Evaluation Report on Effects of Performance-Funding Pilot Project for Florida's School Readiness Providers

Prepared for the Florida Office of Early Learning by

The Florida Center for Reading Research

Florida State University

September 16, 2015

Evaluation Report on Effects of Performance-Funding Pilot Project for Florida's School Readiness Providers

This report describes the impact, or effects, of the Performance Funding Pilot Project for school readiness providers. The goal of the performance funding pilot project evaluation was to determine if provision of a package of quality-enhancement activities resulted in measurably better outcomes for children who participated in Florida's School Readiness Program. The expected model of effectiveness is that participation in specific professional development concerning teacher-child interactions and the provision of additional monetary resources would result in higher-quality teacher-child interaction in school-readiness classrooms and that these interactions would be better aligned with the specific needs of children. These changes, in turn, were expected to result in children gaining more skills in a number of school readiness domains, including socio-emotional development, language and communication skills, and cognitive development and general knowledge.

School-readiness providers that participated in this evaluation were drawn from a pool of qualified volunteer providers throughout the state of Florida. Providers were randomized to a Pilot group, which participated in the complete package of quality-enhancement activities, and a business-as-usual (BAU) Comparison group, which participated in whatever professional development activities that would have otherwise been provided to them but they did not participate in the activities associated with the quality-enhancement package. The quality-enhancement package consisted of two elements:

1. Receiving extensive professional development concerning teacher-child interactions associated with the Classroom Assessment Scoring System (CLASS) including the implementation of an improvement plan.

2. Receipt of additional monetary resources to be used to enhance quality, as means of promoting higher quality teacher-child interactions and improved child outcomes.

The quality enhancement package originally proposed included two additional components; however, logistical issues precluded the provision of these components as part of the quality-enhancement package in this implementation year.

Design of Evaluation

The design of this evaluation involved a cluster-randomized study in which school readiness providers were randomly assigned to a Pilot (i.e., experimental) group or a Comparison (i.e., control) group. Two groups of children, two-year-old children and three- to five-year-old children, were assessed to determine the impact of the quality enhancement program on children's developing pre-academic and socio-emotional skills. These children were selected from up to two classrooms per provider per age group in participating centers. Additionally, classrooms in centers were observed using the Classroom Assessment Scoring System.

Assignment of Providers to Condition

The Florida Office of Early Learning (FL-OEL) recruited School-Readiness Providers to participate in the study. The initial list of providers included 401 eligible participants. Using this initial list of eligible providers, providers were assigned by a lottery (i.e., randomization) to the Pilot group, which was intended to receive the intervention package, and a Comparison group, which was not intended to receive the intervention package. The lottery was designed to allocate more of the providers to the Pilot group than to the Comparison group--in approximately a two-to-one ratio. Prior to lottery assignment to Pilot or Comparison group, providers were matched in triads according to a subset of the selection criteria. To the extent possible providers were matched into triads on variables, including (a) whether or not the provider was in a high-need tract, (b) whether or not the provider had the Gold Seal designation, (c) the provider's licensed capacity (size), and (d) provider type (e.g., Licensed Center, Family Child Care) prior to assignment. From each matched triad, the lottery was applied such that two providers in the triad were assigned to the Pilot group and one provider was assigned to the Comparison group. Some providers were matched into pairs, and one member of the pair was assigned to the Pilot group and the other to the Comparison group by lottery. An additional randomization was conducted for providers assigned to the Pilot group. Of the two providers within the matched triads assigned to the provider group, one was randomly selected to participate in the direct assessment of children

Of the 401 initial eligible providers, 141 were randomized to the Comparison group and 260 were randomized to the Pilot group. Of the 260 providers assigned to the Pilot group, 159 were randomly selected to participate in the direct assessment of children. Between assignment

and the post-intervention assessments, 31 Comparison group providers (22%) and 58 Pilot group providers (22%) opted out of the project. Of the 58 Pilot providers that opted out of the program, 36 (62%) were Pilot providers assigned to participate in the direct assessment of children; hence, 23% of Pilot group providers assigned to the direct assessment of children condition opted out of the program. Although attrition can be problematic in a randomized evaluation, the level of provider attrition in this project was within the level of attrition that is not expected to bias outcomes (e.g., see What Works Clearinghouse, *Procedures and Standards Handbook, version 3*), particularly because there was no differential attrition (i.e., more providers opting out of the Pilot or Comparison groups), which could indicate a differential response to assignment or the activities associated with assignment.

Selection of Children to Participate in Evaluation

From each of the 300 providers that were selected to participate in direct assessments of children, up to eight three- to five-year-old children were randomly selected from among those children whose parents provided consent for participation in the evaluation. For providers with more than two classrooms serving this age group, one or two classrooms within provider were randomly selected, and children were recruited from this (or these) classrooms for the direct-assessment component of the evaluation. In addition, for providers that served two-year-old children, a randomly selected group of up to eight two-year-old children were selected to participate in the direct-assessment component of the evaluation. Again, if a provider had more than two classes serving two-year-old children, random selection of one or two classrooms within provider determined the classroom(s) from which two-year-old children were recruited for direct evaluation.

Figure E1 in Appendix E shows the number of three- to five-year-old children originally recruited for direct assessment from providers in the Pilot and Comparison groups and these children's completion of different assessments during various phases of the project. There were 1,981 three- to five-year-old children consented for the project from the selected classrooms (1,274 from Pilot group, 707 from Comparison group). Of these children, between 1,190 and 1,274 children from the Pilot group (93-100%) completed the pre-intervention direct assessments, and between 657 and 707 children from the Comparison group (93-100%) completed the pre-intervention direct assessments. The rate of return for teacher-completed

ratings of children's socio-emotional development was lower (i.e., 74% of children from the Pilot group, 69% of children from the Comparison group). At the post-intervention assessment, between 1,011 and 1,041 of the 1,274 children originally recruited in the Pilot group completed the direct assessments (79-82% of children), and between 582 and 588 of the 707 children originally recruited in the Comparison group completed the direct assessments (82-83% of children). Again, the rate of return for teachers' ratings of children socio-emotional development was lower (i.e., 45% of Pilot group, 37% of Comparison group). Rates of attrition from the direct assessment conducted by FCRR assessors were due to (a) providers opting out of the project following recruitment of children, (b) children leaving a provider's center prior to completion of assessments, and (c) child absences over the period of assessment. The rate of attrition from the teacher ratings was, in part, the result of providers opting out of the project following recruitment of children, but, in many cases, teachers simply did not return the required rating forms despite repeated requests.

Figure E2 in Appendix E shows the number of two-year-old children originally recruited for direct assessment from providers in the Pilot and Comparison groups and these children's completion of different assessments during various phases of the project. There were 1,067 two-year-old children consented for the project from the selected classrooms (682 from Pilot group, 385 from Comparison group). Of these children, between 624 and 682 children from the Pilot group (92-100%) completed the pre-intervention direct assessments, and between 357 and 385 children from the Comparison group (93-100%) completed the pre-intervention direct assessments. The rate of return for teacher-completed ratings of children's socio-emotional development was lower (i.e., 66% of children from the Pilot group, 64% of children from the Comparison group). At the post-intervention assessment, between 543 and 547 of the 682 children originally recruited in the Pilot group completed the direct assessments (80% of children), and 310 of the 385 children originally recruited in the Comparison group completed the direct assessments (81% of children). Again, the rate of return for teachers' ratings of children socio-emotional development was lower (i.e., 40% of Pilot group, 36% of Comparison group). Rates of attrition from the direct assessment conducted by FCRR assessors were due to (a) providers opting out of the project following recruitment of children, (b) children leaving a provider's center prior to completion of assessments, and (c) child absences over the period of assessments. The rate of attrition from the teacher ratings was, in part, the result of providers

opting out of the project following recruitment of children, but, in many cases, teachers simply did not return the required rating forms despite repeated requests.

For both age groups of children, the rate of attrition was close to the rate of attrition for the providers (i.e., ~80%) for direct assessments conducted by FCRR assessment staff, and the rate of differential attrition was low. This level of attrition was within the level of attrition that is not expected to bias outcomes (e.g., see What Works Clearinghouse, *Procedures and Standards Handbook, version 3*). In contrast, the level of non-response for the teacher-completed measures of children's socio-emotional development (i.e., 63-55% non-response at post-intervention assessment) was high, and there was differential non-response (8% in three- to five-year-old group, 4% in two-year-old group, at post-intervention assessment). Consequently, teacher-rated socio-emotional behavior may not be a trustworthy indicator of children's development in these domains between children who attended centers in the Pilot and the Comparison groups.

Provider Outcomes

Classrooms of providers that participated in the pilot project were observed using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). CLASS data was obtained from FL-OEL. CLASS data was collected by observers who were employed either by the early learning coalitions or by their respective subcontractors. However, because the evaluators were not involved in the collection of CLASS data directly, it is unknown the extent to which CLASS observers met certification standards for use of the CLASS or the degree to which observers were blind to assigned condition of a provider (i.e., Pilot versus Comparison).

Using the CLASS, observers rate classrooms on indicators of 11 dimensions using a 7-point Likert scale. The 11 CLASS dimensions yield three subscales, Emotional Support (includes positive climate, negative climate, teacher sensitivity, and regard for student perspectives dimensions), Instructional Support (includes concept development, quality of feedback, language modeling, and literacy focus dimensions), and Classroom Organization (includes behavior management, productivity, and instructional learning formats dimensions). Versions of the CLASS have been used widely in both prekindergarten and older grades and have strong psychometric characteristics (Burchinal et al., 2008; Cash, Hamre, Pianta, & Myers, 2012).

Each CLASS dimensions is rated on a 1 - 7 scale, with ratings of 1 or 2 indicating low quality, ratings of 3 - 5, indicating mid-range quality, and ratings of 6 or 7 indicating high

quality. The CLASS Domain scores represent the average ratings on each of the relevant dimensions. The CLASS used with classrooms serving two-year-old children includes only the Emotional Support and Instructional Support domains. The CLASS used with classrooms serving three- to five-year-old children included all three domains.

CLASS data were collected on participating providers at two points in time--a pre-intervention observation and a post-intervention observation. Pre-intervention observations were conducted on 1,170 classrooms from October 2014 to February 2015. The majority of observations were conducted in November 2014. Post-intervention observations were conducted on 1,066 classrooms between February 2015 and July 2015. The majority of observations were conducted in May 2015. Only the 1,041 classrooms (472 classrooms serving two-year-old children, 569 classrooms serving three- to five-year-old children) with both pre-intervention and post-intervention observations were used for analyses. Of these 1,041 classrooms, 383 were Comparison classrooms and 658 were Pilot classrooms.

Average CLASS scores for classrooms serving two-year-old children and classrooms serving three to five-year-old children that had CLASS observations at both the pre-intervention and the post-intervention observations are shown in Table 1. As seen in the table, at pre-intervention, both classrooms serving two-year-old children and classrooms serving three- to five-year-old children were rated as having mid-level quality on the Emotional Support domain and low quality on the Instructional Support domain. Classrooms serving three- to five-year-old children were rated as having mid-level quality on the Classroom Organization domain. Statistical analyses using mixed models indicated that there were no differences between Comparison and Pilot classrooms pre-intervention. Effect sizes for the differences were all small and not statistically significant.

Analyses of post-intervention CLASS scores took into account a classroom's pre-intervention CLASS score in the same domain and the number of days between the pre-intervention observations and the post-intervention observation. Analyses were conducted using mixed models (i.e., classrooms were nested within providers), and all covariates were evaluated for homogeneity of regression. Average post-intervention CLASS scores--adjusted for pre-intervention CLASS scores and other covariates--for classrooms serving two-year-old children and classrooms serving three to five-year-old children are shown in Table 2. As seen in the table, at post-intervention, Comparison classrooms serving two-year-old children and Comparison

Table 1

Average CLASS Scores for Comparison and Pilot Classrooms Pre-Intervention

CLASS Domain	Comparison Classrooms			Pilot Classrooms			Hedges g
	Mean	(*SD*)	*N* Classes	Mean	(*SD*)	*N* Classes	
Classrooms with Two-Year-Old Children							
Emotional Support	5.18	(.91)	172	5.19	(.92)	300	.01
Instructional Support	2.72	(1.03)	172	2.65	(1.12)	300	-.06
Classrooms with Three- to Five-Year-Old Children							
Classroom Organization	4.72	(.96)	211	4.78	(.97)	358	.06
Emotional Support	5.38	(.78)	211	5.45	(.81)	358	.09
Instructional Support	2.19	(.88)	211	2.16	(.82)	358	-.04

Note. Hedges g is a standardized effect size of differences between groups; no group difference was statistically significant at $p < .05$.

classrooms serving three- to five-year-old children continued to be rated as having mid-level quality on the Emotional Support domain. Classrooms serving three- to five-year-old children continued to be rated as having low quality on the Instructional Support domain and mid-level quality on the Classroom Organization domain. For classrooms serving two-year-old children, ratings for classrooms in the Pilot group crossed the threshold into mid-level quality range on the Instructional Support domain, but ratings for classrooms in the Comparison group continued to be in the low-quality range on the Instructional Support domain. There were statistically significant and positive impacts of the pilot program on all ratings on the CLASS both for classrooms serving two-year-old children and for classrooms serving three- to five-year-old children, and effect sizes were in the moderate to large range. That is, classrooms in the Pilot condition were rated as having .43 to .68 standard deviation units higher quality than were classrooms in the Comparison condition.

In addition to these main effects of the pilot program, there were a number of significant interactions involving the covariates. For classrooms serving two-year-old children, experimental group interacted with pre-intervention scores on the CLASS Instruction Support scale. For classrooms initially one standard deviation below the mean pre-intervention score, the impact of

Table 2

Adjusted Average CLASS Scores for Comparison and Pilot Classrooms Post-Intervention

CLASS Domain	Comparison Classrooms			Pilot Classrooms			Hedges g
	Mean	(*SD*)	*N* Classes	Mean	(*SD*)	*N* Classes	
Classrooms with Two-Year-Old Children							
Emotional Support	5.17	(.90)	172	5.55	(.87)	300	.43***
Instructional Support	2.74	(1.10)	172	3.54	(1.30)	300	.65***
Classrooms with Three- to Five-Year-Old Children							
Classroom Organization	4.88	(.97)	211	5.33	(.93)	358	.48***
Emotional Support	5.44	(.81)	211	5.91	(.67)	358	.65***
Instructional Support	2.19	(.90)	211	2.91	(1.14)	358	.68***

Note. Hedges g is a standardized effect size of differences between groups; ***pilot versus comparison group differences statistically significant at $p < .001$.

being in the Pilot group (Hedges g = .48, p < .001) was smaller than the impact at the mean, and for classrooms initially one standard deviation above the mean per-intervention score, the impact of being in the Pilot group (Hedges g = .83, p < .001) was larger than the impact at the mean. These interaction results indicate that for classrooms serving two-year-old children that were in the Pilot group, the effect of the intervention was greater when a classroom started out with higher Instruction Support ratings than it was when a classroom started out with lower Instruction Support ratings.

For classrooms serving three to five-year-old children, experimental group interacted with pre-intervention scores on the CLASS Instruction Support scale, and experimental group interacted with days between pre-intervention observation and post-intervention observation for the CLASS Emotional Support scale. For classrooms initially one standard deviation below the mean pre-intervention score on the CLASS Instructional Support scale, the impact of being in the Pilot group (Hedges g = .48, p < .001) was smaller than the impact at the mean, and for classrooms initially one standard deviation above the mean per-intervention score, the impact of being in the Pilot group (Hedges g = .88, p < .001) was larger than the impact at the mean. For classrooms for which the number of days between pre-intervention and post-intervention observations was one standard deviation below the mean, the impact of being in the Pilot group

on the CLASS Emotional Support scale (Hedges g = .48, p < .001) was smaller than the impact at the mean, and for classrooms for which the number of days between pre-intervention and post-intervention observations was one standard deviation above the mean, the impact of being in the Pilot group on the CLASS Emotional Support scale (Hedges g = .81, p < .001) was larger than the impact at the mean. These interaction results indicate that for classrooms serving three to five-year-old children that were in the Pilot group, the effect of the intervention was greater when a classroom started out with higher Instruction Support ratings than it was when a classroom started out with lower Instruction Support ratings, and the effect of the intervention was greater when more time passed between the pre-intervention observation and the post-intervention observation.

Child Outcomes

Measurement of Child Outcomes

Florida's Early Learning and Development Standards outline developmental expectations of young children (e.g., birth to 48 months) and the Florida Voluntary Pre-Kindergarten (VPK) Standards outline developmental expectation of 4-year-old children. The birth to 48-month standards include benchmarks in five domains (subdomains are listed in parentheses): (a) Physical Development (Gross motor, fine motor, self-help, and health), (b) Approaches to Learning (eagerness & curiosity, persistence, creativity & inventiveness, planning & reflection), (c) Socio-emotional (trust & emotional security, self-regulation, self-concept), (d) Language and Communication (Listening & Understanding, Early Reading, Early Writing), and (e) Cognitive Development and General Knowledge (Exploration & discovery, Concept development & memory, Problem-solving & creative expression). The VPK standards also include benchmarks in five domains (subdomains are listed in parentheses): (a) Physical Health (Health & Wellness, Self Help, Gross Motor Development, Fine Motor Development), (b) Approaches to Learning (Eagerness & Curiosity, Persistence, Creativity, Planning & Reflection), (c) Social and Emotional Development (Self-Regulation, Relationships, Social Problem Solving), (d) Language, Communication, and Emergent Literacy (Listening & Understanding, Speaking, Vocabulary, Sentences & Structure, Conversation, Emergent Reading, Emergent Writing), and (e) Cognitive Development and General Knowledge (Mathematical Thinking, Scientific Inquiry, Social Studies, Creative Expression Through The Arts).

Reliable and valid direct measures of children's cognitive/pre-academic development are available across the age group of children included in this study, including measures of language, early literacy, and early math skills. Measurement of other aspects of children's development is typically observational, including teacher and parent report of characteristics associated with the domains of social and emotional competence and approaches to learning. Studies of those aspects of children's school readiness skills that are most predictive of later academic outcomes have identified behaviors associated with attention and self-regulation as those that are strongly related to later outcomes in reading and mathematics (e.g., Duncan et al., 2007). Other aspects of socio-emotional development may be more related to socio-emotional outcomes even though they are not related to academic outcomes. Physical development and health are not typically conceptualized as malleable by general classroom curricular activities. Standards in these areas most likely serve as the metric against which early childhood educators can identify children who are outside of the normative range of development, allowing educators to make appropriate referrals for further evaluation.

Direct Assessments of Children

Oral language. Children's oral language skills form the basis of most other academic domains, including reading and mathematics, and vocabulary development is a central component of oral language skills. In this evaluation, children's vocabulary was assessed using the Expressive One-Word Picture Vocabulary Test (EOWPVT, Brownell, 2010). The EOWPVT involves asking children to name or label pictures of common objects, actions, concepts, and categories. The test is normed for use with individuals from two years of age through adulthood. The test has high reliability for the age group included in this evaluation, including internal consistency reliability (αs = .94 - .95) and test-retest reliability (r = .98). The EOWPVT has good evidence of validity, as indicated by correlations with other measures of vocabulary ranging from .43 - .95. Because the EOWPVT is normed for individuals from age two years through adulthood, all children selected to participate in the evaluation completed this measure. The measure is administered individually to children and takes approximately 10 to 15 minutes to complete, depending on the age of the child and the level of their vocabulary.

Early literacy. The early literacy skills that are most associated with later reading skill (i.e., word decoding) include phonological awareness and print knowledge (Lonigan,

Schatschneider, & Westberg, 2008; Whitehurst & Lonigan, 1998). These skills can be reliably assessed beginning around age three years (Lonigan, Burgess, Anthony, & Barker, 1998). To assess these early literacy skills of the children selected for the evaluation are three years of age and older, the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007) was used. The TOPEL is a nationally normed and standardized measure of preschool children's emergent literacy skills. The measure includes three subtests: Phonological Awareness, Print Knowledge, and Definitional Vocabulary. Each of the three subtests demonstrates high reliability for three- to five-year-old children, including internal consistency reliability (αs > .96), test-retest reliability (rs > .90), and inter-scorer agreement (r = .98). Each subtest also displays good validity, with correlations above .58 with other measures of similar constructs. Because of the overlap in assessed content between the Definitional Vocabulary subtest of the TOPEL and the EOWPVT, only the Phonological Awareness and Print Knowledge subtests of the TOPEL was used in this evaluation. The subtests are administered individually to children and together take approximately 15 to 20 minutes.

Early mathematics. Current conceptualizations of early mathematics skills are rooted in the concepts of formal and informal mathematics skills (Greenes, Ginsburg, & Balfanz, 2004; Starkey, Klein, & Wakeley, 2004). Formal mathematics skills are those skills taught in school that require the use of abstract numerical notation such as writing numerals, place-value tasks, knowledge of the base-ten mathematics system, and decimal knowledge. Informal mathematics skills are the developmental precursors to formal mathematics skills and do not require specific instruction in abstract mathematical notation. Research demonstrates strong continuity between early mathematics skills--including informal mathematics skills and later math outcomes (e.g., Duncan et al., 2007). In this evaluation, children's early mathematics skills were assessed with the Test of Early Mathematics (TEMA; Ginsburg & Baroody, 2003). The TEMA is a nationally normed and standardized measure of preschool children's informal and early mathematics skills. The measure has high reliability for three- to five-year-old children, including internal consistency reliability (αs = .92-.95) and test-retest reliability (rs = .82-.93). The TEMA has evidence of validity, with correlations of .54-.91 with other measures of mathematics skills. The TEMA is administered individually to children and takes approximately 15 to 20 minutes.

Assessment of children whose home language was Spanish. For children in the direct assessment sample who were judged to be Spanish speakers, the direct assessment battery was

augmented. Specifically, for 3- to 5-year-old children who were Spanish speakers, the 20-item language and 20-item code-related screening measures from the Spanish Preschool Early Literacy Assessment (SPELA; Lonigan, 2013) was administered at the pre-intervention and post-intervention assessments. At pre-intervention, 220 children completed the SPELA measures (17% of sample; 91 in Comparison group, 129 in Pilot group). At post-intervention, 175 children completed the SPELA measures (17% of sample that completed post-intervention assessments; 76 in Comparison group, 99 in Pilot group). These children came from 58 different providers at pre-intervention (19% of providers) and 48 providers at post-intervention (21% of participating providers at post-intervention). Consequently, the sample of children who completed the Spanish-language measures does not represent the full sample of participating providers.

Observational Assessments of Children

To measure changes in children's global socio-emotional development, including attention, social competence, as well as behaviors associated with internalizing and externalizing problems, two assessment strategies were be used. First, children's classroom teachers were asked to rate children using two well-validated measures of socio-emotional development, the Social Competence and Behavior Evaluation scales and the short-version of the Strengths and Weakness of ADHD-Symptoms and Normal Behaviors Rating Scale. Second, to provide another measure of children's self-regulation behaviors, assessment staff who completed the direct assessments of children completed the short version of the Strengths and Weakness of ADHD-Symptoms and Normal Behaviors Rating Scale at the end of each assessment session.

Strengths and Weaknesses of ADHD-Symptoms and Normal Behaviors Rating Scale (SWAN). The SWAN (Swanson et al., 2001) is a measure of children's self-regulation in the classroom context. The short version of the SWAN includes 18 items that correspond to the diagnostic criteria for ADHD (i.e., inattention, impulsive, and hyperactivity items). Using the SWAN, children are rated based on comparisons to same-age peers, with scores ranging from -3 to 3 for each item. Each SWAN subscale has strong internal consistency (i.e., $\alpha > .95$ in preschool samples) and test-retest reliability (*r*s =.71-.76; Allan et al., 2014; Lakes, Swanson, & Riggs, 2012). The SWAN is not a normed measure, but it has been used with children similar in age to the children in this sample.

Social Competence and Behavior Evaluation (SCBE-30). The SCBE-30 (Kotler & McMahon, 2002; La Freniere & Dumas, 1996) is a 30-item Likert-type rating scale of young children's prosocial behavior and problem behaviors, including social competence, anger/aggression (externalizing problems), and anxiety (internalizing problems) scales. These subscales of the SCBE-30 have proved reliable and useful in studies of young children's adjustment (e.g., Denham, Caverly, et al., 2002; La Freniere & Dumas, 1996; La Freniere et al., 2002). The SCBE-30 is not a normed measure, but it has been used with children similar in age to the children in this sample.

Preliminary Analyses and Outcomes

Preliminary data analyses were conducted to determine whether children's academic skills increased from pretest to posttest, regardless of whether the center they attended was in the Pilot or Comparison group (i.e. an overall examination of children's growth in skills over time). Results indicated that all academic skills (i.e., expressive vocabulary, print knowledge, phonological awareness, and math) showed significant gains from pretest to posttest beyond the gains expected due to increased age. On average, children's expressive vocabulary scores at posttest were 4.40 standard score points higher than they were at pretest, corresponding to an effect size of .29, $F(1, 2462) = 397.07$, $p < .001$. On average, children's math scores at posttest were 2.10 standard score points higher than they were at pretest, corresponding to an effect size of .14, $F(1, 1533) = 54.43$, $p < .001$. On average, children's phonological awareness scores at posttest were 6.09 standard score points higher than they were at pretest, corresponding to an effect size of .41, $F(1, 1526) = 217.52$, $p < .001$. On average, children's print knowledge scores at posttest were 2.99 standard score points higher than they were at pretest, corresponding to an effect size of .20, $F(1, 1526) = 112.50$, $p < .001$. Overall, these results indicate that, regardless of intervention group (i.e., Pilot or Comparison), children in the school readiness program had language, literacy, and math skills that increased more than what would be expected based on increases in age.

Pretest and posttest scores for the different conditions (i.e., Pilot versus Comparison) for the different age groups of children (i.e., two year olds, three to five year olds, combined) are displayed in Tables 3 through 5. All scores of academic outcomes used in analyses were standard scores adjusted for child age. Descriptive statistics for different age groups of children are

reported in Appendix A. Pretest mean scores reported in Tables 4 through 6 were adjusted for child age at pretest. Posttest mean scores reported in Tables 4 through 6 were adjusted for child age at pretest, scores on the outcome measure at pretest, and expressive vocabulary knowledge at pretest.

Outcomes for Two-Year-Old Children

Results of the impact analyses of the intervention on the academic and socio-emotional outcomes of two-year-old children are reported in Table 4. All analyses were conducted as mixed models (children nested in classrooms), with age and pretest score on the outcome measure used as covariates. Detailed results of specific main effects and interactions are reported in Appendix B. For children's expressive vocabulary knowledge, there was not a significant main effect of the intervention (i.e., Comparison vs. Pilot groups). However, the impact of the intervention differed depending on children's levels of expressive vocabulary knowledge at pretest (see Appendix B, Table B1). This finding indicated that the pilot program had different impacts on children's expressive vocabulary knowledge for children with low initial vocabulary knowledge than it did for children with high initial vocabulary knowledge. Specifically, for children with lower levels of expressive vocabulary knowledge at pretest, the Pilot group scored a non-significant 1.09 points higher on the expressive vocabulary measure at posttest than did the Comparison group. In contrast, for children with higher levels of expressive vocabulary knowledge at pretest, the Pilot group scored a significant 2.63 points lower on the expressive vocabulary measure at posttest than did the Comparison group. There was a significant main effect of expressive vocabulary knowledge at pretest. Children with higher initial levels of vocabulary knowledge also had higher levels of vocabulary knowledge at posttest. Child age did not significantly predict expressive vocabulary scores, and child age did not moderate the impact of the intervention on expressive vocabulary scores.

For examiner ratings of behavior on the SWAN, there were no significant main effects of the intervention. There were significant main effects of examiner ratings of behavior and expressive vocabulary scores at pretest. Children with higher examiner ratings of behavior (i.e., better attention, lower levels hyperactive/impulsive, and lower levels of oppositional defiant behaviors) at pretest also had higher examiner ratings of behavior at posttest. Children with higher levels of expressive vocabulary knowledge at pretest had better attention and fewer hyperactive/impulsive and oppositional defiant behaviors at posttest, according to examiner

Table 3

Pretest and Posttest Means for Comparison and Pilot Groups for Academic and Socioemotional Outcomes for Two-Year-Old Children

	Comparison Group				Pilot Group				
	Pretest		Posttest		Pretest		Posttest		
Outcome	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	ES
				Academic Outcome					
EV	77.86	(16.44)	87.21	(18.05)	80.21	(15.60)	86.38	(15.73)	-.05[a]
				Examiner-Rated Behavior Measures					
SWAN-INT	-.27	(.85)	-.07	(.53)	-.26	(.85)	-.07	(.58)	.00
SWAN-HI	-.19	(.82)	.01	(.55)	-.11	(.82)	.03	(.63)	.03
SWAN-ODD	-.18	(.74)	.00	(.29)	-.11	(.66)	.01	(.43)	.03
				Teacher-Rated Behavior Measures					
SWAN-INT	-.09	(.82)	.08	(.96)	-.11	(.88)	.24	(1.04)	.16
SWAN-HI	-.01	(.86)	-.01	(.94)	-.04	(.95)	.18	(1.02)	.19
SWAN-ODD	-.10	(.94)	-.02	(1.02)	-.05	(1.01)	.09	(1.06)	.11[a]
SCBE-AA	1.48	(1.03)	1.30	(.92)	1.36	(.87)	1.33	(.77)	.04[a]
SCBE-SC	2.57	(.92)	2.79	(.81)	2.49	(.80)	2.81	(.86)	.02
SCBE-ANX	1.14	(.57)	1.12	(.64)	1.12	(.68)	1.00	(.60)	-.20

Note. EV = Expressive Vocabulary. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety. PA = Phonological Awareness. PK = Print Knowledge. ES = Hedge's *g*.

[a] Effect of intervention condition moderated by Time 1 EV scores.

ratings. After controlling for expressive vocabulary knowledge and examiner ratings of behavior at pretest, younger children had fewer oppositional defiant behaviors at posttest than did older children. Age was not significantly related to examiner ratings of inattention or hyperactivity and impulsivity at posttest.

For teacher ratings of behavior on the SWAN, there were no significant main effects of the intervention; however, the effect of intervention on teacher ratings of oppositional defiant behavior was moderated by children's levels of expressive vocabulary knowledge at pretest (see

Appendix B, Table B3). This finding indicated that the pilot program had different impacts on child behavior for children with lower initial language skills than it did for children with high language skills. Specifically, among children with lower initial levels of expressive vocabulary knowledge, the Pilot group was rated a non-significant .34 points higher than was the Comparison group (fewer oppositional behaviors). In contrast, among children with higher initial levels of expressive vocabulary knowledge, the Pilot group was rated a non-significant .11 points lower than was the Comparison group (more oppositional behaviors). Children with higher teacher ratings of behavior at pretest also had higher teacher ratings of behavior at posttest for all subscales of the SWAN. Children with higher levels of expressive vocabulary knowledge at pretest had better attention and fewer hyperactive/impulsive behaviors at posttest; however, expressive vocabulary knowledge was not significantly related to teacher ratings of oppositional defiant behavior. Younger children had better attention than did older children. Child age was not significantly related to teacher ratings of hyperactivity and impulsivity or oppositional defiant behavior.

For teacher ratings of behavior on the SCBE, there were no significant main effects of the intervention; however, children's levels of expressive vocabulary knowledge at pretest significantly moderated the effect of intervention group on teacher ratings of anger and aggression (see Appendix B, Table B4). This finding indicated that the intervention had a different impact on behavior for children with lower initial vocabulary knowledge than it did for children with higher initial vocabulary knowledge. Specifically, among children with lower initial levels of expressive vocabulary knowledge, the Pilot group was rated a non-significant .16 points lower on the Anger and Aggression subscale than was the Comparison group. In contrast, among children with higher initial levels of expressive vocabulary knowledge, the Pilot group was rated a non-significant .22 points higher on the Anger and Aggression subscale than was the Comparison group. Teacher ratings of behavior at pretest were significantly related to teacher ratings of behavior at posttest for all subscales of the SCBE, such that children with higher teacher ratings of behavior at pretest also had higher teacher ratings of behavior at posttest. Children's expressive vocabulary knowledge at pretest significantly predicted teacher ratings of social competence and anxiety at posttest. Specifically, children with higher levels of expressive vocabulary knowledge at pretest had higher teacher ratings of social competence and lower

teacher ratings of anxiety at posttest. No other main effects or interactions were statistically significant.

Outcomes for Three- to Five-Year-Old Children

Analyses of the impact of the intervention for three- to five-year-old children are reported in Table 5. Detailed results of specific main effects and interactions are reported in Appendix C. All analyses were conducted as mixed models (children nested in classrooms), with age and pretest score on the outcome measure used as covariates. For academic outcomes, there was no significant main effect of the intervention; however, the effect of the intervention on children's print knowledge at posttest was moderated by expressive vocabulary knowledge at pretest and child age (see Appendix C, Table C2). This finding indicated that the intervention had a different impact on print knowledge for children with more versus less vocabulary knowledge and for younger children than it did for older children. Specifically, among children with lower levels of vocabulary knowledge at pretest, the Pilot group scored a non-significant 1.64 points lower than the Comparison group, and among children with higher levels of vocabulary knowledge at pretest, the Pilot group scored a non-significant .57 points higher than the Comparison group. Among younger children, the Pilot group scored a non-significant 1.21 points higher at posttest than did the Comparison group. In contrast, among older children, the Pilot group scored a significant 1.84 points lower at posttest than did the Comparison group. Expressive vocabulary knowledge at pretest significantly predicted posttest scores for all academic outcomes, such that children with higher expressive vocabulary knowledge at pretest had higher scores on all academic skills at posttest. For all academic outcomes children's scores on the outcome measure at pretest significantly predicted posttest scores, such that children with higher scores at pretest also had higher scores at posttest. Child age significantly predicted expressive vocabulary and phonological awareness scores at posttest. Specifically, younger children had higher expressive vocabulary scores at posttest than did older children (after accounting for expressive vocabulary scores at pretest); however, older children had higher phonological awareness scores at posttest than did younger children. No other main effects or interactions were statistically significant.

For examiner ratings of behavior on the SWAN, there were no significant main effects of the intervention; however, the effect of the intervention on children's oppositional defiant behavior at posttest was moderated by children's oppositional defiant behavior at pretest (see Appendix C, Table C3). This finding indicated that the intervention had a different impact on

Table 4

Pretest and posttest means for intervention and comparison groups for academic and socioemotional outcomes for 3-5 year old children

	Comparison Group				Pilot Group				
	Pretest		Posttest		Pretest		Posttest		
Outcome	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	*SD*)	ES
					Academic Outcomes				
EV	90.85	(16.99)	94.24	(16.69)	91.01	(16.13)	94.95	(16.15)	.04
Math	85.45	(12.23)	87.87	(13.37)	85.34	(12.04)	87.62	(13.50)	-.02
PA	85.44	(15.84)	91.61	(16.88)	85.33	(16.20)	91.53	(17.14)	-.01
PK	96.71	(14.71)	100.69	(14.75)	97.55	(15.18)	100.23	(15.18)	-.03[a,b]
					Examiner-Rated Behavior Measures				
SWAN-INT	.11	(.73)	.07	(.65)	.05	(.81)	.03	(.61)	-.06
SWAN-HI	.07	(.70)	.01	(.64)	.02	(.76)	.02	(.66)	.02
SWAN-ODD	.06	(.53)	.04	(.36)	.02	(.51)	.03	(.35)	-.03[c]
					Teacher-Rated Behavior Measures				
SWAN-INT	.12	(.94)	.41	(.94)	.18	(1.06)	.39	(1.15)	-.02
SWAN-HI	.07	(.97)	.29	(.95)	.11	(1.08)	.25	(1.12)	-.04[a]
SWAN-ODD	.12	(1.10)	.25	(1.05)	.14	(1.21)	.31	(1.24)	-.05
SCBE-AA	1.25	(.93)	1.16	(.86)	1.24	(.92)	1.25	(.90)	.10
SCBE-SC	2.82	(.88)	3.16	(.83)	2.89	(.97)	3.08	(.95)	-.09
SCBE-ANX	1.06	(.66)	.98	(.61)	.98	(.67)	.95	(.63)	-.05

Note. EV = Expressive Vocabulary. PA = Print Knowledge. PK = Phonological Awareness. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety. PA = Phonological Awareness. PK = Print Knowledge. ES = Hedges' *g*.

*** $p < .0001$; ** $p < .01$; * $p < .05$.

[a] Effect of intervention condition moderated by Time 1 EV scores.

[b] Effect of intervention condition moderated by child age at Time 1.

[c] Effect of intervention condition moderated by examiner ratings at Time 1.

behavior for children who had higher levels of problematic behavior at pretest than it did for children with lower levels of problematic behavior at pretest. Specifically, among children with relatively more oppositional defiant behaviors at pretest, the Pilot group was rated a significant .06 points higher (denoting relatively fewer oppositional defiant behaviors) at posttest than was the Comparison group. In contrast, among children with fewer oppositional defiant behaviors at pretest, the Pilot group was rated a significant .07 points lower (denoting relatively more oppositional defiant behaviors) at posttest than was the Comparison group. For all subscales of the SWAN, children's scores on each subscale at pretest significantly predicted their score on the same subscale at posttest such that children with higher examiner ratings of behavior at pretest also had higher examiner ratings of behavior at posttest. In addition, child age at pretest significantly predicted posttest ratings on all subscales of the SWAN, and children's expressive vocabulary scores at pretest significantly predicted examiner ratings of inattention at posttest. Specifically, older children were rated as having better attention and fewer hyperactive/impulsive and fewer oppositional defiant behaviors than were younger children, and children with higher scores on the expressive vocabulary measure at pretest were rated as having better attention than were children with lower scores on the expressive vocabulary measure at pretest. No other main effects or interactions were statistically significant.

For teacher ratings of behavior on the SWAN, there was no significant main effect of the intervention; however, the effect of the intervention on children's hyperactive and impulsive behaviors at posttest was moderated by children's expressive vocabulary scores at pretest (see Appendix C, Table C4). This finding indicated that the impact of the intervention on children's behavior was different for children with lower initial vocabulary knowledge than it was for children with higher initial vocabulary knowledge. Specifically, among children with lower vocabulary knowledge at pretest, the Pilot group was rated a non-significant .10 points higher (denoting relatively fewer hyperactive and impulsive behaviors) at posttest than was the Comparison group. In contrast, among children with higher vocabulary knowledge at pretest, the Pilot group was rated a non-significant .18 points lower (denoting relatively more hyperactive and impulsive behaviors) at posttest than was the comparison group. For all subscales of the SWAN, children's scores on each subscale at pretest significantly predicted their score on the same subscale at posttest. Specifically, children with higher teacher ratings behavior at pretest also had higher teacher ratings of behavior at posttest. In addition, children's expressive

vocabulary scores at pretest significantly predicted teacher ratings on the Inattention and Hyperactivity/Impulsivity subscales of the SWAN at posttest such that children with higher scores on the expressive vocabulary measure were rated as having better attention and fewer hyperactive and impulsive behaviors than were children with lower scores on the expressive vocabulary measure. No other main effects or interactions were statistically significant.

For teacher ratings of behavior on the SCBE, there were no significant main effects of the intervention. Teacher ratings of behavior at pretest significantly predicted teacher ratings of behavior at posttest for all subscales of the SCBE. Specifically, children with higher teacher ratings of behavior at pretest also had higher teacher ratings of behavior at posttest. Children's levels of expressive vocabulary knowledge at pretest significantly predicted teacher ratings of children's social competence at posttest. Specifically, children with higher levels of expressive vocabulary at pretest had higher teacher ratings of social competence at posttest. No other main effects or interactions were statistically significant.

Outcomes for Combined Sample: All Age Groups

Analyses of the impact of the intervention for the combined sample of two-year-old and three- to five-year-old children are reported in Table 6. All analyses were conducted as mixed models (children nested in classrooms), with age and pretest score on the outcome measure used as covariates. Detailed results of specific main effects and interactions are reported in Appendix D. For children's expressive vocabulary knowledge, there was no significant main effect of the intervention. However, expressive vocabulary knowledge at pretest significantly predicted posttest expressive vocabulary scores, such that children with higher initial expressive vocabulary knowledge at pretest had higher expressive vocabulary knowledge at posttest. Child age also significantly predicted expressive vocabulary scores at posttest. Specifically, younger children had higher expressive vocabulary scores at posttest than did older children (after accounting for expressive vocabulary scores at pretest). No other main effects or interactions were statistically significant.

For examiner ratings of behavior on the SWAN, there were no significant main effects of the intervention; however, the effect of the intervention on children's oppositional defiant behavior at posttest was moderated by children's oppositional defiant behavior at pretest (see Appendix D, Table D2). This finding indicated that the intervention had a different impact on

Table 5

Pretest and Posttest Means for Pilot and Comparison Groups for Academic and Socioemotional Outcomes for the Combined Sample.

	Comparison Group				Pilot Group				
	Pretest		Posttest		Pretest		Posttest		
Outcome	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	ES
					Academic Outcome				
EV	86.36	(17.89)	91.77	(17.58)	87.26	(16.79)	91.91	(16.44)	-.01
					Examiner-Rated Behavior Measures				
SWAN-INT	-.03	(.79)	.03	(.61)	-.05	(.83)	-.01	(.60)	.02
SWAN-HI	-.02	(.75)	.02	(.61)	-.02	(.78)	.02	(.65)	.02
SWAN-ODD	-.03	(.62)	.02	(.34)	-.02	(.57)	.03	(.38)	-.03[a]
					Teacher-Rated Behavior Measures				
SWAN-INT	.06	(.91)	.33	(.95)	.09	(1.02)	.34	(1.12)	-.01
SWAN-HI	.04	(.94)	.22	(.95)	.07	(1.05)	.23	(1.09)	-.01[b]
SWAN-ODD	.05	(1.06)	.19	(1.04)	.07	(1.16)	.25	(1.20)	-.05[b]
SCBE-AA	1.32	(.97)	1.19	(.88)	1.29	(.91)	1.28	(.86)	-.10
SCBE-SC	2.73	(.90)	3.05	(.83)	2.76	(.94)	2.99	(.93)	.07
SCBE-ANX	1.09	(.63)	1.01	(.63)	1.03	(.68)	.96	(.62)	.08

Note. EV = Expressive Vocabulary. PA = Print Knowledge. PK = Phonological Awareness. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety. PA = Phonological Awareness. PK = Print Knowledge. ES = Hedges' *g*.

*** $p < .0001$; ** $p < .01$; * $p < .05$.

[a] Effect of intervention condition moderated by Time 1 examiner ratings.

[b] Effect of intervention condition moderated by child EV scores at Time 1.

behavior for children with higher initial levels of problematic behavior than it did for children with lower initial levels of problematic behavior. Specifically, among children with relatively more oppositional defiant behaviors at Pretest, the pilot group was rated a significant .06 points higher (denoting relatively fewer oppositional defiant behaviors) at posttest than was the Comparison group. In contrast, among children with fewer oppositional defiant behaviors at

pretest, the Pilot group was rated a non-significant .05 points lower (denoting relatively more oppositional defiant behaviors) at posttest than was the comparison group. For all subscales of the SWAN, children's scores on each subscale at pretest significantly predicted their score on the same subscale at posttest such children who had higher examiner ratings of behavior at test also had higher examiner ratings of behavior at posttest. In addition, child age at pretest and children's scores on the expressive vocabulary pretest assessment significantly predicted posttest ratings on the inattention subscale of the SWAN such that older children and children with higher expressive vocabulary scores had higher levels of examiner-rated attention at posttest. No other main effects or interactions were statistically significant.

For teacher ratings of behavior on the SWAN, there was a significant main effect of the intervention on children's scores on the hyperactivity/impulsivity subscale of the SWAN. There were no other significant main effects of the intervention on children's scores on the SWAN; however, the effect of the intervention on children's hyperactivity/impulsivity and on their oppositional defiant behaviors at posttest was moderated by children's initial expressive vocabulary (see Appendix D, Table D3). With regard to their scores on the hyperactivity/impulsivity subscale of the SWAN, among children lower initial vocabulary knowledge, the Pilot group was rated a non-significant .17 points higher (denoting relatively fewer hyperactive and impulsive behaviors) at posttest than was the Comparison group. In contrast, among children with higher initial vocabulary knowledge, the Pilot group was rated a non-significant .11 points lower (denoting relatively more hyperactive and impulsive behaviors) at posttest than was the comparison group. With regard to their scores on the oppositional defiant subscale of the SWAN, among children with lower initial vocabulary knowledge, the Pilot group was rated a non-significant .19 points higher (denoting relatively fewer oppositional defiant behaviors) at posttest than was the Comparison group. In contrast, among children with higher initial vocabulary knowledge, the Pilot group was rated a non-significant .07 points lower (denoting relatively more oppositional defiant behaviors) at posttest than was the comparison group. For the inattention subscale of the SWAN, teacher ratings of inattention at pretest significantly predicted children's scores at posttest such that children with high levels of attention at pretest had high levels of attention at posttest. In addition, children's expressive vocabulary scores at pretest significantly predicted teacher ratings on the inattention and hyperactivity/impulsivity subscales of the SWAN at posttest such that higher expressive

vocabulary scores were associated with better attention and fewer hyperactive and impulsive behaviors than were lower expressive vocabulary scores. No other main effects or interactions were statistically significant.

For teacher ratings of behavior on the SCBE, there were no significant main effects or interactions of the intervention. Teacher ratings of behavior at pretest significantly predicted teacher ratings of behavior at posttest for all subscales of the SCBE. Specifically, children with higher teacher ratings of behavior at pretest also had higher teacher ratings of behavior at posttest. Children's level of expressive vocabulary knowledge at pretest significantly predicted teacher ratings of children's social competence at posttest. Specifically, children with higher expressive vocabulary at pretest had high teacher ratings of social competence at posttest. No other main effects or interactions were statistically significant.

Outcomes for Children who Completed Spanish-Language Measures

Children who completed the Spanish-language measures experienced non-significant increases in scores on both the vocabulary assessment (p = .14) and the code assessment (p = .10). Analyses of the impact of the intervention for the children who completed the Spanish-language assessments are reported in Table 6. All analyses were conducted as mixed models (children nested in classrooms), with age and pretest score on the outcome measure used as

Table 6

Pretest and posttest means for intervention and comparison groups for Spanish-language vocabulary and code-related outcomes

	Comparison Group				Pilot Group				
	Pretest		Posttest		Pretest		Posttest		
Outcome	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	*SD*)	ES
				Academic Outcomes					
SPELA DV	6.16	(5.93)	7.51	(6.32)	5.48	(5.84)	7.24	(6.57)	-.12
SPELA Code	5.39	(4.65)	6.06	(3.90)	4.55	(4.23)	5.90	(4.60)	-.04

Note. SPELA = Spanish Preschool Early Literacy Assessment; DV = definitional vocabulary; Code = print knowledge and phonological awareness; ES = Hedges' *g*.

covariates. For children's Spanish vocabulary knowledge, there was no significant main effect of the intervention. Children's age and initial vocabulary scores significantly predicted vocabulary scores at posttest. Older children and children with higher initial vocabulary scores had higher vocabulary scores at posttest. No interactions were statistically significant. For children's Spanish code-related knowledge, there was no significant main effect of the intervention. Children's age and initial code scores significantly predicted vocabulary scores at posttest. Older children and children with higher initial code scores had higher vocabulary scores at posttest. No interactions were statistically significant.

Secondary Analyses

In addition to the primary questions concerning the impact of the pilot program, there were a number of additional questions concerning the degree to which characteristics of providers may increase or decrease the obtained scores on both child-level and provider-level outcomes. One specific question raised by a stakeholder concerned the effect of prior use of or exposure to the CLASS. Although complete data for these analyses was not provided by FL-OEL with sufficient time to conduct all of the analyses, preliminary analyses indicated that most provider and teacher characteristics did not moderate the impact of the intervention. Additional analyses are ongoing, and updates to this report will be provided when those analyses are completed.

Conclusions

This purpose of this evaluation was to determine if provision of a package of quality-enhancement activities resulted in measurably better quality metrics for classrooms and better outcomes for children who participated in Florida's School Readiness Program. The results of the evaluation demonstrated that providers assigned to the Pilot group (receipt of intervention package) had classrooms that were rated significantly higher in organization, emotional support, and instructional support than did providers that were assigned to the Comparison group (no receipt of intervention package). Despite these significant differences between classroom observations, however, at the post-intervention observation, the average classroom in the Pilot group was rated as providing mid-level quality for classroom organization and emotional support. Classrooms serving three- to five-year-old children were rated as low in quality for

instructional support, but classrooms serving two-year-old children were rated as providing mid-level quality for instructional support. Additionally, there was evidence that classrooms that initially had higher levels of quality on the instructional support domain had a greater increase in quality in this domain than did classrooms that had initially lower levels of quality in this domain. The effect of the intervention on the instructional support domain also appeared to increase with increasing time between the initial observation and the posttest observation.

The evaluation did not reveal any impact of the quality-enhancement program on child outcomes. That is, despite statistically significant and, in some cases, statistically large impacts of the quality enhancement program on observations of classroom quality, these effects did not result in measurable increases in children's language, pre-literacy, math, or socio-emotional skills. There were a small number of statistically significant interactions in which initial child characteristics moderated the impact of the intervention; however, most of these were not statistically significant when examined with follow-up analyses and/or were the result of effects in opposite directions at the extremes of the moderator variable (e.g., impact positive for younger children and impact negative for older children). It is possible that the short time frame between the initial assessment of children and the posttest decreased the likelihood of identifying impacts of the professional development program on children's skills.

As of this date, the final planned secondary analyses could not be completed because of incomplete and late-acquired data on provider, classroom, and teacher characteristics. However, preliminary analyses did not indicate that there these characteristics moderated the impact of the intervention in any meaningful way.

References

Allan, D.,M., Allan, N. P., Lerner, M. D., Farrington, A. L., & Lonigan, C. J. (2014). *Aspects of preschoolers' self-regulatory skills and the continuous performance test: A factor analysis*. Manuscript submitted for publication.

Brownell, R. (2010). *Expressive One-Word Picture Vocabulary Test-Fourth Edition*. Novato, CA: Academic Therapy.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12*, 140-153.

Cash, A. H., Hamre, B. K., Pianta, R. C., & Myers, S. S. (2012). Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly, 27*, 529-542.

Denham, S. A., Caverly, S., Schmidt, M., Blair, K., DeMulder, E., Caal, S., Hamada, H., & Mason, T. (2002). Preschool understanding of emotions: Contributions to classroom anger and aggression. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 901-916.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446.

Ginsburg, H. P. & Baroody, A. J. (2003). *Test of Early Mathematics Ability (3rd ed.)*, Pro-ed, Austin, TX.

Greenes, C., Ginsburg, H. P., & Balfanz, R. (2004). Big math for little kids. *Early Childhood Research Quarterly*, 19, 159 - 166.

Kotler, J. C., & McMahon, R. J. (2002). Differentiating anxious, aggressive and socially competent preschool children: Validation of the Social Competence and Behavior Evaluation-30 (parent version). *Behaviour Research and Therapy, 40*, 947-959.

La Freniere, P., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., Schreiner, S., Montirosso, R., & Frigerio, A. (2002). Cross-cultural analysis of social competence and behavior problems in preschoolers. *Early Education and Development, 13*, 201-219

LaFreniere, P. J. & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3-6 years: The short form (SCBE-30).*Psychological Assessment, 8*, 369-377.

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The reliability and validity of the English and Spanish Strengths and Weaknesses of ADHD and Normal Behavior rating scales in a preschool sample: Continuum measures of hyperactivity and inattention. *Journal of Attention Disorders, 16*, 510-516.

Lonigan, C. J. (2013). *Spanish Preschool Early Literacy Assessment*. Tallahassee, FL: Author.

Lonigan, C. J., Schatschneider, C., & Westberg, L. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. In *Developing Early*

Literacy: Report of the National Early Literacy Panel (pp. 55-106). Washington, DC: National Institute for Literacy.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. (2007). *Test of Preschool Early Literacy*. Austin, TX: ProEd.

Pianta, R. La Paro, K. M. and Hamre, B. (2008). Classroom Assessment Scoring System Manual, Pre-K. Baltimore, MD: Brookes.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19, 99 - 120.

Swanson, J.M., Schuck, S., Mann, M, Carlson, C., Hartman, k., Sergeant, J.,….McCleary, R. (2001). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN rating scales (Available at: http://adhd.net).

Whitehurst, G. J. & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872.

Appendix A

Descriptive statistics for measures used for direct assessment for two-year-old children, three-to-five year-old children, and combined samples at pre-intervention and post-intervention assessments.

Table A1

Descriptive statistics for measures of academic and socioemotional skills at Time 1 for 2-year-old children.

	Mean	*SD*	Range	N
Age	29.82	3.31	24 – 35	1067
EV	79.20	15.94	56 – 145	1067
E-SWAN-INT	-.25	.85	-3 – 3	981
E-SWAN-HI	-.11	.82	-3 – 3	982
E-SWAN-ODD	-.12	.69	-3 – 3	982
T-SWAN-INT	-.12	.86	-3 – 3	695
T-SWAN-HI	-.05	.92	-3 – 3	695
T-SWAN-ODD	-.07	.99	-3 – 3	695
SCBE-AA	1.41	.93	0 – 5	695
SCBE-SC	2.51	.84	0 – 5	695
SCBE-ANX	1.14	.65	0 – 5	695

Note. EV = Expressive Vocabulary. E = Examiner. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety.

Table A2

Descriptive statistics for measures of academic and socioemotional skills at Time 2 for 2-year-old children

	Mean	*SD*	Range	N
Age	33.37	3.35	26 – 40	859
EV	86.58	16.61	55 – 141	853
E-SWAN-INT	-.06	.55	-3 – 3	858
E-SWAN-HI	.03	.59	-3 – 3	857
E-SWAN-ODD	.01	.38	-3 – 3	857
T-SWAN-INT	.14	1.04	-3 – 3	410
T-SWAN-HI	.11	.97	-3 – 3	410
T-SWAN-ODD	.06	1.05	-3 – 3	410
SCBE-AA	1.33	.85	0 – 5	410
SCBE-SC	2.74	.86	0 – 5	410
SCBE-ANX	1.07	.67	0 – 5	410

Note. EV = Expressive Vocabulary. E = Examiner. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety.

Table A3

Descriptive statistics for measures of academic and socioemotional skills at Time 1 for 3- to 5-year-old children

	Mean	*SD*	Range	N
Age	49.03	8.41	36 – 70	1981
EV	90.94	16.44	55 – 137	1981
E-SWAN-INT	.06	.78	-3 – 3	1862
E-SWAN-HI	.03	.74	-3 – 3	1865
E-SWAN-ODD	.03	.52	-3 – 3	1864
T-SWAN-INT	.16	1.02	-3 – 3	1432
T-SWAN-HI	.10	1.05	-3 – 3	1432
T-SWAN-ODD	.15	1.17	-3 – 3	1432
SCBE-AA	1.24	.92	0 – 5	1432
SCBE-SC	2.85	.94	0 – 5	1432
SCBE-ANX	1.00	.67	0 – 5	1432
Math	85.34	12.11	57 – 137	1857
PA	85.06	16.07	54 – 135	1850
PK	97.18	15.02	58 – 145	1851

Note. EV = Expressive Vocabulary. E = Examiner. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety. PA = Phonological Awareness. PK = Print Knowledge.

Table A4

Descriptive statistics for measures of academic and socioemotional skills at Time 2 for 3- to 5-year-old children

	Mean	*SD*	Range	N
Age	52.76	8.43	38 – 73	1630
EV	94.57	16.35	55 – 137	1610
E-SWAN-INT	.04	.63	-3 – 3	1628
E-SWAN-HI	.00	.64	-3 – 3	1629
E-SWAN-ODD	.03	.35	-3 – 3	1629
T-SWAN-INT	.35	1.07	-3 – 3	839
T-SWAN-HI	.26	1.05	-3 – 3	839
T-SWAN-ODD	.27	1.17	-3 – 3	839
SCBE-AA	1.22	.87	0 – 5	839
SCBE-SC	3.07	.91	0 – 5	839
SCBE-ANX	.99	.63	0 – 5	839
Math	87.65	13.38	55 – 132	1577
PA	91.48	17.10	54 – 145	1570
PK	100.22	15.06	64 – 145	1570

Note. EV = Expressive Vocabulary. E = Examiner. T = Teacher. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety. TEMA = Test of Early Mathematics Ability. PA = Phonological Awareness. PK = Print Knowledge.

Table A5

Descriptive statistics for measures of academic and socioemotional skills at Time 1 for the combined sample

	Mean	*SD*	Range	N
Age	42.31	11.56	24 – 70	3048
EV	86.83	17.20	55 – 145	3048
E-SWAN-INT	-.05	.82	-3 – 3	2843
E-SWAN-HI	-.02	.77	-3 – 3	2847
E-SWAN-ODD	-.02	.59	-3 – 3	2846
T-SWAN-INT	.07	.98	-3 – 3	2127
T-SWAN-HI	.05	1.01	-3 – 3	2127
T-SWAN-ODD	.08	1.12	-3 – 3	2127
SCBE-AA	1.30	.93	0 – 5	2127
SCBE-SC	2.74	.92	0 – 5	2127
SCBE-ANX	1.05	.66	0 – 5	2127

Note. EOV = Expressive Vocabulary. E = Examiner. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety.

Table A6

Descriptive statistics for measures of academic and socioemotional skills at Time 2 for the combined sample

	Mean	*SD*	Range	N
Age	46.07	11.63	26 – 73	2489
EV	91.79	16.87	55 – 145	2463
E-SWAN-INT	.00	.61	-3 – 3	2486
E-SWAN-HI	.01	.63	-3 – 3	2486
E-SWAN-ODD	.02	.36	-3 – 3	2486
T-SWAN-INT	.28	1.07	-3 – 3	1249
T-SWAN-HI	.21	1.02	-3 – 3	1249
T-SWAN-ODD	.20	1.13	-3 – 3	1249
SCBE-AA	1.26	.87	0 – 5	1249
SCBE-SC	2.96	.91	0 – 5	1249
SCBE-ANX	1.02	.64	0 – 5	1249

Note. EOV = Expressive Vocabulary. E = Examiner. SWAN = Strengths and Weaknesses of ADHD and Normal Behavior Scale. INT = Inattention. HI = Hyperactivity/Impulsivity. ODD = Oppositional defiant disorder. SCBE = Social Competence and Behavior Evaluation. AA = Anger and aggression. SC = Social competence. ANX = Anxiety.

Appendix B

Detailed results for statistical analyses of impact of pilot program for two-year-old children.

Table B1

Impact of intervention on children's expressive vocabulary abilities at posttest for 2-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Expressive Vocabulary			
EV – T1	831.53	1	836.87	.000		
Age	.32	1	846.98	.569		
Group	.76	1	842.11	.384		
Group*EV	5.29	1	836.87	.022	.07	-.16*
Group*Age	.06	1	846.98	.811		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. EV = Expressive Vocabulary. T1 = Time 1.

* $p < .05$.

Table B2

Impact of intervention on SWAN examiner ratings at posttest for 2-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT – T1	44.77	1	690.57	.000		
Age	.54	1	779.75	.464		
EV – T1	15.70	1	763.60	.000		
Group	1.03	1	775.90	.310		
Group*INT	1.15	1	690.57	.284		
Group*Age	1.49	1	779.75	.223		
Group*EV	.08	1	763.60	.773		
			Hyperactivity/Impulsivitiy			
HI – T1	28.08	1	761.09	.000		
Age	.02	1	754.11	.887		
EV – T1	6.71	1	762.14	.010		
Group	.22	1	765.33	.640		
Group*HI	.64	1	761.09	.424		
Group*Age	.12	1	754.11	.731		
Group*EV	.10	1	762.14	.757		
			Oppositional Defiant			
ODD – T1	15.69	1	779.83	.000		
Age	4.50	1	666.04	.034		
EV – T1	5.97	1	672.57	.015		
Group	.12	1	690.44	.726		
Group*ODD	.05	1	779.83	.832		
Group*Age	.35	1	666.04	.557		
Group*EV	.08	1	672.57	.775		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder. T1 = Time 1.

Table B3

Impact of intervention on SWAN teacher ratings at posttest for 2-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT – T1	46.03	1	292.76	.000		
Age	9.09	1	267.69	.003		
EV – T1	15.12	1	257.83	.000		
Group	1.62	1	272.34	.205		
Group*INT	.00	1	292.76	.959		
Group*Age	.10	1	267.69	.752		
Group*EV	2.54	1	257.83	.112		
			Hyperactivity/Impulsivity			
HI – T1	42.53	1	290.68	.000		
Age	3.70	1	271.86	.056		
EV – T1	4.87	1	259.49	.028		
Group	.16	1	274.76	.688		
Group*HI	.01	1	290.68	.933		
Group*Age	.41	1	271.86	.522		
Group*EV	2.65	1	259.49	.105		
			Oppositional Defiant			
ODD – T1	63.73	1	277.06	.000		
Age	3.71	1	277.98	.055		
EV – T1	3.29	1	267.16	.071		
Group	.01	1	280.10	.908		
Group*ODD	1.01	1	277.06	.317		
Group*Age	1.24	1	277.98	.267		
Group*EV	4.73	1	267.16	.031	.33	-.11

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder. T1 = Time 1.

Table B4

Impact of intervention on SCBE ratings at posttest for 2-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Anger/Aggression			
AA – T1	144.04	1	292.79	.000		
Age	.01	1	290.14	.929		
EOW – T1	3.42	1	282.98	.065		
Group	.01	1	289.37	.937		
Group*AA	3.02	1	292.79	.083		
Group*Age	1.04	1	290.14	.308		
Group*EV	5.96	1	292.79	.015	-.19	.27
			Social Competence			
SC – T1	73.08	1	292.19	.000		
Age	2.27	1	279.31	.133		
EOW – T1	10.99	1	272.27	.001		
Group	4.70	1	283.23	.031		
Group*SC	.19	1	292.19	.660		
Group*Age	2.74	1	279.31	.099		
Group*EV	1.01	1	272.27	.316		
			Anxiety			
ANX – T1	74.94	1	290.23	.000		
Age	2.48	1	284.56	.116		
EOW – T1	6.97	1	281.35	.009		
Group	.22	1	284.92	.639		
Group*ANX	1.77	1	290.23	.184		
Group*Age	.58	1	284.56	.447		
Group*EV	.34	1	281.35	.561		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. AA = Anger and aggression. EV = Expressive Vocabulary. SC = Social Competence. ANX = Anxiety. T1 = Time 1.

Appendix C

Detailed results for statistical analyses of impact of pilot program for three to five-year-old children.

Table C1

Impact of intervention on children's expressive vocabulary and math abilities at posttest for 3- to 5-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Expressive Vocabulary			
EV – T1	2781.36	1	1552.34	.000		
Age	4.13	1	1265.93	.042		
Group	.06	1	1351.70	.815		
Group*EV	.11	1	1552.34	.746		
Group*Age	.13	1	1265.93	.721		
			Math			
Math – T1	725.11	1	1507.75	.000		
Age	.14	1	1323.62	.709		
EOW – T1	29.53	1	1515.78	.000		
Group	.11	1	1445.94	.735		
Group*Math	.01	1	1507.75	.938		
Group*Age	.19	1	1323.62	.659		
Group*EV	1.03	1	1515.78	.310		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. EV = Expressive Vocabulary. T1 = Time 1.

Table C2

Impact of intervention on children's TOPEL phonological awareness and print knowledge abilities at posttest for 3- to 5-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Phonological Awareness			
PA – T1	204.62	1	1517.06	.000		
Age	22.00	1	1287.04	.000		
EV – T1	116.76	1	1511.75	.000		
Group	.07	1	1302.82	.792		
Group*PA	.90	1	1517.06	.344		
Group*Age	2.32	1	1287.04	.128		
Group*EV	.09	1	1511.75	.769		
			Print Knowledge			
PK – T1	1168.09	1	1516.19	.000		
Age	3.39	1	1200.03	.066		
EV – T1	10.28	1	1450.05	.001		
Group	.47	1	1346.62	.494		
Group*PK	.14	1	1516.19	.709		
Group*Age	6.18	1	1200.03	.013	.08	-.12*
Group*EV	3.88	1	1465.41	.050	-.11	.04

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. PA = Phonological Awareness. EV = Expressive Vocabulary. PK = Print Knowledge T1 = Time 1.

* $p < .05$.

Table C3

Impact of intervention on examiner SWAN ratings at posttest for 3- to 5-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT-T1	150.46	1	1520.13	.000		
Age	44.39	1	1269.92	.000		
EV – T1	21.88	1	1482.92	.000		
Group	.47	1	1321.53	.492		
Group*INT	.51	1	1520.13	.473		
Group*Age	.35	1	1269.92	.553		
Group*EV	3.46	1	1482.92	.063		
			Hyperactivity/Impulsivity			
HI – T1	124.79	1	1476.19	.000		
Age	7.85	1	1366.86	.005		
EV – T1	1.15	1	1525.99	.285		
Group	.68	1	1407.47	.409		
Group*HI	2.49	1	1476.19	.115		
Group*Age	.98	1	1366.86	.322		
Group*EV	.02	1	1525.99	.904		
			Oppositional Defiant Behaviors			
ODD – T1	88.97	1	1471.48	.000		
Age	7.04	1	1432.29	.008		
EV – T1	1.50	1	1529.71	.220		
Group	1.16	1	1456.73	.283		
Group*ODD	15.75	1	1471.48	.000	.06*	-.06*
Group*Age	.56	1	1432.29	.456		
Group*EV	.50	1	1529.71	.482		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder.

* $p < .05$.

Table C4

Impact of intervention on teacher SWAN ratings at posttest for 3- to 5-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT-T1	167.49	1	676.78	.000		
Age	1.95	1	611.64	.163		
EV – T1	5.74	1	673.94	.017		
Group	.29	1	626.32	.592		
Group*INT	.03	1	676.78	.853		
Group*Age	.12	1	611.64	.727		
Group*EV	1.68	1	673.94	.195		
			Hyperactivity/Impulsivity			
HI – T1	175.19	1	670.64	.000		
Age	.01	1	628.87	.907		
EV – T1	5.83	1	676.67	.016		
Group	.08	1	641.90	.778		
Group*HI	.10	1	670.64	.752		
Group*Age	1.49	1	628.87	.223		
Group*EV	3.98	1	676.67	.046	.10	-.18
			Oppositional Defiant Behaviors			
ODD – T1	221.79	1	646.58	.000		
Age	.01	1	644.49	.939		
EV – T1	1.56	1	676.81	.213		
Group	.12	1	649.00	.732		
Group*ODD	.04	1	646.58	.849		
Group*Age	2.12	1	644.49	.146		
Group*EV	1.19	1	676.81	.276		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder.

Table C5

Impact of intervention on SCBE ratings at posttest for 3- to 5-year-old children

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Anger/Aggression			
AA – T1	372.76	1	676.45	.000		
Age	.30	1	531.74	.585		
EV – T1	.14	1	638.97	.708		
Group	.01	1	592.51	.924		
Group*AA	.00	1	676.45	.966		
Group*Age	.15	1	531.74	.696		
Group*EV	.09	1	638.97	.768		
			Social Competence			
SC – T1	200.61	1	669.48	.000		
Age	.19	1	641.08	.385		
EV – T1	3.94	1	676.49	.048		
Group	.06	1	658.88	.807		
Group*SC	.19	1	669.48	.660		
Group*Age	.19	1	641.08	.662		
Group*EV	.01	1	676.49	.910		
			Anxiety			
ANX – T1	168.24	1	667.67	.000		
Age	.46	1	594.26	.496		
EV – T1	.52	1	670.06	.471		
Group	.19	1	609.94	.662		
Group*ANX	.77	1	667.67	.381		
Group*Age	.05	1	594.26	.822		
Group*EV	.10	1	670.06	.750		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. AA = Anger and aggression. EV = Expressive Vocabulary. SC = Social Competence. ANX = Anxiety. T1 = Time 1.

Appendix D

Detailed results for statistical analyses of impact of pilot program for combined sample: i.e., two-year-old children and three- to five-year-old children.

Table D1

Impact of intervention on children's expressive vocabulary abilities at posttest for the combined sample

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
EV – T1	3469.69	1	2397.43	.000		
Age	8.08	1	1460.47	.005		
Group	.01	1	1707.39	.941		
Group*EV	.77	1	2397.43	.380		
Group*Age	1.69	1	1460.47	.194		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. EV = Expressive Vocabulary. T1 = Time 1.

Table D2

Impact of intervention on SWAN examiner ratings at posttest for the combined sample

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT-T1	196.24	1	2302.94	.000		
Age	15.80	1	1433.10	.000		
EV – T1	27.76	1	2282.78	.000		
Group	.30	1	1635.32	.583		
Group*INT	.67	1	2302.94	.414		
Group*Age	.47	1	1433.10	.495		
Group*EV	2.07	1	2282.78	.151		
HI – T1	158.05	1	2296.96	.000		
Age	.08	1	1709.19	.772		
EV – T1	1.98	1	2310.80	.160		
Group	.70	1	1860.98	.404		
Group*HI	2.25	1	2296.96	.134		
Group*Age	.42	1	1709.19	.519		
Group*EV	.20	1	2310.80	.652		
ODD – T1	99.93	1	2281.57	.000		
Age	.75	1	1906.52	.386		
EV – T1	3.45	1	2254.04	.063		
Group	.23	1	2016.28	.632		
Group*ODD	13.74	1	2281.57	.000	.14*	-.11
Group*Age	.00	1	1906.52	.995		
Group*EV	.29	1	2254.04	.588		

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder. T1 = Time 1.

* $p < .05$

Table D3

Impact of intervention on SWAN teacher ratings at posttest for the combined sample

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Inattention			
INT-T1	222.99	1	976.37	.000		
Age	.01	1	734.67	.915		
EV – T1	13.27	1	971.85	.000		
Group	3.43	1	810.35	.065		
Group*INT	.00	1	976.37	.987		
Group*Age	.16	1	734.67	.691		
Group*EV	3.53	1	971.85	.060		
HI – T1	217.80	1	837.66	.000		
Age	.02	1	762.46	.899		
EV – T1	9.01	1	963.64	.003		
Group	4.04	1	837.66	.045		
Group*HI	.03	1	964.58	.857		
Group*Age	.00	1	964.58	.958		
Group*EV	6.04	1	963.64	.014	.17	-.12
ODD – T1	300.20	1	934.90	.000		
Age	.00	1	793.42	.984		
EV – T1	3.69	1	959.51	.055		
Group	.73	1	853.76	.394		
Group*ODD	.05	1	934.90	.831		
Group*Age	1.70	1	793.42	.192		
Group*EV	4.28	1	959.51	.039	.17	-.06

Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. INT = Inattention. EV = Expressive Vocabulary. HI = Hyperactivity/Impulsivity. ODD = Oppositional Defiant Disorder. T1 = Time 1.

Table D4

Impact of intervention on SCBE ratings at posttest for the combined sample

	F-Test	Num *df*	Den *df*	*p*-value	ES -1 *SD*	ES +1 *SD*
			Anger/Aggression			
AA – T1	539.85	1	976.19	.000		
Age	.29	1	642.15	.589		
EV – T1	1.75	1	967.93	.187		
Group	.08	1	761.71	.785		
Group*AA	.67	1	976.19	.413		
Group*Age	.01	1	642.15	.943		
Group*EV	.71	1	967.93	.399		
			Social Competence			
SC – T1	274.69	1	969.26	.000		
Age	1.02	1	779.27	.313		
EV – T1	10.89	1	962.08	.001		
Group	.90	1	904.67	.342		
Group*SC	.19	1	969.26	.666		
Group*Age	.24	1	779.27	.628		
Group*EV	.57	1	962.08	.450		
			Anxiety			
ANX – T1	233.38	1	970.58	.000		
Age	.10	1	685.57	.755		
EV – T1	.56	1	977.90	.455		
Group	.01	1	789.79	.910		
Group*ANX	2.22	1	970.58	.136		
Group*Age	.03	1	685.57	.866		
Group*EV	.04	1	977.90	.846		

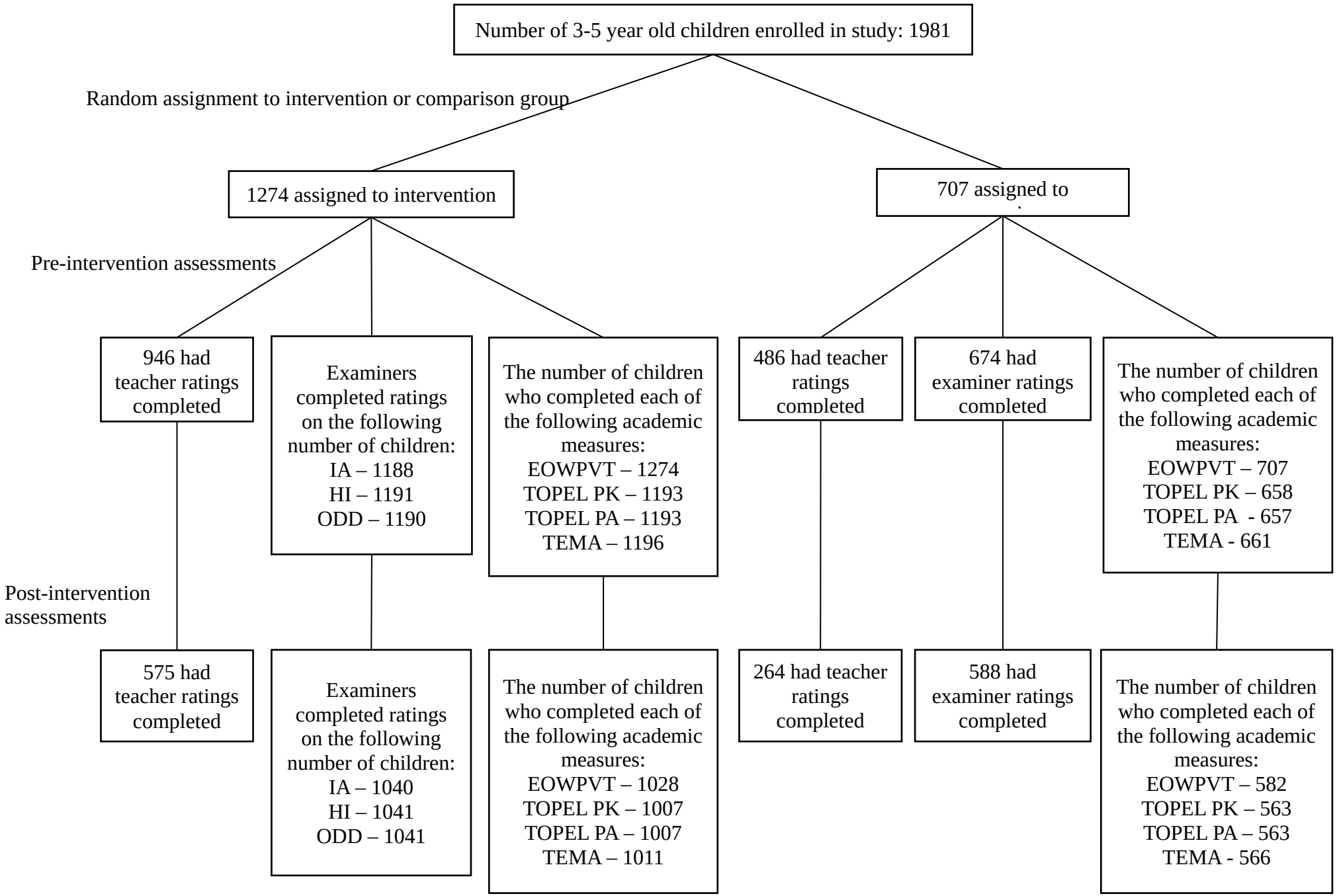
Note. Num = Numerator. Den = Denominator. 1 SD < M = Effect size of intervention group at one standard deviation below the mean of moderator. 1 SD > M = Effect size of intervention group at one standard deviation above the mean of moderator. AA = Anger and aggression. EV = Expressive Vocabulary. SC = Social Competence. ANX = Anxiety. T1 = Time 1.

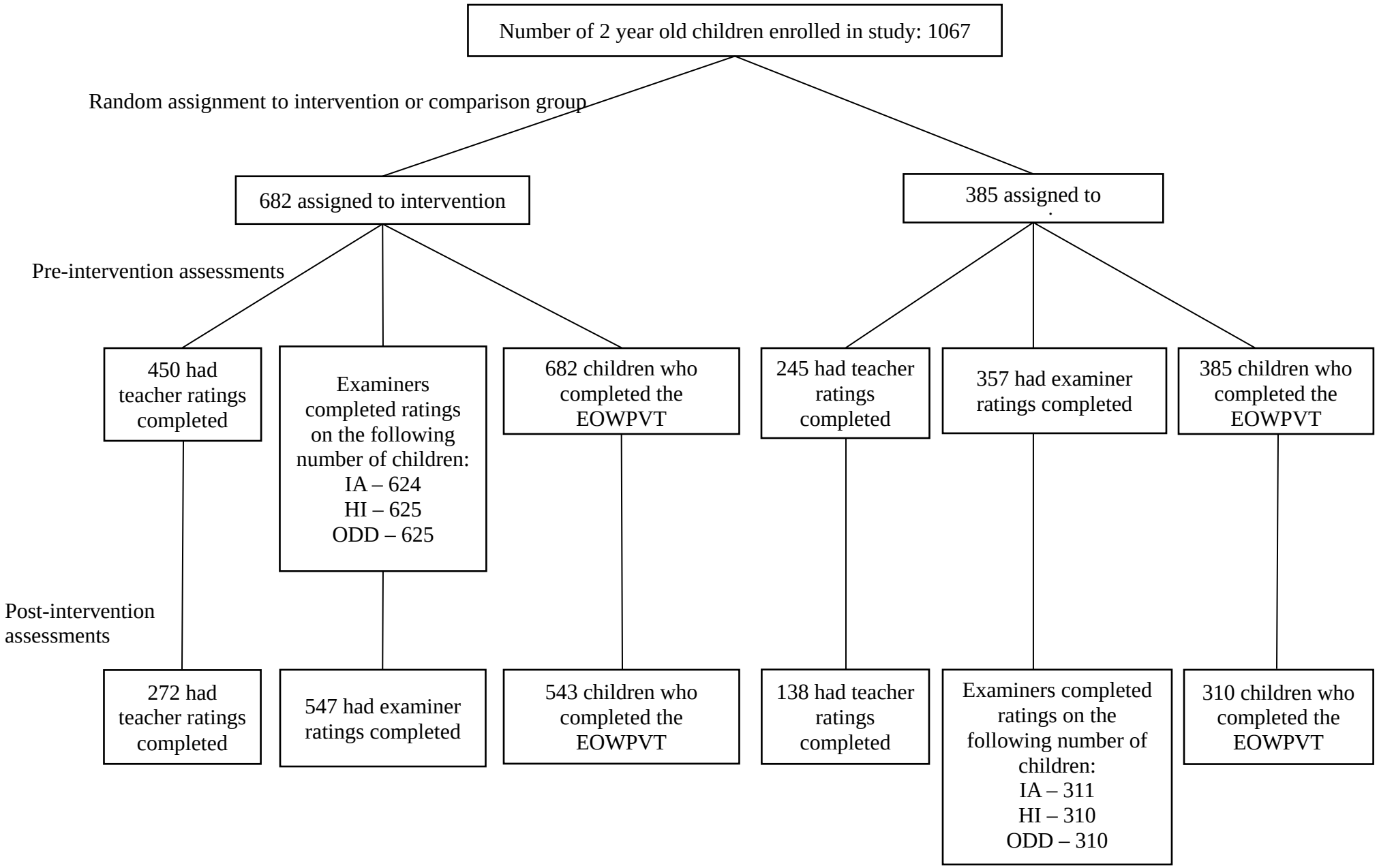
Appendix E

Consort diagram of children recruited and retained in Pilot and Comparison groups at the different phases of the evaluation.

Figure E1. Consort diagram from three- to five-year-old children

Figure E2. Consort diagram from two-year-old children


Number of 3-5 year old children enrolled in study: 1981
Random assignment to intervention or comparison group
1274 assigned to intervention
707 assigned to
Pre-intervention assessments
946 had teacher ratings completed
Examiners completed ratings on the following number of children:
IA – 1188
HI – 1191
ODD – 1190
The number of children who completed each of the following academic measures:
EOWPVT – 1274
TOPEL PK – 1193
TOPEL PA – 1193
TEMA – 1196
486 had teacher ratings completed
674 had examiner ratings completed
The number of children who completed each of the following academic measures:
EOWPVT – 707
TOPEL PK – 658
TOPEL PA - 657
TEMA - 661
Post-intervention assessments
575 had teacher ratings completed
Examiners completed ratings on the following number of children:
IA – 1040
HI – 1041
ODD – 1041
The number of children who completed each of the following academic measures:
EOWPVT – 1028
TOPEL PK – 1007
TOPEL PA – 1007
TEMA – 1011
264 had teacher ratings completed
588 had examiner ratings completed
The number of children who completed each of the following academic measures:
EOWPVT – 582
TOPEL PK – 563
TOPEL PA – 563
TEMA - 566


Number of 2 year old children enrolled in study: 1067
Random assignment to intervention or comparison group
682 assigned to intervention
385 assigned to
Pre-intervention assessments
450 had teacher ratings completed
Examiners completed ratings on the following number of children:
IA – 624
HI – 625
ODD – 625
682 children who completed the EOWPVT
245 had teacher ratings completed
357 had examiner ratings completed
385 children who completed the EOWPVT
Post-intervention assessments
272 had teacher ratings completed
547 had examiner ratings completed
543 children who completed the EOWPVT
138 had teacher ratings completed
Examiners completed ratings on the following number of children:
IA – 311
HI – 310
ODD – 310
310 children who completed the EOWPVT