behaviors	-0.30	-0.05	-0.49	-0.13
Executive Functioning				
Pencil Tapping	0.13	0.29	0.14	0.05
Leiter Examiner Ratings				
Attention	0.23	0.37	0.23	0.10
Activity	0.13	0.30	0.16	-0.00
Sociability	0.15	0.34	0.14	0.05
Family Child Care				
Pre-LAS total language screener score English				
Simon Says Score	0.35	0.48	0.21	0.10
Art Show Score	0.35	0.47	0.22	0.12
	0.28	0.41	0.17	0.06
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.32	0.40	0.20	0.15
Exposición de Arte Score	0.62	0.50	0.41	0.28
	-0.07	0.12	-0.06	-0.02
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	0.05	0.31	0.08	-0.03
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.21	0.45	0.20	-0.01
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.24	0.46	0.25	0.04
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	-0.01	0.40	0.12	0.06
Letter-Word Identification W Score	0.03	0.44	0.18	0.11
Applied Problems Standard Score	0.28	0.51	0.31	0.21
Applied Problems W Score	0.33	0.54	0.39	0.29
Spelling Standard Score	0.25	0.53	0.39	0.39
Spelling W Score	0.28	0.54	0.45	0.44

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	-0.01	0.44	0.14	-0.13
Letter-Word Identification W Score	0.09	0.47	0.23	-0.04
Applied Problems Standard Score	0.08	0.09	0.11	0.03
Applied Problems W Score	0.15	0.13	0.18	0.08
Ortografía Standard Score	-0.06	0.39	0.16	-0.15
Ortografía W Score	0.05	0.47	0.28	-0.07
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.61	0.43	0.73	0.66
Social Interaction	0.84	0.67	0.78	0.63
Social Independence	0.86	0.70	0.71	0.73
Total Positive Social Skills	0.84	0.66	0.81	0.73
Externalizing Problems	-0.19	-0.11	-0.40	-0.26
Internalizing Problems	-0.28	-0.28	-0.27	-0.41
Problem Behaviors	-0.24	-0.14	-0.35	-0.29
Executive Functioning				
Pencil Tapping	0.32	0.45	0.41	0.25
Leiter Examiner Ratings				
Attention	0.31	0.34	0.40	0.24
Activity	0.27	0.20	0.36	0.19
Sociability	0.26	0.15	0.27	0.14

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

Table III.18. Correlations (Pairwise) Between Change Scores on Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010, by Program Type

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
	Center			
Pre-LAS total language screener score English	0.07	0.19	0.16	0.08
Simon Says Score	0.06	0.09	0.13	0.04
Art Show Score	0.05	0.21	0.12	0.09
Pre-LAS total language screener score Spanish	-0.02	0.05	0.08	0.00
Tío Simón Dice Score	-0.07	0.08	0.01	-0.08
Exposición de Arte Score	0.04	-0.01	0.12	0.09
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	-0.07	-0.07	-0.03	-0.03
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.03	0.00	0.06	0.02
Expressive One Word Picture Vocabulary Test—IRT Score	0.00	-0.05	0.04	0.01
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	-0.02	0.10	0.04	-0.01
Letter-Word Identification W Score	-0.04	0.07	0.02	-0.03
Applied Problems Standard Score	-0.14	-0.02	-0.04	-0.10
Applied Problems W Score	-0.10	-0.01	0.00	-0.07
Spelling Standard Score	-0.04	0.07	0.03	0.01
Spelling W Score	-0.04	0.06	0.03	0.01
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	-0.18	-0.07	-0.33	-0.27
Letter-Word Identification W Score	-0.14	0.00	-0.28	-0.16
Applied Problems Standard Score	-0.07	-0.13	0.02	-0.03
Applied Problems W Score	-0.07	-0.14	0.03	-0.02
Ortografía Standard Score	0.03	-0.06	-0.03	0.02
Ortografía W Score	0.04	-0.06	-0.02	0.03

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.14	0.01	0.06	0.00
Social Interaction	0.26	0.17	0.17	0.15
Social Independence	0.17	0.11	0.08	0.03
Total Positive Social Skills	0.22	0.12	0.11	0.05
Externalizing Problems	-0.07	-0.07	-0.09	0.01
Internalizing Problems	-0.07	-0.03	-0.08	-0.03
Problem Behaviors	-0.05	-0.01	-0.04	0.00
Executive Functioning				
Pencil Tapping	-0.04	-0.02	-0.07	-0.08
Leiter Examiner Ratings				
Attention	0.10	-0.03	0.09	0.05
Activity	0.09	0.04	0.07	0.02
Sociability	0.04	-0.04	0.06	-0.01
Family Child Care				
Pre-LAS total language screener score English				
Simon Says Score	0.15	0.32	0.15	0.20
Art Show Score	0.19	0.28	0.18	0.20
	0.04	0.25	0.05	0.12
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.29	0.22	0.28	0.24
Exposición de Arte Score	0.18	0.26	0.33	0.17
	0.19	-0.02	-0.04	0.14
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	-0.10	0.09	-0.16	-0.10
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	-0.06	0.03	-0.09	-0.11
Expressive One Word Picture Vocabulary Test—IRT Score				
	-0.06	0.01	-0.11	-0.11
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.06	0.14	0.13	0.08
Letter-Word Identification W Score	0.04	0.12	0.10	0.06
Applied Problems Standard Score	0.15	0.22	0.00	0.03
Applied Problems W Score	0.18	0.28	-0.02	0.05
Spelling Standard Score	0.17	0.10	0.12	0.04
Spelling W Score	0.15	0.12	0.10	0.03

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Bateria III (WM-III)				
Letter-Word Identification Standard Score	0.34	0.58	0.53	0.59
Letter-Word Identification W Score	0.21	0.58	0.37	0.51
Applied Problems Standard Score	-0.37	0.22	-0.01	-0.07
Applied Problems W Score	-0.37	0.26	0.00	-0.02
Ortografía Standard Score	0.36	0.54	0.15	0.31
Ortografía W Score	0.37	0.54	0.17	0.32
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.19	0.47	0.11	-0.12
Social Interaction	0.39	0.59	0.31	0.39
Social Independence	0.00	0.11	-0.11	-0.10
Total Positive Social Skills	0.31	0.57	0.17	0.15
Externalizing Problems	-0.14	0.03	-0.05	0.16
Internalizing Problems	-0.27	-0.11	-0.25	-0.02
Problem Behaviors	-0.09	0.05	-0.14	0.13
Executive Functioning				
Pencil Tapping	0.11	0.10	0.00	-0.03
Leiter Examiner Ratings				
Attention	-0.05	-0.05	-0.06	-0.04
Activity	-0.13	-0.14	-0.14	-0.15
Sociability	-0.11	-0.14	-0.13	-0.10

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

**Table III.19. Predictive Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscale Scores from Fall 2009 and Standardized Measures from Spring 2010, by Program Type**

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
	Center			
Pre-LAS total language screener score English	0.28	0.44	0.20	0.17
Simon Says Score	0.27	0.37	0.21	0.16
Art Show Score	0.20	0.39	0.14	0.14
Pre-LAS total language screener score Spanish	0.13	0.27	0.04	-0.03
Tío Simón Dice Score	0.12	0.24	0.08	0.01
Exposición de Arte Score	0.09	0.20	-0.01	-0.05
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.21	0.40	0.20	0.12
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.23	0.48	0.18	0.13
Expressive One Word Picture Vocabulary Test—IRT Score	0.26	0.52	0.21	0.15
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.16	0.46	0.19	0.10
Letter-Word Identification W Score	0.21	0.53	0.23	0.13
Applied Problems Standard Score	0.26	0.44	0.33	0.19
Applied Problems W Score	0.32	0.52	0.36	0.24
Spelling Standard Score	0.16	0.42	0.23	0.17
Spelling W Score	0.23	0.52	0.28	0.22
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.15	0.21	0.19	0.09
Letter-Word Identification W Score	0.20	0.26	0.21	0.08
Applied Problems Standard Score	0.11	-0.04	0.11	0.12
Applied Problems W Score	0.14	-0.01	0.13	0.13
Ortografía Standard Score	0.08	0.00	0.09	0.11
Ortografía W Score	0.11	0.04	0.11	0.11

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.26	0.22	0.47	0.27
Social Interaction	0.45	0.27	0.40	0.29
Social Independence	0.34	0.21	0.38	0.25
Total Positive Social Skills	0.40	0.27	0.48	0.31
Externalizing Problems	-0.07	0.01	-0.28	-0.05
Internalizing Problems	-0.06	0.10	-0.11	0.01
Problem Behaviors	-0.08	0.02	-0.23	-0.03
Executive Functioning				
Pencil Tapping	0.26	0.40	0.28	0.18
Leiter Examiner Ratings				
Attention	0.25	0.35	0.27	0.16
Activity	0.13	0.27	0.17	0.13
Sociability	0.18	0.31	0.17	0.12
Family Child Care				
Pre-LAS total language screener score English				
Simon Says Score	0.28	0.45	0.16	0.25
Art Show Score	0.20	0.42	0.10	0.19
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.25	0.42	0.16	0.23
Exposición de Arte Score	0.34	0.04	0.33	0.34
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	0.51	0.17	0.39	0.41
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.08	-0.08	0.16	0.14
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.19	0.36	0.16	0.24
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.31	0.45	0.22	0.30
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.33	0.50	0.27	0.32
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	-0.03	0.47	0.03	0.14
Letter-Word Identification W Score	0.03	0.57	0.11	0.20
Applied Problems Standard Score	0.21	0.44	0.25	0.21
Applied Problems W Score	0.28	0.52	0.35	0.27
Spelling Standard Score	0.10	0.41	0.24	0.24
Spelling W Score	0.17	0.48	0.32	0.29

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	-0.04	0.02	-0.08	-0.29
Letter-Word Identification W Score	-0.01	0.07	-0.02	-0.26
Applied Problems Standard Score	0.30	-0.04	0.21	0.30
Applied Problems W Score	0.31	-0.03	0.24	0.30
Ortografía Standard Score	-0.30	-0.32	-0.12	-0.35
Ortografía W Score	-0.29	-0.28	-0.06	-0.33
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.50	0.32	0.65	0.37
Social Interaction	0.68	0.45	0.61	0.42
Social Independence	0.72	0.62	0.61	0.62
Total Positive Social Skills	0.70	0.51	0.68	0.52
Externalizing Problems	-0.29	-0.27	-0.58	-0.19
Internalizing Problems	-0.40	-0.47	-0.35	-0.51
Problem Behaviors	-0.37	-0.39	-0.45	-0.41
Executive Functioning				
Pencil Tapping	0.15	0.34	0.25	0.19
Leiter Examiner Ratings				
Attention	0.16	0.29	0.29	0.12
Activity	0.24	0.28	0.35	0.21
Sociability	0.24	0.22	0.29	0.19

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Teacher-Administered Measures.



Table III.20. Concurrent Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010, by Concentration of ELLs

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
High Concentrations of ELLs				
Pre-LAS total language screener score English	0.26	0.40	0.14	0.06
Simon Says Score	0.23	0.25	0.17	0.09
Art Show Score	0.20	0.41	0.08	0.01
Pre-LAS total language screener score Spanish	0.11	0.15	0.01	-0.01
Tío Simón Dice Score	0.14	0.20	0.12	0.05
Exposición de Arte Score	0.03	0.04	-0.08	-0.05
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.10	0.33	0.04	0.02
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.23	0.44	0.13	0.01
Expressive One Word Picture Vocabulary Test—IRT Score	0.24	0.46	0.16	0.04
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.09	0.46	0.09	0.05
Letter-Word Identification W Score	0.13	0.50	0.15	0.09
Applied Problems Standard Score	0.08	0.32	0.15	0.09
Applied Problems W Score	0.11	0.38	0.22	0.13
Spelling Standard Score	0.05	0.45	0.23	0.16
Spelling W Score	0.09	0.48	0.29	0.20
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.07	0.27	0.18	-0.09
Letter-Word Identification W Score	0.10	0.31	0.21	-0.05
Applied Problems Standard Score	-0.01	0.01	0.13	-0.06
Applied Problems W Score	0.01	0.04	0.15	-0.05
Ortografía Standard Score	-0.09	0.05	0.09	0.03
Ortografía W Score	-0.06	0.08	0.11	0.06

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.48	0.37	0.62	0.33
Social Interaction	0.73	0.44	0.44	0.27
Social Independence	0.69	0.35	0.54	0.36
Total Positive Social Skills	0.73	0.46	0.61	0.37
Externalizing Problems	-0.26	-0.12	-0.49	-0.21
Internalizing Problems	-0.33	0.03	-0.27	-0.09
Problem Behaviors	-0.33	-0.06	-0.42	-0.17
Executive Functioning				
Pencil Tapping	0.12	0.28	0.23	-0.01
Leiter Examiner Ratings				
Attention	0.25	0.41	0.38	0.16
Activity	0.13	0.31	0.20	0.00
Sociability	0.10	0.29	0.18	0.03
Low Concentrations of ELLs				
Pre-LAS total language screener score English				
Simon Says Score	0.32	0.37	0.17	0.14
Art Show Score	0.27	0.33	0.17	0.13
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.28	0.32	0.12	0.11
Exposición de Arte Score	0.38	0.40	0.24	0.28
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	0.55	0.37	0.30	0.29
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.15	0.31	0.14	0.20
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.16	0.28	0.13	0.03
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.23	0.35	0.16	0.04
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.25	0.39	0.19	0.09
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.20	0.42	0.20	0.09
Letter-Word Identification W Score	0.25	0.50	0.24	0.16
Applied Problems Standard Score	0.31	0.48	0.36	0.19
Applied Problems W Score	0.36	0.56	0.40	0.27
Spelling Standard Score	0.29	0.48	0.31	0.24
Spelling W Score	0.35	0.56	0.36	0.32

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	-0.32	0.23	-0.31	-0.36
Letter-Word Identification W Score	-0.19	0.37	-0.25	-0.26
Applied Problems Standard Score	-0.01	-0.25	-0.11	0.08
Applied Problems W Score	0.05	-0.18	-0.09	0.13
Ortografía Standard Score	-0.24	-0.15	-0.16	-0.31
Ortografía W Score	-0.18	-0.08	-0.13	-0.27
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.37	0.31	0.53	0.24
Social Interaction	0.48	0.33	0.47	0.26
Social Independence	0.50	0.32	0.44	0.38
Total Positive Social Skills	0.49	0.35	0.54	0.32
Externalizing Problems	-0.27	-0.20	-0.56	-0.13
Internalizing Problems	-0.34	-0.21	-0.35	-0.19
Problem Behaviors	-0.33	-0.23	-0.51	-0.17
Executive Functioning				
Pencil Tapping	0.24	0.34	0.23	0.21
Leiter Examiner Ratings				
Attention	0.25	0.26	0.20	0.11
Activity	0.21	0.24	0.24	0.10
Sociability	0.28	0.27	0.20	0.12

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

Table III.21. Correlations (Pairwise) Between Change Scores on Pre-Kindergarten Observation Form (P-KOF) Subscales and Standardized Measures Spring 2010, by Concentration of ELLs

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
High Concentrations of ELLs				
Pre-LAS total language screener score English	0.07	0.29	0.14	0.10
Simon Says Score	0.04	0.16	0.12	0.05
Art Show Score	0.07	0.29	0.10	0.11
Pre-LAS total language screener score Spanish	-0.01	0.09	0.10	0.07
Tío Simón Dice Score	-0.04	0.16	0.09	-0.01
Exposición de Arte Score	0.03	-0.03	0.06	0.12
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.00	-0.03	-0.04	0.00
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.06	0.00	0.03	0.01
Expressive One Word Picture Vocabulary Test—IRT Score	0.04	-0.03	0.03	0.01
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.05	0.19	0.06	0.06
Letter-Word Identification W Score	0.03	0.15	0.04	0.03
Applied Problems Standard Score	-0.05	-0.03	0.06	0.03
Applied Problems W Score	-0.04	-0.04	0.08	0.05
Spelling Standard Score	0.05	0.12	0.18	0.13
Spelling W Score	0.04	0.11	0.17	0.12
Woodcock-Muñoz Bateria III (WM-III)				
Letter-Word Identification Standard Score	0.06	0.16	-0.09	-0.09
Letter-Word Identification W Score	0.02	0.18	-0.10	0.00
Applied Problems Standard Score	-0.04	0.08	0.12	-0.04
Applied Problems W Score	-0.03	0.08	0.13	-0.01
Ortografía Standard Score	0.12	0.15	0.12	0.12
Ortografía W Score	0.12	0.15	0.13	0.12

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.14	0.02	0.05	0.00
Social Interaction	0.33	0.16	0.17	0.21
Social Independence	0.18	0.04	0.01	0.00
Total Positive Social Skills	0.25	0.10	0.08	0.06
Externalizing Problems	-0.12	-0.06	-0.06	0.01
Internalizing Problems	-0.14	-0.01	-0.17	-0.06
Problem Behaviors	-0.11	-0.03	-0.08	-0.01
Executive Functioning				
Pencil Tapping	-0.09	0.01	-0.05	-0.09
Leiter Examiner Ratings				
Attention	0.05	-0.05	0.06	0.08
Activity	0.10	0.00	0.06	0.09
Sociability	-0.02	-0.10	-0.03	0.00
Low Concentrations of ELLs				
Pre-LAS total language screener score English				
Simon Says Score	0.10	0.12	0.15	0.08
Art Show Score	0.15	0.09	0.16	0.08
	-0.02	0.11	0.06	0.05
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.18	0.06	0.16	-0.05
Exposición de Arte Score	0.09	0.03	0.07	-0.10
	0.21	0.06	0.20	0.02
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	-0.14	-0.01	-0.09	-0.08
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	-0.07	0.01	0.02	-0.04
Expressive One Word Picture Vocabulary Test—IRT Score				
	-0.09	-0.03	-0.03	-0.05
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	-0.03	0.07	0.06	-0.02
Letter-Word Identification W Score	-0.05	0.04	0.04	-0.03
Applied Problems Standard Score	-0.09	0.09	-0.11	-0.14
Applied Problems W Score	-0.04	0.14	-0.09	-0.12
Spelling Standard Score	0.00	0.06	0.00	-0.02
Spelling W Score	0.00	0.07	0.00	-0.02

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Bateria III (WM-III)				
Letter-Word Identification Standard Score	-0.25	0.27	-0.17	0.08
Letter-Word Identification W Score	-0.25	0.42	-0.08	0.15
Applied Problems Standard Score	-0.37	-0.38	-0.25	-0.09
Applied Problems W Score	-0.37	-0.40	-0.25	-0.09
Ortografía Standard Score	-0.03	-0.14	-0.33	-0.10
Ortografía W Score	-0.03	-0.14	-0.33	-0.09
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.15	0.14	0.12	-0.02
Social Interaction	0.17	0.29	0.18	0.09
Social Independence	0.14	0.26	0.20	0.06
Total Positive Social Skills	0.17	0.24	0.17	0.03
Externalizing Problems	0.00	-0.08	-0.16	0.03
Internalizing Problems	-0.06	-0.17	-0.11	-0.04
Problem Behaviors	-0.01	-0.04	-0.09	0.00
Executive Functioning				
Pencil Tapping	0.10	0.01	-0.06	-0.05
Leiter Examiner Ratings				
Attention	0.10	-0.03	0.06	-0.02
Activity	-0.01	0.00	-0.01	-0.12
Sociability	0.07	-0.03	0.09	-0.05

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

**Table III.22. Predictive Validity: Correlations (Pairwise) Between Pre-Kindergarten Observation Form (P-KOF) Subscale Scores from Fall 2009 and Standardized Measures from Spring 2010, by Concentration of ELLs**

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
High Concentrations of ELLs				
Pre-LAS total language screener score English	0.31	0.42	0.17	0.17
Simon Says Score	0.29	0.33	0.17	0.16
Art Show Score	0.22	0.39	0.11	0.11
Pre-LAS total language screener score Spanish	0.19	0.15	0.10	0.03
Tío Simón Dice Score	0.22	0.19	0.12	0.11
Exposición de Arte Score	0.09	0.06	0.05	-0.04
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.20	0.37	0.18	0.09
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.28	0.47	0.20	0.11
Expressive One Word Picture Vocabulary Test—IRT Score	0.29	0.51	0.22	0.13
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.12	0.43	0.16	0.13
Letter-Word Identification W Score	0.17	0.53	0.21	0.17
Applied Problems Standard Score	0.19	0.40	0.24	0.14
Applied Problems W Score	0.25	0.53	0.31	0.19
Spelling Standard Score	0.06	0.35	0.13	0.13
Spelling W Score	0.13	0.47	0.20	0.18
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.11	0.15	0.18	-0.04
Letter-Word Identification W Score	0.16	0.20	0.22	-0.00
Applied Problems Standard Score	0.11	-0.01	0.12	0.13
Applied Problems W Score	0.14	0.01	0.15	0.15
Ortografía Standard Score	0.04	-0.03	0.05	0.03
Ortografía W Score	0.08	0.01	0.09	0.06

Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.30	0.33	0.51	0.30
Social Interaction	0.54	0.33	0.47	0.35
Social Independence	0.42	0.34	0.46	0.34
Total Positive Social Skills	0.49	0.38	0.55	0.38
Externalizing Problems	-0.16	-0.21	-0.40	-0.14
Internalizing Problems	-0.14	-0.10	-0.20	-0.09
Problem Behaviors	-0.18	-0.18	-0.34	-0.13
Executive Functioning				
Pencil Tapping	0.22	0.36	0.23	0.13
Leiter Examiner Ratings				
Attention	0.25	0.38	0.29	0.14
Activity	0.13	0.31	0.18	0.10
Sociability	0.21	0.30	0.24	0.11
Low Concentrations of ELLs				
Pre-LAS total language screener score English				
Simon Says Score	0.27	0.42	0.23	0.15
Art Show Score	0.23	0.38	0.21	0.12
Pre-LAS total language screener score Spanish				
Tío Simón Dice Score	0.22	0.36	0.21	0.16
Exposición de Arte Score	0.16	0.40	0.14	0.12
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score				
	0.20	0.32	0.23	0.12
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.07	0.33	0.01	0.09
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.23	0.38	0.20	0.15
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score				
	0.24	0.43	0.20	0.17
Expressive One Word Picture Vocabulary Test—IRT Score				
	0.28	0.47	0.24	0.18
Woodcock-Johnson III (WJ-III)				
Letter-Word Identification Standard Score	0.13	0.46	0.16	0.07
Letter-Word Identification W Score	0.19	0.53	0.20	0.12
Applied Problems Standard Score	0.29	0.45	0.35	0.21
Applied Problems W Score	0.35	0.50	0.39	0.26
Spelling Standard Score	0.20	0.44	0.29	0.20
Spelling W Score	0.26	0.52	0.34	0.26



Standardized Measures	P-KOF			
	Social Expression	General Knowledge	Self-Regulation	Self-Care & Motor Skills
Woodcock-Muñoz Batería III (WM-III)				
Letter-Word Identification Standard Score	0.05	0.24	-0.17	0.08
Letter-Word Identification W Score	0.09	0.31	-0.15	-0.01
Applied Problems Standard Score	0.28	-0.02	0.17	0.32
Applied Problems W Score	0.31	0.01	0.18	0.28
Ortografía Standard Score	-0.11	-0.05	0.01	0.07
Ortografía W Score	-0.09	-0.02	0.03	0.01
Preschool Kindergarten Behavior Scales-2 (PKBS-2)				
Social Cooperation	0.35	0.24	0.49	0.31
Social Interaction	0.36	0.24	0.36	0.18
Social Independence	0.42	0.28	0.37	0.27
Total Positive Social Skills	0.41	0.28	0.46	0.28
Externalizing Problems	-0.11	-0.01	-0.31	-0.05
Internalizing Problems	-0.18	-0.05	-0.14	-0.10
Problem Behaviors	-0.16	-0.07	-0.25	-0.11
Executive Functioning				
Pencil Tapping	0.25	0.40	0.32	0.22
Leiter Examiner Ratings				
Attention	0.23	0.31	0.27	0.16
Activity	0.19	0.26	0.25	0.20
Sociability	0.19	0.28	0.18	0.16

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Teacher-Administered Measures.

**This page has been left blank for double-sided copying.**

**APPENDIX D**

**TABLES OF RESULTS PRESENTED IN CHAPTER IV:  
PROVIDER ADMINISTERED ASSESSMENT**

**This page has been left blank for double-sided copying.**

**Table IV.1 Number of Children with Provider Administered Assessment (PAA) Data Fall 2009 and Spring 2010, by Program Type and Language Group**

<i>Language group</i>	Fall 2009			Spring 2010		
	Center	FCC	Total	Center	FCC	Total
<b>ASQ-3 Communication, English</b>						
English only	97	35	132	82	36	118
English Primarily	63	15	78	52	14	66
Spanish only	22	8	30	22	3	25
Span primarily	76	19	95	63	18	81
Other Language only or Primarily	7	1	8	4	1	5
Missing	1		1	1		1
Total	266	78	344	224	72	296
<b>ASQ-3 Communication, Spanish</b>						
English only	2	1	3	2		2
English Primarily	16	6	22	13	5	18
Spanish only	25	10	35	14	7	21
Span primarily	59	19	78	51	17	68
Other Language only or Primarily	0	0	0	0	0	0
Missing	0	0	0	0	0	0
Total	102	36	138	80	29	109
<b>ASQ-3 Problem Solving</b>						
English only	1	125	43	168	124	165
English Primarily	2	85	29	114	86	108
Spanish only	46	10	56	40	10	50
Span primarily	93	17	110	89	16	105
Other Language only or Primarily	15	3	18	13	3	16
Missing	0	0	0	0	0	0
Total	364	102	466	352	92	444
<b>ASQ-3 Gross Motor</b>						
English only	41	10	51	31	6	37
English Primarily	29	12	41	29	5	34
Spanish only	13	2	15	13	1	14
Span primarily	29	4	33	33	2	35
Other Language only or Primarily	6		6	6		6
Missing	0	0	0	0	0	0
Total	118	28	146	112	14	126
<b>ASQ-3 Fine Motor</b>						
English only	42	10	52	33	6	39
English Primarily	29	14	43	30	5	35
Spanish only	13	2	15	13	1	14
Span primarily	29	4	33	33	2	35
Other Language only or Primarily	6		6	6		6
Missing	0	0	0	0	0	0
Total	119	30	149	115	14	129
<b>Rapid Letter Naming</b>						
English only	94	36	130	92	37	129
English Primarily	68	17	85	58	18	76
Spanish only	36	9	45	27	9	36
Span primarily	72	14	86	56	14	70
Other Language only or Primarily	10	3	13	7	3	10
Missing	2					
Total	282	79	361	240	81	321

**Table IV.2. Means, Standard Deviations, and Cronbach's Alphas for Provider Administered Assessment (PAA) Measures Fall 2009 – Spring 2010**

Outcome	N	Fall 2009		Spring 2010		Cronbach's Alpha	Fall-Spring Change		Effect Size
		Mean	Standard Deviation	Mean	Standard Deviation		Mean	Standard Deviation	
ASQ-3 Subscale Total Score <sup>a</sup>									
Communication, English	245	54.07***	25.46	69.99	21.67	0.85	15.91	21.27	0.62
Communication, Spanish	94	55.61***	25.69	69.69	23.36	0.90	14.07	21.76	0.55
Problem Solving	369	61.38***	22.10	86.86	19.34	0.72	25.49	16.45	1.15
Gross Motor	93	69.61***	19.39	81.95	12.39	0.73	12.34	17.70	0.64
Fine Motor	97	60.14***	24.32	74.57	19.22	0.79	14.43	17.58	0.59
Rapid Letter Naming <sup>b</sup>	284	12.79***	10.81	19.37	9.52	0.96	6.58	8.67	0.61

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-tests to compare fall and spring scores.

<sup>a</sup>Scored based on all items in the subscale.

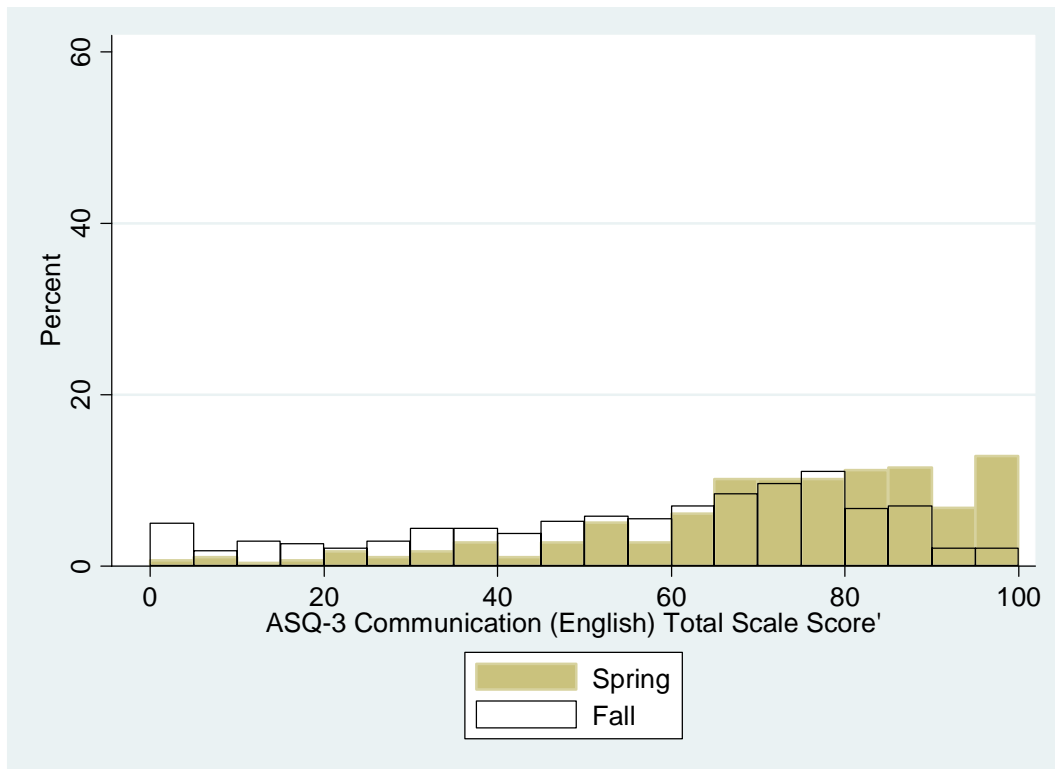
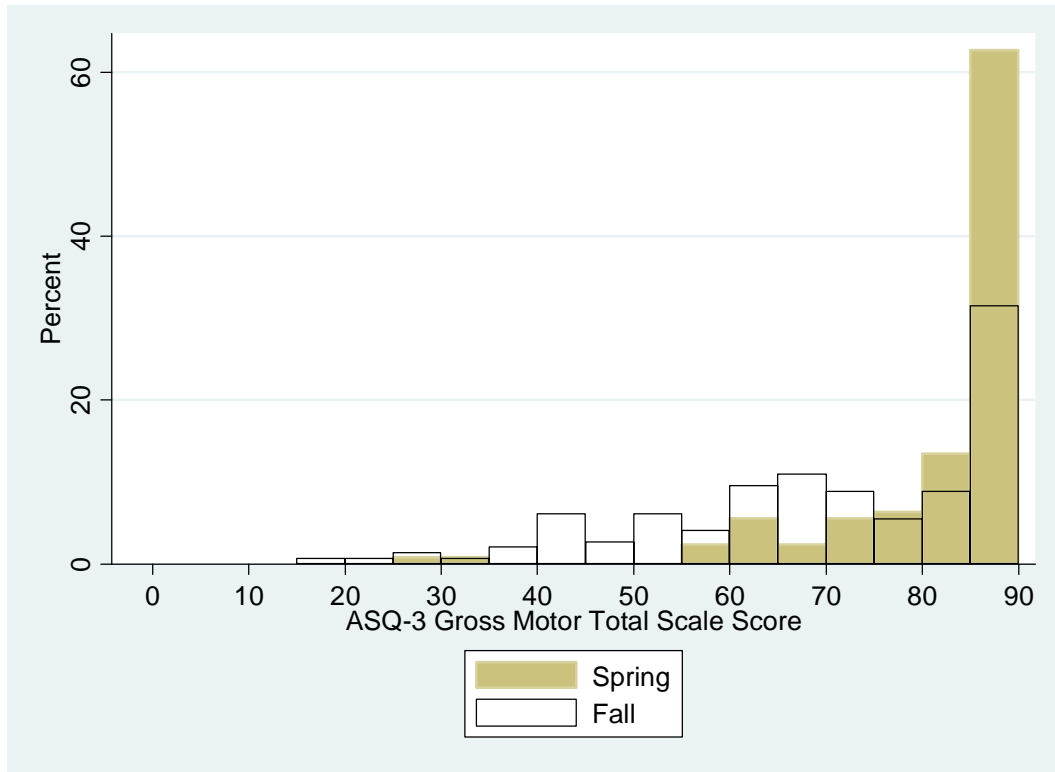
<sup>b</sup>This measure was conceptually scored.

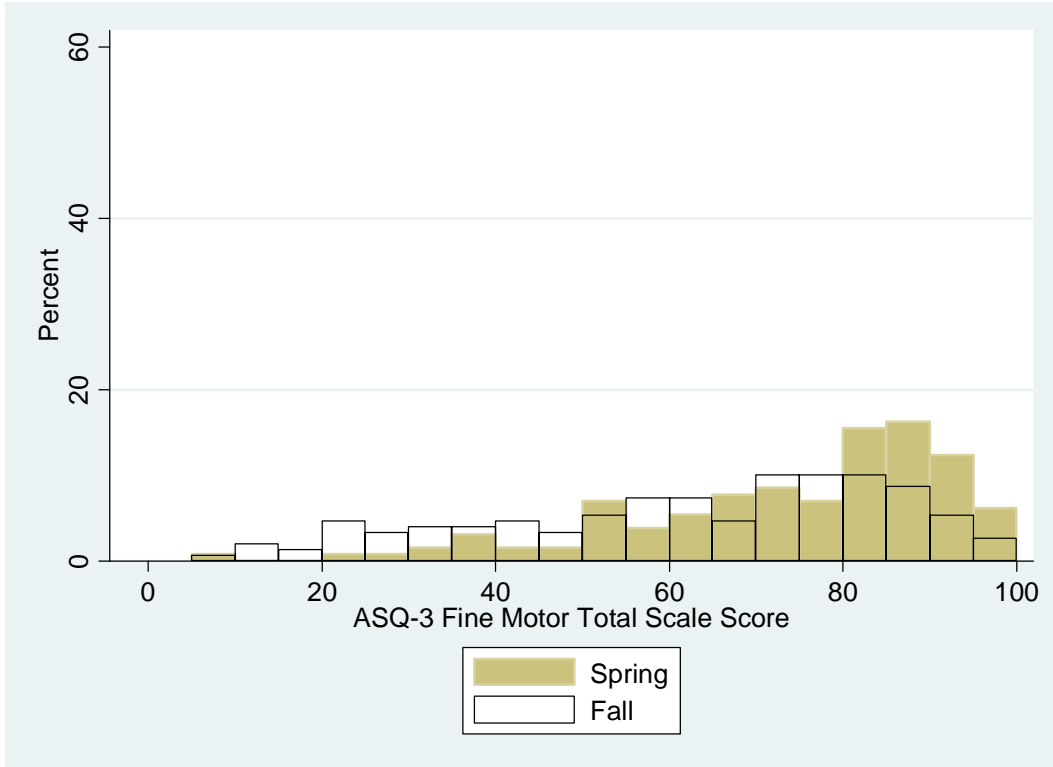
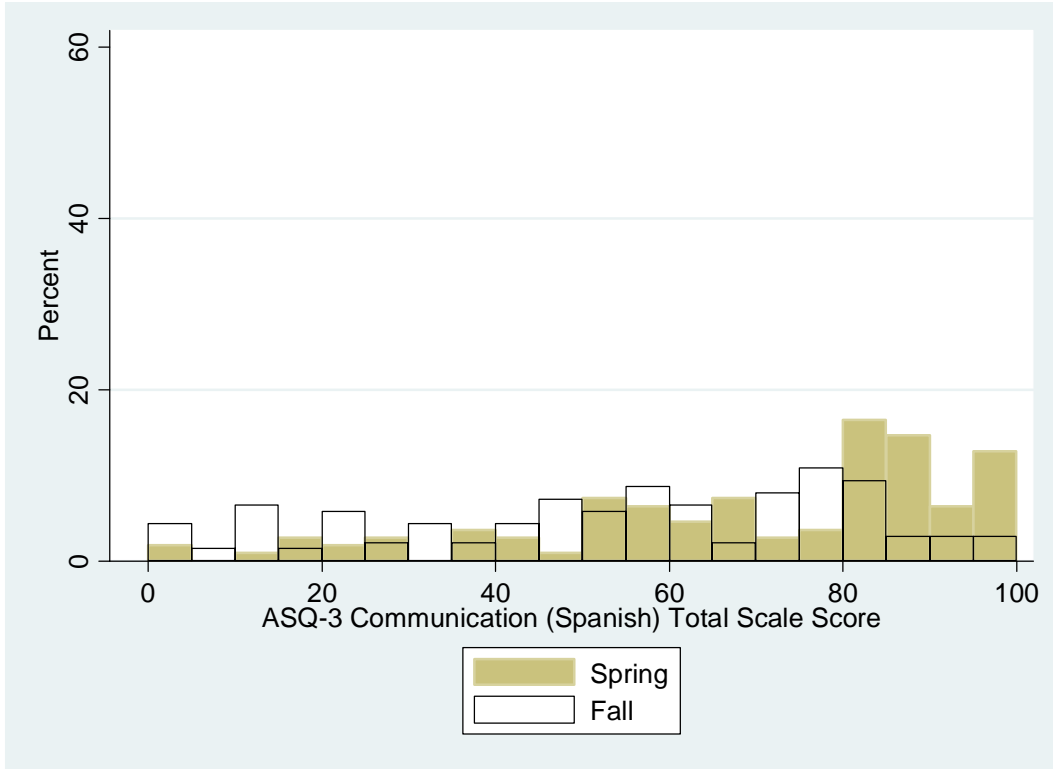
\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

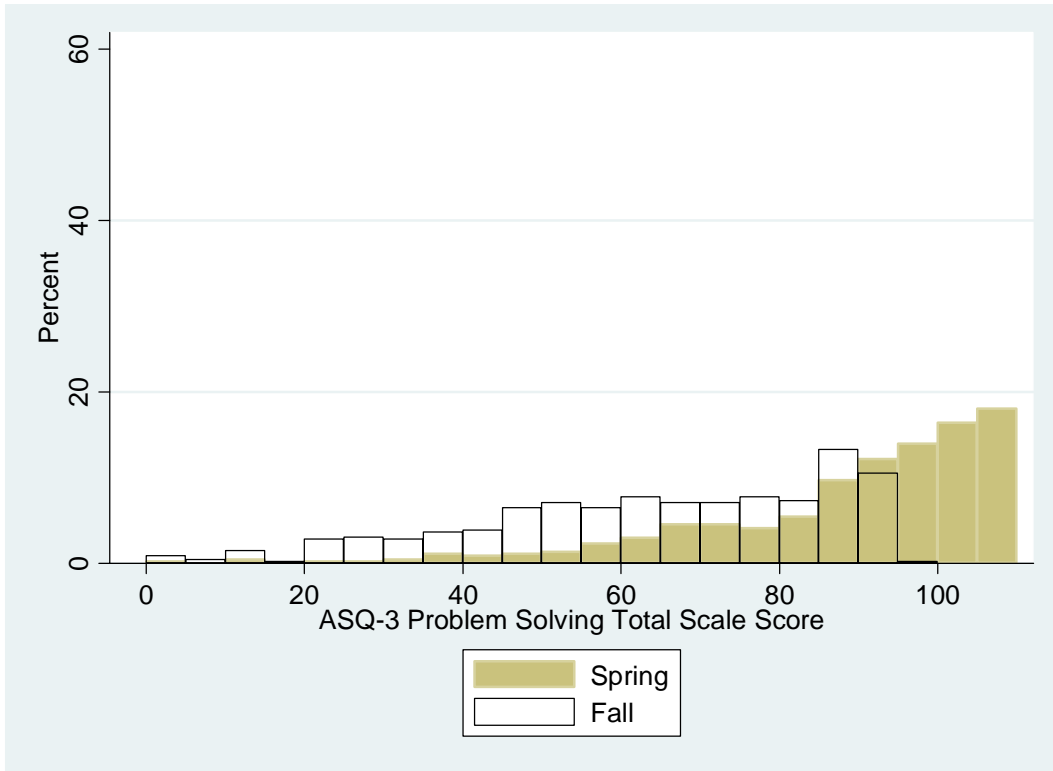
\*\*\*Significantly different from zero at the .01 level, two-tailed test.

Figure IV.1 Distributions of Fall and Spring ASQ-3 Total Scale Scores









**Table IV.3. Means/Percentages, Standard Deviations, and Cronbach's Alphas for ASQ-3 Screening Scores Spring 2010**

Measure	N	Mean	Standard Deviation	Number of Items	Cronbach's Alpha
ASQ-3 Screening Score <sup>a</sup>					
Communication, English	305	39.78	14.19	6	0.72-0.85
Communication, Spanish	111	38.53	15.24	6	0.79-0.85
Problem Solving	448	51.95	10.81	6	0.70-0.71
Gross Motor	126	54.56	8.44	6	0.63-0.70
Fine Motor	129	44.03	10.53	6	0.32-0.53

Source: Spring 2009 Provider-Administered Assessment (PAA).

Note: ASQ-3 = Ages & Stages Questionnaires (Third Edition).

<sup>a</sup>Scored based on each child's age using the appropriate items.

Table IV.4. Percentages for ASQ-3 Cutoff Scores fall 2009 - Spring 2010

Measure	N	Fall	Spring	Fall - Spring Change
		Percentage	Percentage	
ASQ in the At-Risk Range (2SDs below the mean or lower)				
Communication, English	254	39.37***	25.20	-14.20
Communication, Spanish	114	41.23	38.60	-2.63
Problem Solving	375	10.40***	4.80	-5.60
Gross Motor	93	13.98**	5.38	-8.60
Fine Motor	97	11.34*	4.12	-7.22
ASQ in the Monitoring Zone (1 to 2SDs below the mean)				
Communication, English	254	21.65**	31.10	9.45
Communication, Spanish	96	23.96	28.13	4.17
Problem Solving	375	16.27***	9.87	-6.40
Gross Motor	93	15.05***	3.23	-11.8
Fine Motor	97	24.74	17.53	-7.22

Source: Fall 2009 Provider-Administered Assessment (PAA).

Note: ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted chi-square tests to compare fall to spring change.

\*Scored based on each child's age using the appropriate items.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.5 Means and Standard Deviations for Provider Administered Assessment (PAA) Measures Fall 2009 and Spring 2010, by Language Group**

	English Only		English Primarily		Spanish Only		Spanish Primarily	
	Fall 2009 Mean (SD <sup>a</sup> )	Spring 2010 Mean (SD)	Fall 2009 Mean (SD <sup>a</sup> )	Spring 2010 Mean (SD)	Fall 2009 Mean (SD)	Spring 2010 Mean (SD)	Fall 2009 Mean (SD)	Spring 2010 Mean (SD)
Outcome								
ASQ-3 Subscale Total Score <sup>b</sup>								
Communication, English	66.34*** (17.22)	78.27 (14.21)	58.33*** (22.44)	71.94 (18.36)	28.33*** (26.81)	58.04 (29.08)	41.53*** (26.48)	61.62 (25.60)
Communication, Spanish			56.67 (31.09)	64.70 (25.02)	47.96** (24.23)	65.65 (24.25)	57.61*** (24.68)	72.11 (22.71)
Problem Solving	72.31*** (16.43)	93.19 (14.29)	61.00*** (21.19)	86.77 (19.28)	47.31*** (21.46)	75.96 (20.09)	49.78*** (21.58)	81.04 (21.93)
Gross Motor	67.18*** (15.52)	79.93 (11.63)	69.45** (21.76)	80.53 (14.63)	73.30 (21.61)	83.47 (9.43)	71.15** (20.90)	82.96 (13.12)
Fine Motor	61.65** (22.10)	76.79 (20.24)	58.70* (26.58)	71.12 (20.24)	54.18** (23.84)	72.84 (10.69)	59.16** (26.32)	73.46 (20.20)
Rapid Letter Naming <sup>c</sup>	16.50*** (10.31)	21.47 (8.78)	12.89*** (10.71)	20.65 (8.78)	6.44*** (8.07)	13.97 (9.42)	8.35*** (9.49)	16.68 (9.90)

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment.

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-tests to compare fall and spring scores within each subgroup.

<sup>a</sup>SD = Standard deviation.

<sup>b</sup> Scored based on all items in the subscale.

<sup>c</sup> This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.6. Change Scores for Provider Administered Assessment (PAA) Measures Fall 2009 – Spring 2010, by Language Group**

Outcome	English Only			English Primarily			Spanish Only			Spanish Primarily		
	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size	N	Mean (SD)	Effect Size
ASQ-3 Subscale Total Score <sup>a</sup>												
Communication, English	94	11.93*** (17.28)	0.47	54	13.61 (21.64)	0.53	19	29.71 (30.46)	1.17	72	20.09 (20.93)	0.79
Communication, Spanish				15	8.04 (27.67)	0.31	18	17.69 (25.41)	0.69	61	14.49 (18.96)	0.56
Problem Solving	130	20.87*** (13.71)	0.94	86	25.76 (15.65)	1.17	46	28.65 (18.04)	1.30	91	31.26 (18.79)	1.41
Gross Motor	25	12.75 (13.29)	0.66	25	11.08 (22.55)	0.57	11	10.17 (17.15)	0.52	26	11.80 (17.26)	0.61
Fine Motor	27	15.10 (17.92)	0.62	27	12.42 (16.10)	0.51	11	18.66 (19.83)	0.77	26	14.29 (19.40)	0.59
Rapid Letter Naming <sup>b</sup>	104	4.97*** (8.79)	0.46	65	7.75 (8.78)	0.72	36	7.53 (7.40)	0.70	69	8.33 (8.89)	0.77

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment.

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted ANOVA to compare changes across language groups.

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.7. Internal Consistency Reliability of Provider Administered Assessment (PAA) Measures Spring 2010, by Language Group**

Outcome	English Only		English Primarily		Spanish Only		Spanish Primarily	
	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha	N	Cronbach's Alpha
ASQ-3 Subscale Total Score <sup>a</sup>								
Communication, English	118	0.72	66	0.81	25	0.90	81	0.88
Communication, Spanish			18	0.92	21	0.89	68	0.87
Problem Solving	165	0.66	108	0.81	50	0.69	105	0.77
Gross Motor	37	0.75	34	0.73	14	0.25	35	0.77
Fine Motor	39	0.72	35	0.80	14	0.70	35	0.79
Rapid Letter Naming <sup>b</sup>	129	0.96	76	0.96	36	0.95	70	0.96

Source: Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition).

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

**Table IV.8. Means and Standard Deviations for Provider Administered Assessment (PAA) Measures Fall 2009 and Spring 2010, by Program Type**

Outcome	Center				Family Child Care			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
ASQ-3 Subscale Total Score <sup>a</sup>								
Communication, English	52.16***	25.19	69.90	21.79	60.68**	25.52	70.27	21.48
Communication, Spanish	54.52***	25.15	68.00	23.55	58.33**	27.28	73.89	22.76
Problem Solving	60.07***	22.60	85.36	20.38	66.33***	19.46	92.55	13.43
Gross Motor	70.65***	19.27	83.19	12.01	63.22*	19.68	74.33	12.41
Fine Motor	60.11***	23.71	74.38	19.51	60.29	28.67	75.71	18.07
Rapid Letter Naming <sup>b</sup>	12.19***	10.86	18.44	9.60	14.52***	10.54	22.07	8.78

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-tests to compare fall and spring scores within subgroups.

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.9. Change Scores for Provider Administered Assessment (PAA) Measures Fall 2009 – Spring 2010, by Program Type**

Outcome	N	Center			Family Child Care			Effect Size
		Mean	Standard Deviation	Effect Size	N	Mean	Standard Deviation	
ASQ-3 Subscale Total Score <sup>a</sup>								
Communication, English	190	17.74**	20.98	0.70	55	9.60	21.25	0.38
Communication, Spanish	67	13.48	19.01	0.52	27	15.56	27.82	0.61
Problem Solving	292	25.29	16.60	1.14	77	26.22	15.96	1.19
Gross Motor	80	12.54	18.35	0.65	13	11.11	13.58	0.57
Fine Motor	83	14.27	17.09	0.59	14	15.42	20.97	0.63
Rapid Letter Naming <sup>b</sup>	211	6.25	8.65	0.58	73	7.55	8.70	0.70

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-tests to compare change scores between centers and FCCs.

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.



**Table IV.10. Internal Consistency Reliability of Provider Administered Assessment (PAA) Measures Spring 2010, by Program Type**

Outcome	Center		Family Child Care	
	N	Cronbach's Alpha	N	Cronbach's Alpha
ASQ-3 Subscale Total Score <sup>a</sup>				
Communication, English	224	0.85	72	0.84
Communication, Spanish	80	0.89	29	0.90
Problem Solving	352	0.73	92	0.65
Gross Motor	112	0.73	14	0.73
Fine Motor	115	0.80	14	0.66
Rapid Letter Naming <sup>b</sup>	240	0.96	81	0.96

Source: Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition).

<sup>a</sup> Scored based on all items in the subscale.

<sup>b</sup> This measure was conceptually scored.

**Table IV.11. Means and Standard Deviations for Provider Administered Assessment (PAA) Measures Fall 2009 and Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	Fall 2009		Spring 2010		Fall 2009		Spring 2010	
	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation	Mean/ Percentage	Standard Deviation
ASQ-3 Subscale Total Score <sup>a</sup>								
Communication, English	46.11***	27.34	65.87	23.62	62.37***	20.35	74.27	18.59
Communication, Spanish	57.47***	25.60	70.88	23.91	45.00**	24.42	62.90	19.27
Problem Solving	55.49***	22.26	83.17	19.74	66.52***	20.69	90.08	18.45
Gross Motor	72.56***	20.20	86.58	7.37	67.91***	18.88	79.28	13.88
Fine Motor	61.81***	24.56	75.73	16.65	59.20***	24.33	73.92	20.64
Rapid Letter Naming <sup>b</sup>	9.63***	10.30	17.17	10.11	16.33***	10.29	21.83	8.16

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment.

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-test to compare fall and spring scores within subgroups.

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.12. Change Scores for Provider Administered Assessment (PAA) Measures Fall 2009 – Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs				Low Concentration of ELLs			
	N	Mean	Standard Deviation	Effect Size	N	Mean/ Percentage	Standard Deviation	Effect Size
ASQ-3 Subscale Total Score <sup>a</sup>								
Communication, English	125	19.76***	23.18	0.78	120	11.91	18.34	0.47
Communication, Spanish	80	13.40	22.49	0.52	14	17.90	17.22	0.70
Problem Solving	172	27.69**	18.22	1.25	197	23.56	14.51	1.07
Gross Motor	34	14.03	17.60	0.72	59	11.37	17.84	0.59
Fine Motor	35	13.92	17.58	0.57	62	14.73	17.72	0.61
Rapid Letter Naming <sup>b</sup>	150	7.55**	8.70	0.70	134	5.50	8.54	0.51

Source: Fall 2009 and Spring 2010 Provider-Administered Assessment.

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). We conducted t-tests to compare change scores between high and low ELL programs.

<sup>a</sup>Scored based on all items in the subscale.

<sup>b</sup>This measure was conceptually scored.

\*Significantly different from zero at the .10 level, two-tailed test.

\*\*Significantly different from zero at the .05 level, two-tailed test.

\*\*\*Significantly different from zero at the .01 level, two-tailed test.

**Table IV.13. Internal Consistency Reliability of Provider Administered Assessment (PAA) Measures Spring 2010, by Concentration of ELLs**

Outcome	High Concentration of ELLs		Low Concentration of ELLs	
	N	Cronbach's Alpha	N	Cronbach's Alpha
ASQ-3 Subscale Total Score <sup>a</sup>				
Communication, English	142	0.86	154	0.82
Communication, Spanish	87	0.90	22	0.87
Problem Solving	190	0.74	254	0.70
Gross Motor	41	0.69	85	0.74
Fine Motor	41	0.77	88	0.80
Rapid Letter Naming <sup>b</sup>	150	0.96	171	0.96

Source: Spring 2010 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition).

<sup>a</sup> Scored based on all items in the subscale.

<sup>b</sup> This measure was conceptually scored.

**Table IV.14. Inter-factor Correlations for ASQ-3 Subscale Total Scores Spring 2010**

Outcome	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor
ASQ-3 Subscale Total Score					
Communication, English					
Communication, Spanish	0.56				
Problem Solving	0.62	0.48			
Gross Motor			0.34		
Fine Motor			0.58	0.54	

Source: Spring 2010 Provider-Administered Assessment.

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). Some of the inter-factor correlations could not be calculated because children did not have ratings on all subscales as a result of the study design.

**Table IV.15. Inter-Factor Correlations for ASQ-3 Subscale Total Scores Spring 2010, by Program Type**

Measure	Communication, English	Communication, Spanish	Problem Solving	Gross Motor
Center				
ASQ-3 Subscale Total Score				
Communication, English				
Communication, Spanish	0.61			
Problem Solving	0.62	0.54		
Gross Motor			0.40	
Fine Motor			0.60	0.61
Family Child Care				
ASQ-3 Subscale Total Score				
Communication, English				
Communication, Spanish	0.39			
Problem Solving	0.57	0.30		
Gross Motor			0.03	
Fine Motor			0.33	0.10

Source: Fall 2009 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). Some of the inter-factor correlations could not be calculated because children did not have ratings on all subscales as a result of the study design.

**Table IV.16. Inter-Factor Correlations for ASQ-3 Subscale Total Scores Spring 2010, by Concentration of ELLs**

Measure	Communication, English	Communication, Spanish	Problem Solving	Gross Motor
High Concentrations of ELLs				
ASQ-3 Subscale Total Score				
Communication, English	0.71			
Communication, Spanish		0.71		
Problem Solving	0.62		0.31	
Gross Motor			0.36	0.76
Fine Motor				
Low Concentrations of ELLs				
ASQ-3 Subscale Total Score				
Communication, English	-0.03			
Communication, Spanish		-0.23		
Problem Solving	0.52		.	
Gross Motor			0.39	
Fine Motor			0.66	0.50

Source: Fall 2009 Provider-Administered Assessment (PAA).

Note: Reported by teachers. ASQ-3 = Ages & Stages Questionnaires (Third Edition). Some of the inter-factor correlations could not be calculated because children did not have ratings on all subscales as a result of the study design.

Table IV.17. Concurrent Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) and Standardized Measures Spring 2010

Standardized Measures	ASQ-3					RLN
	Communication English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	
Pre-LAS total language screener score English	0.53	-0.03	0.53	0.24	0.35	0.39
Simon Says Score	0.38	-0.05	0.40	0.27	0.32	0.26
(N)	272	100	416	121	124	298
Art Show Score	0.52	-0.06	0.48	0.16	0.28	0.38
(N)	271	99	414	120	123	297
Pre-LAS total language screener score Spanish	0.00	0.25	0.17	0.10	0.35	0.07
Tío Simón Dice Score	0.08	0.15	0.14	0.26	0.28	0.02
(N)	99	82	146	47	47	99
Exposición de Arte Score	-0.06	0.24	0.13	-0.05	0.27	0.09
(N)	99	82	146	47	47	99
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.37	0.12	0.37	0.14	0.14	0.36
(N)	220	53	329	96	99	236
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.46	0.18	0.51	0.12	0.24	0.46
(N)	271	99	413	119	122	297
Expressive One Word Picture Vocabulary Test—IRT Score	0.45	0.19	0.51	0.17	0.29	0.44
(N)	271	99	413	119	122	297
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.16	-0.22	0.39	0.24	0.31	0.59
(N)	220	53	329	96	99	236
Letter-Word Identification W Score	0.18	-0.16	0.46	0.34	0.41	0.62



Standardized Measures	ASQ-3					RLN
	Communication English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	
(N)	220	53	329	96	99	236
Applied Problems Standard Score	0.31	0.13	0.42	0.36	0.41	0.36
(N)	220	53	329	96	99	236
Applied Problems W Score	0.34	0.25	0.49	0.47	0.51	0.37
(N)	220	53	329	96	99	236
Spelling Standard Score	0.28	0.03	0.33	0.23	0.38	0.47
(N)	220	53	329	96	99	236
Spelling W Score	0.30	0.11	0.42	0.32	0.48	0.49
(N)	220	53	329	96	99	236
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.16	-0.12	0.14	0.12	0.24	0.28
(N)	51	46	83	22	22	61
Letter-Word Identification W Score	0.18	-0.10	0.17	0.06	0.25	0.29
(N)	51	46	83	22	22	61
Applied Problems Standard Score	-0.07	0.17	0.02	0.01	-0.13	0.05
(N)	51	46	83	22	22	61
Applied Problems W Score	-0.06	0.18	0.03	-0.04	-0.13	0.07
(N)	51	46	83	22	22	61
Ortografía Standard Score	0.26	0.17	0.09	-0.07	-0.24	0.16
(N)	51	46	83	22	22	61
Ortografía W Score	0.28	0.18	0.12	-0.12	-0.26	0.17
(N)	51	46	83	22	22	61
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.05	0.28	0.39	.	.	0.21
(N)	141	51	148	.	.	152
Social Interaction	0.27	0.56	0.39	.	.	0.22
(N)	141	51	148	.	.	152
Social Independence	0.18	0.44	0.39	.	.	0.20
(N)	141	51	149	.	.	153
Total Positive Social Skills	0.19	0.48	0.45	.	.	0.24
(N)	141	51	148	.	.	152
Externalizing Problems	-0.01	-0.24	-0.09	.	.	-0.07

Standardized Measures	ASQ-3					RLN
	Communication English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	
(N)	133	50	140	.	.	144
Internalizing Problems	0.03	-0.32	-0.04	.	.	-0.05
(N)	137	50	145	.	.	149
Problem Behaviors	0.02	-0.32	-0.10	.	.	-0.08
(N)	139	51	147	.	.	151
Executive Functioning						
Pencil Tapping	0.36	0.23	0.40	0.43	0.44	0.24
(N)	271	99	413	119	122	297
Leiter Examiner Ratings						
Attention	0.30	0.19	0.41	0.23	0.35	0.33
(N)	271	99	414	120	123	297
Activity	0.27	0.23	0.25	0.25	0.33	0.14
(N)	271	99	414	120	123	297
Sociability	0.30	0.07	0.34	0.21	0.36	0.28
(N)	271	99	414	120	123	297

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.18. Correlations (Pairwise) Between Change Scores on Provider Administered Assessment (PAA) and Standardized Measures Spring 2010

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Pre-LAS total language screener score English	0.19	0.19	0.15	0.14	0.18	0.09
Simon Says Score	0.07	0.15	0.07	0.14	0.15	0.01
(N)	225	85	346	88	92	264
Art Show Score	0.26	0.15	0.19	0.10	0.14	0.15
(N)	225	85	345	87	91	264
Pre-LAS total language screener score Spanish	-0.09	0.01	-0.08	-0.09	0.09	-0.23
Tío Simón Dice Score	-0.03	-0.01	-0.05	-0.02	0.06	-0.11
(N)	85	72	128	35	35	97
Exposición de Arte Score	-0.09	0.03	-0.08	-0.13	0.07	-0.22
(N)	85	72	128	35	35	97
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.08	0.23	-0.02	-0.37	-0.08	-0.08
(N)	148	23	222	60	63	164
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.10	0.29	0.15	-0.27	-0.09	0.02
(N)	224	85	342	85	89	263
Expressive One Word Picture Vocabulary Test—IRT Score	0.06	0.36	0.06	-0.18	-0.01	-0.08
(N)	224	85	343	86	90	263
Woodcock-Johnson III (WJ-III) Letter-Word Identification Standard	0.06	0.23	0.18	-0.02	-0.02	0.35

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Score						
(N)	148	23	221	59	62	164
Letter-Word Identification W Score	0.04	0.21	0.17	-0.01	-0.03	0.33
(N)	148	23	221	59	62	164
Applied Problems Standard Score	-0.01	0.17	0.05	0.07	0.19	0.12
(N)	148	23	221	59	62	164
Applied Problems W Score	-0.03	0.19	0.05	0.13	0.23	0.12
(N)	148	23	221	59	62	164
Spelling Standard Score	0.03	0.00	0.12	-0.10	0.15	0.26
(N)	148	23	221	59	62	164
Spelling W Score	0.00	0.02	0.12	-0.08	0.17	0.26
(N)	148	23	221	59	62	164
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard						
Score	0.46	0.33	0.07	0.51	-0.02	0.10
(N)	35	32	50	10	11	40
Letter-Word Identification W Score	0.44	0.30	0.03	0.61	0.22	0.04
(N)	38	36	65	12	13	53
Applied Problems Standard Score	0.11	0.06	0.12	-0.05	0.05	-0.17
(N)	38	36	65	12	13	53
Applied Problems W Score	0.10	0.05	0.11	-0.05	0.05	-0.17
(N)	38	36	65	12	13	53
Ortografía Standard Score	0.27	0.24	0.11	0.22	0.08	0.04
(N)	38	36	65	12	13	53
Ortografía W Score	0.27	0.25	0.10	0.27	0.08	0.03
(N)	38	36	65	12	13	53
Preschool Kindergarten Behavior Scales- 2 (PKBS-2)						
Social Cooperation	0.15	0.03	0.27	.	.	0.05
(N)	118	45	114	.	.	132
Social Interaction	0.32	0.29	0.32	.	.	0.15
(N)	118	45	113	.	.	131
Social Independence	0.32	0.05	0.19	.	.	0.07
(N)	118	45	115	.	.	133
Total Positive Social Skills	0.33	0.13	0.26	.	.	0.08
(N)	118	45	113	.	.	131
Externalizing Problems	0.03	0.11	-0.19	.	.	0.07
(N)	111	44	111	.	.	125
Internalizing Problems	0.01	-0.03	-0.17	.	.	0.06
(N)	114	44	112	.	.	129

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Problem Behaviors (N)	0.08 116	0.01 45	-0.19 114	. .	. .	0.09 131
Executive Functioning						
Pencil Tapping (N)	0.03 224	0.14 85	-0.04 340	0.03 83	-0.03 87	0.07 263
Leiter Examiner Ratings						
Attention (N)	-0.02 224	0.04 85	0.08 342	0.08 84	0.03 88	0.17 264
Activity (N)	0.00 224	0.08 85	0.06 342	-0.15 84	0.05 88	0.13 264
Sociability (N)	-0.04 224	-0.05 85	0.13 342	-0.10 84	0.09 88	0.14 264

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.19. Predictive Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) Measures from Fall 2009 and Standardized Measures from Spring 2010

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Pre-LAS total language screener score English	0.66	0.16	0.62	0.37	0.38	0.42
Simon Says Score (N)	0.52 304	0.15 121	0.48 416	0.37 128	0.27 131	0.33 317
Art Show Score (N)	0.63 303	0.11 120	0.58 414	0.25 127	0.36 130	0.39 316
Pre-LAS total language screener score Spanish	0.06	0.30	0.22	0.34	0.39	0.17
Tío Simón Dice Score (N)	0.22 113	0.30 99	0.26 147	0.51 43	0.27 43	0.15 114
Exposición de Arte Score (N)	-0.11 113	0.17 99	0.10 147	0.04 43	0.33 43	0.12 114
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score (N)	0.43 244	0.08 64	0.50 328	0.25 103	0.23 105	0.43 246
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score (N)	0.54 302	0.19 120	0.62 413	0.24 126	0.36 129	0.47 316
Expressive One Word Picture Vocabulary Test—IRT Score (N)	0.56 302	0.23 120	0.65 413	0.26 126	0.39 129	0.51 316
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score (N)	0.22 244	-0.15 64	0.45 328	0.22 103	0.36 105	0.59 246
Letter-Word Identification W Score (N)	0.28 244	-0.08 64	0.53 328	0.28 103	0.47 105	0.65 246
Applied Problems Standard Score (N)	0.39 244	-0.04 64	0.52 328	0.24 103	0.40 105	0.37 246
Applied Problems W Score (N)	0.48 244	0.08 64	0.59 328	0.27 103	0.48 105	0.40 246
Spelling Standard Score (N)	0.30 244	-0.10 64	0.41 328	0.34 103	0.53 105	0.46 246
Spelling W Score (N)	0.38 244	0.04 64	0.50 328	0.39 103	0.63 105	0.52 246
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score (N)	-0.05 58	-0.09 56	0.15 84	0.13 22	0.45 23	0.12 70
Letter-Word Identification W Score (N)	-0.08 58	-0.06 56	0.22 84	0.17 22	0.50 23	0.14 70
Applied Problems Standard Score (N)	-0.18 58	0.26 56	-0.11 84	-0.23 22	-0.29 23	-0.04 70
Applied Problems W Score	-0.19	0.28	-0.06	-0.23	-0.29	-0.02

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	58	56	84	22	23	70
Ortografía Standard Score	-0.06	0.04	-0.13	-0.41	-0.33	0.05
(N)	58	56	84	22	23	70
Ortografía W Score	-0.08	0.07	-0.08	-0.43	-0.33	0.07
(N)	58	56	84	22	23	70
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.11	0.34	0.18	0.49	0.33	0.13
(N)	123	59	130	10	10	134
Social Interaction	0.34	0.66	0.17	0.49	0.05	0.14
(N)	123	59	130	10	10	134
Social Independence	0.27	0.57	0.19	0.19	-0.14	0.15
(N)	123	59	131	10	10	135
Total Positive Social Skills	0.26	0.59	0.20	0.62	0.19	0.16
(N)	123	59	130	10	10	134
Externalizing Problems	-0.05	-0.16	0.10	-0.33	-0.26	0.02
(N)	117	57	126	10	10	126
Internalizing Problems	-0.10	-0.19	0.15	-0.06	0.46	0.07
(N)	120	58	127	10	10	130
Problem Behaviors	-0.09	-0.20	0.11	-0.25	0.04	0.02
(N)	122	59	129	10	10	132
Executive Functioning						
Pencil Tapping	0.31	0.14	0.49	0.31	0.49	0.23
(N)	302	120	413	126	129	316
Leiter Examiner Ratings						
Attention	0.26	0.14	0.41	0.28	0.37	0.23
(N)	303	120	414	127	130	316
Activity	0.21	0.20	0.29	0.36	0.38	0.11
(N)	303	120	414	127	130	316
Sociability	0.27	0.12	0.36	0.34	0.40	0.22
(N)	303	120	414	127	130	316

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.20. Concurrent Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) and Standardized Measures Spring 2010, by Program Type

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Center						
Pre-LAS total language screener score English	0.50	-0.04	0.53	0.28	0.44	0.38
Simon Says Score	0.34	-0.05	0.40	0.33	0.40	0.25
(N)	209	75	332	107	110	225
Art Show Score	0.50	-0.10	0.48	0.18	0.33	0.39
(N)	208	74	330	106	109	224
Pre-LAS total language screener score Spanish	0.02	0.25	0.19	0.10	0.41	0.09
Tío Simón Dice Score	0.06	0.14	0.12	0.21	0.39	-0.02
(N)	81	61	122	44	44	78
Exposición de Arte Score	-0.02	0.25	0.20	0.00	0.26	0.16
(N)	81	61	122	44	44	78
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.35	0.05	0.38	0.11	0.13	0.42
(N)	169	42	260	85	88	175
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.46	0.11	0.51	0.13	0.24	0.49
(N)	208	74	329	105	108	224
Expressive One Word Picture Vocabulary Test—IRT Score	0.46	0.14	0.52	0.17	0.29	0.48
(N)	208	74	329	105	108	224
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.13	-0.33	0.39	0.22	0.34	0.61
(N)	169	42	260	85	88	175
Letter-Word Identification W Score	0.17	-0.26	0.48	0.30	0.42	0.65



Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	169	42	260	85	88	175
Applied Problems Standard Score	0.25	-0.13	0.43	0.47	0.40	0.40
(N)	169	42	260	85	88	175
Applied Problems W Score	0.29	0.00	0.53	0.55	0.50	0.44
(N)	169	42	260	85	88	175
Spelling Standard Score	0.20	-0.01	0.33	0.22	0.29	0.49
(N)	169	42	260	85	88	175
Spelling W Score	0.25	0.09	0.45	0.30	0.39	0.54
(N)	169	42	260	85	88	175
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.06	-0.23	0.14	0.19	0.21	0.17
(N)	39	32	68	19	19	49
Letter-Word Identification W Score	0.10	-0.16	0.17	0.12	0.25	0.20
(N)	39	32	68	19	19	49
Applied Problems Standard Score	-0.06	0.13	0.01	0.01	-0.17	0.02
(N)	39	32	68	19	19	49
Applied Problems W Score	-0.05	0.17	0.01	-0.04	-0.16	0.05
(N)	39	32	68	19	19	49
Ortografía Standard Score	0.17	0.15	0.05	-0.06	-0.31	0.06
(N)	39	32	68	19	19	49
Ortografía W Score	0.21	0.20	0.07	-0.12	-0.32	0.08
(N)	39	32	68	19	19	49
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.06	0.06	0.42	.	.	0.26
(N)	122	38	129	.	.	130
Social Interaction	0.30	0.45	0.40	.	.	0.24
(N)	122	38	129	.	.	130
Social Independence	0.16	0.29	0.36	.	.	0.17
(N)	122	38	130	.	.	131
Total Positive Social Skills	0.21	0.29	0.46	.	.	0.26
(N)	122	38	129	.	.	130
Externalizing Problems	-0.01	-0.17	-0.13	.	.	-0.08
(N)	116	37	123	.	.	124
Internalizing Problems	0.08	-0.13	-0.04	.	.	-0.02
(N)	120	38	126	.	.	127
Problem Behaviors	0.04	-0.18	-0.12	.	.	-0.07
(N)	120	38	128	.	.	129
Executive Functioning						
Pencil Tapping	0.33	0.16	0.41	0.43	0.41	0.27

Standardized Measures (N)	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	208	74	329	105	108	224
Leiter Examiner Ratings						
Attention (N)	0.32 208	0.21 74	0.44 330	0.27 106	0.38 109	0.39 224
Activity (N)	0.27 208	0.29 74	0.28 330	0.27 106	0.39 109	0.19 224
Sociability (N)	0.32 208	0.08 74	0.39 330	0.22 106	0.41 109	0.35 224
Family Child Care						
Pre-LAS total language screener score English	0.65	0.03	0.54	0.03	-0.07	0.39
Simon Says Score (N)	0.53 63	0.02 25	0.46 84	-0.01 14	-0.14 14	0.31 73
Art Show Score (N)	0.58 63	0.04 25	0.50 84	0.07 14	0.00 14	0.35 73
Pre-LAS total language screener score Spanish	-0.09	0.25	0.12	-0.40	0.25	0.02
Tío Simón Dice Score (N)	0.23 18	0.22 21	0.38 24	0.09 3	-0.24 3	0.15 21
Exposición de Arte Score (N)	-0.22 18	0.19 21	-0.12 24	-0.85 3	0.76 3	-0.07 21
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score (N)	0.45 51	0.31 11	0.31 69	0.60 11	0.21 11	0.16 61
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score (N)	0.43 63	0.32 25	0.55 84	0.17 14	0.25 14	0.34 73
Expressive One Word Picture Vocabulary Test—IRT Score	0.41	0.30	0.52	0.21	0.35	0.32

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	63	25	84	14	14	73
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.33	0.23	0.45	0.53	0.13	0.57
(N)	51	11	69	11	11	61
Letter-Word Identification W Score	0.30	0.30	0.46	0.66	0.27	0.58
(N)	51	11	69	11	11	61
Applied Problems Standard Score	0.53	0.70	0.46	0.72	0.69	0.30
(N)	51	11	69	11	11	61
Applied Problems W Score	0.53	0.75	0.45	0.71	0.77	0.29
(N)	51	11	69	11	11	61
Spelling Standard Score	0.56	0.17	0.37	0.28	0.88	0.43
(N)	51	11	69	11	11	61
Spelling W Score	0.53	0.20	0.35	0.34	0.91	0.41
(N)	51	11	69	11	11	61
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.39	0.11	0.27	-0.98	1.00	0.55
(N)	12	14	15	3	3	12
Letter-Word Identification W Score	0.37	0.01	0.32	-1.00	0.97	0.52
(N)	12	14	15	3	3	12
Applied Problems Standard Score	-0.12	0.22	0.10	-1.00	0.97	0.15
(N)	12	14	15	3	3	12
Applied Problems W Score	-0.11	0.15	0.14	-0.50	0.28	0.13
(N)	12	14	15	3	3	12
Ortografía Standard Score	0.63	0.12	0.45	-0.95	1.00	0.66
(N)	12	14	15	3	3	12
Ortografía W Score	0.58	-0.01	0.51	-1.00	0.97	0.61
(N)	12	14	15	3	3	12
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.08	0.81	-0.06	.	.	-0.06
(N)	19	13	19	.	.	22
Social Interaction	0.22	0.92	0.41	.	.	0.18
(N)	19	13	19	.	.	22
Social Independence	0.28	0.93	0.67	.	.	0.31
(N)	19	13	19	.	.	22
Total Positive Social Skills	0.20	0.92	0.44	.	.	0.17
(N)	19	13	19	.	.	22
Externalizing Problems	-0.09	-0.49	0.11	.	.	-0.09
(N)	17	13	17	.	.	20

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Internalizing Problems (N)	-0.28 17	-0.79 12	-0.13 19	.	.	-0.23 22
Problem Behaviors (N)	-0.13 19	-0.69 13	-0.08 19	.	.	-0.18 22
Executive Functioning Pencil Tapping (N)	0.48 63	0.44 25	0.51 84	0.37 14	0.67 14	0.22 73
Leiter Examiner Ratings						
Attention (N)	0.20 63	-0.14 25	0.24 84	-0.09 14	0.27 14	0.09 73
Activity (N)	0.25 63	-0.07 25	0.09 84	0.21 14	0.05 14	-0.04 73
Sociability (N)	0.15 63	-0.16 25	0.08 84	0.15 14	0.16 14	0.02 73

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.21. Correlations (Pairwise) Between Change Scores on Provider Administered Assessment (PAA) and Standardized Measures Spring 2010, by Program Type

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
	Center					
Pre-LAS total language screener score English	0.18	0.08	0.14	0.22	0.28	0.06
Simon Says Score	0.05	0.15	0.07	0.19	0.23	-0.03
(N)	175	62	274	75	78	198
Art Show Score	0.27	-0.06	0.16	0.17	0.22	0.15
(N)	175	62	273	74	77	198
Pre-LAS total language screener score Spanish	-0.10	-0.02	-0.12	-0.05	0.11	-0.28
Tío Simón Dice Score	-0.07	-0.06	-0.13	-0.01	0.19	-0.24
(N)	69	52	106	32	32	76
Exposición de Arte Score	-0.09	0.03	-0.06	-0.08	-0.02	-0.20
(N)	69	52	106	32	32	76
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.10	0.45	-0.02	-0.32	-0.09	-0.12
(N)	110	16	170	50	53	115
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.16	0.40	0.17	-0.24	-0.09	0.01
(N)	174	62	270	72	75	197
Expressive One Word Picture Vocabulary Test—IRT Score	0.08	0.51	0.07	-0.14	0.00	-0.09
(N)	174	62	271	73	76	197
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.01	-0.10	0.19	-0.01	-0.05	0.32
(N)	110	16	169	49	52	115
Letter-Word Identification W Score	-0.02	-0.11	0.17	0.01	-0.06	0.29

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	110	16	169	49	52	115
Applied Problems Standard Score	-0.01	0.41	0.08	0.08	0.08	0.08
(N)	110	16	169	49	52	115
Applied Problems W Score	-0.04	0.42	0.08	0.19	0.13	0.07
(N)	110	16	169	49	52	115
Spelling Standard Score	-0.02	0.14	0.15	-0.03	0.19	0.36
(N)	110	16	169	49	52	115
Spelling W Score	-0.05	0.13	0.15	0.01	0.21	0.35
(N)	110	16	169	49	52	115
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.36	0.24	-0.03	0.49	0.56	-0.20
(N)	27	21	38	9	9	30
Letter-Word Identification W Score	0.38	0.19	-0.06	0.57	0.56	-0.21
(N)	29	23	50	10	10	41
Applied Problems Standard Score	0.11	-0.02	0.06	-0.06	-0.07	-0.29
(N)	29	23	50	10	10	41
Applied Problems W Score	0.09	-0.04	0.05	-0.07	-0.07	-0.30
(N)	29	23	50	10	10	41
Ortografía Standard Score	0.18	0.21	0.01	0.18	0.13	-0.12
(N)	29	23	50	10	10	41
Ortografía W Score	0.17	0.21	0.01	0.23	0.13	-0.13
(N)	29	23	50	10	10	41
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.17	0.07	0.23	.	.	-0.03
(N)	101	32	98	.	.	112
Social Interaction	0.30	0.27	0.28	.	.	0.13
(N)	101	32	98	.	.	112
Social Independence	0.36	0.15	0.19	.	.	-0.01
(N)	101	32	99	.	.	113
Total Positive Social Skills	0.34	0.15	0.22	.	.	0.01
(N)	101	32	98	.	.	112
Externalizing Problems	0.02	-0.03	-0.17	.	.	0.07
(N)	96	31	96	.	.	106
Internalizing Problems	0.01	-0.16	-0.19	.	.	0.07
(N)	99	32	96	.	.	109
Problem Behaviors	0.09	-0.11	-0.19	.	.	0.09
(N)	99	32	98	.	.	111
Executive Functioning						
Pencil Tapping	-0.10	0.07	-0.06	0.05	-0.07	0.02

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	174	62	268	70	73	197
Leiter Examiner Ratings						
Attention	0.01	0.20	0.12	0.17	0.14	0.16
(N)	174	62	270	71	74	198
Activity	0.03	0.22	0.10	-0.15	0.02	0.17
(N)	174	62	270	71	74	198
Sociability	-0.03	0.06	0.18	-0.02	0.07	0.14
(N)	174	62	270	71	74	198
Family Child Care						
Pre-LAS total language screener score English	0.13	0.45	0.21	-0.23	-0.13	0.19
Simon Says Score	0.02	0.21	0.06	-0.15	-0.10	0.14
(N)	50	23	72	13	14	66
Art Show Score	0.23	0.56	0.37	-0.36	-0.16	0.17
(N)	50	23	72	13	14	66
Pre-LAS total language screener score Spanish	-0.11	0.09	0.10	-0.54	-0.69	-0.04
Tío Simón Dice Score	-0.10	0.10	0.23	-0.09	-0.95	0.21
(N)	16	20	22	3	3	21
Exposición de Arte Score	-0.06	0.01	-0.17	-0.96	0.49	-0.27
(N)	16	20	22	3	3	21
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.07	0.12	-0.04	-0.69	0.02	0.01
(N)	38	7	52	10	10	49
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	-0.12	0.11	0.05	-0.60	-0.06	0.07
(N)	50	23	72	13	14	66
Expressive One Word Picture Vocabulary Test—IRT Score	-0.04	0.12	0.05	-0.59	-0.06	-0.02

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	50	23	72	13	14	66
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.17	0.59	0.17	-0.11	0.17	0.41
(N)	38	7	52	10	10	49
Letter-Word Identification W Score	0.17	0.57	0.16	-0.12	0.15	0.42
(N)	38	7	52	10	10	49
Applied Problems Standard Score	0.04	-0.17	-0.04	-0.01	0.73	0.22
(N)	38	7	52	10	10	49
Applied Problems W Score	0.03	-0.08	-0.05	-0.14	0.73	0.24
(N)	38	7	52	10	10	49
Spelling Standard Score	0.12	-0.25	0.01	-0.49	-0.13	0.09
(N)	38	7	52	10	10	49
Spelling W Score	0.12	-0.14	0.02	-0.52	-0.11	0.09
(N)	38	7	52	10	10	49
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.54	0.45	0.24	.	.	0.59
(N)	8	11	12	1	2	10
Letter-Word Identification W Score	0.46	0.40	0.25	1.00	-0.63	0.54
(N)	9	13	15	2	3	12
Applied Problems Standard Score	0.12	0.18	0.37	1.00	-0.99	0.21
(N)	9	13	15	2	3	12
Applied Problems W Score	0.17	0.18	0.39	1.00	-0.99	0.26
(N)	9	13	15	2	3	12
Ortografía Standard Score	0.63	0.48	0.58	1.00	-0.93	0.62
(N)	9	13	15	2	3	12
Ortografía W Score	0.67	0.48	0.58	1.00	-0.94	0.64
(N)	9	13	15	2	3	12
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.03	0.00	0.60	.	.	0.70
(N)	17	13	16	.	.	20
Social Interaction	0.50	0.42	0.64	.	.	0.20
(N)	17	13	15	.	.	19
Social Independence	-0.07	-0.08	0.22	.	.	0.63
(N)	17	13	16	.	.	20
Total Positive Social Skills	0.25	0.23	0.61	.	.	0.58
(N)	17	13	15	.	.	19
Externalizing Problems	0.24	0.33	-0.34	.	.	0.09
(N)	15	13	15	.	.	19



Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Internalizing Problems (N)	0.10 15	0.23 12	-0.12 16	. .	. .	-0.04 20
Problem Behaviors (N)	0.15 17	0.26 13	-0.18 16	. .	. .	0.05 20
Executive Functioning						
Pencil Tapping (N)	0.50 50	0.37 23	0.05 72	-0.42 13	0.52 14	0.27 66
Leiter Examiner Ratings						
Attention (N)	-0.08 50	-0.30 23	-0.09 72	-0.47 13	-0.31 14	0.19 66
Activity (N)	-0.14 50	-0.21 23	-0.06 72	-0.20 13	0.19 14	0.01 66
Sociability (N)	0.00 50	-0.40 23	-0.07 72	-0.61 13	0.21 14	0.13 66

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.22. Predictive Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) Measures from Fall 2009 and Standardized Measures from Spring 2010, by Program Type

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Center						
Pre-LAS total language screener score English	0.62	0.09	0.62	0.39	0.37	0.41
Simon Says Score	0.47	0.08	0.47	0.40	0.27	0.32
(N)	237	89	327	106	107	247
Art Show Score	0.60	0.04	0.58	0.21	0.32	0.38
(N)	236	88	325	105	106	246
Pre-LAS total language screener score Spanish	0.14	0.36	0.31	0.41	0.48	0.22
Tío Simón Dice Score	0.23	0.26	0.26	0.52	0.31	0.17
(N)	88	73	123	38	38	93
Exposición de Arte Score	0.01	0.30	0.23	0.09	0.36	0.18
(N)	88	73	123	38	38	93
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.43	-0.15	0.55	0.26	0.22	0.44
(N)	194	51	257	86	87	188
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.56	0.09	0.65	0.26	0.34	0.47
(N)	235	88	324	104	105	246
Expressive One Word Picture Vocabulary Test—IRT Score	0.58	0.13	0.68	0.27	0.36	0.52
(N)	235	88	324	104	105	246
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.23	-0.22	0.46	0.26	0.42	0.62
(N)	194	51	257	86	87	188
Letter-Word Identification W Score	0.30	-0.16	0.54	0.29	0.52	0.69

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	194	51	257	86	87	188
Applied Problems Standard Score	0.39	-0.13	0.52	0.28	0.37	0.42
(N)	194	51	257	86	87	188
Applied Problems W Score	0.49	-0.02	0.60	0.27	0.43	0.50
(N)	194	51	257	86	87	188
Spelling Standard Score	0.27	-0.21	0.42	0.36	0.49	0.49
(N)	194	51	257	86	87	188
Spelling W Score	0.38	-0.10	0.54	0.39	0.61	0.60
(N)	194	51	257	86	87	188
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.05	-0.02	0.19	0.14	0.45	0.08
(N)	41	37	66	17	17	58
Letter-Word Identification W Score	0.01	0.08	0.25	0.17	0.47	0.12
(N)	41	37	66	17	17	58
Applied Problems Standard Score	-0.17	0.19	-0.06	-0.14	-0.11	-0.04
(N)	41	37	66	17	17	58
Applied Problems W Score	-0.20	0.24	-0.03	-0.15	-0.13	0.00
(N)	41	37	66	17	17	58
Ortografía Standard Score	-0.10	0.07	-0.08	-0.35	-0.15	0.07
(N)	41	37	66	17	17	58
Ortografía W Score	-0.13	0.13	-0.04	-0.37	-0.17	0.10
(N)	41	37	66	17	17	58
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.12	0.25	0.19	0.49	0.33	0.17
(N)	106	46	110	10	10	113
Social Interaction	0.36	0.61	0.16	0.49	0.05	0.15
(N)	106	46	110	10	10	113
Social Independence	0.25	0.51	0.15	0.19	-0.14	0.14
(N)	106	46	111	10	10	114
Total Positive Social Skills	0.27	0.52	0.19	0.62	0.19	0.18
(N)	106	46	110	10	10	113
Externalizing Problems	-0.04	-0.07	0.13	-0.33	-0.26	0.00
(N)	102	44	107	10	10	106
Internalizing Problems	-0.06	-0.01	0.22	-0.06	0.46	0.10
(N)	105	46	107	10	10	109
Problem Behaviors	-0.08	-0.07	0.17	-0.25	0.04	0.02
(N)	105	46	109	10	10	111
Executive Functioning						
Pencil Tapping	0.36	0.09	0.52	0.26	0.45	0.29

Standardized Measures (N)	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
	235	88	324	104	105	246
Leiter Examiner Ratings						
Attention (N)	0.30 236	0.14 88	0.47 325	0.23 105	0.35 106	0.30 246
Activity (N)	0.20 236	0.11 88	0.34 325	0.36 105	0.45 106	0.17 246
Sociability (N)	0.30 236	0.08 88	0.41 325	0.29 105	0.47 106	0.27 246
Family Child Care						
Pre-LAS total language screener score English Simon Says Score (N)	0.80 0.72 67	0.44 0.49 32	0.62 0.50 89	0.34 0.27 22	0.41 0.30 24	0.46 0.35 70
Art Show Score (N)	0.74 67	0.34 32	0.59 89	0.38 22	0.49 24	0.43 70
Pre-LAS total language screener score Spanish Tío Simón Dice Score (N)	-0.29 0.23 25	0.04 0.44 26	-0.13 0.31 24	0.11 0.56 5	0.24 0.18 5	-0.15 0.01 21
Exposición de Arte Score (N)	-0.52 25	-0.28 26	-0.41 24	-0.34 5	0.24 5	-0.17 21
Expressive One Word Picture Vocabulary Test— English Edition (EOWPVT) Standard Score (N)	0.36 50	0.41 13	0.25 71	0.29 17	0.27 18	0.34 58
Expressive One Word Picture Vocabulary Test— Spanish Bilingual Edition (EOWPVT-SBE) Standard Score (N)	0.46 67	0.36 32	0.45 89	0.18 22	0.45 24	0.45 70
Expressive One Word Picture Vocabulary Test—IRT Score (N)	0.47 67	0.40 32	0.49 89	0.26 22	0.57 24	0.45 70
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score (N)	0.24 50	0.03 13	0.42 71	0.12 17	0.09 18	0.50 58
Letter-Word Identification W Score (N)	0.28 50	0.27 13	0.50 71	0.25 17	0.26 18	0.50 58
Applied Problems Standard Score (N)	0.36 50	0.06 13	0.51 71	0.23 17	0.54 18	0.26 58
Applied Problems W Score (N)	0.43 50	0.34 13	0.57 71	0.39 17	0.72 18	0.18 58

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Spelling Standard Score (N)	0.44 50	0.34 13	0.39 71	0.30 17	0.67 18	0.36 58
Spelling W Score (N)	0.47 50	0.62 13	0.45 71	0.40 17	0.78 18	0.32 58
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score (N)	-0.24 17	-0.22 19	0.20 18	0.06 5	0.49 6	0.22 12
Letter-Word Identification W Score (N)	-0.21 17	-0.29 19	0.29 18	0.18 5	0.57 6	0.19 12
Applied Problems Standard Score (N)	-0.25 17	0.39 19	-0.31 18	-0.62 5	-0.83 6	-0.08 12
Applied Problems W Score (N)	-0.23 17	0.36 19	-0.23 18	-0.63 5	-0.83 6	-0.11 12
Ortografía Standard Score (N)	0.06 17	-0.09 19	-0.54 18	-0.61 5	-0.76 6	0.15 12
Ortografía W Score (N)	0.06 17	-0.17 19	-0.47 18	-0.63 5	-0.77 6	0.08 12
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation (N)	0.14 17	0.70 13	0.14 20	.	.	-0.09 21
Social Interaction (N)	0.35 17	0.87 13	0.24 20	.	.	0.13 21
Social Independence (N)	0.35 17	0.86 13	0.47 20	.	.	0.24 21
Total Positive Social Skills (N)	0.29 17	0.83 13	0.32 20	.	.	0.12 21
Externalizing Problems (N)	-0.20 15	-0.52 13	-0.16 19	.	.	0.03 20
Internalizing Problems (N)	-0.48 15	-0.70 12	-0.26 20	.	.	-0.09 21
Problem Behaviors (N)	-0.31 17	-0.63 13	-0.27 20	.	.	-0.08 21
Executive Functioning						
Pencil Tapping (N)	0.18 67	0.19 32	0.45 89	0.46 22	0.61 24	0.11 70
Leiter Examiner Ratings						
Attention (N)	0.02 67	-0.29 32	0.21 89	0.41 22	0.46 24	-0.04 70

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Activity (N)	0.19 67	0.37 32	0.15 89	0.42 22	0.19 24	-0.07 70
Sociability (N)	0.13 67	0.11 32	0.17 89	0.51 22	0.24 24	0.02 70

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.23. Concurrent Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) and Standardized Measures Spring 2010, by Concentration of ELLs

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
High Concentration of ELLs						
Pre-LAS total language screener score English	0.47	0.16	0.43	0.10	0.27	0.35
Simon Says Score	0.28	0.08	0.25	0.14	0.17	0.16
(N)	131	79	182	40	40	143
Art Show Score	0.50	0.12	0.44	0.04	0.28	0.39
(N)	130	78	181	40	40	142
Pre-LAS total language screener score Spanish	-0.02	0.16	0.09	0.18	0.37	-0.03
Tío Simón Dice Score	0.08	0.10	0.02	0.46	0.45	-0.07
(N)	86	69	98	19	19	79
Exposición de Arte Score	-0.08	0.14	0.11	0.00	0.21	0.02
(N)	86	69	98	19	19	79
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.35	0.28	0.39	0.31	0.19	0.41
(N)	85	39	117	26	26	92
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.42	0.32	0.55	0.22	0.25	0.48
(N)	130	78	180	39	39	142
Expressive One Word Picture Vocabulary Test—IRT Score	0.44	0.35	0.54	0.28	0.34	0.44
(N)	130	78	180	39	39	142
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.01	-0.19	0.41	0.43	0.48	0.69
(N)	85	39	117	26	26	92
Letter-Word Identification W Score	0.04	-0.11	0.46	0.49	0.58	0.70

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	85	39	117	26	26	92
Applied Problems Standard Score	0.23	0.04	0.48	0.41	0.39	0.43
(N)	85	39	117	26	26	92
Applied Problems W Score	0.27	0.18	0.56	0.61	0.60	0.43
(N)	85	39	117	26	26	92
Spelling Standard Score	0.14	-0.13	0.29	0.68	0.63	0.56
(N)	85	39	117	26	26	92
Spelling W Score	0.18	0.00	0.36	0.70	0.68	0.59
(N)	85	39	117	26	26	92
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.15	-0.08	0.22	0.53	0.63	0.37
(N)	45	39	62	12	12	50
Letter-Word Identification W Score	0.16	-0.05	0.24	0.53	0.68	0.39
(N)	45	39	62	12	12	50
Applied Problems Standard Score	-0.01	0.15	-0.10	0.02	-0.37	-0.02
(N)	45	39	62	12	12	50
Applied Problems W Score	-0.02	0.16	-0.10	0.04	-0.33	0.01
(N)	45	39	62	12	12	50
Ortografía Standard Score	0.25	0.23	0.07	-0.08	-0.48	0.20
(N)	45	39	62	12	12	50
Ortografía W Score	0.25	0.24	0.09	-0.06	-0.46	0.22
(N)	45	39	62	12	12	50
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.08	0.20	0.45	.	.	0.30
(N)	73	39	74	.	.	75
Social Interaction	0.28	0.66	0.38	.	.	0.25
(N)	73	39	74	.	.	75
Social Independence	0.16	0.43	0.36	.	.	0.23
(N)	73	39	74	.	.	75
Total Positive Social Skills	0.22	0.52	0.47	.	.	0.31
(N)	73	39	74	.	.	75
Externalizing Problems	0.06	-0.06	-0.21	.	.	-0.22
(N)	71	39	67	.	.	68
Internalizing Problems	0.17	-0.25	0.02	.	.	-0.06
(N)	71	39	71	.	.	72
Problem Behaviors	0.14	-0.18	-0.15	.	.	-0.19
(N)	72	39	72	.	.	73
Executive Functioning						
Pencil Tapping	0.29	0.33	0.31	0.53	0.60	0.19



Standardized Measures (N)	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Leiter Examiner Ratings						
Attention (N)	0.35 130	0.13 78	0.50 181	0.44 40	0.47 40	0.37 142
Activity (N)	0.29 130	0.22 78	0.42 181	0.44 40	0.50 40	0.24 142
Sociability (N)	0.30 130	0.00 78	0.39 181	0.57 40	0.56 40	0.30 142
Low Concentration of ELLs						
Pre-LAS total language screener score English (N)	0.55 141	-0.61 21	0.59 234	0.34 81	0.40 84	0.28 155
Simon Says Score (N)	0.44 141	-0.51 21	0.52 233	0.37 80	0.39 83	0.29 155
Art Show Score (N)	0.51 141	-0.57 21	0.48 233	0.26 80	0.30 83	0.18 155
Pre-LAS total language screener score Spanish (N)	-0.08 13	0.67 13	0.36 48	0.07 28	0.35 28	0.54 20
Tío Simón Dice Score (N)	-0.02 13	0.40 13	0.46 48	0.18 28	0.22 28	0.54 20
Exposición de Arte Score (N)	-0.11 13	0.73 13	0.17 48	-0.05 28	0.32 28	0.40 20
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score (N)	0.37 135	-0.31 14	0.33 212	0.08 70	0.11 73	0.28 144
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score (N)	0.41 141	-0.29 21	0.42 233	0.13 80	0.24 83	0.29 155
Expressive One Word Picture Vocabulary Test—IRT Score	0.40	-0.38	0.44	0.18	0.28	0.32

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	141	21	233	80	83	155
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.22	0.54	0.37	0.21	0.25	0.50
(N)	135	14	212	70	73	144
Letter-Word Identification W Score	0.24	0.19	0.46	0.32	0.34	0.54
(N)	135	14	212	70	73	144
Applied Problems Standard Score	0.33	0.50	0.39	0.33	0.42	0.28
(N)	135	14	212	70	73	144
Applied Problems W Score	0.37	0.41	0.46	0.43	0.50	0.31
(N)	135	14	212	70	73	144
Spelling Standard Score	0.34	0.50	0.33	0.13	0.28	0.38
(N)	135	14	212	70	73	144
Spelling W Score	0.37	0.36	0.44	0.24	0.39	0.42
(N)	135	14	212	70	73	144
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	-0.14	0.37	-0.11	-0.10	0.01	0.13
(N)	6	7	21	10	10	11
Letter-Word Identification W Score	0.15	0.25	-0.09	-0.22	-0.02	0.16
(N)	6	7	21	10	10	11
Applied Problems Standard Score	-0.60	0.27	0.40	0.06	0.03	0.40
(N)	6	7	21	10	10	11
Applied Problems W Score	-0.35	0.23	0.42	0.00	0.02	0.38
(N)	6	7	21	10	10	11
Ortografía Standard Score	0.43	-0.43	0.14	-0.08	-0.12	0.05
(N)	6	7	21	10	10	11
Ortografía W Score	0.73	-0.36	0.17	-0.19	-0.19	0.06
(N)	6	7	21	10	10	11
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.06	0.61	0.34	.	.	0.14
(N)	68	12	74	.	.	77
Social Interaction	0.20	0.36	0.42	.	.	0.20
(N)	68	12	74	.	.	77
Social Independence	0.21	0.60	0.45	.	.	0.20
(N)	68	12	75	.	.	78
Total Positive Social Skills	0.15	0.55	0.45	.	.	0.21
(N)	68	12	74	.	.	77
Externalizing Problems	-0.16	-0.68	-0.04	.	.	-0.01
(N)	62	11	73	.	.	76
Internalizing Problems	-0.21	-0.61	-0.18	.	.	-0.21

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	66	11	74	.	.	77
Problem Behaviors	-0.13	-0.70	-0.14	.	.	-0.12
(N)	67	12	75	.	.	78
Executive Functioning						
Pencil Tapping	0.39	-0.05	0.45	0.41	0.38	0.24
(N)	141	21	233	80	83	155
Leiter Examiner Ratings						
Attention	0.22	0.40	0.30	0.17	0.31	0.21
(N)	141	21	233	80	83	155
Activity	0.25	0.12	0.12	0.16	0.26	0.00
(N)	141	21	233	80	83	155
Sociability	0.25	0.28	0.28	0.10	0.30	0.13
(N)	141	21	233	80	83	155

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

Table IV.24. Correlations (Pairwise) Between Change Scores on Provider Administered Assessment (PAA) and Standardized Measures Spring 2010, by Concentration of ELLs

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
High Concentration of ELLs						
Pre-LAS total language screener score English	0.24	0.13	0.14	0.03	0.01	0.06
Simon Says Score	0.08	0.10	0.04	0.03	-0.02	-0.02
(N)	115	71	163	34	35	140
Art Show Score	0.34	0.10	0.18	0.01	0.06	0.12
(N)	115	71	163	34	35	140
Pre-LAS total language screener score Spanish	-0.01	0.11	-0.03	-0.41	0.11	-0.28
Tío Simón Dice Score	0.01	0.05	-0.05	-0.19	-0.11	-0.17
(N)	73	62	87	14	14	78
Exposición de Arte Score	-0.02	0.12	0.00	-0.38	0.26	-0.24
(N)	73	62	87	14	14	78
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.23	0.21	-0.13	-0.49	-0.30	-0.07
(N)	52	18	76	21	21	61
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.19	0.30	0.21	-0.32	-0.22	0.06
(N)	114	71	162	33	34	140
Expressive One Word Picture Vocabulary Test—IRT Score	0.16	0.37	0.12	-0.32	-0.20	-0.07
(N)	114	71	162	33	34	140
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.03	0.22	0.24	-0.03	-0.24	0.45
(N)	52	18	76	21	21	61
Letter-Word Identification W Score	-0.01	0.20	0.22	-0.01	-0.21	0.42

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	52	18	76	21	21	61
Applied Problems Standard Score	-0.05	-0.09	-0.03	-0.04	0.35	-0.02
(N)	52	18	76	21	21	61
Applied Problems W Score	-0.05	-0.10	-0.04	0.08	0.45	-0.07
(N)	52	18	76	21	21	61
Spelling Standard Score	0.05	-0.37	0.01	-0.30	0.00	0.29
(N)	52	18	76	21	21	61
Spelling W Score	0.01	-0.34	0.00	-0.27	0.03	0.28
(N)	52	18	76	21	21	61
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.46	0.39	0.08	0.46	-0.44	0.16
(N)	30	28	40	7	8	34
Letter-Word Identification W Score	0.43	0.35	0.05	0.51	-0.39	0.13
(N)	33	32	50	7	8	44
Applied Problems Standard Score	0.17	0.01	0.12	-0.22	0.12	-0.03
(N)	33	32	50	7	8	44
Applied Problems W Score	0.16	0.00	0.11	-0.23	0.13	-0.05
(N)	33	32	50	7	8	44
Ortografía Standard Score	0.25	0.29	0.13	0.32	-0.04	0.11
(N)	33	32	50	7	8	44
Ortografía W Score	0.25	0.30	0.13	0.39	-0.07	0.10
(N)	33	32	50	7	8	44
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.09	-0.01	0.30	.	.	-0.01
(N)	63	38	67	.	.	74
Social Interaction	0.32	0.32	0.49	.	.	0.13
(N)	63	38	66	.	.	73
Social Independence	0.24	0.00	0.21	.	.	0.00
(N)	63	38	67	.	.	74
Total Positive Social Skills	0.31	0.15	0.36	.	.	0.03
(N)	63	38	66	.	.	73
Externalizing Problems	0.25	0.01	-0.32	.	.	-0.05
(N)	62	38	65	.	.	68
Internalizing Problems	-0.01	-0.05	-0.41	.	.	0.01
(N)	61	38	66	.	.	72
Problem Behaviors	0.14	-0.01	-0.36	.	.	-0.01
(N)	62	38	67	.	.	73

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Executive Functioning						
Pencil Tapping (N)	0.04 114	0.19 71	-0.15 162	-0.25 33	-0.22 34	-0.06 140
Leiter Examiner Ratings						
Attention (N)	-0.10 114	-0.07 71	0.13 163	0.13 34	-0.01 35	0.10 140
Activity (N)	0.02 114	0.05 71	0.13 163	-0.22 34	0.11 35	0.13 140
Sociability (N)	-0.08 114	-0.12 71	0.13 163	-0.11 34	0.11 35	0.08 140
Low Concentration of ELLs						
Pre-LAS total language screener score English	0.01	0.64	0.15	0.21	0.27	0.10
Simon Says Score (N)	-0.04 110	0.71 14	0.08 183	0.22 54	0.26 57	0.02 124
Art Show Score (N)	0.07 110	0.40 14	0.18 182	0.16 53	0.19 56	0.19 124
Pre-LAS total language screener score Spanish	-0.78	-0.67	-0.25	0.05	0.04	-0.06
Tío Simón Dice Score (N)	-0.47 12	-0.50 10	-0.02 41	0.06 21	0.11 21	0.00 19
Exposición de Arte Score (N)	-0.71 12	-0.56 10	-0.44 41	0.02 21	-0.06 21	-0.09 19
Expressive One Word Picture Vocabulary Test— English Edition (EOWPVT) Standard Score (N)	0.01 96	0.26 5	0.03 146	-0.37 39	0.04 42	-0.07 103
Expressive One Word Picture Vocabulary Test— Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	-0.05	0.17	0.04	-0.30	0.01	-0.04

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	110	14	180	52	55	123
Expressive One Word Picture Vocabulary Test—IRT Score	-0.06	0.21	-0.02	-0.12	0.13	-0.08
(N)	110	14	181	53	56	123
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.06	0.78	0.16	0.04	0.08	0.27
(N)	96	5	145	38	41	103
Letter-Word Identification W Score	0.06	0.73	0.14	0.06	0.05	0.26
(N)	96	5	145	38	41	103
Applied Problems Standard Score	0.02	0.91	0.10	0.10	0.14	0.24
(N)	96	5	145	38	41	103
Applied Problems W Score	-0.02	0.94	0.12	0.14	0.16	0.27
(N)	96	5	145	38	41	103
Spelling Standard Score	0.03	0.88	0.20	0.20	0.28	0.26
(N)	96	5	145	38	41	103
Spelling W Score	0.01	0.91	0.21	0.22	0.29	0.26
(N)	96	5	145	38	41	103
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.26	0.48	-0.08	0.70	-0.66	0.40
(N)	5	4	10	3	3	6
Letter-Word Identification W Score	0.26	0.45	-0.14	0.78	0.78	0.15
(N)	5	4	15	5	5	9
Applied Problems Standard Score	-0.54	0.96	-0.02	0.19	0.33	-0.53
(N)	5	4	15	5	5	9
Applied Problems W Score	-0.55	0.97	-0.04	0.00	0.15	-0.51
(N)	5	4	15	5	5	9
Ortografía Standard Score	0.93	-0.72	0.02	0.04	0.31	-0.19
(N)	5	4	15	5	5	9
Ortografía W Score	0.94	-0.78	0.02	0.06	0.34	-0.22
(N)	5	4	15	5	5	9
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.30	-0.42	0.10	.	.	0.11
(N)	55	7	47	.	.	58
Social Interaction	0.34	-0.12	-0.07	.	.	0.14
(N)	55	7	47	.	.	58
Social Independence	0.46	-0.49	0.07	.	.	0.16
(N)	55	7	48	.	.	59

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Total Positive Social Skills (N)	0.41 55	-0.42 7	-0.03 47	. .	. .	0.11 58
Externalizing Problems (N)	-0.13 49	-0.21 6	0.03 46	. .	. .	0.23 57
Internalizing Problems (N)	-0.01 53	-0.10 6	0.00 46	. .	. .	0.08 57
Problem Behaviors (N)	0.02 54	0.05 7	0.00 47	. .	. .	0.19 58
Executive Functioning Pencil Tapping (N)	0.01 110	-0.34 14	0.07 178	0.20 50	0.08 53	0.22 123
Leiter Examiner Ratings						
Attention (N)	0.08 110	0.72 14	-0.03 179	0.00 50	0.07 53	0.25 124
Activity (N)	-0.04 110	0.28 14	-0.07 179	-0.14 50	-0.02 53	0.10 124
Sociability (N)	0.01 110	0.36 14	0.11 179	-0.13 50	0.07 53	0.23 124

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Teacher-Administered Measures.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.



**Table IV.25. Predictive Validity: Correlations (Pairwise) Between Provider Administered Assessment (PAA) Measures from Fall 2009 and Standardized Measures from Spring 2010, by Concentration of ELLs**

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
High Concentration of ELLs						
Pre-LAS total language screener score English	0.68	0.22	0.59	0.06	0.22	0.32
Simon Says Score	0.53	0.20	0.42	0.14	0.11	0.22
(N)	150	98	188	43	44	167
Art Show Score	0.65	0.16	0.57	-0.01	0.25	0.31
(N)	149	97	187	43	44	166
Pre-LAS total language screener score Spanish	-0.04	0.17	0.18	0.30	0.72	0.14
Tío Simón Dice Score	0.18	0.19	0.18	0.49	0.39	0.13
(N)	94	83	100	18	18	90
Exposición de Arte Score	-0.19	0.07	0.10	0.05	0.60	0.10
(N)	94	83	100	18	18	90
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.45	-0.02	0.53	0.37	0.34	0.37
(N)	99	48	124	31	31	109
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.52	0.19	0.62	0.17	0.38	0.40
(N)	149	97	186	42	43	166
Expressive One Word Picture Vocabulary Test—IRT Score	0.55	0.22	0.66	0.19	0.47	0.43
(N)	149	97	186	42	43	166
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.24	-0.14	0.49	0.58	0.60	0.62
(N)	99	48	124	31	31	109
Letter-Word Identification W Score	0.32	-0.07	0.57	0.56	0.70	0.68

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	99	48	124	31	31	109
Applied Problems Standard Score	0.43	-0.10	0.49	0.45	0.32	0.35
(N)	99	48	124	31	31	109
Applied Problems W Score	0.52	0.02	0.59	0.42	0.49	0.42
(N)	99	48	124	31	31	109
Spelling Standard Score	0.23	-0.31	0.40	0.66	0.71	0.55
(N)	99	48	124	31	31	109
Spelling W Score	0.34	-0.14	0.50	0.58	0.76	0.66
(N)	99	48	124	31	31	109
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	-0.04	-0.08	0.22	-0.33	0.72	0.09
(N)	50	49	61	10	11	57
Letter-Word Identification W Score	-0.07	-0.06	0.28	-0.29	0.76	0.11
(N)	50	49	61	10	11	57
Applied Problems Standard Score	-0.24	0.25	-0.04	-0.02	-0.23	-0.09
(N)	50	49	61	10	11	57
Applied Problems W Score	-0.25	0.26	-0.01	-0.02	-0.20	-0.07
(N)	50	49	61	10	11	57
Ortografía Standard Score	-0.11	0.06	-0.05	-0.37	-0.30	0.03
(N)	50	49	61	10	11	57
Ortografía W Score	-0.13	0.07	-0.01	-0.38	-0.27	0.05
(N)	50	49	61	10	11	57
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.17	0.31	0.14	.	.	0.19
(N)	64	48	69	0	0	75
Social Interaction	0.40	0.71	0.10	.	.	0.15
(N)	64	48	69	.	.	75
Social Independence	0.34	0.55	0.11	.	.	0.15
(N)	64	48	69	.	.	75
Total Positive Social Skills	0.34	0.60	0.13	.	.	0.20
(N)	64	48	69	.	.	75
Externalizing Problems	-0.16	-0.12	0.05	.	.	-0.13
(N)	63	47	66	.	.	68
Internalizing Problems	-0.14	-0.20	0.19	.	.	0.04
(N)	62	48	67	.	.	72
Problem Behaviors	-0.15	-0.17	0.08	.	.	-0.10
(N)	63	48	68	.	.	73

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Executive Functioning						
Pencil Tapping	0.27	0.13	0.47	0.22	0.63	0.19
(N)	149	97	186	42	43	166
Leiter Examiner Ratings						
Attention	0.34	0.15	0.45	0.04	0.46	0.23
(N)	149	97	187	43	44	166
Activity	0.23	0.23	0.37	0.27	0.47	0.11
(N)	149	97	187	43	44	166
Sociability	0.29	0.15	0.36	0.22	0.47	0.19
(N)	149	97	187	43	44	166
Low Concentration of ELLs						
Pre-LAS total language screener score English	0.48	-0.01	0.59	0.52	0.45	0.45
Simon Says Score	0.38	-0.02	0.50	0.47	0.34	0.39
(N)	154	23	224	81	83	150
Art Show Score	0.44	0.00	0.53	0.42	0.43	0.37
(N)	154	23	223	80	82	150
Pre-LAS total language screener score Spanish	0.56	0.72	0.42	0.38	0.37	0.29
Tío Simón Dice Score	0.52	0.72	0.53	0.55	0.29	0.30
(N)	19	16	46	24	24	24
Exposición de Arte Score	0.43	0.51	0.13	-0.01	0.26	0.20
(N)	19	16	46	24	24	24
Expressive One Word Picture Vocabulary Test—English Edition (EOWPVT) Standard Score	0.39	0.23	0.46	0.20	0.18	0.39
(N)	145	16	202	70	72	137
Expressive One Word Picture Vocabulary Test—Spanish Bilingual Edition (EOWPVT-SBE) Standard Score	0.43	0.19	0.54	0.28	0.34	0.43
(N)	153	23	223	80	82	150

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
Expressive One Word Picture Vocabulary Test—IRT Score	0.47	0.24	0.57	0.30	0.35	0.48
(N)	153	23	223	80	82	150
Woodcock-Johnson III (WJ-III)						
Letter-Word Identification Standard Score	0.16	-0.12	0.42	0.00	0.26	0.55
(N)	145	16	202	70	72	137
Letter-Word Identification W Score	0.23	-0.11	0.51	0.10	0.36	0.62
(N)	145	16	202	70	72	137
Applied Problems Standard Score	0.33	0.26	0.52	0.17	0.43	0.31
(N)	145	16	202	70	72	137
Applied Problems W Score	0.44	0.29	0.57	0.23	0.48	0.33
(N)	145	16	202	70	72	137
Spelling Standard Score	0.32	0.27	0.40	0.08	0.43	0.34
(N)	145	16	202	70	72	137
Spelling W Score	0.41	0.31	0.50	0.21	0.57	0.41
(N)	145	16	202	70	72	137
Woodcock-Muñoz Bateria III (WM-III)						
Letter-Word Identification Standard Score	0.19	-0.52	0.06	0.19	0.36	0.29
(N)	8	7	21	10	10	13
Letter-Word Identification W Score	0.26	-0.24	0.19	0.30	0.42	0.33
(N)	8	7	21	10	10	13
Applied Problems Standard Score	0.13	0.77	-0.30	-0.40	-0.34	0.58
(N)	8	7	21	10	10	13
Applied Problems W Score	0.20	0.89	-0.24	-0.41	-0.37	0.55
(N)	8	7	21	10	10	13
Ortografía Standard Score	0.56	-0.42	-0.41	-0.50	-0.37	0.29
(N)	8	7	21	10	10	13
Ortografía W Score	0.58	-0.02	-0.36	-0.52	-0.41	0.31
(N)	8	7	21	10	10	13
Preschool Kindergarten Behavior Scales-2 (PKBS-2)						
Social Cooperation	0.14	0.47	0.30	0.49	0.33	0.11
(N)	59	11	61	10	10	59
Social Interaction	0.21	0.37	0.38	0.49	0.05	0.16
(N)	59	11	61	10	10	59
Social Independence	0.20	0.63	0.43	0.19	-0.14	0.23
(N)	59	11	62	10	10	60
Total Positive Social Skills	0.20	0.54	0.42	0.62	0.19	0.19

Standardized Measures	ASQ-3					
	Communication, English	Communication, Spanish	Problem Solving	Gross Motor	Fine Motor	RLN
(N)	59	11	61	10	10	59
Externalizing Problems	-0.06	-0.15	-0.03	-0.33	-0.26	0.11
(N)	54	10	60	10	10	58
Internalizing Problems	-0.14	-0.04	-0.22	-0.06	0.46	-0.08
(N)	58	10	60	10	10	58
Problem Behaviors	-0.11	-0.22	-0.15	-0.25	0.04	-0.02
(N)	59	11	61	10	10	59
Executive Functioning						
Pencil Tapping	0.30	0.12	0.48	0.37	0.45	0.22
(N)	153	23	223	80	82	150
Leiter Examiner Ratings						
Attention	0.21	-0.01	0.34	0.41	0.33	0.17
(N)	154	23	223	80	82	150
Activity	0.23	-0.08	0.24	0.44	0.34	0.08
(N)	154	23	223	80	82	150
Sociability	0.24	-0.07	0.33	0.40	0.38	0.17
(N)	154	23	223	80	82	150

Source: UPCOS-3 Spring 2010 Direct Child Assessment and Fall 2009 Provider-Administered Assessment.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); RLN = Rapid Letter Naming Task.

**This page has been left blank for double-sided copying.**

---

**MATHEMATICA**  
Policy Research, Inc.

---

[www.mathematica-mpr.com](http://www.mathematica-mpr.com)

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research