

Contract No.: Contract 07110
MPR Reference No.: 6344-300

MATHEMATICA
Policy Research, Inc.

**Reliability and Validity of
Child Outcome Measures
with Culturally and
Linguistically Diverse
Preschoolers:**

**The First 5 LA Universal
Preschool Child
Outcomes Study Spring
2007 Pilot Study**

March 28, 2008

***Cheri Vogel
Nikki Aikens
Sally Atkins-Burnett
Emily Sama Martin
Margaret Caspe
Susan Sprachman
John M. Love***

Submitted to:

First 5 LA
750 N. Alameda Street, Suite 300
Los Angeles, CA 90012
Telephone: 213.482.9489
Facsimile: 213.482.5903

Project Officer:
Katie Fallin

Submitted by:

Mathematica Policy Research, Inc.
600 Alexander Park
Princeton, NJ 08540
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
John M. Love

CONTENTS

Section	Page
A INTRODUCTION.....	1
B METHODOLOGY.....	3
C SAMPLE	7
1. Programs.....	7
2. Teachers.....	7
3. Children.....	9
D MEASURES AND ADMINISTRATION PROCEDURES.....	11
E ANALYTIC APPROACH	13
F DESCRIPTIVE INFORMATION FOR SPRING MEASURES	15
1. Psychometric Evidence for Utility of Measures	18
a. Cognitive and General Knowledge Measures	18
b. Language Measures	24
c. Social-Emotional and Approaches to Learning Measures	29
d. Physical and Motor Development Measures.....	33
2. Summary of Spring 2007 Pilot Test Results	33
REFERENCES	35
APPENDIX A: ENHANCING THE VALUE OF THE PARENT INTERVIEW IN THE UNIVERSAL PRESCHOOL CHILD OUTCOMES STUDY: METHODS AND RESULTS OF FOCUS GROUPS AND COGNITIVE INTERVIEWS.....	39

TABLES

Table		Page
1	CHARACTERISTICS OF PROGRAMS IN PHASE 1 SAMPLE	7
2	TEACHER DEMOGRAPHIC CHARACTERISTICS AND LANGUAGE USAGE	8
3	TEACHERS' EXPERIENCE.....	9
4	TEACHER EDUCATION AND CREDENTIALS.....	9
5	CHILD DEMOGRAPHIC CHARACTERISTICS.....	10
6	PHASE 1 STUDY MEASURES, BY DATA SOURCE AND DEVELOPMENTAL DOMAIN	12
7	COGNITIVE AND LANGUAGE RAW AND STANDARD SCORES FOR ENTIRE PILOT SAMPLE	16
8	SOCIAL-EMOTIONAL DEVELOPMENT, RAW AND STANDARD SCORES FOR ENTIRE PILOT SAMPLE	17
9	PHYSICAL/MOTOR DEVELOPMENT SCORES FOR ENTIRE PILOT SAMPLE.....	18
10	LETTER NAMING SCORES AND RELIABILITIES	19
11	BIVARIATE CORRELATIONS OF DRDP-R MEASURES WITH COGNITIVE MEASURES.....	24
12	TEACHER REPORTED SOCIAL-EMOTIONAL MEASURES RELIABILITY	30
13	BIVARIATE CORRELATIONS OF LANGUAGE AND COGNITIVE MEASURES WITH APPROACHES TO LEARNING	31

Table		Page
A.1	SUMMARY OF FOCUS GROUP CARDS: AVERAGE, MAXIMUM, AND MINIMUM VALUES FOR ALL GROUPS	45
A.2	SUMMARY OF FOCUS GROUP CARDS: FILIPINO, KOREAN, AND CHINESE OR CAMBODIAN GROUPS	48
A.3	SUMMARY OF FOCUS GROUP CARDS: LATINO/SPANISH-SPEAKING, LATIN/ENGLISH-SPEAKING, AND AFRICAN AMERICAN GROUPS	52
A.4	SUMMARY OF FOCUS GROUP CARDS: FILIPINO, KOREAN, AND CHINESE OR CAMBODIAN GROUPS	56

FIGURES

Figure		Page
1	ITEM-PERSON MAP OF LETTER NAMING TASK FOR FORM 1.....	20
2	ITEM-PERSON MAP OF LETTER NAMING TASK FOR FORM 2.....	21
3	LANGUAGE ROUTING USED SPRING 2007.....	25
4	CURVE OF PRELAS SCORES USING PPVT 1.5 SD CUT-OFF SCORE.....	26
5	RECOMMENDED ROUTING PROCEDURES FOLLOWING PILOT STUDY.....	27

A. INTRODUCTION

First 5 LA contracted with Mathematica Policy Research, Inc. (MPR) and its partners to study the implementation of a newly funded universal preschool program operating in Los Angeles County. The First 5 LA Universal Preschool Child Outcomes Study (UPCOS) has an overall goal of learning about how children enrolled in Los Angeles Universal Preschool (LAUP) programs, including the Power of Preschool (PoP) demonstration, are faring. In this two-phase study, MPR, along with its partners (Juárez and Associates, American Institutes for Research, Berkeley Evaluation and Assessment Research Center, and Deanna Gomby, a senior consultant), are working with First 5 LA, its Research Advisory Committee (RAC), and LAUP to design and implement the study.

Because Los Angeles County residents are diverse in language, race, ethnicity, and country of origin, the focus of the spring 2007 pilot study (Phase 1) was to identify child assessments appropriate for the population. We examined the feasibility, reliability, and validity of child development measures. We conducted an in-depth review of measures of children's developmental outcomes and recommended ones for inclusion in the Phase 1 battery (see Atkins-Burnett, Caspe et al. 2007). The second phase of the study (fall 2007 and spring 2008) will answer questions about the quality, intensity, and overall implementation of LAUP programs, exploring how those factors relate to children's comprehensive developmental outcomes that may contribute to their school readiness by the end of their preschool experience. This formative evaluation will help programs early in their implementation and will contribute to the larger early childhood program evaluation field through increased understanding of issues in assessing young children from diverse linguistic and cultural backgrounds.

This report describes the implementation of Phase 1 of the study and its findings. We focus exclusively on the Phase 1 portion of the study, the methods we used, and the results of our analysis of the cultural and linguistic appropriateness of individual items and measures. These findings, already reported in a preliminary version to First 5 LA, LAUP, and the RAC, helped to inform selection of the measures for the battery that is being used in the second phase of the study. We conclude with a discussion of the changes to the final Phase 2 battery based on Phase 1 findings and discussion with First 5 LA and the RAC.

As already noted, the primary goal of the Phase 1 study was to examine the reliability and validity of child development measures for diverse groups of children. The overarching research question for Phase 1, described in the research design and analysis plan (Vogel et al. 2007), is as follows: What are the reliability and validity of measures of child development in a culturally and linguistically diverse group of children who have different home languages and different levels of English proficiency? Several additional sub-questions were directed at specific elements of the Phase 1 assessment battery (including reliability and validity of the child assessments and of parent and teacher ratings of children's behavior). Next, we describe our rationale for the sample used, the assessment instruments used, and how we implemented the data collection.

B. METHODOLOGY

The Phase 1 study employed a purposive sample of programs and children. This was appropriate because the nature of the research questions concerned the reliability and validity of measures with children from a variety of backgrounds. Most important for Phase 1 was to have a sample that included the diverse racial/ethnic and language groups that are enrolled in all LAUP/PoP programs. Phase 2 will use a more complex sampling of programs and children to ensure representation of the full range of programs that will allow us to generalize results to all LAUP/PoP programs.

We worked with First 5 LA, and with LAUP coaches (technical assistance and support staff employed by LAUP to assist programs) to identify centers and programs that (1) might be willing to participate in the pilot study and (2) had children enrolled who were in racial/ethnic groups that were relatively less prevalent in the overall LAUP program. Our recruitment efforts sought to represent the range of racial/ethnic and language groups to facilitate analysis of item functioning as well as overall reliability and validity of measures. For example, we made a special effort to recruit programs that enrolled White children as they would comprise a reference group for analysis of item functioning. In addition to these less prevalent racial/ethnic groups (Asians, Blacks, Whites), we also asked the coaches to recommend programs that served both English- and Spanish-speaking Latinos. Study coordinators from MPR, AIR, and Juárez and Associates contacted the nominated programs to obtain cooperation. Of the 25 programs nominated, 14 agreed to participate. The study coordinators worked with the programs to schedule the site visits, obtain rosters of eligible children, and secure consent to participate from the children's parents.¹ The site visits occurred between May 7, 2007 and June 4, 2007. Of the 513 children enrolled, 11 were ineligible, providing an eligible sample size of 502.² We received an 86.1 percent consent rate (432 children) and were able to conduct assessments with 418 children, which was 96.8 percent of the children with parental consent and 83.3 percent of those eligible. Teacher Child Report surveys (TCR, see below) were received for 383 children. We also received 513

¹ All children funded by LAUP were eligible. Some programs also served 3-year-old children. Using birthdates, we confirmed that the children in the study would be eligible for kindergarten the following year.

² Children were deemed ineligible if they had left the program. The 11 ineligible children were in 4 of the 14 participating programs; two programs had 2 ineligible children, one program had 3 ineligible children, and the remaining program had 4 ineligible children.

teacher ratings using the Desired Results Developmental Profiles-Revised (DRDP-R) for 259 children (reports from two teachers per child). A DRDP-R form was received from at least one teacher for 60 percent of the children.

We hired field staff to conduct child assessments and trained them to administer the assessments reliably. Of 11 field staff hired for the Phase 1 study, 9 were English-Spanish bilingual, 8 were female, and 3 were male. All participated in a four-day training held at the Los Angeles offices of Juárez and Associates in May 2007 immediately preceding the start of the data collection period. After three days of training and practice, each assessor administered the assessment battery to a 4-year-old child and was certified as able to administer the assessment.³ MPR trainers used a certification form that tracked errors in administrative procedures such as inappropriate hand gestures, coaching, and straying from the script, as well as scoring errors. To be certified, an assessor could not make more than 25 errors out of a total of more than 400 possible errors. After the assessor completed the certification assessment, the certifier provided detailed feedback. All assessors were observed and certification was rechecked at least once by a gold standard trainer while conducting actual pilot assessments.

Assessors were assigned to programs in teams of two to four and spent up to one week at each program conducting the assessments. All child assessment instruments were administered through a computer-assisted personal interview (CAPI) system that provided item wording and assessor instructions, recorded responses entered by assessors, and calculated basals and ceilings. Children were routed through the assessment in either an English- or Spanish-language path based on a routing algorithm developed in collaboration with First 5 LA and LAUP staff. The routing algorithm used data from the parent provided on the consent form as well as the child's performance on the English screener (Simon Says and Art Show of the PreLAS; Duncan and DeAvila 1998) and the initial receptive vocabulary measure (described in greater detail below). We used "conceptual scoring" on the majority of the measures.⁴ This procedure allowed Spanish-speaking children to have the opportunity of hearing the questions in English when needed, and they could respond in either English or Spanish. Assessors recorded the language in which the child responded. Children who did not speak or understand English or Spanish at a threshold level were routed out of the assessment.

Two teachers from each classroom participated in the pilot study. Teachers received paper versions of the Teacher Background Questionnaire, TCR, and DRDP-R. Teachers

³ Juárez and Associates recruited parents of children who are part of their focus group database as well as from other sources. The recruited parents' children were not from LAUP programs, but were preschool age.

⁴ Conceptual scoring allows credit for correct answers given in different dialects of Spanish, as well as in English. Prompts to the child can be given in either language. The use of conceptual scoring on the majority of the measures was adapted for this study. Only one measure, the Expressive One-Word Picture Vocabulary Test - Spanish Bilingual Edition (EOWPVT-SBE; Brownell, 2000), was standardized by the publisher with the use of conceptual scoring. In addition to the EOWPVT, we applied a conceptual scoring procedure to the ECLS-B Mathematics assessment and a letter naming task.

were randomly given one of two forms of the TCR, each version containing different social-emotional scales; thus, each teacher completed forms for half the children on version A and half on version B (194 teachers completed Form A and 189 completed Form B). This allowed examination of the appropriateness of a greater number of social-emotional measures.

C . S A M P L E

1. PROGRAMS

We recruited 14 programs with the assistance of LAUP coaches and staff. All programs were Early Launch, non-PoP programs (Table 1). Three were located in defined areas of greatest need (AGN) within Los Angeles County.⁵

Table 1. Characteristics of Programs in Phase 1 Sample

Characteristic	Number	Percentage
Area of Greatest Need (AGN)		
AGN	3	21.4
Non-AGN	11	78.6
Type of Program		
Early Launch, non-POP	14	100.0
Sample Size	14	

2. TEACHERS

Within programs, 34 teachers participated in the study by providing ratings of study children's development, and 26 provided us with basic information about themselves.⁶ The teachers were mainly female, and more than half were of Latino origin (Table 2). Almost 60 percent of teachers were bilingual, mainly speaking Spanish and English, although 4 percent spoke another language in addition to English. More than half of teachers reported reading to children in both English and Spanish, and 46 percent made instructional presentations in class in both languages.

⁵ Area of greatest need (AGN) is defined by need for preschool spaces. Thirty-four zip codes in Los Angeles County are classified as AGN.

⁶ Both lead and assistant teachers participated in the pilot study by providing ratings of children's development on the TCR and by completing DRDP-R forms (n=34). Any teacher who completed these ratings was also asked to complete the brief teacher questionnaire (n=26). Differences in sample sizes are due to non-response.

Table 2. Teacher Demographic Characteristics and Language Usage

Characteristic	Number	Percentage
Gender		
Female	25	96.2
Male	1	3.8
Race/Ethnicity ^a		
Latino		
Mexican American, Chicano	9	34.6
Other Hispanic/Spanish/Latino	6	23.1
Black, non-Latino	8	30.8
White, non-Latino	2	7.7
Asian, non-Latino	1	3.9
Language Spoken at Home ^a		
English	10	40.0
Spanish	1	4.0
English and Spanish	13	52.0
English and other language	1	4.0
Language Used to Read to Children		
English	11	42.3
English and Spanish	15	57.7
Language Used for Presentations ^a		
English	14	53.9
English and Spanish	12	46.2
Sample Size	26	

^aCategories are mutually exclusive.

Of the 26 teachers who provided information about themselves, all had substantial experience working with children, nearly 15 years on average (median 12.5 years; Table 3). Most of that time was spent in preschool settings, about 10 years on average, with a median of almost 8 years. Many of the teachers held associate's degrees (42 percent), and 31 percent held bachelor's degrees (Table 4). Fifteen percent had some college experience and about 8 percent had some post-baccalaureate graduate or professional school. Few (less than 4 percent) had only a high school diploma or equivalent. Among teachers with college degrees, more than 60 percent had majored in early childhood education or development. Nearly all teachers reported having taken six or more college classes in early childhood development, whether they held a degree or not.

Table 3. Teachers' Experience

	Mean (SD)	Minimum	Maximum	Median
Years Working with Children	14.5 (7.7)	5	29	12.5
Years Teaching Preschool	10.2 (7.2)	1	29	7.5
Sample Size	26			

Table 4. Teacher Education and Credentials

Education and Credentials	Percentage
Highest Grade of School Completed	
High school diploma/equivalent	3.9
Some college but no degree	15.4
Associate's degree	42.3
Bachelor's degree	30.8
Graduate or professional school, no degree	7.7
Field in Which Obtained Highest Degree ^a	
Child Development/Developmental Psychology	30.8
Early Childhood Education	30.8
Other ^b	15.4
College Courses Included 6 or More Classes in EC or CD	92.3
Has a CDA Credential	46.2
Has a State-Awarded Preschool Certificate	73.1
Has a Teaching Certificate or License ^c	57.7
Sample Size	26

^aIncludes only those teachers who have an Associate's degree or higher (n=21).

^bTwo teachers majored in liberal studies, one in sociology and one in human services.

^cSix teachers did not respond to this item, five without a college degree.

3. CHILDREN

The team obtained written consent from parents and completed assessments with 418 children. Our objective was to represent the range of racial/ethnic and language groups to facilitate analysis of item functioning as well as overall reliability and validity of measures. The group was about evenly split between boys and girls (Table 5). In keeping with the

diversity of LAUP enrollment, children represented a variety of combinations of racial/ethnic heritage and backgrounds. Most children were of Latino descent, primarily from Mexico, followed by Black, White, Asian, and other groups. Because we did not conduct parent interviews, we relied on teacher reports of the child’s racial and ethnic background. Teachers were allowed to choose more than one racial or ethnic category, and many of the children in the study were from more than one background. If the teacher identified the child as being part Latino we classified the child as Latino for analysis. Sixty percent of children were Latino, 24 percent Black non-Latino, 9 percent White non-Latino, and 5 percent Asian non-Latino. English and Spanish languages predominated, with 51 percent of children speaking English only, 8 percent Spanish only, 20 percent English and Spanish about equally, 16 percent English and some Spanish, and 3 percent a combination of English and another language.⁷ Less than 2 percent spoke other languages, including Armenian and Farsi. Children’s ages averaged 59.5 months, or 5 years, which was expected given that the assessments were held at the end of their preschool year (not shown in table).

Table 5. Child Demographic Characteristics

Characteristic	Number	Percentage
Gender		
Female	208	49.8
Male	210	50.2
Race/Ethnicity^a		
Latino (any race)	247	60.4
Black, non-Latino	99	24.2
White, non-Latino	35	8.6
Asian, non-Latino	20	4.9
Other, non-Latino	8	2.0
Language		
English only	215	51.4
Spanish only	34	8.1
Spanish primarily	84	20.1
English primarily (Spanish home language)	67	16.0
English primarily (other home language)	11	2.6
Other only or primarily	7	1.7
Sample Size	418	

^aCategories are mutually exclusive.

⁷ We determined children’s language using four questions asked of parents at the time they gave consent. Parents reported their home language, the language they use in speaking to the child, the language the child uses with the parent, and the language the child uses with other children. Parents responded on a 5-point scale ranging from “only English” to “mostly English but sometimes my home language,” “both equally,” “mostly my primary language but some English,” to “only my primary language (not English).” Based on these responses, the child was assigned to one of the six groups described above. We used these categories to route children through the assessment in the appropriate language.

D. MEASURES AND ADMINISTRATION PROCEDURES

As described in detail in Atkins-Burnett, Caspe et al. (2007), we undertook a comprehensive review of measures of children's development that tap five domains (cognition and general knowledge, language development, approaches to learning, social-emotional development, and physical and motor development). We examined each measure for its evidence of prior use with diverse populations (including various linguistic, racial/ethnic, socioeconomic, and ability groups), evidence of reliability and validity, sensitivity to interventions, age-appropriateness, and availability in both Spanish and English. The majority of the assessments that we used were scored conceptually; that is, the child could provide the correct answer in either English or Spanish and be scored correct. Table 6 lists the measures that were the sources of our data and the developmental domain assessed by each instrument.

We collected data in three ways: from direct assessments of children, from teachers' self reports and ratings of children, and from parents' reports. Trained assessors completed child assessments using CAPI. All children were administered language screening, and language and executive functioning assessments. Children in each class were then randomly assigned to receive either mathematics or literacy assessments (rapid letter naming and phonological awareness). Teachers completed paper questionnaires, answering questions about their own background and experience and about children's social-emotional development. Teachers also completed the DRDP-R to rate children's skills, knowledge, and behaviors in language and literacy, mathematics, cognition/approaches to learning, self-regulation, social-emotional development, and physical development/health and safety. We asked both classroom teachers to complete the DRDP-R for each sample child to enable us to examine interrater reliability). In addition, all teachers were asked to complete the positive social scales from the Preschool Kindergarten Behavior Scales-2 (PKBS-2; Merrell 2002), problem behavior items from the Social Skills Rating System (SSRS; Gresham and Elliott 1990), and the Approaches to Learning Scale from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-1999 (ECLS-K) study. Half of the teachers also rated positive and problem behaviors using the Strengths and Difficulties Questionnaire (SDQ; Goodman 2001) and the other half used the Devereux Early Childhood Assessment (DECA; LeBuffe and Naglieri 1999). Thus, not all measures were administered to every child, nor does every child have a score on each of the teacher-reported social-emotional measures. Parents' reports were obtained as part of focus groups and cognitive interviews to better understand the meaning that parents inferred from each item (see the appendix for a summary of

findings from the parent focus groups and cognitive interviews). Here we report the results of the direct child assessments and the teacher reports.

Table 6. Phase 1 Study Measures, by Data Source and Developmental Domain

Data Source	
Direct Child Assessment	
Language	<ul style="list-style-type: none"> PreLAS English (Simon Says, Art Show) PreLAS Spanish (Tío Simón Dice, Exposición de Arte) Peabody Picture Vocabulary Test-4 (PPVT-4) Test de Vocabulario en Imágenes Peabody (TVIP) Expressive One Word Picture Vocabulary – English and Spanish Bilingual Edition (EOWPVT -SBE)
Literacy	<ul style="list-style-type: none"> DELSS Phonological Awareness – English DELSS Phonological Awareness – Spanish Letter Naming
Mathematics	<ul style="list-style-type: none"> Woodcock-Johnson Applied Problems Woodcock-Muñoz Applied Problems ECLS-B Mathematics
Executive Functioning and Motor Development	<ul style="list-style-type: none"> Pencil Tapping Draw a Circle (includes observer ratings)^a Walk a Line Slowly (includes observer ratings)^a
Approaches to Learning	<ul style="list-style-type: none"> Leiter Examiner Ratings-Cognitive Social
Teacher Report	
Approaches to Learning	<ul style="list-style-type: none"> ECLS-K Approaches to Learning Scale
Social-Emotional	<ul style="list-style-type: none"> Strengths and Difficulties Questionnaire (SDQ) Devereaux Early Childhood Assessment (DECA) Preschool Kindergarten Behavior Scale (PKBS)
Desired Results Developmental Profile-Revised (DRDP-R)	<ul style="list-style-type: none"> Language and Literacy Mathematics Cognition and Approaches to Learning Self and Social Regulation, Safety and Health Motor
Parent Report^b	
Social-Emotional	<ul style="list-style-type: none"> Strengths and Difficulties Questionnaire (SDQ) Devereaux Early Childhood Assessment (DECA) Preschool Kindergarten Behavior Scale (PKBS)

^aAssessors also coded information about the child's balance and motor control during these tasks to provide a direct measure of balance and visual-motor coordination.

^bThese instruments were administered to parents in focus groups. See Appendix A for more details about the methods and results of the focus groups and cognitive interviews with parents.

INFORMATION NOT RELEASABLE TO THE PUBLIC UNLESS AUTHORIZED: This information has not been publicly disclosed and is privileged and confidential. It is for internal First 5 LA use only and must not be disseminated, distributed, or copied to persons not authorized to receive the information.

E. ANALYTIC APPROACH

Given the diverse population in Los Angeles County, First 5 LA wanted this pilot study to inform the final decision making regarding the appropriateness of the measures for the range of children enrolled in LAUP programs. We examined the reliability, means and standard deviations, and differential item functioning (DIF) of items for the measures that were administered to all or most of the children. DIF is a statistical analysis to determine whether, for children of the same ability or achievement level, the difficulty of an item is the same for children from different subgroups. In this case, we examined differential item functioning by language group, comparing children who are primarily English speakers with children who are primarily Spanish speakers.⁸

In addition to examining the standard scores from the respective publisher's normative data, we also estimated scores based on the local sample. We applied Item Response Theory (IRT, a one-parameter Rasch model) to the analysis of tests administered to both Spanish and English speakers: the Peabody Picture Vocabulary Test (PPVT), Expressive One Word Vocabulary Test (EOWPVT), the Woodcock-Johnson-III Applied Problems, the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B) Mathematics assessments, and the Rapid Letter-Naming task. In addition, we applied the Rasch rating scale model to the measures of social-emotional competence. The Rasch analysis of the social-emotional measures allowed us to examine the face validity of the measures (that is, whether the ordering of item difficulty matches the theoretical construct) as well as the item characteristics (item-to-total correlation, item fit, and presence of multiple dimensions). Finally, we computed measures of internal consistency (coefficient alpha and IRT model reliability) and examined the convergent/divergent validity of the measures for the local Los Angeles County sample included in this pilot study.

⁸ A DIF analysis of measures by race/ethnicity was not conducted because many of the children in the study were reported by teachers as being from more than one racial/ethnic group.

F. DESCRIPTIVE INFORMATION FOR SPRING MEASURES

We examined scores for each of the measures calculated according to instructions in each of the publisher's user's manual for computing standard scores. As expected because of the target population of the programs, children from this selective sample of LAUP programs scored below national norms on most measures of cognitive and language development; LAUP scores were similar to children from low-income families (such as those enrolled in Head Start; Zill, Sorongon, Kim, Clark, and Woolverton 2006).

Table 7 presents scores on each cognitive measure for the entire Phase 1 sample. Variations in sample sizes are the result of missing data, language routing rules, and matrix sampling of some measures. For example, only half the sample was assessed in literacy and the other half in mathematics; among those assessed in literacy, half the children (approximately 25 percent of the total sample) received one version of the letter naming task, and the remainder were assessed with an alternate version (another 25 percent of the sample).

Internal consistency reliability (coefficient alpha) was quite high for all measures with the exception of the phonological awareness measure in Spanish, indicating strong correlations among items within each measure. Similar information is presented for the social-emotional and approaches to learning total scores in Table 8 and for motor scores in Table 9. Variations in samples sizes are due to assignment to the different child-report forms.

Table 7. Cognitive and Language Raw and Standard Scores for Entire Pilot Sample

Developmental Domain	N	Mean	Standard Deviation	Minimum Score	Maximum Score	Alpha
Language						
Expressive One Word Picture Vocabulary Test-SBE Standard Score	387	84.1	14.6	56	144	0.95
PPVT-4 Standard Score	339	84.6	17.5	20	122	0.98
PPVT-4 W Score	339	110.9	19.5	26	150	0.98
TVIP-R Standard Score	103	77.0	16.5	0	110	0.97
Literacy						
DELSS Phonological Awareness Raw Score—English	109	6.3	4.1	1	15	0.90 ^a
DELSS Phonological Awareness Raw Score—Spanish	50	3.6	2.3	0	10	0.53 ^a
PreLAS English ^b	384	16.6	3.9	0	20	0.95
PreLAS Spanish ^b	118	11.7	4.3	0	20	0.97
Letter Naming Raw Score (Form 1) ^{b,c}	99	15.8	10.2	0	30	0.97
Letter Naming Raw Score (Form 2) ^{b,c}	95	17.8	10.2	0	30	0.97
DRDP-R Scale Score - Language and Literacy ^d	207	555.6	80.8	334.5	739.9	0.92
Mathematics:						
Numeracy/General Knowledge						
Woodcock Johnson III Applied Problems (AP) Standard Score	177	92.9	13.7	50	139	0.84
Woodcock-Johnson III AP W Score	177	402.3	19.5	332	458	0.84
Woodcock-Muñoz-III AP Standard Score	38	80.3	13.1	46	102	0.89
Woodcock-Muñoz-III AP W Score	39	382.5	22.6	318	411	0.89
ECLS-B Mathematics	222	500.0	52.0	373	687	0.80
DRDP-R Scale Score—Mathematics ^d	207	557.2	83.1	327.2	746.6	0.88

^aReliability estimate is based on IRT analyses due to amount of missing data.

^bThese scores are raw counts of children's correct responses on Simon Says (Tío Simón Dice) and Art Show (Exposición de Arte).

^cMeans are average number of letters named correctly.

^dDRDP-R subscales were rescaled to have a mean of 500.

Table 8. Social-Emotional Development, Raw and Standard Scores for Entire Pilot Sample

Domain	N	Mean	Standard Deviation
Executive Functioning			
Pencil Tapping ^a	416	65.5%	0.3
Draw a Circle ^b	416	17.8 sec.	15.6
Walk a Line Slowly ^c	416	1.1 sec.	0.4
Approaches to Learning			
ECLS-K Approaches to Learning Scale Raw Score (Form A)	206	2.5	0.5
ECLS-K Approaches to Learning Scale Raw Score (Form B)	196	2.5	0.5
Leiter Examiner Ratings Cognitive Social Standard Score	417	95.6	16.6
Strengths and Difficulties (SDQ) Scale Raw Scores	191	2.2	2.6
Conduct Problems	191	1.8	1.8
Prosocial	191	7.0	2.6
Emotional Symptoms	191	1.5	1.7
Hyperactivity and Attention	191	3.9	2.9
Devereaux Early Childhood Assessment (DECA) T Scores			
Initiative	178	53.4	11.8
Self-Control	178	55.1	11.9
Attachment	178	38.6	6.6
Behavioral Concerns	178	53.1	8.8
Total Protective Factors	178	48.8	11.1
PKBS Standard Scores			
Social Cooperation	190	71.4	16.1
Social Interaction	179	93.1	16.1
Social Independence	191	67.7	14.0
DRDP-R Scale Scores ^c			
Approaches to Learning and Cognition	207	556.5	81.7
Regulation and Safety	207	556.6	76.6
Self and Social	207	555.6	78.9

^aPencil tapping is percentage of the time the child responded correctly.

^bMean scores for Draw a Circle and Walk a Line tasks are in seconds.

^cDRDP-R subscales were rescaled to have a mean of 500.

Table 9. Physical/Motor Development Scores for Entire Pilot Sample

Domain	N	Mean	Standard Deviation
Draw a Circle—did child stay within lines?	416	68.4%	0.24
Walk a Line Slowly—did child stay on line?	416	93.8%	0.47
DRDP-R Scale Score ^a			
Motor development	207	561.1	83.3

^aDRDP-R subscales were rescaled to have a mean of 500.

1. PSYCHOMETRIC EVIDENCE FOR UTILITY OF MEASURES

In this section we review the results of psychometric analyses measure by measure. We present the measures' reliability and validity, and the equivalence of the measures for different groups, including examination of item functioning. We first describe cognitive and general knowledge measures, followed by measures of language and of social-emotional development and approaches to learning, including executive functioning. We conclude with a discussion of children's physical/motor development.

a. Cognitive and General Knowledge Measures

The measures of cognitive and general knowledge outcomes address letter knowledge, phonological awareness, mathematics and logical reasoning, conceptual understanding, and problem solving.

Rapid Letter Naming Task. In the pilot study we used a criterion reference measure of the number of letters that a child could name quickly and easily. Letter naming (Moats 1998; Snow, Burns, and Griffin 1998) and, in particular, rapid letter naming (O'Connor and Jenkins 1999; Rouse and Fantuzzo 2006; Uhry 2002), is among the strongest predictors of later success in school. We developed two different forms with some overlap of items. Each form had some uppercase and some lowercase items. If an item was uppercase on one form, it was lowercase on the other form. All children were asked about either the uppercase or lowercase form of every letter of the alphabet. Very common letters (such as vowels) were presented as both uppercase and lowercase on both forms to help provide overlap in scaling the measure. Therefore, each form contained 30 letters, allowing for a maximum score of 30 on the letter naming task.

The letter naming task had high reliability ($\alpha = .97$) and good fit to the model. The order of item difficulty of the letters was logical, consistent with other studies, and indicated that children recognized uppercase letters more frequently than lowercase letters (see Figures 1 and 2). Because this task was developed for UPCOS, we do not have standard scores. We compared the raw scores on this measure with the average number of letters preschoolers in Head Start named (Administration for Children and Families 2005). Slightly less than half (46 percent) of the children in Head Start named 17 or more letters (ACF 2005). On Form 1, about half of the children named 19 letters and on Form 2 (with more uppercase letters), 56 percent of the children named 22 letters. In the spring pilot, 4 percent of the children were able to name all 30 letters administered to them (see number of maximum extreme scores in Table 10), and many (about 35 percent) named more than 25 letters. This suggests a potential ceiling problem for this measure if used in spring 2008. We therefore recommended adding some word identification items to the battery for the spring. Only 1 percent of children had a minimum extreme score (that is, they did not name any letters and had a score of 0).

Table 10. Letter Naming Scores and Reliabilities

30 Items Each Form	N	IRT Reliability	Item Reliability	Coefficient Alpha	Minimum Extreme ^a	Maximum Extreme ^b
Letter Naming Form 1	99	0.93	0.95	0.97	0	2
Letter Naming Form 2	97	0.91	0.89	0.97	2	6

^aNumber of children who did not identify any letters correctly, that is, who had a score of 0.

^bNumber of children who identified all of the letters correctly, that is, who named all 30 letters and had the maximum score.

The item reliabilities are based on separate calibrations of the forms. The common items ordered in similar ways across forms with the “O” and “X” as the easiest items and lowercase letters as the most difficult items (see Figures 1 and 2). The “O” did show some misfit. It was the easiest item on both forms, yet some children were not able to name it correctly, often calling it “circle” or “zero.”

The letter naming task is reliable, has strong face validity, and measures an important predictor of literacy achievement in schools.⁹ We recommended that it be included in the fall 2007 battery.

⁹ Letter naming is a stronger predictor when it is a rapid letter naming task, which is how the task was administered in this study (Georgiou, Parrila, and Krrby 2006; O’Connor and Jenkins 1999; Rouse and Fantuzzo 2006; Uhry 2002)

Figure 1. Item-Person Map of Letter Naming Task for Form 1

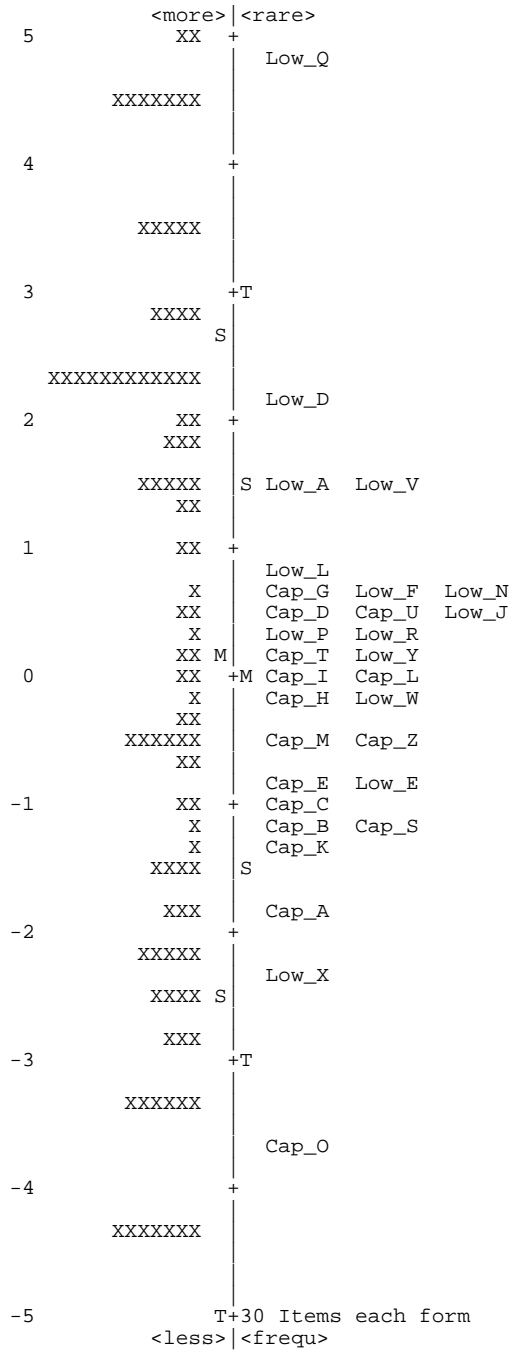
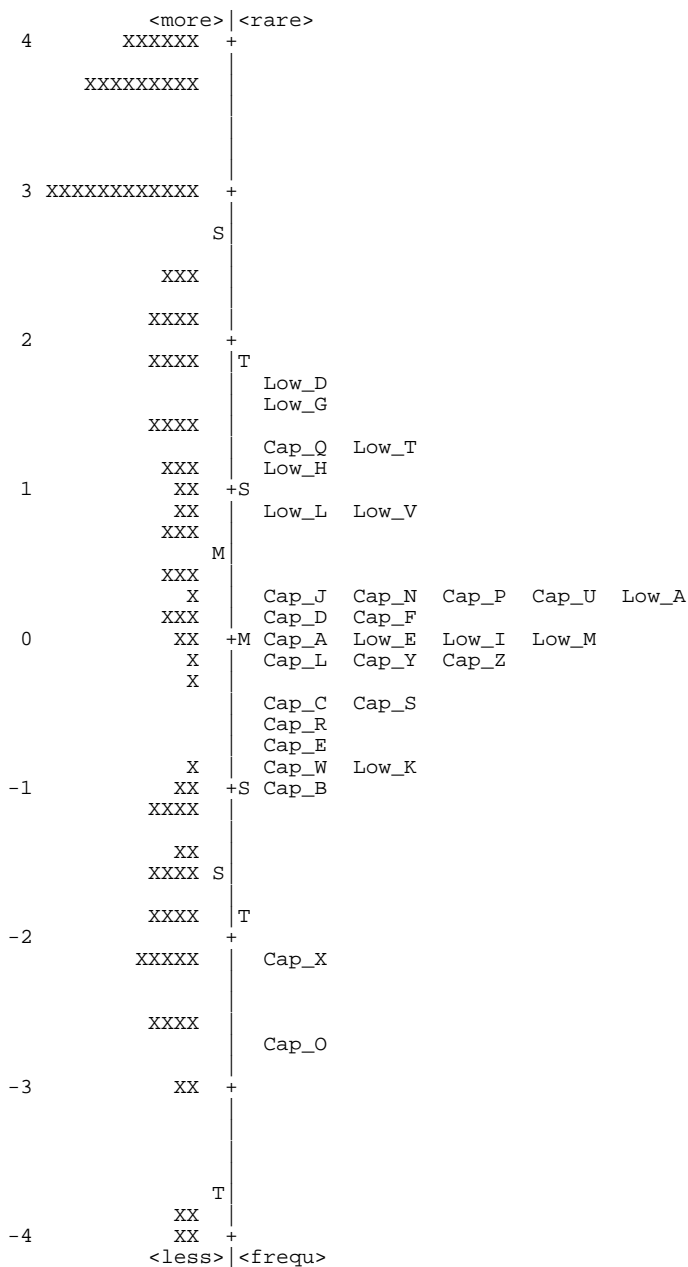


Figure 2. Item-Person Map of Letter Naming Task for Form 2



Development of English Literacy in Spanish-Speaking Children—Phonological Awareness Test I (DELSS Phonological Awareness I; Harvard University and Center for Applied Linguistics 2002). Although this relatively new measure had limited reliability data, it was included for piloting based on the importance of phonological skills and the strength of having been evaluated previously with both English- and Spanish-speaking children. Each version of the measure included 18 items that ask children to identify words that rhyme, produce a rhyme, and identify words that begin with the same sound (alliteration). The rhyme production task is administered only if the child passes the rhyme recognition scale. The child had to respond correctly to four of six rhymes in order to be administered the rhyme production subtask. Children routed into English and Spanish paths received either the English or Spanish version of the DELSS as appropriate. We examined this measure with a Rasch model as well as with classical methods and report results separately for the two versions.

When we administered the rhyme production task, we included a code for rhyming nonsense words. Of the one-half of sampled children who were assigned to the literacy assessment and took the rhyme production task, one-quarter to one-half of the responses on each word were nonsense words that rhymed with the stimulus word. The original DELSS accepted only actual rhyming words as the correct response. However, that requires vocabulary knowledge and more working memory. In our sample, the English version had ceiling problems (13 percent of children scored correct on all items) and the Spanish version had floor problems (8 percent did not get any items correct). The Spanish rhyming items were much more difficult than the English rhymes. Only the most difficult English rhyming items were multisyllabic, whereas all of the Spanish rhyming items were multisyllabic. As a result of low performance on rhyme recognition, children who participated in the Spanish assessment were routed out of rhyme production. In addition, the reliability of the measure with the Spanish sample was poor ($\alpha = .53$), although it was much stronger for the English sample ($\alpha = .90$; see Table 7).¹⁰ The IRT reliability was very poor for the English ($\alpha = .45$) and the Spanish measures ($\alpha = .22$). Given the problems with this measure in the pilot, a decision was made to drop the DELSS from the fall 2007 battery. We recommended including the Spelling/Ortografía subtest of the Woodcock-Johnson/Woodcock–Muñoz III (WJ-III; Woodcock, McGrew, and Mather 2001). This test provides information about children’s ability to encode sounds and words in written form. It is used in the FACES 2006 and provides standard scores based on a nationally representative sample.

Woodcock-Johnson Psycho-Educational Battery (and Bateria III Woodcock-Muñoz) Third Edition (WJ-III; Woodcock, McGrew, and Mather 2001). We administered the Applied Problems test from the English and Spanish versions of this battery. This subtest measures a child’s skill in analyzing and solving practical problems in mathematics. The reliability of the scales in English and Spanish was acceptable ($\alpha = .84$ and $.89$, respectively). The bivariate correlation of the WJ-III with the receptive language measure (PPVT-4; $r = .64$) was as strong as the correlation with the ECLS Mathematics measure ($r =$

¹⁰ Due to the amount of missing data on the DELSS, this reliability estimate is based on IRT analyses.

.63), indicating a strong influence of language on the WJ-III Applied Problems. It also showed a slightly stronger relationship to measures of social skills than the ECLS Mathematics measure. Only one item showed differential item functioning. The Spanish version of the WJ-III Applied Problems had floor problems (10 percent of children got one or no items correct), and a smaller percentage (4 percent) of children scored at the floor on the English version. To manage burden on children, we will include only one measure of mathematics in the fall 2007 battery and so we recommended dropping the WJ-III Applied Problems.

Early Childhood Longitudinal Study Mathematics. We supplemented the WJ-III Applied Problems with items from the ECLS-B and ECLS-K Mathematics assessments that tap additional areas of mathematics skills, such as spatial abilities and measurement. Items in Spanish and English were identical, and children could respond in either language.¹¹ These items were developed for this age group and target the range of mathematical concepts typically developing among 4-year-old children. The reliability of the scale was adequate (Rasch scale reliability = .80) and no differential item functioning was identified. No floor or ceiling effects were noted. We recommended including the ECLS Mathematics in the fall 2007 battery.

Desired Results Developmental Profile-Revised (DRDP-R). The DRDP-R is implemented statewide as the required assessment tool for all care programs funded by the California Department of Education, Child Development Division to help guide and individualize instructional practices. This teacher-report observational assessment is designed to measure children's development in all areas of child outcomes: social skills, language, literacy, mathematics, self-regulation, approaches to learning, cognitive and general knowledge, health, and physical development. All the scales within the DRDP-R showed strong reliability ($\alpha = .89$ to $.96$). A multidimensional Rasch model was used to estimate scores, with the scale projected on a common axis for the sake of comparability of dimensions. This method of calibration combined with the shared method variance and an adjustment of correlations for the measurement error increased the inter-factor correlations ($r = .82 - .99$). Separate calibrations of the subscales indicated lower correlations, although there appears to be a strong higher order factor (see memorandum by Atkins-Burnett, Vogel et al. 2007).

The other dimensions on the DRDP-R are also moderately correlated with direct measures of vocabulary (Table 11).¹² This suggests that the teachers are strongly influenced by children's verbal skills when rating them on all the domains, as much of what teachers know about children's skills and knowledge is based on what children say. Inter-rater reliability was relatively weak for these scales, ranging from $r = .58$ to $r = .66$. We

¹¹ The use of conceptual scoring on the ECLS-B Mathematics assessment was adapted for this study so that regardless of whether the child was routed into English or Spanish, he or she could answer in Spanish, as well as in English. Assessors could prompt the child in either language.

¹² Table 11 displays all DRDP-R items, although some are discussed in later sections related to the domains assessed by those items.

recommended that before administering the DRDP-R in the fall teachers receive additional training in how to observe and rate children's skills, knowledge, and behavior.

Table 11 . Bivariate Correlations of DRDP-R Measures with Cognitive Measures

	EOWPVT	PPVT-4	WJ-III	ECLS
EOWPVT Theta	1			
PPVT-4 Theta	0.756	1		
WJ III Applied Problems Theta	0.543	0.637	1	
ECLS Mathematics Theta	0.503	0.537	0.632	1
DRDP-R				
Self and Social	0.409	0.406	0.271	0.253
Regulation, Safety and Health	0.325	0.327	0.236	0.300
Language and Literacy	0.505	0.480	0.385	0.379
Cognition and Approaches to Learning	0.428	0.414	0.278	0.288
Mathematics	0.488	0.469	0.395	0.410
Motor	0.388	0.384	0.294	0.332

b. Language Measures

Language Screener and Language of Administration. The language screener used the PreLAS Simon Says and Art Show, as well as parent reports on the language(s) a child speaks at home and to other children, to determine whether a child should receive the English or the Spanish version of the child assessment battery or not be tested at all. Figure 3 illustrates the flow of the screening and how it informed the language of administration. The reliability of the PreLAS scale was very strong for both the English ($\alpha = .95$) and Spanish versions ($\alpha = .97$). We further examined the optimal cut-score for the PreLAS using receiver-operating curves (ROC). The ROC helped to establish a new decision rule for the Phase 2 study (Figure 4).¹³ The empirically derived stop rule for the PreLAS is such that children who make six or more errors across both components of the two English-language PreLAS subtests, and are from Spanish-speaking households, are routed into the Spanish assessment. This will serve as an initial language screener for children in the main study, with additional procedures in other components of the assessment that allow children to continue in Spanish or to be re-routed back into English if necessary. See Figure 5 for the recommended routing procedures.

¹³ The goal of the ROC analyses is to determine an optimal cut point that maximizes sensitivity while also maximizing specificity, or to have both values equal to or approaching 1. Sensitivity in the current context refers to correctly identifying children who should be assessed in Spanish rather than English. High specificity means that we are not over-identifying children as needing to be tested in Spanish. Using the English PPVT standard scores more than 1.5 of a standard deviation below the mean as a cut point, the best-fitting model captured and correctly routed 89 percent of the children.

English Language Instructions

Spanish Language Instructions

*Conceptual Scoring

Parent Report of Child's Primary Language Using 3 items

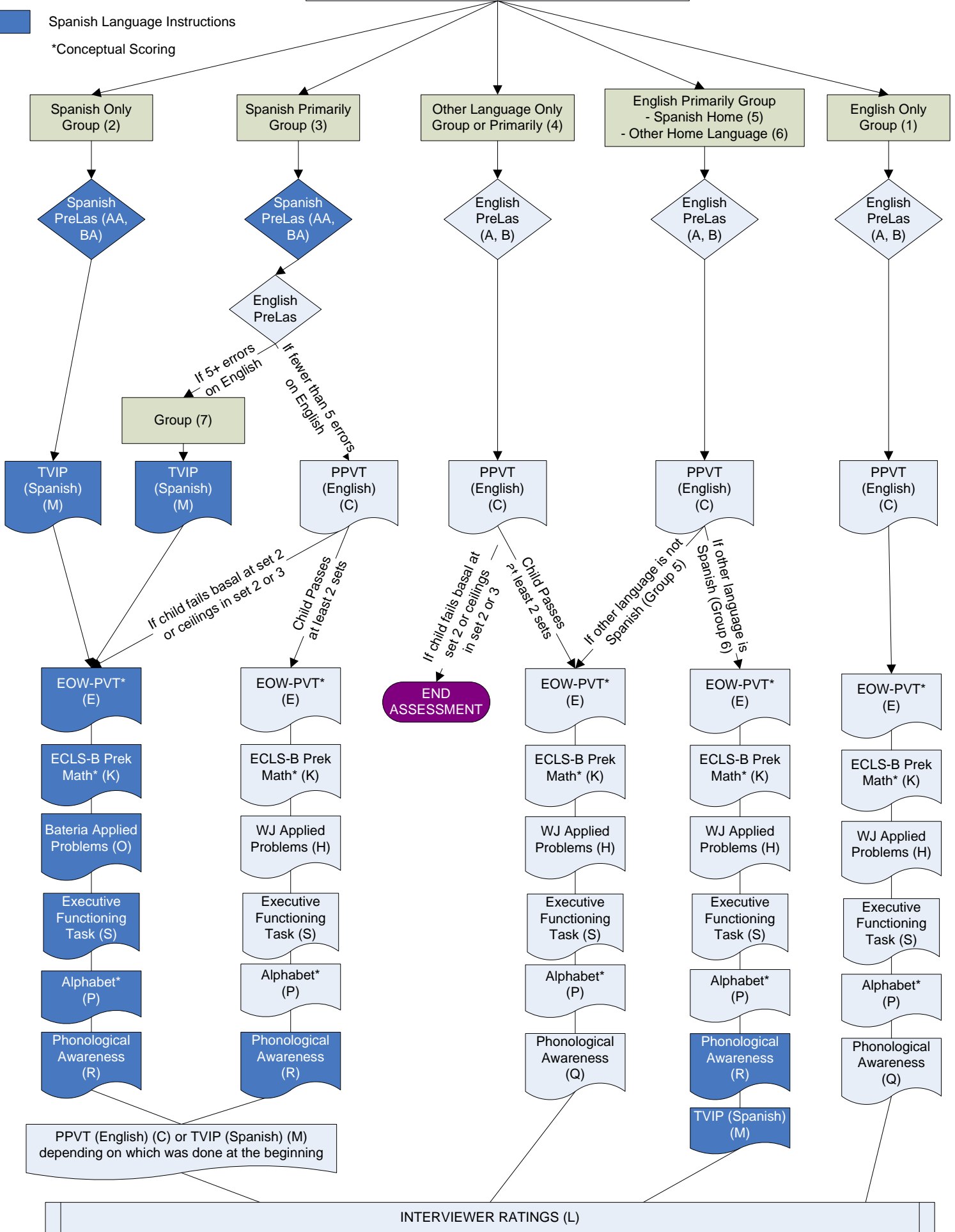
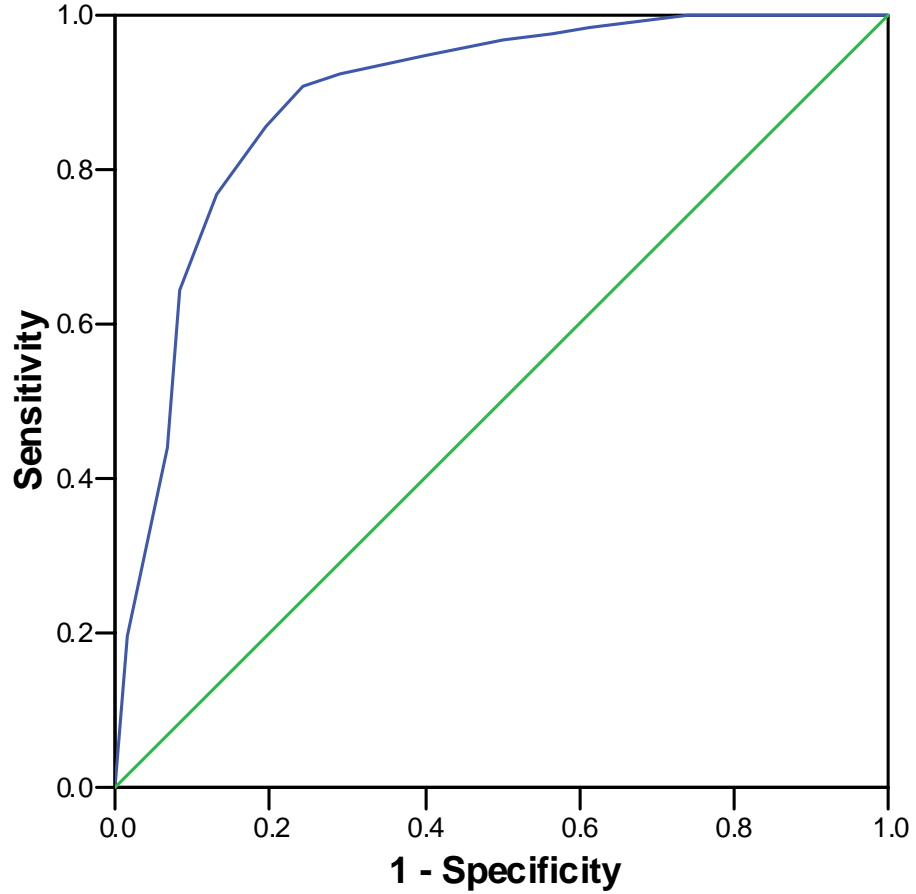
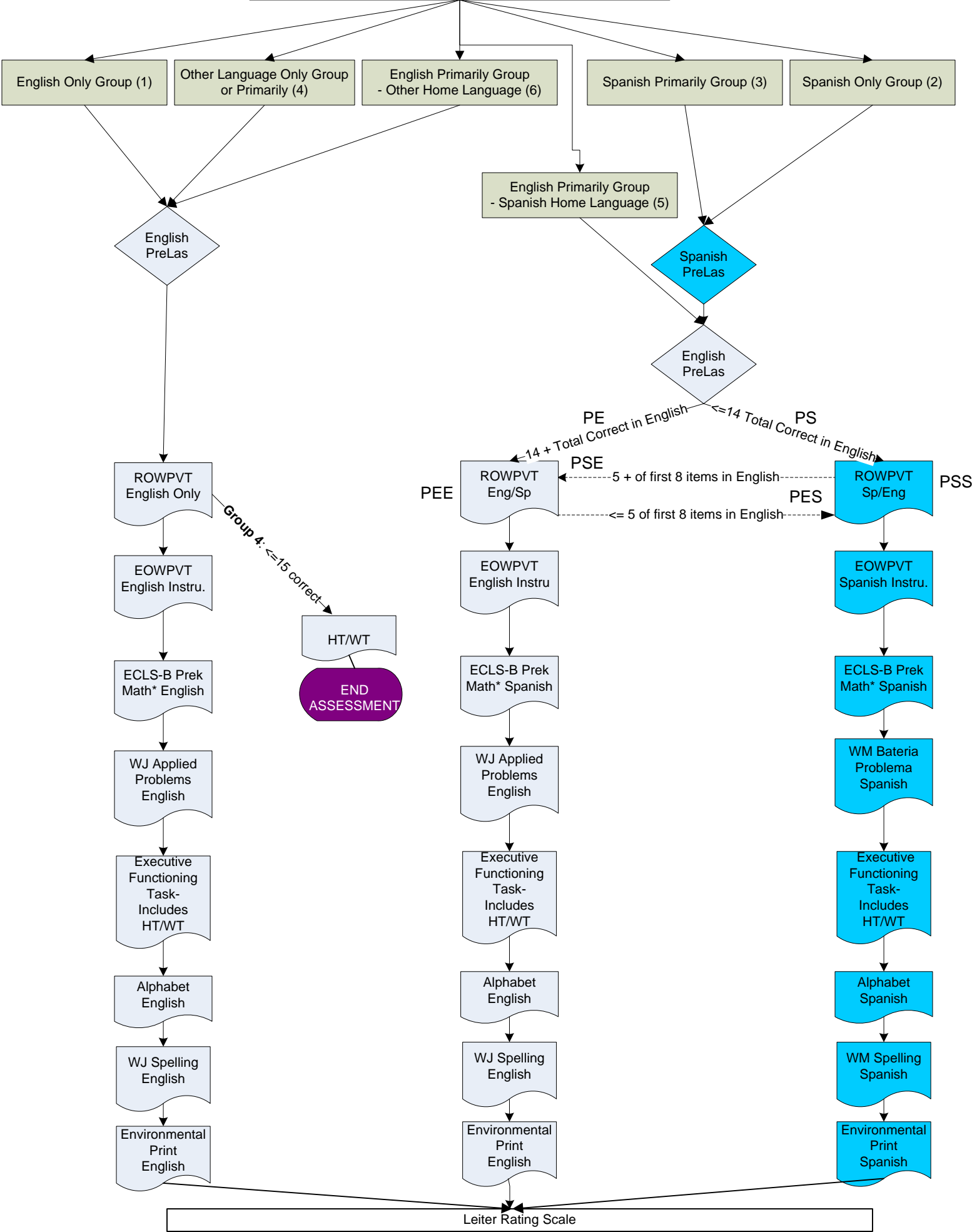


Figure 4. ROC Curve of Prelas Scores Using PPVT 1.5 SD Cut-Off Score



Diagonal segments are produced by ties.



Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4). The PPVT-4 (Dunn and Dunn 2006) is the measure of receptive vocabulary piloted in the spring with children who passed the English screener. The Spanish-language version of the PPVT, the Test de Vocabulario en Imagenes Peabody (TVIP; Dunn et al. 1986), was used with children whose primary language was Spanish. The TVIP was also administered to Spanish-speaking children who were able to complete the assessments in English in order to describe their overall receptive vocabulary proficiency and examine potential cross-language transfer of language abilities. Scale reliability was very high for the PPVT ($\alpha = .98$) and the TVIP ($\alpha = .97$). However, the analysis of the PPVT-4 with our sample identified 7 misfitting items, and 10 items with differential item functioning (7 favoring English speakers and 3 favoring Spanish speakers). The distribution of the items with DIF was not balanced (items favoring and not favoring Spanish speakers were at different difficulty levels). In our sample, the TVIP had floor problems (9 percent of children got one or no items correct). No floor effects were noted for the PPVT. The potential problems with DIF and misfit in the PPVT combined with the floor problems in the TVIP led to a search for another measure of receptive language to use in the Phase 2 study. As seen in Figure 5, the Receptive One-Word Picture Vocabulary Test: English and Spanish-Bilingual Edition (ROWPVT; ROWPVT-SBE; Brownell 2001¹⁴) will replace the PPVT and TVIP, respectively, in the fall. Both measures assess the child’s receptive vocabulary, with the Spanish-Bilingual Edition for use with individuals who are bilingual in English and Spanish. The ROWPVT-SBE and the EOWPVT-SBE were standardized on the same sample and so standard scores on expressive and receptive vocabulary can be easily compared.

Expressive One-Word Picture Vocabulary Test: English and Spanish-Bilingual Edition (EOWPVT; EOWPVT-SBE; Brownell 2000). The EOWPVT-SBE was administered to all children and scored conceptually according to the standardized scoring. The reliability was very high ($\alpha = .95$). The analysis identified seven items with differential item functioning (three items that favored English speakers, and four that favored Spanish speakers); however, the items were balanced. That is, for a set of items at a given difficulty level, one might favor English speakers and one would favor Spanish speakers. Examination of the items suggested that some of these may have been identified by chance (with so many comparisons and a smaller sample of Spanish speakers, this is likely). No floor effects were noted for either subgroup. We recommended retaining this measure for fall 2007.

Desired Results Developmental Profile-Revised. The language and literacy scale of the DRDP-R includes ratings of children’s ability to comprehend language, follow complex directions, express themselves through language, demonstrate phonological awareness and emergent writing, understand concepts of print, and recognize sight vocabulary. The scale had high internal consistency reliability, with $\alpha = .92$ (Table 7). We recommended inclusion of the DRDP-R for fall 2007.

¹⁴ The ROWPVT and the ROWPVT-SBE are the same measure. The ROWPVT-SBE excludes items that had differential item functioning from the standard scoring. Twenty out of 170 items are not administered in the ROWPVT-SBE edition.

c. Social-Emotional and Approaches to Learning Measures

Ratings of Social-Emotional Development. Teachers rated children's social-emotional competence using the positive social scales from the Preschool Kindergarten Behavior Scales-2 (PKBS-2; Merrell 2002) and rated problem behaviors with the Social Skills Rating System (SSRS; Gresham and Elliott 1990). Half of the teachers also rated positive and problem behaviors using the Strengths and Difficulties Questionnaire (SDQ; Goodman 2001) and the other half used the Devereux Early Childhood Assessment (DECA; LeBuffe and Naglieri 1999). Except for the SDQ Problem behaviors scale, all of these scales exhibited good reliability (alpha ranged from $\alpha = .84$ to $.96$). Problem behaviors on all scales had the lowest person reliability (ranging from $\alpha = .61$ to $.90$). The DECA Social Skills scale had the strongest reliability ($\alpha = .96$).

Teachers were also asked how appropriate these measures were for the children in their classroom and which measure they thought was most appropriate. The teachers were positive about the PKBS-2 social scales, and some teachers commented that it captured more typical behaviors and was more closely aligned with common preschool behaviors. Teachers were concerned about the high number of problem behaviors included. We recommended using the PKBS-2 social scales in the fall 2007 battery and replacing the lengthy problem behavior scale with a shorter problem behavior scale.

Standard scores on some of the social-emotional and approaches to learning measures indicated some areas of risk or potential need for children in the social-emotional area (see Atkins-Burnett, Aikens et al. 2007). As is usually found with measures of social-emotional functioning, the distribution of the data was skewed (see Table 12). For the DECA social skills subscales, at least 5 percent of children were at the maximum score. There was a strong floor effect for the SDQ problem behaviors subscales, with 45 percent of children at minimum for emotional problems, 42 percent at minimum for conduct problems, 18 percent at minimum for hyperactivity and attention, and 30 percent at minimum for peer problems. Conversely, 23 percent of children were at maximum for prosocial, suggesting a strong ceiling effect. Compared to boys, teachers were more likely to rate girls at the minimum for the conduct problems and hyperactivity and attention subscales. They were also more likely to rate girls at the maximum on the prosocial subscale and children in non-AGN programs at the minimum on the peer problems subscale. No consistent floor or ceiling effects for the other subgroups were evident.

Table 12. Teacher Reported Social-Emotional Measures Reliability

	N	Person Reliability	Item Reliability	Alpha	Percent at Ceiling	Percent at Floor	Number of Items
Teacher Report Form A							
Strengths and Difficulties Questionnaire							
Problem behaviors	206	0.61	0.95	0.79	0	16	14
Social skills	206	0.82	0.97	0.88	29	0	10
Preschool Kindergarten Behavior Scales							
Problem behaviors	206	0.85	0.95	0.91	0	9	19
Social skills	206	0.88	0.93	0.94	12	0	24
ECLS-K Approaches to Learning Scale	206	0.80	0.93	0.89	44	2	7
Teacher Report Form B							
Devereaux Early Childhood Assessment							
Problem behaviors	197	0.77	0.98	0.84	12	1	10
Social skills	197	0.93	0.96	0.96	6	0	27
Preschool Kindergarten Behavior Scales							
Problem behaviors	197	0.90	0.94	0.94	0	10	23
Social skills	197	0.89	0.97	0.91	7	0	13
ECLS-K Approaches to Learning Scale	196	0.81	0.91	0.91	52	3	6

Note: NA = not applicable.

ECLS-K Approaches to Learning. Teachers were asked to rate each child on the six items that comprise the Approaches to Learning Scale from the Early Childhood Longitudinal Study – Kindergarten Class of 1998-1999 (ECLS-K) study. The items tap the child’s motivation, attention, organization, persistence, and independence in learning. The reliability of the scale in our sample was strong ($\alpha = .91$), although there was a strong ceiling problem (27 percent at maximum score). The bivariate correlations with measures of mathematics ($r = .35 - .37$) and language ($r = .21 - .24$) were not high but statistically significant (Table 13). Despite the ceiling problem, we recommended retaining this measure for providing comparability to a national sample.

Table 13. Bivariate Correlations of Language and Cognitive Measures With Approaches to Learning

	EOWPVT	PPVT	WJ-III	ECLS
EOWPVT Theta	1			
PPVT Theta	0.756	1		
Applied Problems Theta	0.543	0.637	1	
ECLS Mathematics Theta	0.503	0.537	0.632	1
Approaches Form A Theta	0.200	0.180	0.292	0.386
Approaches Form B Theta	0.238	0.211	0.353	0.371

$p \leq .01$ for all coefficients

Leiter Examiner Rating Scales. At the conclusion of the direct assessment, assessors rated children on subtests from the Leiter Examiner Rating Scales (Roid and Miller 1997) to obtain a more comprehensive picture of the child’s attention, activity level, and sociability. In the pilot, we included the items from the Leiter-R subscale for self-regulation. The four subscales comprise the Cognitive-Social Scale. The reliability of the Cognitive-Social Scale in our sample was very strong ($\alpha = .98$). Twenty-eight percent of the sample obtained the highest possible score, indicating a ceiling effect. Assessors reported difficulty rating the items on the self-regulation subscale since it was not aligned with our testing protocol.¹⁵ With such high reliability in the overall Cognitive-Social Scale, eliminating four items should not affect reliability of the scale, although it may increase the ceiling effect. However, given the dubious validity of the ratings of self-regulation, we recommended eliminating that subscale, but retaining the other three subscales.

¹⁵ The Leiter scales for which these ratings were developed include manipulatives that children are to resist touching or playing with during parts of the assessment. Since our assessment did not include these materials, it was not possible to accurately rate how well a child would inhibit the urge to touch manipulatives.

Executive Functioning Tasks: Pencil Tapping, Balance Beam (Walk a Line Slowly), Draw a Circle. Executive functioning (EF) involves “processes that are integral to the emerging self-regulation of behavior and developing social and cognitive competence in young children” (Blair, Zelazo, and Greenberg 2005, p. 561). These important skills are developing during the preschool years and are strong indicators of children’s future success both socially and academically. However, issues related to measurement of EF have constrained research in this area until recently.

These “cool” executive functioning tasks were easy to administer and provide a direct measure of children’s executive functioning and motor skills. Each of the measures assesses inhibitory control; in addition, the pencil tapping task also requires working memory and more attention to auditory and visual stimuli.

Pencil Tapping (Smith-Donald et al. in press; Blair 2002; Diamond and Taylor 1996). The pencil tapping task is an adaptation of a peg-tapping task (Blair 2002; Diamond and Taylor 1996). In our sample, children effectively inhibited their initial impulse and responded correctly 66 percent of the time on the pencil tapping task (that is, followed the rule and tapped once even when the assessor tapped two times). A similar percentage of children responded correctly across trials (59 percent) in a sample of low-income 3- and 4-year-olds in Head Start (Smith-Donald et al. in press). Eighteen percent of the children scored at the ceiling (that is, tapped correctly every time), while 4 percent scored at the floor (that is, never tapped correctly). The reliability of the 16 trials in this task was strong ($\alpha = .93$). We recommended retaining this measure.

Walk-a-Line Slowly (Murray and Kochanska 2002; Smith-Donald et al. in press). This is the large muscle (gross motor) version of the Draw a Circle task. The child is asked to walk on a 6-foot line, and then is asked to walk the line as slowly as he or she can. Each of the child’s attempts is timed. The difference between the first attempt and the slow attempt is the child’s score. When used in a study in Chicago Head Start classrooms (Smith-Donald et al. in press), children took approximately 2 seconds more when asked to walk the line slowly. Children in the pilot took on average 1.1 seconds more. Reliability for this measure was strong ($\alpha = .96$).

Draw a Circle (Kochanska and Murray 2002; Murray and Kochanska 2002). This task involves having the child control fine motor movements. The child is asked to draw a circle between lines with a crayon after the assessor says, “Go!”. The child is then asked to do this as fast as they can and finally as slowly as they can. Each trial is timed. The objective is to see if the child can slow down motor activity. According to the developers’ studies, with typically developing Caucasian children, there was a large increase in control from preschool to kindergarten (with mean difference scores of 11.9 at preschool and 39.6 at kindergarten age), with higher scores indicating greater control (Murray and Kochanska 2002). The mean score for children in our sample was 17.8. Given concern about child burden and the availability of two other measures of executive functioning, we recommended that the Walk-a-Line Slowly and Pencil Tapping tasks be retained and the Draw a Circle task be dropped.

Desired Results Developmental Profile-Revised. The self-regulation scale of the DRDP-R includes ratings of children's ability to control impulses, take turns with peers, and share space and materials with peers. The self and social scale includes ratings of children's self-awareness, positive self-concept, empathy, cooperation, friendships and cooperative play skills, and conflict negotiation. The scale had high internal consistency reliability, with $\alpha = .87$. Means and standard deviations for these scales are shown in Table 8. We recommend collecting further information on the DRDP-R by including in the fall 2007 battery. We plan to collect this information in the PoP sites in the sample.

d. Physical and Motor Development Measures

Draw a Circle and Walk the Line Tasks. The assessment battery included limited direct observation of children's balance and visual motor coordination. Because we were already administering the Draw a Circle and Walk the Line tasks as executive functioning measures, assessors were asked to code additional information about the child's balance and fine motor control during these tasks. The first step in each of these executive functioning tasks provided basic information about the child's motor competence. The assessor rated how closely the child stayed within the lines and how he or she gripped the pencil in the Draw a Circle task, and indicated how many times the child stepped off the line in his or her first attempt to walk the line. On the Walk the Line task, 94 percent of the children were able to stay on the line "almost all of the time" on their first attempt. Sixty-eight percent of the children were able to stay within the lines "almost all of the time" on Draw a Circle. Inter-rater reliability was high for this task ($\alpha = .97$ to 1.00).

Desired Results Developmental Profile-Revised (DRDP-R). This measure, as noted earlier, includes scales that allow the teacher to assess the child's fine and gross motor development. Reliability for this measure was strong ($\alpha = .93$). Table 9 presents the physical/motor development scores for the pilot sample. For fall 2007, DRDP-R data will be collected in PoP program sites.

2. SUMMARY OF SPRING 2007 PILOT TEST RESULTS

The data from the pilot test were analyzed for the appropriateness of the measures' use with children in LAUP programs. Preliminary results of these analyses were discussed with staff of First 5 LA and LAUP and with First 5's RAC to make decisions about the child assessment battery that would be used in Phase 2 (to be conducted in the 2007-2008 program year). Based on these analyses and discussions, some measures were dropped from the assessment battery. For example, for Spanish speakers we found poor reliability and floor problems with the DELSS, while the English sample had ceiling problems. Thus, for Phase 2, we dropped the DELSS and used the time this made available to administer the Spelling/Ortografía subtest of the Woodcock-Johnson/Woodcock-Muñoz III. Similarly, analyses of the PPVT-4 identified misfitting items and items with differential item functioning between English speakers and Spanish speakers. As a result, the PPVT was dropped and replaced by the Receptive One-Word Picture Vocabulary Test (ROWPVT; Brownell 2001)

In other instances, measures were dropped to reduce the burden on children. While the Woodcock-Johnson Applied Problems measure had acceptable reliability, and only one item with DIF, we dropped this measure in order to maintain a reasonable testing time. The ECLS-B Mathematics assessment is used in its place in Phase 2, as it provides a measure of mathematics that can be conceptually scored and provides information on more extensive mathematical content (for example, items on data analysis). The Draw a Circle task was also dropped from the executive functioning battery to reduce the burden on children. Finally, in response to teachers' concerns about the number of problem behaviors included in the TCR, a shorter 10-item problem behavior scale from the Social Skills Rating System (SSRS; Gresham and Elliott 1990) was recommended to be used in Phase 2. To continue to collect information on the performance of the DRDP-R and to balance the data collection burden of the study on programs, we will collect DRDP-R data in the Power of Preschool (PoP) program sites only. These decisions and greater detail about the Phase 2 child assessment measures can be found in Atkins-Burnett, Caspe et al. 2007).

REFERENCES

- Administration for Children and Families. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impact of Early Head Start." Washington, DC: U.S. Department of Health and Human Services, 2002.
- Administration for Children and Families. "Report to Congress: Assessment of Children in Head Start Programs." Washington, DC: U.S. Department of Health and Human Services, 2005.
- Atkins-Burnett, Sally, Maggie Caspe, Cheri Vogel, Susan Sprachman, and John Love. "Fall Child Assessment Battery Decisions." Memorandum to Katie Fallin, First 5 LA, October 19, 2007.
- Atkins-Burnett, Sally, Cheri Vogel, John M. Love, Stephen Moore, and Mark Wilson. "Reanalysis of the Desired Results Developmental Profile-Revised (DRDP-R)." Memorandum to Katie Fallin, First 5 LA, in press, 2008.
- Atkins-Burnett, Sally, Nikki Aikens, Cheri Vogel, John M. Love, and Susan Sprachman. "Spring Data on Children with Suspected Special Needs." Memorandum to Katie Fallin, First 5 LA, August 20, 2007.
- Blair, C. "School Readiness: Integrating Cognition and Emotion in a Neurobiological Conceptualization of Children's Functioning at School Entry." *American Psychologist*, vol. 57, 2002, pp. 111-127.
- Blair, Clancy, Philip David Zelazo, and Mark Greenberg. "The Measurement of Executive Function in Early Childhood." *Developmental Neuropsychology*, vol. 28, no. 2, 2005, pp. 561-571.
- Brownell, Rick. "Expressive One-Word Picture Vocabulary Tests." San Antonio, TX: Harcourt Assessment, Inc., 2000.
- Brownell, Rick. "Receptive One-Word Picture Vocabulary Tests." San Antonio, TX: Harcourt Assessment, Inc., 2001.
- Diamond, Adele, and Colleen Taylor. "Development of an Aspect of Executive Control: Development of the Abilities to Remember What I Said and to "Do as I Say, Not as I Do." *Developmental Psychobiology*, vol. 29, 1996, pp. 315-334.

- Duncan, S.E., and E. DeAvila. "Preschool Language Assessment Survey 2000 Examiner's Manual." Monterey, CA: CTB/McGraw-Hill, 1998.
- Dunn, Lloyd, and Doug Dunn. "Peabody Picture Vocabulary Test—Fourth Edition." Circle Pines, MS: American Guidance Service, 2006.
- Dunn, Lloyd M., Delia E. Lugo, Eligio R. Padilla, and Leota M. Dunn. "Test de Vocabulario en Imágenes Peabody (TVIP)". Circle Pines, MN: AGS Publishing, 1986.
- Georgiou, George K., Rauno Parrila, and John Kirby. "Rapid Naming Speed Components and Early Reading Acquisition." *Scientific Studies of Reading* vol. 10, no. 2, 2006, pp. 199-220.
- Goodman, R. "Psychometric Properties of the Strengths and Difficulties Questionnaire (SDQ)." *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 40, 2001, pp. 1337-1345.
- Gresham, Frank M., and Stephen N. Elliott. "Social Skills Rating System." Circle Pines, MN: American Guidance Service, 1990.
- Harvard University and Center for Applied Linguistics. "Development of English Literacy in Spanish-Speaking Children—Phonological Awareness I." Cambridge, MA: Harvard University and Center for Applied Linguistics, 2002.
- Kochanska, Grazyna, and Kathleen Murray. "Parent-Child Study: Effortful Control Battery." Iowa City, IA: University of Iowa, 2002.
- LeBuffe, Paul, and Jack Naglieri. "The Devereux Early Childhood Assessment (DECA)". Lewisville, NC: Kaplan Early Learning Company, 1999.
- Merrell, Kenneth. "Preschool Kindergarten Behavior Scale-2." Austin, TX: PRO-ED, 2002.
- Moats, Louisa. "Teaching Reading Is Rocket Science: What Expert Teachers of Reading Should Know and Be Able to Do." Washington, DC: American Federation of Teachers, 1998.
- Murray, Kathleen T., and Grazyna Kochanska. "Effortful Control: Factor Structure and Relation to Externalizing and Internalizing Behaviors." *Journal of Abnormal Child Psychology*, vol. 30, 2002, pp. 503-514.
- O'Connor, Rollanda E., and Joseph R. Jenkins. "Prediction of Reading Disabilities in Kindergarten and First Grade." *Scientific Studies of Reading* vol. 3, no 2, 1999, pp. 159–197.
- Roid, Gale H., and Lucy J. Miller. "Leiter-R Performance Scale—Revised." Wood Dale, IL: Stoelting Co., 1997.
- Rouse, Heather L., and John W. Fantuzzo. "Validity of the Dynamic Indicators for Basic Early Literacy for Urban Kindergarten Children." *School Psychology Review*, vol. 35, 2006, pp. 341–355.

-
- 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.
- Snow, Catherine E., M. Susan Burns, and Peg Griffin (eds.). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press, 1998.
- Uhry, J.K. "Finger-Point Reading in Kindergarten: The Role of Phonemic Awareness, One-to-One Correspondence, and Rapid Serial Naming." *Scientific Studies of Reading* vol. 6, no. 4, 2002, pp. 319–342.
- Vogel, Cheri, John Love, Sally Atkins-Burnett, and Susan Sprachman. "UPCOS Research Design and Data Analysis Plan." Memorandum to Katie Fallin, First 5 LA, Spring 2007.
- Woodcock, Richard W., Kevin S. McGrew, and Nancy Mather. "Woodcock-Johnson III." Itasca, IL: Riverside Publishing, 2001.
- Zill, Nicholas, Alberto Sorongon, Kwang Kim, Cheryl Clark, and Maria Woolverton. "FACES 2003 Research Brief: Children's Outcomes and Program Quality in Head Start." Washington, DC: Administration for Children and Families U.S. Department of Health and Human Services, 2006.

APPENDIX A

ENHANCING THE VALUE OF THE PARENT INTERVIEW IN THE UNIVERSAL PRESCHOOL CHILD OUTCOMES STUDY: METHODS AND RESULTS OF FOCUS GROUPS AND COGNITIVE INTERVIEWS

Phase 2 of UPCOS will include a parent interview to add to the richness of information provided to First 5 LA about children's parents, their home environments, and the parents' perceptions of children's physical and social-emotional development. Selected parent interview items and scales build on the foundation of the Head Start Family and Child Experiences Survey (FACES), in which many of the items were used with great success. Items that were not used in previous cohorts of FACES also will be included to respond to the particular characteristics of children enrolled in LAUP programs. For example, items will assess parents' perspectives on programs' support of home culture. In gathering information about the child's social-emotional development, the UPCOS parent interview requires a measure that provides a culturally sensitive assessment of children's social-emotional competence. Parents from different socioeconomic and cultural backgrounds vary in their perceptions of behaviors and the importance of behaviors (Powless and Elliott 1993; Taylor, Campbell, and Burchinal 1997). Differences in ratings related to the parent's employment status, child's gender, and language level of the child have also been found in a study with preschoolers (Elliott, Barnard, and Gresham 1989). In addition, the reliability of parent scales is often lower than teacher scales (Rock and Pollack 2002; U. S. Department of Health and Human Services 2003). These findings suggest that more work examining the reliability, stability, and validity of parent ratings of children's social and emotional development is needed. They also suggest a need to examine how culturally and linguistically diverse parents perceive the importance of different social skills and problem behaviors, and how they understand interview items relating to children's social and emotional behaviors.

To address these gaps and inform the final selection of UPCOS parent interview items related to children's social-emotional outcomes and parents' perception of the programs' support of home culture, two activities were conducted: (1) focus group discussions with approximately 57 parents in May 2007,¹⁶ and (2) cognitive interviews with 53 parents in early June 2007. Final piloting of the full parent interview took place in late June 2007. We describe each of these activities in this appendix.

Focus Group Discussions

MPR and its subcontractor, Juárez and Associates, conducted 6 separate focus groups (of 7 to 13 participants each) with parents of preschool-aged children from varying racial/ethnic backgrounds. Due to a short timeline for initiating the focus groups and the burden that participating in the focus groups would place on programs during the data collection period, we did not recruit parents of children in LAUP programs to participate in the groups. First 5 LA assisted in locating organizations with Asian families who might be willing to participate. Juárez and Associates recruited African-American and Latino parents of preschool children from a database of individuals willing to participate in focus groups. Parents received an honorarium of \$50 for their participation. Individual focus groups

¹⁶ Due to scheduling difficulties, a focus group with 10 parents was conducted in July 2007. These parents were from Chinese, Indonesian, and Cambodian ethnic backgrounds.

consisted of parents who were Filipino (n=13), Korean (n=8), Chinese (n=10),¹⁷ African American (n=9), Latino Spanish-speaking monolingual (n=10), and Latino bilingual or English-speaking monolingual (n=7). We provided interpreters for members of the Filipino, Korean, and Chinese focus groups as necessary. The Latino Spanish-speaking monolingual focus group was conducted in Spanish, and the other Latino focus group was conducted in English. The majority (63.2 percent) of parents participating in the focus group discussions had at least some college education. Nearly a quarter (24.6 percent) were born in the United States. Most parents spoke English (50.9 percent) and/or Spanish (22.8 percent) as the main language at home.

The purpose of the focus groups was to allow parents from diverse racial/ethnic backgrounds to highlight and discuss the social skills and problem behaviors that were of importance and concern to them, and to learn what parents thought was most important for preschool to do to help their child's development in these areas. All focus group discussions, including the card-sorting task, were audio-recorded. Following each of the focus groups, separate coders reviewed the audiotapes and highlighted the predominant themes and behaviors that emerged as important to families. This process was intended to highlight behaviors and concerns that were particularly important to parents of varying racial/ethnic backgrounds, as well as identify consistent themes across parents in all focus groups. For example, members of several focus groups cited "respect for adults" and "manners" as important skills for children to have. Predominant themes across focus groups were considered when selecting the set of items and areas for discussion for subsequent cognitive interviews.

The focus groups included a card-sorting task in which parents were asked to look at index cards with a single item from selected parent-report measures of children's social and emotional behaviors. Fifty-one items were selected for the card-sorting task from among five standardized parent questionnaire measures having evidence of reliability and validity with nationally representative populations: the Devereux Early Childhood Assessment (DECA; LeBuffe, and Naglieri 1999), the Preschool and Kindergarten Behavior Scales – 2 (PKBS-2; Merrell 2002), the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997), the Social Skills Rating System (SSRS; Gresham and Elliot 1990), and the SRS Approaches to Learning scale from the ECLS-K (Rock and Pollack 2002). The purpose of the card-sorting task was to assess whether parents understood the intended meaning of the selected questions and whether they felt comfortable answering them. More broadly, the task was intended to assess the acceptability of different ways of asking about the child's behavior. Parents were instructed to sort each item into groups as follows: (1) items that were easy to answer, (2) items that were confusing and/or hard to answer, and (3) items that were uncomfortable to answer.

The focus groups included a general discussion of questions such as: "What does it mean to you when you say your child is well behaved?" "What social skills do you think

¹⁷ Some of the parents participating in the Chinese focus group were also from Indonesian (n=1) and Cambodian (n=4) ethnic backgrounds.

going to preschool will help your child have?” and “Are there questions about your child’s behavior or well-being that you do not want to be asked about in an interview?” In addition to a general discussion, the moderator asked the parents to focus specifically on the items from the cards that were confusing or made them uncomfortable.

The last activity in the focus groups were discussions about how preschools fostered or conflicted with the family’s values. During this activity, parents discussed their families’ traditions and customs, perceived barriers for passing on their home culture to children, and ways preschool programs could support home culture. For example, parents answered questions such as: “What traditions and customs do you think are important for you and your family to share with your child?” “How can the preschool program support you in sharing your culture with your child?” “What challenges are there to passing on the important parts of your culture to your child?” and “What traditions or customs do you have at home that you think will be more difficult to continue now that your child is spending more time with other children?”

After the focus groups were completed, we tabulated the card sorting activity to identify the differences between the groups in naming various items as easy, hard to understand, or uncomfortable to answer. The results are shown in Tables A1, A2, and A3. From among the five standardized social-emotional measures, PKBS-2 items were selected as the final set of questions to review for the cognitive interviews because this scale provided a more comprehensive measure of social-emotional competence than many other commercially available measures, which often emphasize either positive behaviors or clinical problems. In addition, the publisher was willing to consider adaptations to the measure for use with this diverse sample. Parents’ grouping of PKBS-2 items as easy to answer, hard to answer, and uncomfortable to answer, along with publisher-reported factor loadings of items, were considered when selecting the set of items and areas for discussion for subsequent cognitive interviews. Specifically, items that had comparatively low factor loadings on subscales or that were cited by multiple parents as being hard to answer or uncomfortable to answer were excluded. For example, among the focus group participants, 80 percent of English-speaking Latino, 57 percent of Spanish-speaking Latino, 44 percent of Chinese, 33 percent of African American, and 21 percent of Filipino parents felt that the item “Gives in or compromises with peers when necessary” was difficult to answer. Some parents expressed that “giving in” was not a positive trait, and they therefore had difficulty responding to this item. As a result, this item was excluded from subsequent cognitive interviews and the final parent interview.

Cognitive Interviewing

MPR conducted an iterative series of cognitive interviews to understand participants’ process and rationale for answering items about their child’s social skills, problem behaviors, and how cultural differences are addressed in the preschool (Tourangeau, Rips, and Rasinski 2000). The cognitive interviews were led by two senior Mathematica staff with assistance from the Juárez and Associates staff. Interviewers took extensive notes during the cognitive interviews and revised drafts of the survey according to parents’ comments.

Cognitive interview respondents were selected from parents of children who attended LAUP programs in the pilot study and had given consent for their children to participate. Fifty-three parents participated in the cognitive interviews, the majority of whom (60.4 percent) had at least some college education. Many were born in the United States (43.4 percent). In addition, most parents spoke English (66.0 percent) or Spanish (24.5 percent) as the main language at home.

Cognitive Interviewing for Social-Emotional Items. During the cognitive interviews, parents filled out a sample questionnaire containing demographic, PKBS-2 social-emotional, and home culture questions. We used both English-only and bilingual versions of the questionnaire so that if a bilingual parent did not understand a term in English she or he could refer to the Spanish translation. After completing the questionnaire, questions similar to those posed during focus groups were asked. For example, parents indicated whether any of the PKBS-2 questions were difficult to understand or were otherwise confusing. Parents were also asked to explain their perception of the meaning of individual items and to provide examples of observed behaviors that informed their responses to items. Interviewers probed items that parents rated at either extreme of the response scale.

From these discussions we identified a handful of items from the PKBS-2 whose meaning was ambiguous to parents in English or Spanish, or that used Spanish words that were not commonly used in Los Angeles. Once these problems were identified, we added additional items to the questionnaires that tested alternative wordings and explored their interpretation with participating parents. Only a handful of PKBS-2 items required changes to the wording or translation prior to the final pilot. These typically involved language that had a more technical or clinical meaning. One example of an item that posed difficulty for parents is “clings to parent or caregiver.” Some parents interpreted this as positive behavior (offering “when my child is excited” as an example of when it occurs). This item was replaced with two items in the final pilot: “Holds onto a parent or caregiver when strangers are around” and “Stays very close to a parent or caregiver for a long time in a new place.”

Cognitive Interviewing for Fostering Home Culture. Based on the focus groups, we developed a set of parent questions that were designed to explore the ways in which parents perceived preschool programs as fostering home culture and values that they considered important. The first draft of these questions focused on how often programs had certain activities or emphasized certain behaviors. Early cognitive interviews suggested that many parents were unfamiliar with whether or how frequently programs were engaging in specific activities to support the family’s home culture. Parents would rate the program positively (that is, parents rated practices as occurring “often”), but then were unable to provide an example of when it occurred. In addition, the variability in parents’ *frequency* reports of these items was limited. We also asked parents to report on *how important* it was for programs to engage in these activities. Once again, there was limited variability in parents’ responses, with most parents citing all activities as very important for programs to do.

After probing parents in the initial days of cognitive interviewing, MPR staff created alternative questions focused on observable child behaviors that would result from positive practices in diverse cultural settings. This revised scale for assessing programs’ support for

home culture, named the Respect for Differences Scale, asks parents to respond to items such as, “Compared to when CHILD started preschool, how willing is he to play with children from different cultural backgrounds now? Is he a lot more willing, a little more willing, less willing, or has there been no change?” During subsequent cognitive interviews, the wording of the items was refined and resulted in parent-provided examples indicative of the behaviors being probed as well as greater variability in response categories. The final version of this scale, along with the full parent interview battery, was tested in a small pilot in late June (n=12) and again resulted in variability in parents’ responses.

To allow us to look at change over the course of the preschool year, and to understand what “no change” means (is it no change because the child was always willing to play with children from different cultural backgrounds, or because she or he did not do so at the start of preschool and still does not?), we developed a new version of these questions that uses a frequency response scale. The scale has the response categories “always,” “very often,” “sometimes,” “rarely,” and “never,” with most questions phrased as “How willing is...?” or “Does he do ...?”. The complete scale is shown in Table A.4. After the fall data collection we will evaluate the psychometric properties of this scale.

TABLE A.1

SUMMARY OF FOCUS GROUP CARDS: AVERAGE, MAXIMUM, AND MINIMUM VALUES FOR ALL GROUPS

Label	Card	Average for All Groups (Percentage)				Easy (Percentage)		Hard (Percentage)		Uncomfortable (Percentage)	
		Easy	Hard	Uncomfortable	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	
PKBS—Problem Behavior Scale 31	1	57	25	17	75	40	40	13	29	0	
PKBS—Problem Behavior Scale 26	2	78	9	13	100	56	14	0	22	0	
PKBS—Problem Behavior Scale 04	3	58	26	16	86	29	57	0	22	11	
PKBS—Problem Behavior Scale 39	4	60	27	13	79	44	56	0	21	0	
PKBS—Problem Behavior Scale 15	5	68	20	11	86	56	33	7	22	0	
PKBS—Problem Behavior Scale 22	6	52	21	27	75	30	33	0	40	21	
PKBS—Social Skills Scale 14	7	84	10	8	93	67	22	0	14	0	
SSRS—Item 45	8	61	16	22	88	33	33	0	44	7	
DECA—Item 1	9	79	16	7	89	56	30	0	22	0	
PKBS—Social Skills Scale 17	10	76	20	4	93	56	44	7	14	0	
PKBS—Social Skills Scale 27	11	81	15	4	100	44	33	0	22	0	
DECA—Item 30	12	70	21	11	100	40	40	0	25	0	
PKBS—Problem Behavior Scale 38	13	50	23	27	75	29	44	0	40	11	
PKBS—Problem Behavior Scale 9	14	54	23	24	75	36	57	0	40	7	
SDQ—Item V	15	52	22	27	88	29	44	0	43	0	
PKBS—Problem Behavior Scale 12	16	73	16	10	88	44	56	0	20	0	
PKBS—Problem Behavior Scale 36	18	60	26	14	75	29	56	0	30	0	
SDQ—Item J	19	60	29	11	75	43	56	14	43	0	
DECA—Item 3	20	79	12	11	100	50	33	0	30	0	
PKBS—Social Skills Scale 12	21	83	11	7	100	60	22	0	20	0	
PKBS—Problem Behavior Scale 42	22	66	11	22	88	56	22	0	33	11	
PKBS—Problem Behavior Scale 28	23	56	31	13	71	33	78	10	30	0	
SSRS—Item 48	24	59	22	19	75	43	44	0	30	11	
PKBS—Problem Behavior Scale 35	25	66	20	14	79	50	33	7	20	11	

TABLE A.1 (continued)

Label	Card	Average for All Groups (Percentage)			Easy (Percentage)		Hard (Percentage)		Uncomfortable (Percentage)	
		Easy	Hard	Uncomfortable	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Appropriately Uses a Variety of Words to Describe Feelings (e.g., Excited, Mad)	26	64	27	9	78	30	50	14	21	0
Uses a Variety of Ways to Resolve Fights or Conflicts, Such As Bargaining	27	57	32	10	86	40	44	0	29	0
DECA—Item 9	28	52	19	29	75	29	50	0	64	10
PKBS—Problem Behavior Scale 11	29	60	20	20	75	43	40	7	43	0
PKBS—Social Skills Scale 32	30	82	14	6	100	60	33	0	14	0
PKBS—Problem Behavior Scale 30	31	66	16	18	88	44	33	0	43	0
DECA—Item 14	32	58	22	20	88	33	56	7	44	0
SSRS—Item 25	33	87	10	5	100	70	22	0	20	0
PKBS—Social Skills Scale 34	34	85	10	7	100	67	22	0	14	0
SDQ—Item S	35	59	25	17	88	43	40	13	43	0
PKBS—Social Skills Scale 28	36	50	39	13	88	10	80	0	22	7
Keeps Working at Something Even When It Is Difficult or Boring	37	73	14	13	93	57	33	0	29	0
ECLS-K—Teases or Makes Fun of Other Children	38	60	15	23	89	33	22	7	40	0
DECA—Item 18	39	50	18	30	75	30	30	7	43	13
ECLS-K—Stays on Tasks Even with Background Noise	40	74	17	11	89	57	30	0	14	7
PKBS—Social Skills Scale 31	41	69	23	8	100	40	50	0	22	0
SDQ—Item P	42	56	26	18	71	33	67	0	33	10
PKBS—Problem Behavior Scale 1	43	53	21	24	79	33	56	0	57	0
PKBS—Social Skills Scale 33	44	63	20	17	88	43	29	13	43	0
PKBS—Social Skills Scale 26	45	92	5	4	100	80	11	0	13	0
SSRS—Item 17	46	88	7	5	100	67	22	0	20	0
DECA—Item 28	47	79	13	8	100	63	22	0	25	0
DECA—Item 15	48	39	20	38	63	14	44	7	71	22
PKBS—Problem Behavior Scale 29	49	55	19	24	88	22	33	0	33	0

TABLE A.1 (continued)

Label	Average for All Groups (Percentage)				Easy (Percentage)		Hard (Percentage)		Uncomfortable (Percentage)	
	Card	Easy	Hard	Uncomfortable	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
PKBS—Social Skills Scale 13	50	61	22	15	71	44	38	0	29	7
PKBS—Problem Behavior Scale 21	51	49	23	26	75	22	44	10	40	13
SSRS—Item 29	52	82	10	8	93	60	22	0	20	0
Total		66	19	15						

Note: Card 17 is missing.

PKBS = Preschool Kindergarten Behavior Scale

DECA = Devereux Early Childhood Assessment

SSRS = Social Skills Rating Scale

SDQ = Strengths and Difficulties Questionnaire

ECLS-K = Early Childhood Longitudinal Study Approaches to Learning

TABLE A.2
SUMMARY OF FOCUS GROUP CARDS: FILIPINO, KOREAN, AND CHINESE OR CAMBODIAN GROUPS

Label	Card	Filipino ^a			Korean ^a			Chinese/Cambodian ^a		
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
PKBS—Problem Behavior Scale 31	1	9 (64)	2 (14)	3 (21)	6 (75)	1 (13)	1 (13)	5 (56)	3 (33)	2 (22)
PKBS—Problem Behavior Scale 26	2	12 (86)	1 (7)	1 (7)	7 (88)	0 (0)	1 (13)	9 (100)	1 (11)	0 (0)
PKBS—Problem Behavior Scale 04	3	12 (86)	0 (0)	2 (14)	6 (75)	1 (13)	1 (13)	4 (44)	4 (44)	2 (22)
PKBS—Problem Behavior Scale 39	4	11 (79)	0 (0)	3 (21)	4 (50)	3 (38)	1 (13)	4 (44)	5 (56)	1 (11)
PKBS—Problem Behavior Scale 15	5	11 (79)	1 (7)	2 (14)	6 (75)	2 (25)	0 (0)	5 (56)	3 (33)	2 (22)
PKBS—Problem Behavior Scale 22	6	9 (64)	2 (14)	3 (21)	6 (75)	0 (0)	2 (25)	5 (56)	3 (33)	2 (22)
PKBS—Social Skills Scale 14	7	13 (93)	1 (7)	0 (0)	7 (88)	0 (0)	1 (13)	8 (89)	2 (22)	0 (0)
SSRS—Item 45	8	11 (79)	2 (14)	1 (7)	7 (88)	0 (0)	1 (13)	6 (67)	3 (33)	1 (11)
DECA—Item 1	9	12 (86)	1 (7)	1 (7)	7 (88)	1 (13)	0 (0)	8 (89)	2 (22)	0 (0)
PKBS—Social Skills Scale 17	10	13 (93)	1 (7)	0 (0)	7 (88)	1 (13)	0 (0)	6 (67)	4 (44)	0 (0)
PKBS—Social Skills Scale 27	11	14 (100)	0 (0)	0 (0)	7 (88)	1 (13)	0 (0)	7 (78)	3 (33)	0 (0)
DECA—Item 30	12	14 (100)	0 (0)	0 (0)	6 (75)	0 (0)	2 (25)	7 (78)	2 (22)	1 (11)
PKBS—Problem Behavior Scale 38	13	9 (64)	0 (0)	5 (36)	6 (75)	1 (13)	1 (13)	5 (56)	4 (44)	1 (11)
PKBS—Problem Behavior Scale 9	14	5 (36)	8 (57)	1 (7)	6 (75)	0 (0)	2 (25)	5 (56)	3 (33)	2 (22)
SDQ—Item V	15	6 (43)	2 (14)	6 (43)	7 (88)	1 (13)	0 (0)	5 (56)	4 (44)	1 (11)

TABLE A.2 (continued)

Label	Card	Filipino ^a				Korean ^a				Chinese/Cambodian ^a			
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
PKBS—Problem Behavior Scale 12	16	12 (86)	0 (0)	2 (14)	7 (88)	1 (13)	0 (0)	4 (44)	5 (56)	1 (11)			
PKBS—Problem Behavior Scale 36	18	9 (64)	3 (21)	2 (14)	6 (75)	1 (13)	1 (13)	5 (56)	5 (56)	0 (0)			
SDQ—Item J	19	10 (71)	2 (14)	2 (14)	6 (75)	2 (25)	0 (0)	5 (56)	5 (56)	0 (0)			
DECA—Item 3	20	13 (93)	1 (7)	0 (0)	6 (75)	0 (0)	2 (25)	7 (78)	3 (33)	0 (0)			
PKBS—Social Skills Scale 12	21	12 (86)	2 (14)	0 (0)	7 (88)	0 (0)	1 (13)	8 (89)	1 (11)	1 (11)			
PKBS—Problem Behavior Scale 42	22	8 (57)	2 (14)	4 (29)	7 (88)	0 (0)	1 (13)	6 (67)	1 (11)	3 (33)			
PKBS—Problem Behavior Scale 28	23	10 (71)	3 (21)	1 (7)	5 (63)	3 (38)	0 (0)	3 (33)	7 (78)	0 (0)			
SSRS—Item 48	24	9 (64)	2 (14)	3 (21)	6 (75)	1 (13)	1 (13)	5 (56)	4 (44)	1 (11)			
PKBS—Problem Behavior Scale 35	25	11 (79)	1 (7)	2 (14)	6 (75)	1 (13)	1 (13)	6 (67)	3 (33)	1 (11)			
Appropriately Uses a Variety of Words to Describe Feelings (e.g., Excited, Mad)	26	9 (64)	2 (14)	3 (21)	6 (75)	2 (25)	0 (0)	7 (78)	3 (33)	0 (0)			
Uses a Variety of Ways to Resolve Fights or Conflicts, Such As Bargaining	27	12 (86)	0 (0)	2 (14)	5 (63)	3 (38)	0 (0)	6 (67)	4 (44)	0 (0)			
DECA—Item 9	28	4 (29)	1 (7)	9 (64)	6 (75)	1 (13)	1 (13)	6 (67)	3 (33)	1 (11)			
PKBS—Problem Behavior Scale 11	29	10 (71)	1 (7)	3 (21)	6 (75)	2 (25)	0 (0)	6 (67)	2 (22)	2 (22)			
PKBS—Social Skills Scale 32	30	13 (93)	1 (7)	0 (0)	8 (100)	0 (0)	0 (0)	7 (78)	3 (33)	0 (0)			
PKBS—Problem Behavior Scale 30	31	10 (71)	1 (7)	3 (21)	7 (88)	0 (0)	1 (13)	7 (78)	3 (33)	0 (0)			
DECA—Item 14	32	10 (71)	1 (7)	3 (21)	7 (88)	1 (13)	0 (0)	5 (56)	5 (56)	0 (0)			

INFORMATION NOT RELEASABLE TO THE PUBLIC UNLESS AUTHORIZED: This information has not been publicly disclosed and is privileged and confidential. It is for internal Fir st 5 LA use only and must not be disseminated, distributed, or copied to persons not authorized to receive the information.

TABLE A.2 (continued)

Label	Card	Filipino ^a				Korean ^a				Chinese/Cambodian ^a			
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
SSRS—Item 25	33	14 (100)	0 (0)	0 (0)	8 (100)	0 (0)	0 (0)	8 (89)	2 (22)	0 (0)	8 (89)	2 (22)	0 (0)
PKBS—Social Skills Scale 34	34	13 (93)	0 (0)	1 (7)	8 (100)	0 (0)	0 (0)	9 (100)	1 (11)	0 (0)	9 (100)	1 (11)	0 (0)
SDQ—Item S	35	10 (71)	2 (14)	2 (14)	7 (88)	1 (13)	0 (0)	5 (56)	3 (33)	2 (22)	5 (56)	3 (33)	2 (22)
PKBS—Social Skills Scale 28	36	10 (71)	3 (21)	1 (7)	7 (88)	0 (0)	1 (13)	4 (44)	4 (44)	2 (22)	4 (44)	4 (44)	2 (22)
Keeps Working at Something Even When It Is Difficult or Boring	37	13 (93)	0 (0)	1 (7)	6 (75)	1 (13)	1 (13)	7 (78)	3 (33)	0 (0)	7 (78)	3 (33)	0 (0)
ECLS-K—Teases or Makes Fun of Other Children	38	8 (57)	1 (7)	5 (36)	6 (75)	1 (13)	1 (13)	8 (89)	2 (22)	0 (0)	8 (89)	2 (22)	0 (0)
DECA—Item 18	39	9 (64)	1 (7)	4 (29)	6 (75)	1 (13)	1 (13)	5 (56)	2 (22)	3 (33)	5 (56)	2 (22)	3 (33)
ECLS-K—Stays on Tasks Even with Background Noise	40	12 (86)	1 (7)	1 (7)	7 (88)	0 (0)	1 (13)	8 (89)	1 (11)	1 (11)	8 (89)	1 (11)	1 (11)
PKBS—Social Skills Scale 31	41	12 (86)	2 (14)	0 (0)	8 (100)	0 (0)	0 (0)	6 (67)	4 (44)	0 (0)	6 (67)	4 (44)	0 (0)
SDQ—Item P	42	9 (64)	3 (21)	2 (14)	5 (63)	2 (25)	1 (13)	3 (33)	6 (67)	1 (11)	3 (33)	6 (67)	1 (11)
PKBS—Problem Behavior Scale 1	43	11 (79)	0 (0)	3 (21)	4 (50)	2 (25)	2 (25)	5 (56)	5 (56)	0 (0)	5 (56)	5 (56)	0 (0)
PKBS—Social Skills Scale 33	44	9 (64)	4 (29)	1 (7)	7 (88)	1 (13)	0 (0)	6 (67)	2 (22)	2 (22)	6 (67)	2 (22)	2 (22)
PKBS—Social Skills Scale 26	45	13 (93)	1 (7)	0 (0)	7 (88)	0 (0)	1 (13)	9 (100)	1 (11)	0 (0)	9 (100)	1 (11)	0 (0)
SSRS—Item 17	46	14 (100)	0 (0)	0 (0)	8 (100)	0 (0)	0 (0)	8 (89)	1 (11)	1 (11)	8 (89)	1 (11)	1 (11)
DECA—Item 28	47	11 (79)	3 (21)	0 (0)	5 (63)	1 (13)	2 (25)	8 (89)	2 (22)	0 (0)	8 (89)	2 (22)	0 (0)
DECA—Item 15	48	6 (43)	1 (7)	7 (50)	5 (63)	1 (13)	2 (25)	4 (44)	4 (44)	2 (22)	4 (44)	4 (44)	2 (22)
PKBS—Problem Behavior Scale 29	49	10 (71)	0 (0)	4 (29)	7 (88)	1 (13)	0 (0)	4 (44)	3 (33)	3 (33)	4 (44)	3 (33)	3 (33)
PKBS—Social Skills Scale 13	50	9 (64)	4 (29)	1 (7)	4 (50)	3 (25)	1 (13)	6 (67)	3 (33)	1 (11)	6 (67)	3 (33)	1 (11)

INFORMATION NOT RELEASABLE TO THE PUBLIC UNLESS AUTHORIZED: This information has not been publicly disclosed and is privileged and confidential. It is for internal use only and must not be disseminated, distributed, or copied to persons not authorized to receive the information.

TABLE A.2 (continued)

Label	Card	Filipino ^a			Korean ^a			Chinese/Cambodian ^a		
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
PKBS—Problem Behavior Scale 21	51	8 (57)	3 (21)	3 (21)	6 (75)	1 (13)	1 (13)	4 (44)	4 (44)	2 (22)
SSRS—Item 29	52	13 (93)	1 (7)	0 (0)	7 (88)	1 (13)	0 (0)	8 (89)	2 (22)	0 (0)
Total		524 (75)	74 (11)	103 (15)	324 (79)	47 (12)	37 (9)	307 (60)	158 (31)	45 (9)

Note: Card 17 is missing.

^aThe value in parentheses denotes the percentage of cards that are easy, hard, or uncomfortable for each item.

PKBS = Preschool Kindergarten Behavior Scale

DECA = Devereux Early Childhood Assessment

SSRS = Social Skills Rating Scale

SDQ = Strengths and Difficulties Questionnaire

ECLS-K = Early Childhood Longitudinal Study Approaches to Learning

TABLE A.3
SUMMARY OF FOCUS GROUP CARDS: LATINO/SPANISH-SPEAKING, LATIN/ENGLISH-SPEAKING, AND AFRICAN AMERICAN GROUPS

Label	Card	Latino/Spanish-Speaking ^a			Latino/English-Speaking ^a			African American ^a		
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
PKBS—Problem Behavior Scale 31	1	4 (40)	4 (40)	2 (20)	3 (43)	2 (29)	2 (29)	6 (67)	2 (22)	0 (0)
PKBS—Problem Behavior Scale 26	2	7 (70)	1 (10)	2 (20)	5 (71)	1 (14)	1 (14)	5 (56)	1 (11)	2 (22)
PKBS—Problem Behavior Scale 04	3	7 (70)	1 (10)	2 (20)	2 (29)	4 (57)	1 (14)	4 (44)	3 (33)	1 (11)
PKBS—Problem Behavior Scale 39	4	5 (50)	3 (30)	2 (20)	5 (71)	2 (29)	0 (0)	6 (67)	1 (11)	1 (11)
PKBS—Problem Behavior Scale 15	5	6 (60)	2 (20)	2 (20)	6 (86)	1 (14)	0 (0)	5 (56)	2 (22)	1 (11)
PKBS—Problem Behavior Scale 22	6	3 (30)	3 (30)	4 (40)	3 (43)	2 (29)	2 (29)	4 (44)	2 (22)	2 (22)
PKBS—Social Skills Scale 14	7	8 (80)	1 (10)	1 (10)	6 (86)	0 (0)	1 (14)	6 (67)	2 (22)	1 (11)
SSRS—Item 45	8	6 (60)	1 (10)	3 (30)	3 (43)	2 (29)	2 (29)	3 (33)	1 (11)	4 (44)
DECA—Item 1	9	7 (70)	3 (30)	0 (0)	6 (86)	0 (0)	1 (14)	5 (56)	2 (22)	2 (22)
PKBS—Social Skills Scale 17	10	8 (80)	2 (20)	0 (0)	5 (71)	1 (14)	1 (14)	5 (56)	2 (22)	1 (11)
PKBS—Social Skills Scale 27	11	9 (90)	1 (10)	0 (0)	6 (86)	1 (14)	0 (0)	4 (44)	2 (22)	2 (22)
DECA—Item 30	12	4 (40)	4 (40)	2 (20)	5 (71)	2 (29)	0 (0)	5 (56)	3 (33)	1 (11)
PKBS—Problem Behavior Scale 38	13	3 (30)	3 (30)	4 (40)	2 (29)	3 (43)	2 (29)	4 (44)	1 (11)	3 (33)
PKBS—Problem Behavior Scale 9	14	4 (40)	2 (20)	4 (40)	5 (71)	1 (14)	1 (14)	4 (44)	1 (11)	3 (33)
SDQ—Item V	15	4 (40)	3 (30)	3 (30)	2 (29)	2 (29)	3 (43)	5 (56)	0 (0)	3 (33)

TABLE A.3 (continued)

Label	Card	Latino/Spanish-Speaking ^a				Latino/English-Speaking ^a				African American ^a			
		Easy	Hard	Uncomfortable	Uncomfortable	Easy	Hard	Uncomfortable	Uncomfortable	Easy	Hard	Uncomfortable	Uncomfortable
PKBS—Problem Behavior Scale 12	16	8 (80)	0 (0)	2 (20)	2 (20)	5 (71)	1 (14)	1 (14)	1 (14)	6 (67)	1 (11)	0 (0)	0 (0)
PKBS—Problem Behavior Scale 36	18	7 (70)	0 (0)	3 (30)	3 (30)	2 (29)	3 (43)	2 (29)	2 (29)	6 (67)	2 (22)	0 (0)	0 (0)
SDQ—Item J	19	6 (60)	4 (40)	0 (0)	0 (0)	3 (43)	1 (14)	3 (43)	3 (43)	5 (56)	2 (22)	1 (11)	1 (11)
DECA—Item 3	20	5 (50)	2 (20)	3 (30)	3 (30)	7 (100)	0 (0)	0 (0)	0 (0)	7 (78)	1 (11)	1 (11)	1 (11)
PKBS—Social Skills Scale 12	21	6 (60)	2 (20)	2 (20)	2 (20)	7 (100)	0 (0)	0 (0)	0 (0)	7 (78)	2 (22)	0 (0)	0 (0)
PKBS—Problem Behavior Scale 42	22	6 (60)	2 (20)	2 (20)	2 (20)	5 (71)	0 (0)	2 (29)	2 (29)	5 (56)	2 (22)	1 (11)	1 (11)
PKBS—Problem Behavior Scale 28	23	6 (60)	1 (10)	3 (30)	3 (30)	3 (43)	2 (29)	2 (29)	2 (29)	6 (67)	1 (11)	1 (11)	1 (11)
SSRS—Item 48	24	5 (50)	2 (20)	3 (30)	3 (30)	3 (43)	3 (43)	1 (14)	1 (14)	6 (67)	0 (0)	2 (22)	2 (22)
PKBS—Problem Behavior Scale 35	25	5 (50)	3 (30)	2 (20)	2 (20)	4 (57)	2 (29)	1 (14)	1 (14)	6 (67)	1 (11)	1 (11)	1 (11)
Appropriately Uses a Variety of Words to Describe Feelings (e.g., Excited, Mad)	26	3 (30)	5 (50)	2 (20)	2 (20)	5 (71)	1 (14)	1 (14)	1 (14)	6 (67)	2 (22)	0 (0)	0 (0)
Uses a Variety of Ways to Resolve Fights or Conflicts, Such As Bargaining	27	4 (40)	4 (40)	2 (20)	2 (20)	3 (43)	2 (29)	2 (29)	2 (29)	4 (44)	4 (44)	0 (0)	0 (0)
DECA—Item 9	28	4 (40)	5 (50)	1 (10)	1 (10)	4 (57)	0 (0)	3 (43)	3 (43)	4 (44)	1 (11)	3 (33)	3 (33)
PKBS—Problem Behavior Scale 11	29	5 (50)	4 (40)	1 (10)	1 (10)	3 (43)	1 (14)	3 (43)	3 (43)	5 (56)	1 (11)	2 (22)	2 (22)
PKBS—Social Skills Scale 32	30	6 (60)	3 (30)	1 (10)	1 (10)	5 (71)	1 (14)	1 (14)	1 (14)	8 (89)	0 (0)	1 (11)	1 (11)
PKBS—Problem Behavior Scale 30	31	6 (60)	2 (20)	2 (20)	2 (20)	4 (57)	0 (0)	3 (43)	3 (43)	4 (44)	3 (33)	1 (11)	1 (11)
DECA—Item 14	32	6 (60)	3 (30)	1 (10)	1 (10)	3 (43)	1 (14)	3 (43)	3 (43)	3 (33)	1 (11)	4 (44)	4 (44)
SSRS—Item 25	33	7 (70)	1 (10)	2 (20)	2 (20)	6 (86)	1 (14)	0 (0)	0 (0)	7 (78)	1 (11)	1 (11)	1 (11)
PKBS—Social Skills Scale 34	34	8 (80)	1 (10)	1 (10)	1 (10)	5 (71)	1 (14)	1 (14)	1 (14)	6 (67)	2 (22)	1 (11)	1 (11)
SDQ—Item S	35	5	4	1	1	3	1	3	3	4	3	1	1

INFORMATION NOT RELEASABLE TO THE PUBLIC UNLESS AUTHORIZED: This information has not been publicly disclosed and is privileged and confidential. It is for internal Fir st 5 LA use only and must not be disseminated, distributed, or copied to persons not authorized to receive the information.

TABLE A.3 (continued)

Label	Card	Latino/Spanish-Speaking ^a				Latino/English-Speaking ^a				African American ^a			
		Easy	Hard	Uncomfortable	(%)	Easy	Hard	Uncomfortable	(%)	Easy	Hard	Uncomfortable	(%)
PKBS—Social Skills Scale 28	36	1 (10)	8 (80)	1 (10)	(10)	2 (29)	4 (57)	1 (14)	(43)	5 (56)	3 (33)	1 (11)	(11)
Keeps Working at Something Even When It Is Difficult or Boring	37	7 (70)	1 (10)	2 (20)	(20)	4 (57)	1 (14)	2 (29)	(29)	6 (67)	1 (11)	1 (11)	(11)
ECLS-K—Teases or Makes Fun of Other Children	38	5 (50)	1 (10)	4 (40)	(40)	4 (57)	1 (14)	2 (29)	(29)	3 (33)	2 (22)	2 (22)	(22)
DECA—Item 18	39	3 (30)	3 (30)	4 (40)	(40)	3 (43)	1 (14)	3 (43)	(43)	3 (33)	2 (22)	2 (22)	(22)
ECLS-K—Stays on Tasks Even with Background Noise	40	6 (60)	3 (30)	1 (10)	(10)	4 (57)	2 (29)	1 (14)	(14)	6 (67)	2 (22)	1 (11)	(11)
PKBS—Social Skills Scale 31	41	4 (40)	5 (50)	1 (10)	(10)	4 (57)	2 (29)	1 (14)	(14)	6 (67)	0 (0)	2 (22)	(22)
SDQ—Item P	42	6 (60)	3 (30)	1 (10)	(10)	5 (71)	0 (0)	2 (29)	(29)	4 (44)	1 (11)	3 (33)	(33)
PKBS—Problem Behavior Scale 1	43	6 (60)	1 (10)	3 (30)	(30)	3 (43)	0 (0)	4 (57)	(57)	3 (33)	3 (33)	1 (11)	(11)
PKBS—Social Skills Scale 33	44	5 (50)	2 (20)	3 (30)	(30)	3 (43)	1 (14)	3 (43)	(43)	6 (67)	2 (22)	0 (0)	(0)
PKBS—Social Skills Scale 26	45	8 (80)	1 (10)	1 (10)	(10)	7 (100)	0 (0)	0 (0)	(0)	8 (89)	0 (0)	0 (0)	(0)
SSRS —Item 17	46	7 (70)	1 (10)	2 (20)	(20)	7 (100)	0 (0)	0 (0)	(0)	6 (67)	2 (22)	0 (0)	(0)
DECA—Item 28	47	8 (80)	1 (10)	1 (10)	(10)	7 (100)	0 (0)	0 (0)	(0)	6 (67)	1 (11)	1 (11)	(11)
DECA—Item 15	48	5 (50)	1 (10)	4 (40)	(40)	1 (14)	1 (14)	5 (71)	(71)	2 (22)	3 (33)	2 (22)	(22)
PKBS—Problem Behavior Scale 29	49	5 (50)	2 (20)	3 (30)	(30)	4 (57)	1 (14)	2 (29)	(29)	2 (22)	3 (33)	2 (22)	(22)
PKBS—Social Skills Scale 13	50	7 (70)	2 (20)	1 (10)	(10)	5 (71)	0 (0)	2 (29)	(29)	4 (44)	1 (11)	2 (22)	(22)

INFORMATION NOT RELEASABLE TO THE PUBLIC UNLESS AUTHORIZED: This information has not been publicly disclosed and is privileged and confidential. It is for internal use only and must not be disseminated, distributed, or copied to persons not authorized to receive the information.

TABLE A.3 (continued)

Label	Card	Latino/Spanish-Speaking ^a			Latino/English-Speaking ^a			African American ^a		
		Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable	Easy	Hard	Uncomfortable
PKBS—Problem Behavior Scale 21	51	5 (50)	1 (10)	4 (40)	3 (43)	2 (29)	2 (29)	2 (22)	2 (22)	3 (33)
SSRS—Item 29	52	6 (60)	2 (20)	2 (20)	6 (86)	0 (0)	1 (14)	7 (78)	0 (0)	1 (11)
Total		287 (56)	120 (24)	103 (20)	216 (61)	61 (17)	80 (22)	255 (62)	83 (20)	72 (18)

Note: Card 17 is missing.

^aThe value in parentheses denotes the percentage of cards that are easy, hard, or uncomfortable for each item.

PKBS = Preschool Kindergarten Behavior Scale

DECA = Devereux Early Childhood Assessment

SSRS = Social Skills Rating Scale

SDQ = Strengths and Difficulties Questionnaire

ECLS-K = Early Childhood Longitudinal Study Approaches to Learning

TABLE A.4
RESPECT FOR DIFFERENCES SCALE

W3a. Please tell me how often each of these statements is true of [CHILD]. [CHILD] (INSERT ITEM)? . . .

Show
Card

	Always	Very often	Sometimes	Rarely	Never	Don't Know	NA
a. is willing to try foods from different cultures..... Is (he/she) always, very often, sometimes, rarely, or never willing to try foods from different cultures?	4	3	2	1	0	d	N/A
b. is respectful to adults in your family..... Is (he/she) always, very often, sometimes, rarely, or never respectful to adults?	4	3	2	1		d	
c. is willing to play with children from different backgrounds Is (he/she) always, very often, sometimes, rarely, or never willing?	4	3	2	1		d	
d. talks disrespectfully to adults in your home..... Does (he/she) talk disrespectfully always, very often, sometimes, rarely, or never?	4	3	2	1		d	
e. is willing to play with children who have special needs or physical or mental disabilities Is (he/she) always, very often, sometimes, rarely, or never willing?	4	3	2	1		d	NA
f. is interested in different cultures Is (he/she) always, very often, sometimes, rarely, or never interested in different cultures?	4	3	2	1		d	
g. is interested about why different people celebrate different holidays Is (he/she) always, very often, sometimes, rarely, or never interested in why different people celebrate different holidays?	4	3	2	1		d	

Show
Card

Table A.4 (continued)

	Always	Very often	Sometimes	Rarely	Never	Don't Know	NA
h. avoids playing with children from different cultural backgrounds? Does (he/she) avoid children from different cultural backgrounds always, very often, sometimes, rarely, or never?	4	3	2	1		d	
i. {IF D7 = 1 AND D10 NOT = 25, r} is willing to speak your home language Is (he/she) always, very often, sometimes, rarely, or never willing to speak your home language?	4	3	2	1		d	NA
j. is willing to take part in events that are important to your family's cultural traditions? Is (he/she) always, very often, sometimes, rarely, or never a willing to take part in these events?	4	3	2	1		d	
k. is interested in words from other languages Is (he/she) always, very often, sometimes, rarely, or never a interested in words from other languages?	4	3	2	1		d	
l. is respectful of people from different cultures or backgrounds Is (he/she) always, very often, sometimes, rarely, or never respectful of people from different cultures of backgrounds?	4	3	2	1		d	

Table A.4 (continued)

{SPRING 2008}

W3. Please think of how [CHILD] is now compared to how (he/she) was when (he/she) started attending this preschool. Compared to when [CHILD] started [PRESCHOOL], how (INSERT ITEM)?

Show Card

	A lot more	A little more	Less	No change	Don't Know	NA
a. willing is [CHILD] to try foods from different cultures now? Is (he/she) a lot more willing, a little more willing, less willing, or has there been no change (since (he/she) started attending this preschool)?	4	3	2	1	d	N/A
b. respectful is [CHILD] to adults in your family now? Is (he/she) a lot more respectful, little more respectful, less respectful to adults or has there been no change?	4	3	2	1	d	N/A
c. willing is [CHILD] to play with children from different cultures and backgrounds now? Is (he/she) a lot more willing, a little more willing, less willing, or has there been no change?	4	3	2	1	d	N/A
d. much does [CHILD] talk disrespectfully to adults in your home now? Does (he/she) talk disrespectfully a lot more often, a little more often, less often, or has there been no change?	4	3	2	1	d	N/A
e. willing is [CHILD] to play with children who have special needs or physical or mental disabilities now? Is (he/she) a lot more willing, a little more willing, less willing, or has there been no change?	4	3	2	1	d	N/A
f. interested is [CHILD] in different cultures now? Is (he/she) a lot more interested, a little more interested, less interested, or has there been no change?	4	3	2	1	d	N/A
g. interested is [CHILD] about why people celebrate different holidays now? Is (he/she) a little more interested, less knowledgeable, or has there been no change?	4	3	2	1	d	N/A

Show Card

Table A.4 (continued)

	A lot more	A little more	Less	No change	Don't Know	NA
h. much does (CHILD) avoid playing with children from different cultural backgrounds now? Does (he/she) avoid children from different cultural backgrounds a lot more often, a little more often, less often, or has there been no change?	4	3	2	1	d	N/A
i. {IF D7 = 1 AND D10 NOT = 25, r} willing is [CHILD] to speak your home language now?..... Is (he/she) a lot more willing, a little more willing, less willing, or has there been no change?	4	3	2	1	d	N/A
j. willing is [CHILD] to take part in events that are important to your family's cultural traditions now? Is (he/she) a lot more willing, a little more willing, less willing, or has there been no change?	4	3	2	1	d	N/A
k. interested is [CHILD] in words from other languages now? Is (he/she) a lot more interested, a little more interested, less interested, or has there been no change?	4	3	2	1	d	N/A
l. respectful is [CHILD] of people from different cultures or backgrounds now? Is (he/she) a lot more respectful, a little more respectful, less respectful, or has there been no change?	4	3	2	1	d	N/A