Head Start Performance Measures Center
Family and Child Experiences Survey
(FACES 2000)
Technical Report

February 2006

Head Start Performance Measures Center
Family and Child Experiences Survey (FACES 2000)
Technical Report

February 2006

Prepared for:	Maria Woolverton Office of Planning, Research and Evaluation Administration for Children and Families U.S. Department of Health and Human Services Washington, DC

Prepared by:

Westat:	Nicholas Zill, Gary Resnick, Kwang Kim, Kevin O’Donnell, Alberto Sorongon, Yair Ziv, Souma Alva
Xtria:	Ruth Hubbell McKey, Shefali Pai-Samant, Cheryl Clark
The CDM Group, Inc:	Robert O'Brien, Mary Ann D'Elio

This report was prepared for the Administration for Children and Families, U.S. Department of Health and Human Services (DHHS) under contract HHS-105-96-1912, Head Start Quality Research Consortium's Performance Measures Center.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

List of Figures

EXECUTIVE SUMMARY

In 1997, Head Start launched the Family and Child Experiences Survey (FACES), a study with a nationally representative sample of Head Start programs to describe the characteristics, experiences, and outcomes for children and families served by Head Start. It also explores the relationships among family and program characteristics and outcomes. In 1997, the FACES design included a nationally representative sample of 3,200 3- and 4-year old children and their families in 40 programs. The FACES 2000 sample includes 2,800 children and their families in 43 different Head Start programs across the nation.

Each cohort of FACES employs a nationally stratified sample of Head Start programs, centers, classrooms, children, and parents. FACES 2000 features four phases of data collection and follows 3- and 4-year-old children from program entry, through one or two years of program experience, with followup in the Spring of kindergarten. The FACES 2000 battery has four main components: the direct child assessment, parent interview, teacher and staff interviews, and classroom observations. Although there is no non-Head Start comparison group in FACES, the use of assessment measures with national norms permits comparisons between the skills of children in the sample and children of the same ages in the norming samples.

ES.1 Study Findings

FACES provides information about the knowledge and skills that children have when they enter the Head Start program and the gains they make during the Head Start year and the first year of elementary school. It also describes the quality of Head Start classrooms, and factors that help explain variations in quality across Head Start classrooms. In addition, FACES 2000 data provide insight into the relationship of program and classroom characteristics to children's outcomes, as well as the relationship of family and parental characteristics to children's outcomes.

ES.2 Who is Head Start Serving?

- According to data from FACES 2000, Head Start is serving a diverse population of children. Slightly more than one-third (35 percent) of Head Start children were White followed closely by African American (32 percent) and Hispanic (28 percent). The majority of children (62 percent) were four and five years of age and 38 percent of the children were three years of age. Slightly more than two-fifths (41 percent) of the families were located in the southern region of the country and the vast majority of the families (70 percent) were situated in urban/suburban areas.

- Head Start families face numerous risks and challenges that are related to children's well being. Parents (mother or father) of Head Start children had a high school diploma or GED (41 percent), had a vocational/technical diploma or an Associate's degree (30 percent), or had a Bachelor's degree or higher (6 percent). Parents reported working on a full-time basis (39 percent) or on a part-time basis (14 percent), and 34 percent of parents were not in the labor force.

- The majority of Head Start families reported monthly family incomes of either $500 to $999 (27 percent) or $1,000 to $1,499 (25 percent). Nearly one-half (48 percent) of the children resided in dual-parent families whereas more than one-half (52 percent) of children resided in other family arrangements. One-half (50 percent) of the Head Start families were comprised of 4 to 5 persons.

ES.3 Head Start Children's Cognitive and Social-Emotional Outcomes

- Cognitive assessment data from FACES 2000 showed that most children entered Head Start with early literacy and math skills well below national averages. However, there was considerable diversity in skill levels among Head Start children. At Head Start entry, the highest quarter of Head Start children were at or above the national average (50^{th} percentile) in early language and number skills, while the lowest quarter of children ranked in the lowest 2 percent of all U.S. preschoolers in these areas.

- Children made gains toward national averages during the Head Start year, especially with respect to vocabulary knowledge and early writing skills. Children who entered the program with lower levels of knowledge and skill showed larger gains during the program year, yet still lagged considerably behind national averages. Children who started with higher assessment scores in the Fall wound up with higher scores in the Spring, but did not show the same level of gains as those with lower baseline scores.

- Language-minority children in Head Start showed significant gains in English vocabulary skills without declines in their Spanish vocabulary skills. They did not gain in letter recognition skills.

- Children who graduated after two years of Head Start made greater gains from entry to graduation than those who spent only one year in Head Start. Typically, children who graduate Head Start after 2 years enter with lower skills than one-year graduates. However, by the time both groups leave Head Start, they achieve similar levels of performance.

- Head Start graduates also showed further progress toward national averages during kindergarten. Gains of between one-quarter to more than three-quarters of a standard deviation were observed in vocabulary, early math, and early writing skills during kindergarten. Most Head Start graduates could identify the letters of the alphabet by the end of kindergarten and more than half could recognize beginning sounds of words. Further, Head Start graduates achieved early writing scores near the national norm by the end of kindergarten. Nevertheless, Head Start graduates remained behind their more advantaged peers in vocabulary and early math.

- FACES 2000 findings indicate that teachers reported significant positive changes on measures of social competence from Head Start entry to graduation.

- The findings also indicate that children who spent 2 years in Head Start were reported by teachers to display higher levels of cooperative classroom behavior and social skills and less hyperactive behavior at the end of Head Start compared to children who spent only 1 year in Head Start.

- Children entering Head Start with the lowest levels of social skills, and those who entered Head Start with the highest levels of problem behaviors, gained the most during their participation in Head Start. Although these children did not close the gap between themselves and other children, they did narrow it significantly.

- At the beginning of the Head Start year children’s disability status was highly associated with both social-emotional and cognitive developmental outcomes. On average, children with disabilities did not perform as well as other children on all baseline social-emotional and cognitive outcomes with the exception of book knowledge. White children were more likely than African American children to have a diagnosed disability, and boys were more likely than girls to have a disability, consistent with findings from other studies. Speech and language impairments constituted the largest group within the disabilities categorization.

- In addition, children with disabilities did not progress as well as other children on more than half (9 of 16) of the cognitive and social-emotional outcomes, including parent- and teacher-reported hyperactivity, withdrawn behavior, total problem behaviors, early math skills, perceptual–motor skills, and book knowledge. There were no statistically significant differences on the remaining seven outcomes between the two groups.

ES.4 Head Start Program's Use of Curricula

- Findings from FACES 2000 indicate that the great majority of Head Start programs use a curriculum, with the goal of providing a planned, developmentally appropriate early childhood program for children. Most teachers reported using a single curriculum. The most frequently used curricula are the Creative Curriculum (39.1 percent) and High/Scope curriculum (20 percent). Further, 93 percent of teachers reported receiving training in their curriculum, and 92 percent reported receiving ongoing support.

- Data from FACES 2000 also indicate a relationship between program characteristics and the type of curriculum used. Specifically, classrooms using a curriculum other than the Creative Curriculum or High/Scope curriculum served the poorest Head Start families and the highest percentage of non-White children. There was also regional variation in curricula used by classrooms. More teachers from the Northeast (78.7 percent), Midwest (59.4 percent), and West (69.4 percent) used either the Creative Curriculum or High/Scope curriculum compared to any other type of curricula, while the majority of teachers from the South (54.4 percent) used a curriculum other than the Creative Curriculum or High/Scope curriculum.

- Findings also demonstrate a relationship between type of curriculum and classroom quality. Classrooms using Creative Curriculum or High/Scope had significantly higher scores on three measures of classroom quality (the Early Childhood Environment Rating Scales-Revised—ECERS-R—Total score, ECERS-R Language subscale score, and a Quality Composite score) than classrooms using curricula categorized as "Other." Analyses were also conducted separating the "Other" curricula category into Widely Available Curricula and All Other Curricula. Classrooms using Creative Curriculum, High/Scope curriculum, and Widely Available Curricula had similar average ECERS-R Total scores and ECERS-R Language subscale scores, with classrooms using All Other Curricula receiving lower average scores on these instruments. However, on the Quality Composite measure, classrooms using Creative Curriculum or High/Scope had similar scores, with classrooms using either Widely Available Curricula or All Other Curricula receiving lower marks. Multivariate analyses also showed significant differences between classrooms using Creative Curriculum or High/Scope compared to All Other Curricula on all three classroom quality scores.

ES.5 Quality in Head Start Classrooms

- Head Start classrooms continue to show good levels of quality, based on the indicators of quality measured in FACES 2000 in the Fall 2000, Spring 2001, and Spring 2002. These levels of quality are consistent from FACES 1997, and the consistency is evident across a wide variety of the indicators, including child:adult ratio, teacher-child interactions, and classroom activities and materials. In fact, FACES shows that Head Start has a better, more limited range of quality than that seen in child care centers and preschools in several other national studies, with few Head Start classrooms scoring below minimal quality.

- Head Start teachers are qualified and experienced (although as a group they do not have the same level of credentials as public school teachers), and there appear to be substantially more teachers with higher educational attainment in this cohort compared with the 1997 FACES cohort. Head Start teachers in 2000 are younger, compared with those in 1997-1998, and more of them have been teaching in Head Start for two years or less. These newer teachers are also the ones most likely to have a graduate school degree.

- Classrooms with higher levels of quality are those whose teachers have higher levels of education, experience, and knowledge and attitudes of early childhood education practices. The relationship between teacher education and classroom quality is explained by teacher's attitudes and knowledge of early childhood education practices. Teachers who are more educated have more positive attitudes and knowledge, which translates into higher levels of classroom quality.

- Programs using one of the two most widely used integrated early childhood curricula (i.e., High/Scope and Creative Curriculum) also have teachers with positive attitudes and knowledge about early childhood education practice. Both of these factors appear to have the strongest effect on teacher sensitivity and responsiveness compared with other indicators of quality.

- Multilevel analyses suggest that variations in the quality of Head Start classrooms may be explained by characteristics of the families and children they serve, by the curriculum used in the program, and by teacher attitudes and knowledge about early childhood education practice. The results suggest that Head Start classroom quality may be affected by factors beyond the classroom door that are characteristics of the program and the families who participate.

ES.6 Relationship of Program and Classroom Characteristics to Children's Outcomes

- Program- and classroom-level variables provided unique and significant contributions to the prediction of children's assessment scores at Head Start entry and graduation, as well as gains from entry to graduation. Analysis of data from FACES 2000 showed that children's cognitive gains in Head Start were significantly related to use of the High/Scope, higher teacher salaries, teachers' educational credentials (Bachelor's degrees or Associate's degrees or higher), and provision of preschool services for a longer period each day.

- For each cognitive measure, children who attended Head Start for two years showed greater graduation scores and greater entry to graduation gains compared to their peers who attended Head Start for only one year.

- Surprisingly, within the narrow range of child:staff ratios in Head Start, higher child:staff ratios were associated with greater entry-graduation achievement gains in some of the cognitive outcomes. Further, teachers' scores on a scale assessing their knowledge of early childhood education practices were negatively linked with children's cognitive outcomes.

- The strongest predictors of children's social behaviors were predictors at the child and family level, mainly the child's age, gender, disability status, and number of years spent in Head Start, as well as the presence of both biological or adoptive parents in the house and the existence of books in the home. Most child and family characteristics had stronger links with the social behavior variables than program or classroom characteristics.

- The program- and especially the classroom-level characteristics showed relatively low power to predict children's social behaviors. Nonetheless, the addition of the program and classroom levels to the child- and family-level set of characteristics did result in higher explanatory power of the combined model over and above the child- and family-level model by itself. This addition was significantly higher in almost half of the variables examined. Thus, the inclusion of program- and classroom-level sets of variables in the combined models increased the models' overall fit to the data. This finding supports the expectation that the Head Start experience may serve to compensate for child and family disadvantages.

- Differences in cognitive gains between programs and classrooms possessing certain quality indicators (such as the use of High/Scope, higher teacher salaries, teachers with Bachelor's or Associate's degrees or higher, and provision of preschool services for a longer period each day) and programs and classrooms without those features, while statistically significant, were relatively modest in magnitude. By itself, each of the differences was not large enough to close the gap between where Head Start children typically end up at the end of the program year and the average achievement levels of American children at the start of elementary school.

- The ECERS-R Language subscale was not associated with greater gains on any of the cognitive measures. The failure to find significant links between children's cognitive gains in Head Start and class-level scores on the ECERS-R Language subscale may have to do with the generally good quality of Head Start classrooms and the limited range of variation in classroom quality that FACES found in its national samples of programs and classes. Studies encompassing broader ranges of quality of childcare and early education facilities have shown greater variations in classroom quality measures and significant relationships between quality measures and children's gains.

ES.7 Changes in Head Start Families Over Time

- Data from FACES 2000 indicate that the condition of Head Start children's families and especially their mothers has generally remained the same or marginally improved as children move from Head Start to kindergarten. Although not all mothers reside in the household of their children, the percentage remains consistently over 90 percent. A high percentage of mothers are also single, never married mothers. Over 40 percent of households have only the mother (and no father) residing in the child's household.

- The longitudinal analysis shows evidence of an improvement in the socioeconomic status of households over the years. The data show that a greater percentage of households report higher levels of monthly income in Spring 2002 and Spring 2003. Moreover, a smaller percentage of households remains in poverty or depends on welfare as children move through the Head Start program and on to kindergarten.

- Examining the relationship between these mothers' and household characteristics and children's cognitive outcomes using the FACES 2000 data shows that a change in the family situation represented in terms of a change in mothers' marital status or coresidence with parents bears little relationship to children's cognitive outcomes. Instead, household characteristics such as parents' education and household poverty status are more important. While greater parents' education results in better cognitive outcomes, the poor household economic situation is associated with lower assessment scores.

ES.8 Relationship of Family and Parental Characteristics to Children's Outcomes

- One-quarter of the parents were classified as moderately or severely depressed. Parents who were more depressed reported that their children had more problem behaviors and fewer positive social behaviors, a finding supported by the teachers' reports of children's behavior. Their children also had lower scores on one-to-one counting, creativity, design copying, early writing, letter identification, and early math assessments.

- More than one-fifth of the parents reported they had witnessed violent crime. Five percent reported being victims of violent crime in the neighborhood, while a similar percentage reported being victims of violence in their homes. Almost 10 percent of the children were reported to have witnessed domestic violence during the previous year. Less than 2 percent of the children were reported to have been victims of violent crime or victims of domestic violence. Positive correlations were found between increased exposure to neighborhood violence and reports of child problem behavior, while children in more violent neighborhoods had lower assessment scores on the color naming and book knowledge assessments.

- Almost 13 percent of the parents indicated that they have been victims of domestic violence. Teachers and parents reported children in these families had more overall problem behaviors.

- Almost one-half of the Head Start children lived in households with at least one individual who smoked cigarettes and about two-fifths of the households reported having at least one individual who drank alcohol. Living in a household with someone who drinks increased the risk of parental depression, while the children in these homes were reported to have more overall problem behavior and scored lower on vocabulary, color naming, and social awareness assessments.

- Almost one-fifth of the parents reported that someone in their household had been arrested and charged with a crime. Children in these families were more than 3 times more likely to have been a witness to violent crime or domestic violence in the past year. These children were also 3 times more likely to have been a victim of domestic violence or violent crime. These children had lower vocabulary scores, and were reported by both parents and teachers to be more aggressive and have more overall problem behaviors.

- At least one of a set of selected risk factors was evident in over 90 percent of the families. Nearly one-quarter of the families had four or more risk factors. Children in the families with four or more risk factors had lower parent ratings on emergent literacy and higher teacher and parent ratings of problem behavior. In the assessments, these children scored lower on design copying, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification.

- Analyses of child-level data show that Head Start children whose parents read to them more often show greater gains and graduation scores on some measures.

- Families engaged their children in a number of weekly and monthly activities. The number of activities was positively correlated with positive child behaviors and emergent literacy and negatively correlated with problem behaviors. In particular, the weekly activities had positive correlations with scores on the social awareness, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification tasks.

- Higher authoritative parenting style scores were significantly positively correlated with children’s social awareness, but not with any of the other cognitive measures. On the other hand, higher authoritarian scores were significantly negatively correlated with comprehension, color naming, vocabulary, and early math assessments.

- More than two-thirds of parents had attended parent-teacher conferences, observed in their children’s classrooms for at least 30 minutes, or met with a Head Start staff member in their homes. Parent involvement in Head Start was positively correlated with parental reports of positive social behavior and higher emergent literacy skills and negatively correlated with aggressive and overall problem behavior. Children whose parents were more involved in Head Start scored higher on vocabulary, book knowledge, early writing, early math, and letter identification tasks.

- Parent involvement at Head Start, parent reports that they and their children had positive experiences at Head Start, or parent satisfaction with the program significantly moderated relationships between risk factors (e.g., parental depression, exposure to violence and domestic violence, substance use, and involvement with the criminal justice system) and many negative child behavior and lower cognitive outcomes. These findings suggest that Head Start may play an important role in protecting families and children from the challenges that low-income families face.

ES.9 Predictive Validity of the FACES Cognitive and Behavioral Measures

- The FACES measures have strong predictive validity with outcomes at the end of kindergarten. As an indicator of early literacy skills, the cognitive measures show strong associations with reading ability at the end of the kindergarten year. As an indicator of school adjustment and social competence, the behavior ratings demonstrate ability to predict kindergarten behaviors that promote learning. The instruments used in FACES also predict whether a child gets promoted to first grade.

- The instruments used in FACES may tap different types of abilities ("inside-out" or decoding skills vs. "outside-in" or comprehension skills) that are important for children's future reading proficiency and academic achievement.

- In the assessment of children's social competencies, the use of parent and teacher ratings provides data on children's behavior skills in different situations and provides a comprehensive picture of their behavior. Equally important, both parent and teacher ratings significantly contribute to the prediction of social skills at the end of kindergarten. The parent and teacher ratings also significantly predict reading skills and general knowledge at the end of kindergarten. Ratings of problem behaviors were negatively correlated with kindergarten Reading and General Knowledge scale scores, suggesting that behaviors that may impede learning are associated with lower reading skills in kindergarten. High ratings of behaviors that enhance learning, positive approaches to learning and cooperative classroom behavior were positively correlated with kindergarten outcomes.

INTRODUCTION
HEAD START FACES 2000 TECHNICAL REPORT

In 1997, as part of the Head Start Program Performance Measures Initiative, the Head Start Bureau launched the Family and Child Experiences Survey (FACES), a study with a nationally representative sample of 3,200 children and their families, to describe the characteristics, experiences, and outcomes for children and families served by Head Start. FACES also observed the relationship among family and program characteristics and outcomes. In the Fall of 2000, Head Start began data collection on a new national cohort of FACES, called FACES 2000. The FACES 2000 sample included 2,800 children and their families in 43 different Head Start programs across the nation.

FACES complements other Head Start data in many ways. It provides Head Start with the capacity to report on important aspects of outcomes and quality and practices beyond aggregated administrative data. FACES provides the ability for Head Start to examine all facets of key outcomes and children’s school readiness on an ongoing basis. Prior to FACES, Head Start had no national data on child or family outcomes.

This introductory chapter provides a brief review of the research literature on early intervention programs and the link between program quality and children’s outcomes. In addition, it discusses the history, conceptual framework, and terminology of the Head Start Program Performance Measures Initiative, of which FACES is an integral component. The chapter also describes the research design of FACES, and closes with an overview of the report.

Research Context: Brief Review of the Effectiveness and Quality of Early Intervention Programs

A broad base of research exists demonstrating the effectiveness of high-quality, intensive early childhood education programs for improving children's cognitive and social outcomes. For example, the most influential studies demonstrating the positive effects of early childhood education for children living in poverty comprised the Consortium for Longitudinal Studies, including Ira Gordon's Parent Education Program, Francis Palmer's Harlem Training Project, and the High/Scope Perry Preschool Project, among others (Consortium for Longitudinal Studies, 1983; Lazar and Darlington, 1982). Across these studies, the children showed large positive immediate benefits from the programs and even years later, children who had attended preschool were less likely to have failed a grade in school or to have been assigned to special education classes.

Other program evaluations also indicate that high quality early interventions enhance children's cognitive and language development. For example, children participating in the Abecedarian Project had significantly better outcomes in reading achievement and grade retention rates than children in control groups (Seitz, 1990). Children participating in the Infant Health and Development Project also had enhanced cognitive and language development (IHDP, 1990). Recently, a national evaluation of the Early Head Start program (Administration for Children and Families, 2002) showed a consistent pattern of modest yet significant favorable impacts across a range of child and parent outcomes when children were 2- and 3-years-old, with larger impacts in several subgroups. Specifically, positive effects were found in the areas of children's cognitive, language, and social-emotional development.

Other early intervention research highlights that attaining positive outcomes for children can be elusive, particularly for programs that do not directly focus services on children. For example, an evaluation of the Comprehensive Child Development Program (CCDP), a program designed to help low-income families with children from birth to age 5 attain self-sufficiency and enhance child outcomes, showed that CCDP did not produce any important positive effects on families (St. Pierre, Layzer, Goodson, and Bernstein, 1997). Similarly, an evaluation of the New Chance Demonstration, a two-generation program designed to assist low-income young mothers achieve self-sufficiency and enhance the development of their children, found that the children's preschool readiness scores did not improve (Quint, Bos, and Polit, 1997).

Research on the Effectiveness of Head Start

While these evaluated programs share many common features with Head Start, such as the population served, goals, and program strategies, little large-scale research on Head Start's quality or outcomes of participating children and families has been conducted. However, numerous smaller scale studies have examined questions of Head Start's efficacy. In 1985, the Administration on Children, Youth and Families published the final report of the Head Start Evaluation, Synthesis and Utilization Project (McKey, Condelli, Ganson, Barrett, McConkey, and Plantz, 1985). This synthesis provides the most comprehensive assessment of Head Start's effectiveness, reporting the results of a meta-analysis of over

75 Head Start research studies. Overall, the analysis showed that Head Start produced immediate, meaningful gains in all areas of cognitive development, as well as in social behavior, achievement motivation, and health status. However, the cognitive and socio-emotional gains of Head Start children appeared to fade over time.

In addition, a study using data from the National Longitudinal Survey of Youth (NLSY) from 1979 through 1992 compared the outcomes of Head Start children to those of siblings who did not attend Head Start but either attended another type of preschool or did not attend preschool. The study found large and significant gains in receptive vocabulary for both the White and African American children attending Head Start over their siblings, as well as gains for Hispanic children (Currie and Thomas, 1995; Currie and Thomas, 1999). However, the gains of African American children faded by age 8. For White children, the effects of Head Start were greater than the effects of other preschool programs. Children who attended Head Start also had better health outcomes as measured by immunization status and height-for-age (Currie and Thomas, 1995). A second study using a subsample of children from the National Educational Longitudinal Survey who had attended Head Start explored possible reasons that the gains of African American children faded more quickly than those of White children. The results indicated that African American eighth grade students who had attended Head Start were enrolled in schools of significantly lower quality than African American students who did not attend Head Start. However, little difference was found in the quality of schools between White children who did and did not attend Head Start (Currie and Thomas, 1998).

Unfortunately, the data used in these studies is dated. To provide current information on Head Start's quality and effectiveness, the Administration for Children and Families (ACF) has undertaken the national Head Start Impact Study, a longitudinal study that involves approximately 5,000 3- and 4-year-old preschool children across an estimated 75 nationally representative grantee/delegate agencies in communities where there were more eligible children and families than could be served by the program. Data collection for this randomized study began in Fall 2002. The study's goals were to answer two important questions: How does Head Start affect the school readiness of children participating in the program as compared to children not enrolled in Head Start? Secondly, under which conditions does Head Start work best and for which children? More information about the Head Start Impact Study can be found at http://www.acf.hhs.gov/programs/opre/hs/impact_study/index.html.

The Link Between Program Quality and Outcomes for Children and Families

Research in the early childhood education field has pointed to the necessity of high-quality programming to achieve both short- and long-term positive results (Lazar and Darlington, 1982; Weikart and Schweinhart, 1991). While the success of programs such as the Perry Preschool Project and the Abecedarian Project is impressive, the intensity and duration of such programs are difficult and expensive to implement on a wide scale. Indeed, a review of the research found that "the general level of child care quality available to millions of young children is one that many experts consider to be inadequate, mediocre, and perhaps even detrimental to children's well-being. Furthermore, children from low socioeconomic status families under stress are more likely to receive lower-quality out-of-home care" (Love, Schochet, and Meckstroth, 1996, p. 12).

In light of such findings, researchers have continued to study the aspects of high-quality programs that lead to positive outcomes in a cost-effective manner. For example, Bryant, Burchinal, Lau, and Sparling (1994) found that Head Start classroom ratings on the Early Childhood Environment Rating Scale (ECERS) significantly predicted children's cognitive skills and school readiness. In addition, analyses from the NICHD Study of Early Child Care indicated that the overall quality of child care, and language stimulation in particular, were consistently, but modestly related to cognitive and language outcomes at ages 15, 24, and 36 months (NICHD Early Child Care Research Network, 2000). Furthermore, in the Cost, Quality, and Child Outcomes in Child Care Centers study, researchers from Colorado, California, North Carolina, and Connecticut found that children in higher quality classrooms had more advanced pre-mathematical skills than those in lower-rated classrooms (Cost, Quality and Child Outcomes Study Team, 1995). This study also found that staff:child ratios were the most significant determinant of center quality and that children at risk of not doing well in school were more sensitive to the negative effects of poor quality child care and received more benefits from high quality child care. Center quality also increased as the percentage of center staff with a high level of education increased, and with an increase in the center administrator's prior experience.

Another study, using a measure of classroom quality based on the National Association for the Education of Young Children's (NAEYC) Guidelines for Developmentally Appropriate Practice and the High/Scope curriculum, reported that children from classrooms with high quality scores scored significantly higher on a battery of cognitive skills in first grade (Frede, Austin, and Lindauer, 1993).

Howes, Phillips, and Whitebook (1992) reported that children in classrooms with better teacher: child ratios were more likely than children in classrooms with worse ratios to experience both appropriate caregiving and developmentally appropriate classroom activities. Children in classrooms rated higher on "appropriate caregiving" were more likely to be classified as secure, according to the Waters and Deane Attachment Q-Set. Similarly, children who attended preschool classrooms with 18 or fewer children were more likely than children in classrooms not meeting these standards to experience developmentally appropriate activities. These findings support earlier results by Ruopp, Travers, Glantz, and Coelen (1979), who demonstrated that smaller group sizes were consistently associated with better care and more socially active children. Additionally, children in classrooms where caregivers exhibited high levels of social interaction made greater gains in receptive language abilities.

The existing research base documenting the potential benefits of high-quality early intervention programs for low-income preschoolers, coupled with the dearth of recent national research on Head Start, made it critical for the Head Start Bureau to conduct a study of the program's quality and outcomes, particularly as the program underwent rapid expansion and entered an era of increased program accountability.

The Head Start Program Performance Measures Initiative

FACES was launched as a key component of the Head Start Program Performance Measures Initiative. In 1995, Head Start embarked on an initiative to design and implement an outcome-oriented accountability system. This initiative combines the best attributes of scientific research with program-level reporting and monitoring and is based on a consensus-driven set of criteria for program accountability.

The Head Start Program Performance Measures Initiative is a response to a specific legislative mandate, strategic planning for Head Start, and broader public emphasis on accountability and the general movement toward results-oriented evaluation. Specifically, the Program Performance Measures were developed in accordance with the recommendations of the Advisory Committee on Head Start Quality and Expansion, the mandate of Section 641A (b) of the Head Start Act (42 USC 9831 et seq.) as reauthorized in 1994, and the Government Performance and Results Act (GPRA)(Public Law 103-62).

The Head Start Act defines Program Performance Measures as "methods and procedures for measuring, annually and over longer periods, the quality and effectiveness of programs operated by Head Start agencies" that will be used to identify strengths and weaknesses in the Head Start program—both nationally and by region—and pinpoint areas requiring additional training and technical assistance.

In 1995, Head Start undertook a consensus-building process to develop the Head Start Program Performance Measures that drew on the opinions of Head Start program staff and parents; early childhood organization representatives; researchers; experts in the education, child development, and early intervention fields; and Head Start Bureau officials (Administration on Children, Youth and Families, 1995). In 1996 it began the FACES study, with data collection commencing in 1997. In 2000, Head Start collected data on the second cohort of FACES (FACES 2000).

Conceptual Framework

The conceptual framework for the Program Performance Measures Initiative unifies and organizes the Measures to display the linkages between process and outcome measures for Head Start children and families.[1] (See Figure 1 for the graphical representation of the framework.) The framework is based on the ultimate goal of Head Start, which is to promote the school readiness of children.

Head Start has adopted the "whole child" view of school readiness that was recommended by the Goal One Technical Planning Group of the National Education Goals Panel (Goal One Technical Planning Group, 1991, 1993). This view sees school readiness as a multifaceted phenomenon comprising five developmental domains that are important to the child's readiness for school: physical well-being and motor development, social and emotional development, approaches to learning, language usage and emerging literacy, and cognition and general knowledge. Each of these domains is represented in the

[1] See Head Start FACES: Longitudinal findings on program performance. Third progress report (ACYF, 2001) for a list of the 24 Program Performance Measures.

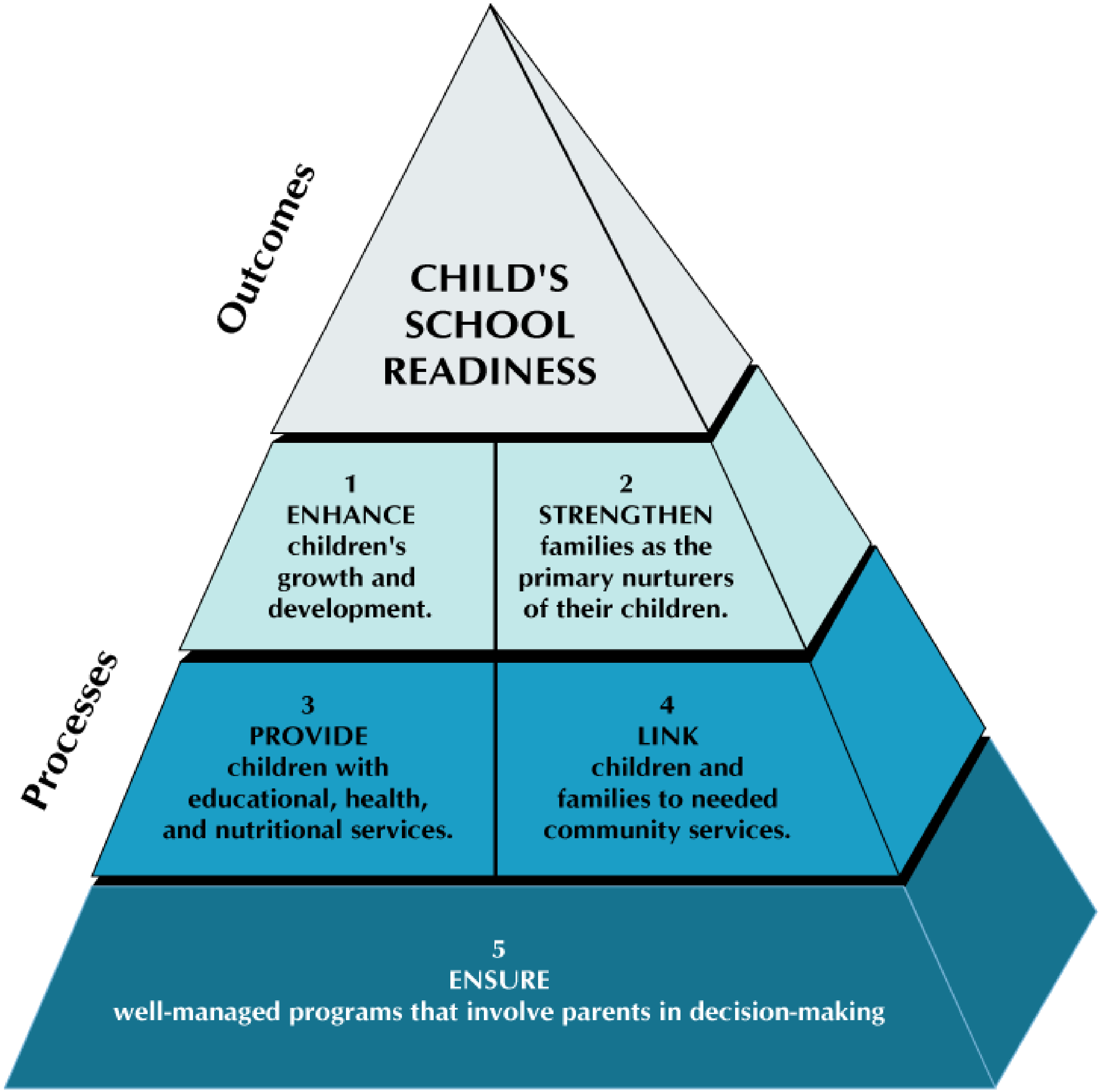


Figure 1. Head Start program performance measures conceptual framework

battery of measures being used in FACES to assess how well Head Start programs are performing. It takes into account the interrelatedness of cognitive, emotional, and social development; physical and mental health; and nutritional needs. School readiness is depicted at the top of the pyramid, with five objectives supporting it:

- Objective 1. Enhance children's healthy growth and development;
- Objective 2. Strengthen families as the primary nurturers of their children;
- Objective 3. Provide children with educational, health, and nutritional services;
- Objective 4. Link children and families to needed community services; and
- Objective 5. Ensure well-managed programs that involve parents in decision-making.

Each of these objectives is critical to helping children of low-income families attain their full potential. They also represent the cornerstones of the Head Start program. Objectives 1 and 2 represent outcomes or results that the program is designed to produce. Achieving both of these objectives is critical to the ultimate success of Head Start. As parent involvement and family support are key tenets of Head Start, both child and family-oriented outcome measures are included here. Objectives 3, 4, and 5 comprise the lower tiers of the pyramid and contain the process measures that are the basis for the attainment of Objectives 1 and 2 and the ultimate goal of enhancing children's social competence. An important aspect of the pyramid is the strong empirical connection between the provision of quality services (process measures) and improvements in child development (outcome measures).

The Head Start Family and Child Experiences Survey

The Head Start Family and Child Experiences Survey (FACES) is a central part of Head Start's Program Performance Measures Initiative. FACES is gathering comprehensive data on the cognitive, social, emotional, and physical development of Head Start children; the characteristics, well-being, and accomplishments of families; the quality of Head Start classrooms; and the characteristics, needs, and opinions of Head Start teachers and other program staff.

Each cohort of FACES employs a nationally representative sample of Head Start programs, centers, classrooms, children, and parents. The sample is stratified by three variables: region of the country (northeast, midwest, south, or west); urbanicity (urban vs. rural); and percentage of minority families in the program (50 percent or more vs. less than 50 percent).

The first cohort of FACES had six phases of data collection:

- Phase One involved a Spring 1997 field test in which approximately 2,400 children and parents were studied in a nationally stratified random sample of 40 Head Start programs. The field test was an opportunity to assess the feasibility of interviewing parents and assessing children on a large scale using the selected instruments. Although it was a field test, it provided valuable information on the status of Head Start programs, children, and families;

- Phases Two and Three of FACES occurred in Fall 1997 and Spring 1998 when data were collected on a sample of 3,200 children and families in the same 40 Head Start programs. Spring 1998 data collection included assessments of both Head Start children completing the program and Head Start graduates completing kindergarten (kindergarten field test), as well as interviews with their parents and ratings by their kindergarten teachers;

- Phase Four occurred in Spring 1999 with data collection in the 40 Head Start programs, plus a kindergarten followup for former Head Start children;

- Phase Five, in Spring 2000, completed the kindergarten followup for the children completing Head Start in Spring 1999 and first grade followup for children who completed Head Start in Spring 1998; and

- Phase Six, in Spring 2001, completed the first grade followup for the children who completed Head Start in Spring 1999.

These phases allow for pre-post comparisons, examining children and parents before their exposure to Head Start and determining their status at the end of the program and in the early school years.

Because Head Start is committed to regular, longitudinal, ongoing accountability measurements and program improvement, a new national cohort of FACES was launched in 2000. FACES 2000 sampled 2,800 children and their families from a new sample of 43 Head Start programs across the nation. This cohort features four phases of data collection and followed 3- and 4-year-old Head Start children from program entry through the Spring of their kindergarten year. Figure 2 presents the FACES 2000 study design.

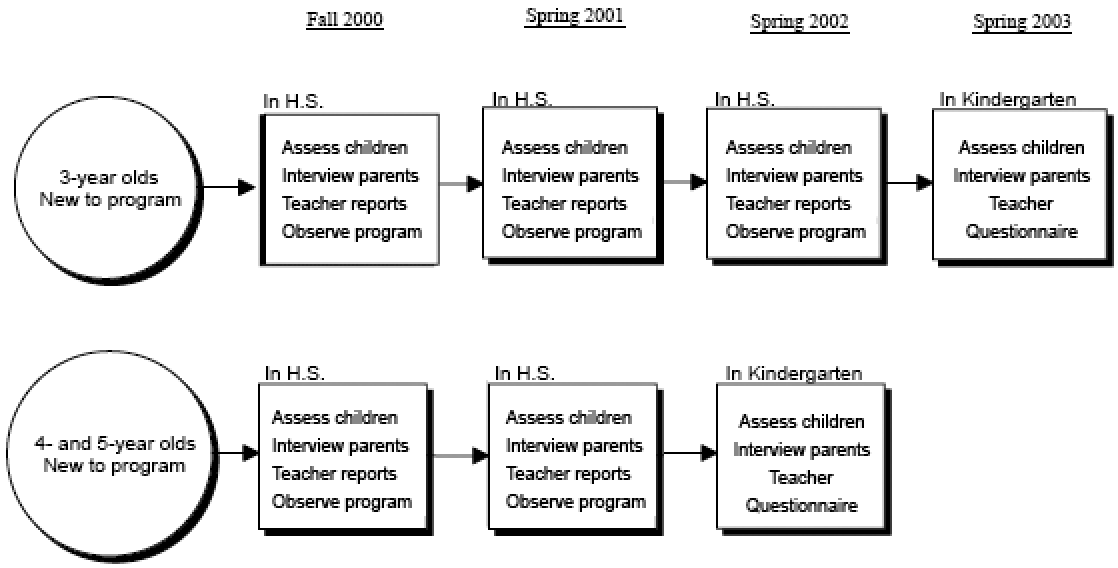


Figure 2. FACES 2000 study design

Survey Measures and Instruments

The FACES 2000 battery has four main components: the child assessment, parent interview, teacher and staff interviews, and classroom observations. The child outcomes include the major components of school readiness, and are collected through direct child assessments and rating scales completed by parents and teachers. Parent interviews are conducted with the primary caregiver of the Head Start child, and tap parenting behaviors, the socioeconomic characteristics of the family, and parental health and well-being. Interviews are administered to classroom teachers, center directors, program directors, and component coordinators to collect data on staff experience, education, and training, as well as knowledge and beliefs about child development and educational activities with children and parents. Classroom observations collect data on both the structure of the classroom and classroom processes, such as teacher-child interactions. This battery has remained largely the same since 1997, with some minor revisions based on field experiences and newly released versions of instruments.[2]

FACES Response Rates

Through the full cooperation of the Head Start programs studied, FACES has been able to achieve high instrument completion rates, over 85 percent return on all survey measures. A detailed list of response rates by study phase and respondent is available in the appendix.

The FACES Embedded Case Study

An additional feature of the 1997 cohort of FACES was the embedded case study of a longitudinal sample of 120 randomly selected families from the larger FACES sample (Vaden-Kiernan, D'Elio, and Sprague, 1999). The goal of the case study was to provide a more complete profile of Head Start families and children, their neighborhoods, and the nature of their interactions with Head Start. The FACES case study provides in-depth cross-sectional and longitudinal descriptive data, both qualitative and quantitative, over a two-year period. Findings from the case study are described in previous FACES reports, including *A Descriptive Study of Head Start Families: FACES Technical Report I,*

[2] See the appendix for details on data collection instruments.

(Administration on Children, Youth and Families, 2002), which integrates extensive data from the case studies with parent and staff interviews from the FACES national sample.

Dissemination of Study Findings

The FACES team strives to disseminate study findings in a timely fashion to the Head Start community as well as the general public. Several reports have been published detailing the development of Head Start's Program Performance Measures and findings from FACES from the Spring 1997 field test and the 1997 and 2000 cohorts. The following reports can be accessed at the website www.acf.hhs.gov/programs/opre/ongoing_research/faces/faces_pubs_reports.html:

- Head Start FACES (2000): A Whole Child Perspective on Program Performance - Fourth Progress Report (June 2003);
- A Descriptive Study of Head Start Families: FACES Technical Report I (January 2002);
- Head Start FACES: Reaching Out to Families: Head Start Recruitment and Enrollment Practices (January 2001);
- Head Start FACES (1997): Longitudinal Findings on Program Performance–Third Progress Report (January 2001);
- FACES Findings: New Research on Head Start Program Quality and Outcomes (June 2000); and
- Head Start FACES (Pilot): Program Performance Measures–Second Progress Report (June 1998).

The website also houses numerous papers presented at national conferences as well as details about the FACES study design and data collection instruments.

Overview of Report

This current document is the final report for FACES 2000. Subsequent chapters describe:

- Cognitive gains of Head Start children and their achievement in kindergarten;

- Changes in social skills and problem behavior in Head Start;
- Relationship between curricula and family, program, and classroom characteristics;
- Quality in Head Start programs and the factors that help explain variations in quality across Head Start classrooms;
- The relationship of program, classroom, and child characteristics and cognitive gains;
- The relationship of program, classroom, child, and family characteristics and social behavior in Head Start;
- The social-emotional and cognitive development of children with disabilities;
- Longitudinal changes in family structure and other household characteristics of Head Start children;
- Family risk factors, parental involvement, and Head Start's protective role; and
- Predictive validity of the FACES 2000 battery.

The technical appendix details the FACES study design, field test, data collection and processing, sample weights and analysis considerations, survey measures and derived variables, and data tables.

CHAPTER 1.
DEMOGRAPHIC CHARACTERISTICS OF HEAD START CHILDREN AND THEIR FAMILIES

In Fall 2000, slightly more than one-third (35 percent) of Head Start children were White, non-Hispanic followed by African American, non-Hispanic (32 percent) and Hispanic/Latino (28 percent) (see Table 1-1). Equal percentages of children were male and female (50 percent). The majority of children (62 percent) were 4 and 5 years of age and 38 percent of the children were 3 years of age.

Parents of Head Start children had a high school diploma or GED (41 percent), had vocational/technical diploma or Associate's degree (30 percent), or a Bachelor's degree or higher (6 percent). Parents reported working on a full-time basis (39 percent) or on a part-time basis (14 percent), and 34 percent of the parents were not in the labor force.

Twenty percent of the Head Start families reported family monthly incomes of $2,000 or more (see Table 1-1). However, the majority of Head Start families reported family monthly incomes of either $500 to $999 (27 percent) or $1,000 to $1,499 (25 percent). Nearly one-half (48 percent) of the children resided in dual-parent families whereas more than one-half (52 percent) of children resided in other family arrangements. One-half (50 percent) of the Head Start families were comprised of 4 to 5 persons. Slightly more than two-fifths (41 percent) of the families were located in the southern region of the country and the vast majority of the families (70 percent) lived in urban/suburban areas.

Table 1-1. Demographic characteristics of Head Start children and families – Fall 2000

Category	Percent
Child's Race/Ethnicity:	
White, non-Hispanic	35
African American, non-Hispanic	32
Hispanic/Latino	28
Asian or Pacific Islander	1
Multiracial/Biracial	3
Other Race	1
Gender:	
Male	50
Female	50
Child's Age:	
3 Years Old	38
4 Years Old	52
5 Years Old or Older	10
Parents' Highest Education Level:*	
Less than High School	23
High School Diploma or GED	41
Voc/Tech Diploma, Some College, or AA	30
Bachelor's Degree or Higher	6
Parent's Employment Status:**	
Working Full-Time	39
Working Part-Time	14
Looking for Work	7
Not in Labor Force	34
No Mother or Father in Household	6
Family Monthly Income:	
Less than $500	13
$500 to $999	27
$1,000 to $1,499	25
$1,500 to $1,999	15
$2,000 or More	20
Family Structure:	
Mother and Father	48
Mother Only	45
Father Only	2
Neither Mother nor Father	5
Family Size:	
2 or 3	26
4 or 5	50
6 or More	24

Table 1-1. Demographic characteristics of Head Start children and families – Fall 2000 (continued)

Category	Percent
Region:	
Northeast	14
South	41
Midwest	25
West	20
Urbanicity:	
Urban	70
Rural	30

Note: Column categories sum to 100 percent.

*The highest level of education for either the mother or father is reported here. If no education information is collected for either the mother or father and the respondent is someone else other than the mother or father, then the education level for the respondent is reported here.

**The employment level of either the mother respondent or father respondent is reported here. If there is no mother or father in the household, i.e., someone else is the respondent, then this is reported as no mother or father in the household.

The demographic characteristics were similar for children and their families who enrolled in Head Start for 1 year versus those who enrolled in Head Start for 2 years (see Table 1-2). However, some noteworthy differences pertain to characteristics such as race/ethnicity. White, non-Hispanic children represented 37 percent of children who enrolled in Head Start for one year; while 39 percent of children who enrolled in Head Start for 2 years were African American, non-Hispanic children. Twenty-five percent of parents whose children attended Head Start for 1 year did not have a high school diploma or GED as compared to 20 percent of parents whose children attended Head start for 2 years did not. Not surprisingly, 65 percent of children who had been in Head Start for 1 year were 4 years of age in the Fall of 2000 as compared to 68 percent of children who had been in Head Start for 2 years were 3 years of age in the Fall of 2000.

Table 1-2. Demographic characteristics of Head Start children and families by years of enrollment in Head Start program – Fall 2000

Category	Percent: Enrolled in Head Start for one year	Percent: Enrolled in Head Start for two years
Child's Race/Ethnicity:		
White, non-Hispanic	37	31
African American, non-Hispanic	27	39
Hispanic/Latino	31	25
Asian or Pacific Islander	1	1
Multiracial/Biracial	3	3
Other Race	1	1
Gender:		
Male	50	51
Female	50	49
Child's Age:		
3 Years Old	19	68
4 Years Old	65	32
5 Years Old or Older	16	NA
Parents' Highest Education Level:*		
Less than High School	25	20
High School Diploma or GED	40	43
Voc/Tech Diploma, Some College, or AA	29	31
Bachelor's Degree or Higher	6	6
Parent's Employment Status:**		
Working Full-Time	39	40
Working Part-Time	14	15
Looking for Work	7	6
Not in Labor Force	34	33
No Mother or Father in Household	6	6
Family Monthly Income:		
Less than $500	12	14
$500 to $999	26	28
$1,000 to $1,499	24	26
$1,500 to $1,999	16	15
$2,000 or More	22	17
Family Structure:		
Mother and Father	50	47
Mother Only	43	47
Father Only	2	2
Neither Mother nor Father	5	4
Family Size:		
2 or 3	25	26
4 or 5	50	49
6 or More	25	25

Table 1-2. Demographic characteristics of Head Start children and families by years of enrollment in Head Start program – Fall 2000 (continued)

Category	Percent: Enrolled in Head Start for one year	Percent: Enrolled in Head Start for two years
Region:		
Northeast	14	12
South	36	51
Midwest	26	22
West	24	15
Urbanicity:		
Urban	69	71
Rural	31	29

Note: Column categories sum to 100 percent.

Note: Number of years of attendance in Head Start is determined using parent interview, child assessment, and teacher-child report data through Spring 2002 though all reported percentages are based on Fall 2000 parent interview data.

*The highest level of education for either the mother or father is reported here. If no education information is collected for either the mother or father and the respondent is someone else other than the mother or father, then the education level for the respondent is reported.

**The employment level of either the mother respondent or father respondent is reported here. If there is no mother or father in the household, i.e., someone else is the respondent, then this is reported as no mother or father in the household.

CHAPTER 2.
COGNITIVE GAINS MADE BY HEAD START CHILDREN AND THEIR ACHIEVEMENT IN KINDERGARTEN

The Head Start Family and Child Experiences Survey (FACES) provides information about the knowledge and skills that children who attend Head Start have when they enter the program and the gains they make during the Head Start year and the first year of elementary school. The information is helpful in assessing how well the Head Start program is performing, and what changes and reforms may be needed to improve program performance. This information is gained through direct, one-on-one assessments of nationally representative samples of Head Start students in the Fall and Spring of the program and at the end of their kindergarten year. Although there is no non-Head Start comparison group in FACES, the use of assessment measures with national norms permits comparisons between the skills of children in the sample and children of the same ages in the norming samples. Analyses are also presented using criterion-referenced measures without national norms.

2.1 Research Questions

Assessment data from FACES 2000 were used to address the following research questions with respect to children's cognitive development:

1. What skills and knowledge do children have when they enter Head Start programs?
2. Do children make significant gains in knowledge and skills during the Head Start year? During the kindergarten year?
3. How do these gains vary across skill areas and among children who enter the program with lower or higher knowledge levels? and
4. Are there differences in the gains in knowledge and skills between children who attend Head Start for one year and those who attend for two years?

In making these comparisons, this chapter focuses on children who were assessed in English in both the Fall and Spring of the Head Start year. Information is also presented about the skills and knowledge of children who were initially assessed in Spanish because they came from Spanish-speaking homes and their knowledge of English was insufficient for testing in English in the Fall.

2.1.1 Composition of the Child Assessment

The child outcomes instruments are a series of tasks designed to appraise the children's cognitive and perceptual-motor development in areas such as word knowledge, letter recognition, early math, and early writing. These tasks have been shown to be predictive of later school achievement, especially of later reading proficiency and general knowledge (ACF, 2003). Instruments included:

- Peabody Picture Vocabulary Test, Third Edition (PPVT-III), (Dunn and Dunn, 1997), a nationally-normed test that measures children's word knowledge through asking children to show the meaning of spoken words by pointing to one of four pictures that best illustrates the meaning of the word;

- Letter-Word Identification, Applied Problems, and Dictation tasks of the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R), (Woodcock and Mather, 1989). These nationally-normed scales ask children to identify letters and words, solve simple addition and subtraction problems, and trace letters and write their own name; and

- McCarthy Draw-A-Design from the McCarthy Scales of Children's Abilities (McCarthy, 1972), a test with national norms in which children copy simple designs, such as a circle, a right angle, and a star.

The children were also asked to tell their own names, ages, and birthdays; recognize colors by name; count objects in one-to-one correspondence; and show familiarity with story books, understanding of print conventions, and comprehension of a simple story. Spanish-speaking children in the FACES sample were assessed in Spanish unless their teachers reported they had sufficient command of English to be assessed in that language. By Spring 2001, language-minority children in most Head Start programs were assessed in English. All child assessors were trained and monitored periodically by research staff. The assessment required 30 to 40 minutes per child.[1]

Cognitive measures with national norms available for comparison with the Head Start results included tests in vocabulary, early writing, letter identification, and early math. Raw scores are available for tests in book knowledge, design copying, one-to-one counting, color naming, and social awareness. In all analyses, statistical tests were carried out on weighted data using the WesVar analytic procedure. This allows for interpretation of the results for the population of Head Start children as a whole. To assist in interpreting the weighted results, the Technical Appendix contains supporting information on the weighting procedures.

---

[1] Please refer to the Technical Appendix for a complete listing of the instruments, their psychometric properties, and procedures used in FACES.

2.2 Findings

2.2.1 Emergent Literacy and Mathematics Skills of Head Start Children Compared with those of the General Population of Preschoolers

These analyses were conducted on children who were assessed in English at both Fall and Spring of the 2000-2001 Head Start program year.

2.2.1.1 Majority Entered Head Start with Academic Skills Below National Norms

A primary focus of FACES was to measure the knowledge and skills children brought with them when they entered the Head Start program, and how this varied across academic skill areas. The majority of children who entered Head Start in Fall 2000 came into the program with early literacy and numeracy skills that were less developed than those of most children of the same age (Table 2-1). FACES found that Head Start entrants had a mean standard score of 85.3 on the PPVT-III. They had mean standard scores of 92.4 on the Letter-Word Identification (LWI) pre-reading task, 85.1 on the dictation (early writing) task, and 87.9 on the Applied Problems (early math) task of the WJ-R achievement battery. Standard scores are constructed to have an overall mean of 100 and a standard deviation of 15, and are based on a sample of children of a given age, across all income levels. Thus, the literacy and numeracy skills that the average Head Start child brought to the program were from half a standard deviation to a full standard deviation below national averages.

On the criterion-referenced tasks, children scored in the middle range of the One-to-One Counting task (2.5 out of a possible 5), indicating that they could count fairly well, with few mistakes. They also scored in the middle range on the Color Naming and Social Awareness tasks. However, they scored in the low range on Book Knowledge (1.6 out of a possible 5) and on the Draw-A-Design task (2.9 out of a possible 19).

Table 2-1. Fall 2000 mean scores (SD) for FACES assessment measures on entry into Head Start

Skill area	Sample size	Mean (SD)		
		Overall	Bottom quartile	Top quartile
Norm referenced measures				
Vocabulary	1,801	85.3 (14.6)	67.0 (8.3)	102.8 (7.1)
Letter-word identification*	833	92.4 (9.6)	82.9 (4.4)	104.0 (8.8)
Early writing	799	85.1 (13.6)	70.8 (12.0)	100.5 (8.4)
Early math	859	87.9 (15.2)	69.0 (11.6)	104.7 (7.3)
Criterion referenced measures				
One-to-one counting	1,848	2.5 (1.2)	1.68 (.5)	3.8 (.9)
Color naming	1,893	11.3 (7.3)	1.73 (1.4)	19.7 (.5)
Social awareness	1,893	3.4 (1.7)	1.26 (.8)	5.6 (.5)
Book knowledge	1,876	1.6 (1.3)	.53 (.5)	3.4 (.6)
Draw-a-design	1,914	2.9 (1.3)	1.51 (.6)	4.5 (1.3)

* Sample sizes for the Letter-Word Identification, Early Writing (Dictation), and Early Math (Applied Problems) tasks of the Woodcock-Johnson Revised achievement battery are smaller than those of the other FACES measures because it was only administered to children who were 4-years old or older. Please see the Technical Appendix for details.

2.2.1.2 Diversity in Skills at Program Entry

Though most children had below-average literacy skills, FACES 2000 found considerable diversity in the Head Start population (Table 2-1). For example, mean standard scores for the highest quarter of children entering Head Start were at national averages: 102.8 in vocabulary, 104.0 in letter recognition, 104.7 in early math, and 101.1 in early writing skills. Thus, these students would rank above the 50th percentile when compared to all U.S. preschoolers. On the other hand, mean standard scores for the lowest quarter of Head Start children were two standard deviations or more below national averages: (e.g., 67.0 in vocabulary and 70.8 in early writing skills). These scores would rank the bottom quarter of Head Start students in the lowest 2 percent of all U.S. preschoolers.

2.2.2 Change in Knowledge and Skills Over the Head Start Year

A primary focus of FACES was to measure the extent of change in children's knowledge and skills from the Fall to the Spring of the Head Start year, especially in comparison to national averages for all children of the same ages. These changes were meant to serve as key indicators of the extent to which programs were enhancing children's school readiness. To determine the children's gains during their Head Start year, differences between the Fall and Spring mean scores were calculated for the norm-referenced word vocabulary, letter-word identification, early writing, early math measures; as well as the criterion-referenced book knowledge, one-to-one counting, social awareness, color naming, design copying, and color naming measures. That is, the gain scores were comprised of subtracting the Spring score from the Fall score, for each child in the study.[2] Fall-to-Spring gain scores were tested for significance using dependent samples t-tests. Effect sizes are provided to estimate how meaningful these gains are. In studies of large samples, it is often the case that differences may be statistically significant, but the size of the effect may be too small to be meaningful from an educational programming perspective. In this case, researchers use the effect size for determining the meaningfulness of statistical results.[3] When sample sizes are small, effect sizes need to be large in order to detect significant differences, but when sample sizes are large, as they are in this study, a very small effect size can be statistically significant but not meaningful from a policy or program perspective (see Cohen, 1988 or Rosenthal and Rosnow, 1984 for details on effect size).

Table 2-2 displays the Fall, Spring, and gain scores from Fall to Spring for the FACES sample, and provides the relevant effect sizes for these differences.

2.2.2.1 Gains in Vocabulary Knowledge and Writing Skills

Children in Head Start showed significant expansion of their vocabularies between the beginning and end of the program year. By the Spring of the Head Start year, mean standard scores were 89.1 for the PPVT-III and 87.1 on the WJ-R Dictation writing task. The mean standard score on the vocabulary test went up by 3.8 points ($p < .001$), or more than one quarter of a standard deviation. The

[2] Much has been written about the unreliability of difference scores derived from subtracting the Time 2 score from the Time 1 score when trying to estimate individual differences in growth. However, according to Willett (1988), the reliability in the difference score can actually be quite high under some conditions, particularly those favored by large-scale population-based survey methods. Difference scores for measuring change over two data points are not intrinsically as unreliable as the psychometric literature has argued and the difference score can be "an appealing and unbiased measure" of individual growth (Willett, 1988).

[3] The effect size is defined as an estimate of the magnitude of the relationship or difference between two or more variables. In this study, effect sizes were calculated as “Cohen’s d,” which was defined as the mean difference divided by the standard deviation for the underlying distribution, and is expressed in standard score units. Effect sizes of approximately .20 are considered “small,” such as the difference in mean IQ between twins and non-twins (see Cohen, 1988 for details). Moderate effect sizes typically approximate .50 and are conceived as those differences large enough to be clearly visible in behavior or attitudes. Large effect sizes are those above .80, and are represented, for example, by the mean IQ difference between holders of the Ph.D. degree and typical college freshman. However, the delineation of thresholds for considering an effect size to be small, moderate, or large may vary depending on the research question being asked, and the methodological rigor of the study (Cohen, 1988). In this study, the weighted means and standard deviations were used, so that they are already adjusted for the underlying FACES population, reflecting the probability sample that was drawn.

mean standard score on the writing task increased by 2.0 points ($p < .05$), or .13 of a standard deviation (Table 2-1).[4]

Table 2-2. Mean scores (SD) for FACES measures from Fall 2000, Spring 2001, and Fall-Spring difference

Skill area	Sample size	Mean (SD) Fall 2000	Spring 2001	Fall-Spring difference	Effect size
Vocabulary	1,801	85.3 (14.6)	89.1 (14.3)	3.8	.26
Letter-word identification	833	92.4 (9.6)	92.9 (12.0)	.5 (ns)	.05
Early writing	799	85.1 (13.6)	87.1 (14.3)	2.0*	.13
Early math	859	87.9 (15.2)	89.0 (17.1)	1.2*	.08
Book knowledge	1,876	1.6 (1.3)	2.5 (1.3)	0.9	.67
Color naming	1,893	11.3 (7.3)	15.7 (5.9)	4.4	.60
Draw-A-Design	1,914	2.9 (1.3)	3.6 (1.7)	0.6	.47
One-to-one counting	1,848	2.5 (1.2)	3.3 (1.4)	0.7	.59
Social awareness	1,893	3.4 (1.7)	4.1 (1.5)	0.6	.38

Note: All Fall-Spring difference scores are significant at $p<.0001$ except where indicated. $*p < .05$.

While the gains shown by children in the FACES 2000 sample from Fall to Spring were relatively modest, they fell within the range that has been deemed “educationally meaningful” (Rosenthal and Rosnow, 1984). “Educationally meaningful” results are those that explain a sufficiently large amount of the variation in scores to have meaningful programmatic or policy implications. They were in line with earlier findings on the immediate effects of Head Start on children’s intellectual performance (Haskins, 1989, p. 277; McKey et al., 1985). On the other hand, the vocabulary gains found in this sample were about half the size of standard-score gains in IQ and achievement that have been obtained in some earlier studies of more intensive interventions with children from disadvantaged families (Barnett, 1998, pp. 13-14). Most children left Head Start with vocabulary knowledge and early writing skills that were still below national averages.

[4] The Fall and Spring mean scores and gains are those for children who were assessed in English in both Fall and Spring. For information about the mean scores and gains of Spanish-speaking children from language-minority families, see below.

In recent years, the national Head Start program has been stressing the importance of early literacy skills like letter recognition. In Fall 2000, the typical child was found to enter the program knowing about four letters (3.9) of the alphabet and to leave knowing about nine (8.9).[5] However, the Fall-to-Spring gain in LWI standard scores was not statistically significant.

Children in FACES 2000 showed very slight gains in early math skills with respect to national averages. Their mean standard scores on the WJ-R Applied Problems task went from 87.9 in the Fall to 89.0 in the Spring. While the gain of 1.2 standard score points was statistically significant ($p < .05$), it amounted to only .08 of a standard deviation.

2.2.2.2 Gains in Book Knowledge, Color Naming, Design Copying, One-to-One Counting, and Social Awareness

Head Start children made significant Fall-to-Spring gains in the criterion-referenced tasks which cover areas of basic knowledge, verbal, mathematical, and perceptual-motor skills that children typically learn in the preschool years (Table 2-1). Effect sizes ranged from .6 of a standard deviation (for color naming) to .38 of a standard deviation (for social awareness).

2.2.2.3 Greater Gains for those Who Entered with Lower Skills

FACES 2000 found that children who came to Head Start with lower early literacy and math skills made greater gains in the program than those who came with average skills (Table 2-3). For example, whereas the average gain in vocabulary standard scores was 3.8 points, the mean gain for those in the lowest quarter of the distribution on this measure was 8.4 points. This amounted to more than half of a standard deviation. By contrast, children who were in the highest quarter of the distribution in the Fall showed no gain (-0.5) in standard scores by Spring. (Differences between mean gain in lowest quartile and overall mean gain, in highest quartile and overall mean gain, and between mean gains in lowest and highest quartiles were all statistically significant, $p < .001$.) This suggests that the greater

[5] The mean number of letters known out of the entire English alphabet are estimates based on the mean scores children received on the WJ-R Letter-Word Identification (LWI) task. Estimates were derived using a known relationship between scores on the WJ-R LWI and a Letter Naming task in which children are shown all the letters of the alphabet. There are no national norms on the number of letters of the alphabet the typical 4-year-old can name. For further details, see the Technical Appendix.

gains by children with lower skills is less likely to be a statistical artifact, such as regression to the mean, and more likely to be a meaningful contrast in relative growth in skills.

Table 2-3. Fall 2000 – Spring 2001 mean gain scores (SD) for FACES assessment measures

		Mean gain score (SD)		
Skill area	Sample size	Overall	Bottom quartile	Top quartile
Vocabulary	1,801	3.8 (10.5)	8.4 (11.5)	-.5 (9.2)
Letter-word identification	833	.5 (9.2)	1.3 (10.6)	-.6 (8.1)
Early writing	799	2.0 (13.3)	7.0 (15.6)	-2.5 (11.9)
Early math	859	1.2 (13.9)	5.9 (17.8)	-3.2 (11.7)
Book knowledge	1,876	0.9 (1.4)	1.5 (1.2)	-.2 (1.1)
Color naming	1,893	4.4 (5.5)	7.8 (6.4)	-.4 (1.6)
Draw-a-design	1,914	0.6 (1.6)	1.3 (1.2)	.1 (2.0)
One-to-one counting	1,848	0.7 (1.4)	1.2 (1.3)	.1 (1.4)
Social awareness	1,893	0.6 (10.5)	1.9 (1.6)	-.4 (1.1)

FACES 2000 found a similar picture with respect to changes in early writing and math skills. With respect to letter-word identification, there was evidence of a slight but significant increase in standard scores (1.3, or .08 of a standard deviation) for children in the lowest quartile in the Fall.

Despite the greater gains shown by children who entered in the lowest quarter of the Head Start population, these children ended the year with skills that were still well below average.

2.2.3 Emergent Literacy Gains of Language-Minority Children

In FACES 2000, children whose English-language skills were not sufficient to enable them to be assessed fully in English, and whose parents spoke Spanish at home, were given a Spanish-language version of the FACES battery in the Fall. In the Spring, these children received the full assessment battery in English. However, these English-language learners also received two additional tests that enabled the research team to track their language development and pre-reading skills in both English and Spanish. In the Fall, they received two components of the battery—vocabulary and letter identification—in English as well as Spanish. In the Spring, they received the same two components in Spanish as well as English. Thus, the levels of proficiency achieved and the gains made in vocabulary and letter identification could be determined in both languages.[6]

[6] English-language learners whose home language was something other than Spanish were not given any direct assessment in their native languages. They received only the full assessment battery in English in the Spring. For further details, see the Technical Appendix.

The dual assessment procedure made it possible to investigate the following research questions:

1. How do the literacy levels and gains of language-minority children in Head Start compare with those of language-majority children? and

2. How do the literacy levels and gains of these children vary across the two languages?

Of course, in comparing children to test norms, a different set of norms has to be used for test performance in Spanish. The Spanish-language knowledge of Head Start children from Spanish-speaking homes was compared to norms based on samples of children drawn in Mexico and Puerto Rico combined (Dunn, Padilla, Lugo, and Dunn 1986).

2.2.3.1 Knowing the Meaning of English Words

Spanish-speaking children in Head Start entered with English vocabulary skills considerably behind those of children who came from households where English was the primary language spoken in the home. They made greater gains over the course of the Head Start year. But their English vocabulary knowledge remained behind that of other children at the end of the year. The mean standard score on the PPVT-III for Spanish-speaking children in Fall 2000 was 59.7, more than two standard deviations below the national norm. By the Spring, their mean score had risen to 66.6, a gain of 7 points. But this was still more than 20 points (or one and a third standard deviations) lower than the mean score in vocabulary for language-majority children in Head Start (Table 2-4).

Table 2-4. Fall 2000, Spring 2001, and Fall-to-Spring gain vocabulary scores for language-minority, language-majority, and combined groups of children

	Vocabulary mean score (SD)		
	Language-minority (N = 306)	Language-majority (N = 1,801)	Language-minority and majority combined
Fall 2000	59.7 (13.3)	85.3 (14.6)	81.4 (17.1)
Spring 2001	66.7 (14.1)	89.1 (14.3)	85.7 (16.4)
Fall-Spring difference	7.0	3.8	4.3

All mean differences significant at $p < .0001$ level.

When the vocabulary scores of language-minority and language-majority children were added together, the combined Head Start student population began the year with a mean standard score of

81.4, and ended the year with a mean standard score of 85.7. This represented a gain of 4.3 standard score points.

2.2.3.2 Identifying Letters in English

Spanish-speaking children in Head Start entered with English letter identification skills that were only slightly behind those of language-majority children. However, like the English-speaking children, they did not make gains in these skills over the course of the year, at least not in comparison to national averages. The mean standard score of Spanish-speaking children tested in English on the WJ-R LWI test for Spanish-speaking children in Fall 2000 was 89.5, about two thirds of a standard deviation below the national norm. By the Spring, their mean score was 87.5, which was not significantly different from the Fall score (Table 2-5).

Table 2-5. Fall 2000, Spring 2001, and Fall-to-Spring gain letter-word identification scores for language-minority, language-majority, and combined groups of children

	Letter-word identification mean score (SD)		
	Language-minority (N = 178)	Language-majority (N =833)	Language-minority and majority combined
Fall 2000	89.5 (6.1)	92.4 (9.6)	91.9 (9.2)
Spring 2001	87.5 (7.9)	92.9 (12.0)	91.9 (11.6)
Fall-Spring difference	-2.0	.5	0.0

*$p < .01$.

When the letter identification scores of language-minority and language-majority children were added together, the combined Head Start student population began and ended the Head Start year with a mean standard score of 91.9.

2.2.3.3 Changes in English Language Skills Versus Changes in Spanish Skills

The Spanish vocabulary knowledge of Spanish-speaking Head Start children was quite comparable, in standard score terms, to the knowledge levels of English-speaking Head Start children in English. While the Spanish children made gains in English vocabulary knowledge during the year, they did not lose ground in Spanish vocabulary knowledge. The mean standard score for Spanish vocabulary knowledge was 84.9 in the Fall and 84.4 in the Spring (Table 2-6).

Table 2-6. Vocabulary skills in English and Spanish of Head Start children from Spanish-speaking language-minority families

	Mean score (SD)	
	English vocabulary task (N = 306)	Spanish vocabulary task (N = 300)
Fall 2000	59.7 (13.3)	84.9 (11.4)
Spring 2001	66.7 (14.1)	84.4 (12.8)
Fall-Spring difference	7.0*	-.6

*$p < .0001$.

The ability of Spanish-speaking Head Start children to identify letters of the alphabet in Spanish was comparable, in standard score terms, to the ability of English-speaking Head Start children to identify letters in English. It was also comparable to their own ability to identify letters in English. However, the Spanish-speaking children showed no gains against norms in their ability to identify letters in either language. The mean standard score for Spanish letter identification was 89.6 in the Fall and 86.2 in the Spring (Table 2-7).

Table 2-7. Letter-word identification skills in English and Spanish of Head Start children from Spanish-speaking language-minority families

	Mean score (SD)	
	English letter-word ID task (N = 178)	Spanish letter-word ID task (N = 300)
Fall 2000	89.5 (6.1)	89.6 (5.6)
Spring 2001	87.5 (7.9)	86.2 (6.7)
Fall-Spring difference	-2.0	-3.4**

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

2.2.4 Growth of Knowledge and Skills in Kindergarten

From assessments in the kindergarten followup, Head Start children continue to make advances against national norms during their first year in elementary school. For example, vocabulary standard scores went from 90.6 at the end of Head Start to 94.5 at the end of kindergarten. This was an increase of 3.9 points, or .37 of a standard deviation. Similarly, early math scores went from a mean standard score of 88.8 at the end of Head Start to a mean of 94.6 at the end of kindergarten. This was a gain of 5.8 points, or .34 of a standard deviation. Early writing scores showed a gain in kindergarten of

.80 of a standard deviation. They went from 85.9 at the end of Head Start to 98.0 at the end of kindergarten, an increase of 12.1 points (Table 2-8).

Table 2-8. Norm-referenced scores at end of Head Start and end of kindergarten and Head Start-to-kindergarten gain scores

		Mean (SD)			
	Sample size	End of Head Start	End of kindergarten	Head Start to kindergarten gain score	Effect size of gain score
Vocabulary	1,433	90.6 (14.2)	94.5 (12.3)	3.9	.37
Early math	1,387	88.8 (17.3)	94.6 (18.1)	5.8	.34
Early writing	1,340	85.9 (14.8)	98.0 (13.9)	12.1	.80

Note: All Head Start to kindergarten gain scores significant at $p < .0001$.

2.2.5 Differences in Gains in Knowledge and Skills Between Children Who Attend One Year of Head Start Versus those Who Attend for Two Years

These analyses were conducted on children who were assessed in English at both Fall and Spring of their respective Head Start program years. All children entered Head Start for the first time in Fall 2000 and either graduated in Spring 2001 (i.e., one-year graduates) or in Spring 2002 (i.e., two-year graduates).

2.2.5.1 Children Who Attend Head Start for Two Years Show Greater Entry to Graduation Gains than those Who Attend for One Year

Analyses were conducted comparing the child assessment scores of children who attended Head Start for 2 years with those who attended Head Start for only 1 year. Three-way Analyses of Covariance (ANCOVA) were conducted with length of Head Start attendance (one year vs. two years) predicting gains in child assessment scores from “entry in” to “graduation from” Head Start, while controlling for child’s age and assessment score in Fall 2000 (i.e., at baseline).[7]

[7] Because the Woodcock-Johnson-Revised subtests (Letter Word Identification, Applied Problems, and Dictation) are only administered to children 4-years-old or older, very few children who attended Head Start for two years have baseline data from Fall 2000. As a result, gain scores were not calculated for these subtests.

From Table 2-9, results show that children who attended Head Start for 2 years showed greater progress than their peers who attended for 1 year in all measures for which gain scores were calculated. On the PPVT-III, two-year Head Start graduates showed greater gains towards national norms in vocabulary than one-year graduates. Children who attended Head Start for only 1 year increased their vocabulary skills from a baseline score of 87.4 to 90.8 when they graduated from Head Start in Spring 2001. This represents a statistically significant gain of 3.4 points. In contrast, children who attended for 2 years increased their vocabulary skills from a baseline score of 82.9 to 90.6 when they graduated from Head Start in Spring 2002. This represents a statistically significant gain of 7.7 points, which was significantly greater than the gains in vocabulary scores for those who attended Head Start for only 1 year. It is important to note that while children who participate for 2 years have higher cognitive gains than those who participate for only 1 year, they achieve similar levels of performance at the end of Head Start (90.8 for one-year graduates; 90.6 for two-year graduates). However, the children who attend for 2 years start at a lower level, which accounts for their greater gain scores.

Table 2-9. Child assessment "entry-graduation" means (SD) for Fall, Spring and Fall-Spring gain scores for one-year Head Start graduates and two-year Head Start graduates

	One-year graduate			Two-year graduate			
	Fall	Spring	F-S gain	Fall	Spring	F-S gain	F-value
Vocabulary	87.4 (14.7)	90.8 (13.9)	3.4 (9.6)	82.9 (13.9)	90.6 (14.4)	7.7 (12.1)	F (1, 1388) = 12.4*
Book knowledge	1.9 (1.3)	2.8 (1.3)	.8 (1.4)	1.3 (1.1)	2.9 (1.4)	1.6 (1.5)	F (1, 1450) = 68.4**
Color naming	13.0 (7.0)	16.9 (5.1)	3.9 (5.3)	9.1 (7.0)	18.5 (3.1)	9.4 (6.6)	F (1, 1463) = 202.9**
Draw-A-Design	3.2 (1.5)	4.0 (2.0)	.8 (1.8)	2.5 (1.1)	4.4 (1.8)	1.9 (1.8)	F (1, 1474) = 128.9**
One-to-one counting	2.8 (1.3)	3.6 (1.4)	.8 (1.5)	2.1 (1.3)	3.8 (1.3)	1.7 (1.4)	F (1, 1425) = 92.5**
Social awareness	3.7 (1.6)	4.3 (1.4)	.6 (1.5)	3.1 (1.7)	4.8 (1.2)	1.7 (1.7)	F (1, 1450) = 134.7**

*$p < .001$; **$p < .0001$.

Children who attended Head Start for 2 years also showed greater "entry to graduation" gains than their peers who attended for only 1 year in each of the criterion-referenced measures, for which only raw scores are available. Gains demonstrated by two-year graduates ranged from approximately two times greater (e.g., Book Knowledge, One-to-One Counting) to approximately three times greater (e.g., Social Awareness, Draw-A-Design) than the gains demonstrated by one-year graduates.

2.3 Conclusions

Cognitive assessment data from FACES 2000 showed that most children entered Head Start with early literacy and math skills well below national averages. However, there was considerable diversity in skill levels among Head Start children. At Head Start entry, the highest quarter of Head Start children were at or above the national average (50^{th} percentile) in early language and number skills, while the lowest quarter of children ranked in the lowest 2 percent of all U.S. preschoolers in these areas.

Children made gains toward national averages during the Head Start year, especially with respect to vocabulary knowledge and early writing skills. Children who entered the program with lower levels of knowledge and skill showed larger gains during the program year, yet still lagged considerably behind national averages. Children who started with higher assessment scores in the Fall wound up with higher scores in the Spring, but did not show the same level of gains as did those with lower baseline scores. Language-minority children in Head Start showed significant gains in English vocabulary skills without declines in their Spanish vocabulary skills. They did not gain in letter recognition skills. Children who graduated after 2 years of Head Start made greater gains from entry to graduation than those who spent only 1 year in Head Start. Typically, children who graduate Head Start after 2 years enter with lower skills than one-year graduates. However, by the time both groups leave Head Start, they achieve similar levels of performance.

Head Start graduates also showed further progress toward national averages during kindergarten. Gains of between a quarter to more than three quarters a standard deviation were observed in vocabulary, early math, and early writing skills during kindergarten. Most Head Start graduates could identify the letters of the alphabet by the end of kindergarten and more than half could recognize beginning sounds of words. Further, Head Start graduates achieved early writing scores near the national norm by the end of kindergarten. Nevertheless, Head Start graduates remained behind their more advantaged peers in vocabulary and early math.

CHAPTER 3.
CHANGES IN SOCIAL COMPETENCE THROUGH THE HEAD START YEARS

FACES 2000 provides information about desirable and undesirable social behaviors that children who attend Head Start display when they enter the program and at the end of the program year. Examining social competence of preschool children is important because of evidence that children's academic achievement and adjustment once they go to school depend not only on their intellectual capabilities, but also on their acquisition of social skills and positive approaches to learning, and on the improvement of negative behavior patterns (Pianta and McCoy, 1997; Zill and West, 2001).

In a previous cohort (FACES 1997), significant positive changes were found in participants' social behaviors from the beginning to the end of Head Start. It was also found that social behaviors of Head Start children could predict social adjustment to elementary school, and, to a lesser extent, could predict academic performance in elementary school. Improvements in social behavior were particularly strong for children who entered Head Start with the lowest levels of social skills and those who entered Head Start with the highest levels of problem behaviors. In this chapter, we wanted to examine whether the significant positive changes in social behaviors found in 1997 are replicated in FACES 2000. The examination of the predictive power of the social behavior ratings in relation to kindergarten outcomes can be found in chapter 10 of this report. Information regarding children's social behavior in Head Start and beyond was obtained through behavior ratings provided by parents, Head Start teachers, and kindergarten teachers.

3.1 Research Questions

1. Do children show significant gains in social skills and significant declines in problem behavior during the Head Start year?

2. Do children graduating from Head Start programs after 2 years demonstrate more social skills and less problem behavior than children who had 1 year of Head Start? and

3. How do gains vary across behavior areas among children who enter the program with different levels of social skills and problem behavior?

3.2 Measures

3.2.1 Teacher Report

3.2.1.1 Cooperative Classroom Behavior of Head Start Children

Lead teachers in the FACES 2000 national sample of Head Start classes (as well as kindergarten teachers) were asked to rate the cooperative classroom behavior of children in their classes. These ratings were obtained early in the program year (October-November) and toward the end of the Head Start year (April-June 2001 and 2002), with followup in kindergarten (April-June 2002 or 2003). The scale was partly composed of items dealing with helpful and compliant behavior like, "Follows the teacher's directions," "Joins an activity or group without being told to do so," and "Helps in putting work materials or center property away." Other items dealt with the child's maturity and skill in interacting with other children. Examples include, "Invites others to join in activities," "Waits her or his turn in games or other activities," and "Accepts classmates' ideas for sharing and playing." For each item, the teacher was asked to judge whether the child behaved in the indicated way, "never," "sometimes," or "very often." There were 12 rating items in all, and the scale score could range from zero (all items marked "never") to 24 (all items marked "very often"). For a complete description of this measure, see the Technical Appendix of this report.

3.2.1.2 Problem Behavior of Head Start Children

Lead teachers in the FACES 2000 national sample of Head Start classes (as well as kindergarten teachers) were also asked to rate the undesirable or problem behavior of children in their classes. Three scales were composed from this measure: the Aggressive Behavior scale was partly composed of items such as, "Hits or fights with others," "Disrupts ongoing activities," and "Has temper tantrums or hot temper." The Hyperactive Behavior scale was composed of items such as "Can't concentrate, can't pay attention for long," and "Is very restless, fidgets all the time, can't sit still." The Withdrawn Behavior scale was composed of items such as "Keeps to herself or himself, tends to withdraw," "Lacks confidence in learning new things or trying new activities," and "Often seems unhappy, sad, or depressed." For each item, the teacher was asked to judge whether the behavioral description was "not true," "somewhat or sometimes true," or "very true or often true" of the child. The Aggressive Behavior scale contained four items and could range in value from zero to eight. The

Hyperactive Behavior scale was composed of three items and could range in value from zero to six. The Withdrawn Behavior scale contained seven items and could range in value from zero to 14. For a complete description of the measures see the Technical Appendix of this report.

3.2.2 Parent Report

3.2.2.1 Social Skills and Approaches to Learning

Parents in the FACES 2000 national sample of Head Start classes were asked to rate the social skills and approaches to learning of their children. These ratings were obtained early in the program year (October-November) and toward the end of the Head Start year (April-June 2001 and 2002), with a followup in kindergarten (April-June 2002 or 2003). Social skills items rated by parents were similar to those presented to teachers (e.g., “Makes friends easily,” “Comforts or helps others,” and “Accepts friends’ ideas in sharing and playing”). Examples of the approaches to learning items included, "Enjoys learning,” “Likes to try new things,” and “Shows imagination in work and play.” There were seven items in this scale, and scores could range from zero (meaning all the items were rated “not true” of the child) to 14 (meaning all the items were rated “very true or often true” of the child). The kindergarten scale was composed of different items and yielded slightly different scores. For a complete description of this measure, see the Technical Appendix of this report.

3.2.2.2 Behavior Problems

Parents were also asked to rate their children on negative behaviors that are relatively common among preschool children and that are associated with adjustment problems in elementary school. Parents were asked to rate their children’s problem behavior on three scales: the Aggressive Behavior scale (example item: “has temper tantrums or hot temper”), the Hyperactive Behavior scale (example item: “Can’t concentrate, Can’t pay attention for long”), and the Withdrawn Behavior scale (example item: “Feels worthless or inferior”). The Aggressive Behavior scale contained four items and could range in value from zero to eight. The Hyperactive Behavior scale was composed of three items and could range in value from zero to six. The Withdrawn Behavior scale contained five items and could range in value from zero to 10. The kindergarten scales were composed of different items and yielded slightly different scores. For a complete description of the problem behavior measures, see the Technical Appendix of this report.

3.3 Findings

3.3.1 Do Children Show Significant Gains in Cooperative Classroom Behavior and Social Skills and Significant Declines in Problem Behavior During the Head Start Years?

Table 3-1 presents the results of a repeated measure ANOVA for each of the eight (4 teacher-rated and 4 parent-rated) social competency variables after controlling for the child's age and the amount of time spent in Head Start between entry and graduation (one year vs. two years). There were major differences between the teachers' and the parents' reports of change in social behaviors: Teachers reported significant improvements in three out of the four social behaviors scales (cooperative classroom behavior, hyperactive behavior, and withdrawn behavior). On the other hand, no significant improvements in children's social behaviors from the beginning to end of Head Start were reported by parents.

Table 3-1. Means and standard deviations of social competence scores at Head Start entry and graduation

Social behavior	Data collection point: Entry	Data collection point: Graduation	*d*	*F*
Teacher report				
Cooperative behavior	14.77 (4.79)	17.80 (5.78)	.57	42.10***
Aggressive behavior	1.73 (1.99)	1.60 (1.38)		.94
Hyperactive behavior	1.37 (1.53)	1.06 (1.18)	.22	10.46**
Withdrawn behavior	2.45 (2.63)	2.10 (2.10)	.14	15.71***
Parent report				
Social skills	12.18 (1.70)	12.32 (2.12)	.07	3.60†
Aggressive behavior	3.18 (1.74)	2.75 (2.38)		1.24
Hyperactive behavior	1.89 (1.54)	1.59 (1.62)		.03
Withdrawn behavior	.61 (.96)	.66 (1.33)		1.07

Note: Standard deviations are given in parenthesis. *F* values are based on a repeated measure ANOVA controlling for child's age and length of exposure to Head Start. Effect sizes (*d*) are given only for significant differences. Sample sizes for teacher reported behaviors ranged from n = 2,057 (for aggressive behavior) to n = 2,169 (for cooperative classroom behavior). Sample sizes for parent reported behaviors ranged from n = 2,102 (for withdrawn behavior) to n = 2,135 (for social skills).

† $p < .1$; ** $p < .01$; *** $p < .001$.

3.3.2 Do Children Graduating from Head Start Programs After Two Years Show More Cooperative Classroom Behavior and Social Skills and Less Problem Behavior Than Children Who had One Year of Head Start?

Table 3-2 presents the results of a two-way ANOVA comparing the group of children who spent only 1 year in Head Start to the group of children who spent 2 years in Head Start on the eight social behavior variables (after controlling for age). The results indicate that children who spent more time in Head Start were reported to display higher levels of cooperative classroom behavior and social skills and less hyperactive behavior at the end of Head Start even after controlling for age. Moreover, the differences between one- and two-year graduates were consistent across reporters: both teachers and parents reported significantly more social skills and significantly less hyperactive behavior for the children who spent 2 years in Head Start compared to children who spent 1 year in Head Start. No significant differences were reported in aggressive and withdrawn behaviors.

Table 3-2. Means and standard deviations of social competence scores at Head Start graduation as a function of length of Head Start participation (1 vs. 2 years)

	Length of Head Start participation			
Social behavior	One year	Two years	*d*	*F*
Teacher report				
Cooperative behavior	17.44 (4.52)	18.26 (4.20)	.19	17.81***
Aggressive behavior	1.62 (2.03)	1.58 (1.58)		.27
Hyperactive behavior	1.12 (1.45)	.98 (1.43)	.10	4.78*
Withdrawn behavior	2.11 (2.46)	2.09 (2.44)		.04
Parent report				
Social skills	12.17 (1.73)	12.53 (1.67)	.21	21.17***
Aggressive behavior	2.80 (1.72)	2.67 (1.81)	.07	2.99†
Hyperactive behavior	1.66 (1.50)	1.50 (1.50)	.11	5.95*
Withdrawn behavior	.68 (.94)	.64 (1.00)		.67

Note: Standard deviations are given in parenthesis. *F* values are based on a three-way ANOVA controlling for child's age and baseline behavior ratings. Effect sizes (*d*) are given only for significant differences. Sample sizes for one-year graduates ranged from $n = 612$ (for teacher-reported withdrawn behavior) to $n = 820$ (for all parent-reported behaviors). Sample sizes for one-year graduates ranged from $n = 889$ (for parent-reported withdrawn behavior) to $n = 954$ (for teacher-reported cooperative classroom behavior).

† $p < .1$; * $p < .05$; *** $p < .001$.

3.3.3 How do Gains Vary Across Behavior Areas Among Children Who Enter the Program with Different Levels of Cooperative Classroom Behavior, Social Skills, and Problem Behavior?

Table 3-3 presents the mean scores of children in the most and least desirable quartiles on measures of social competence at the beginning and end of Head Start participation. Table 3-4 presents the comparisons between the positive gains in social behaviors made during the Head Start years for children at the most versus the least socially desirable behavior quartiles of each of the eight (4 teacher-rated and 4 parent-rated) social behaviors ratings.[1] As can be seen in Table 3-4, the differences between the most and least desirable groups in terms of social behavior gains achieved in Head Start were highly significant. Children who entered Head Start with high social skills and minimal problem behavior did not seem to improve their already high social performance (in fact, their behavior seemed to slightly decline during the Head Start years). In contrast, children who entered Head Start with the lowest levels of social skills and children who had the highest levels of problem behavior made significant improvements in all reported social behaviors, even after controlling for age and amount of time (1 or 2 years) spent in Head Start. Note that although children who started with the least desirable levels of social competency closed some of the gap between them and children in the most desirable quartiles, their scores in all the rating scales remained considerably lower (see Table 3-3).

Table 3-3. Mean scores of children in quartiles representing most and least desirable social competence scores at the beginning and end of Head Start participation

	Most desirable		Least desirable	
Social behavior	Entry	Graduation	Entry	Graduation
Teacher report				
Cooperative behavior	20.47	20.00	8.72	15.40
Aggressive behavior	0	.66	4.33	3.02
Hyperactive behavior	0	.43	2.99	1.75
Withdrawn behavior	0	1.07	5.93	3.46
Parent report				
Social skills	13.52	12.89	10.05	11.47
Aggressive behavior	1.33	1.84	4.89	3.59
Hyperactive behavior	.50	.93	3.79	2.52
Withdrawn behavior	0	.43	1.64	1.05

Note: Children who scored highest on the measures of cooperative behavior and social skills, or lowest on measures of aggressive, hyperactive, or withdrawn behaviors comprise the “most desirable” quartiles. Children who scored lowest on measures of cooperative behavior and social skills, or highest on measures of aggressive, hyperactive, or withdrawn behaviors comprise the “least desirable” quartiles. Comparisons between the different most and least desirable groups were independent from each other. Thus, children who belonged to one of the extreme quartiles on one behavior rating were not necessarily part of the extreme quartiles of any other behavioral ratings.

[1] Comparisons between the different most and least desirable groups were independent from each other. Thus, children who belonged to one of the extreme quartiles in one behavior rating were not necessarily part of the extreme quartiles of any other behavioral ratings.

Table 3-4. Means and standard deviations of social competence gain scores (from entry to graduation) of children scoring in the most and least desirable quartiles

Social behavior	Most desirable	Least desirable	*d*	*F*
Teacher report				
Cooperative behavior	-.47 (3.71)	6.68 (4.91)	1.64	659.49***
Aggressive behavior	.66 (1.27)	-.31 (2.42)	.50	306.45***
Hyperactive behavior	.43 (.81)	-1.24 (1.75)	1.22	498.41***
Withdrawn behavior	1.07 (1.78)	-2.47 (3.14)	1.39	451.67***
Parent report				
Social skills	-.64 (1.33)	1.42 (2.06)	1.19	503.57***
Aggressive behavior	.51 (1.47)	-1.30 (1.83)	1.09	369.77***
Hyperactive behavior	.43 (1.18)	-1.27 (1.53)	1.24	495.34***
Withdrawn behavior	.43 (.81)	-.59 (1.29)	.95	394.49***

Note: Numbers in table represent mean gain scores (Δ). Standard deviations are given in parenthesis. *F* values are based on a repeated measure ANOVA controlling for child's age and length of exposure to Head Start. Sample sizes for the bottom quartile ranged from n = 433 (teacher-reported cooperative classroom behavior) to n = 1,016 (parent-reported withdrawn behavior). Sample sizes for the top quartile ranged from n = 433 (teacher-reported aggressive and withdrawn behavior) to n = 826 (parent-reported social skills).

*** $p < .001$.

3.4 Conclusions

This chapter examined whether social behaviors, as reported by Head Start parents and teachers, change during the Head Start years. The findings indicate that teachers in particular report significant positive changes on measures of social competence from Head Start entry to graduation. The findings also indicate that children who spent 2 years in Head Start were reported to display higher levels of cooperative classroom behavior and social skills and less hyperactive behavior at the end of Head Start compared to children who spent only 1 year in Head Start. Children entering Head Start with the lowest levels of social skills, and those who entered Head Start with the highest levels of problem behaviors, gained the most from participation in Head Start. Although these children did not close the gap between themselves and other children, they did narrow it significantly.

CHAPTER 4.
CURRICULUM, FAMILY, PROGRAM, AND CLASSROOM CHARACTERISTICS

There is a lack of substantial evidence about the relative efficacy of various types of standardized curricula available to preschool programs and their relationship to children's school readiness. Accordingly, the Head Start Program Performance Standards require that programs have a curriculum, and delineate the areas that must be covered by it, but do not prescribe one. Programs may use curricula from a variety of sources, develop one of their own, or use a combination of curricula. In the 2001-2002 program year, the Head Start Program Information Report (PIR) queried local programs about the curricula they use. In descending order of frequency, center-based programs indicated that they used the Creative Curriculum, High/Scope curriculum, and locally designed curriculum. This chapter examines the curricula Head Start programs in the FACES sample are using; the training and ongoing support teachers receive in the use of their curricula; teacher satisfaction with their curricula; and the relationships between the type of curricula used and family, program, and classroom characteristics.

4.1 Research Questions

This chapter addresses the following research questions:

1. What percentage of Head Start programs use a curriculum?
2. What types of curricula are used in Head Start programs?
3. Do Head Start teachers receive training and ongoing support in the use of their curriculum and from whom?
4. What percentage of Head Start teachers have access to a copy of their curriculum?
5. What aspects of the curricula do Head Start teachers like?
6. What is the relationship between the type of curricula used and the characteristics of children and families served?
7. Are there regional and rural-urban differences in the type of curricula used by Head Start programs? and
8. What is the relationship between the type of curricula used and classroom quality?

4.2 Methods

The sample for this chapter includes 231 center-based Head Start teachers from 43 Head Start programs in the FACES 2000 study. Data reported in this chapter are from teacher interviews conducted by field staff in Fall 2000. The monthly family income data and child ethnicity data come from parent interviews conducted in Fall 2000. These parent interview numbers are less than the total number of parents interviewed in Fall 2000 because of non-response or missing data on monthly family income or child ethnicity variables.

Weighted percentages, correlations, independent-sample t-tests and multivariate analyses of variance were used to answer the research questions. Data presented in this chapter are weighted to represent Head Start programs nationally.[1]

4.3 Findings

4.3.1 Head Start Teachers Report Using a Specific Curriculum, Receiving Training, and Having Access to their Curriculum

4.3.1.1 Types of Curricula

Head Start teachers were asked if they used a single specific curriculum, a combination of curricula, or no curriculum. About 70 percent of the teachers used a single curriculum, 21 percent used a combination of curricula, and 9 percent did not use a curriculum. Teachers who reported using a single curriculum or a combination of curricula were asked to name their principal curriculum. The majority (59.1 percent) said that they used either Creative Curriculum or High/Scope curriculum (see Table 4-1). Almost 41 percent used a curriculum other than Creative Curriculum or High/Scope curriculum, which will be referred to as other curricula. Other curricula mentioned by teachers were High Reach, Scholastic, Los Cantos Los Ninos, R.E.A.L, Newport, Global Curriculum, Creating Child Centered Classrooms – Step by Step, Building Bridges, Northern Kentucky Curriculum, Montessori, Teacher Planning Wellbook, Therapeutic Intervention Program, and Kid College Curriculum. Some others include the “Head Start” curriculum or “theme units” although there is no “official” Head Start curriculum.

[1] FACES 2000 employs a nationally representative sample of 43 Head Start programs. It does not include migrant or American Indian/Alaskan Native programs, Early Head Start programs or programs in Puerto Rico and the other territories.

Table 4-1. The majority of teachers used either Creative Curriculum or High/Scope curriculum

Curricula	Percent of teachers
Creative Curriculum	39.1%
High/Scope curriculum	20.0%
Other	40.9%

4.3.1.2 Curricular Training and Support

Teachers were asked if they received training in their curriculum and who provided this training (teachers could respond with only one source). Ninety-three percent reported receiving training in their curricula. Of those that received training, most had received it from their own program staff (58.5 percent), followed by curricula developers (14 percent), a Head Start Quality Improvement Center or HSQIC (10.3 percent), another Head Start program (5 percent), a university school of education (4.4 percent), or another source (7.9 percent).

Almost 92 percent of the teachers received ongoing support such as consultation or mentoring in the use of their curriculum and this support could come from several sources. Of those who received support, it most often came from their supervisor or the education coordinator (70.3 percent), other teachers (23.9 percent), the HSQIC (20 percent), curriculum developers (19.3 percent), or a mentor/master teacher (14 percent). Support also came from other Head Start programs (10.8 percent), schools of education (6.6 percent), the Disability Services Quality Improvement Center or DSQIC (4.1 percent), and other sources (8.6 percent).

4.3.1.3 Teacher Access to Curricula

Of teachers who used a curriculum, 97 percent responded that teachers and assistant teachers in their program had access to a copy of their curriculum.

4.3.1.4 Teacher Satisfaction with Curricula

The vast majority of the teachers (92.3 percent) said that they liked their curriculum. Table 4-2 indicates that an overwhelming majority of the teachers said they liked their curriculum because it addressed multiple domains of learning (99.1 percent), was easy to use and adapt (98.3 percent), involved parents (96.6 percent), had room for teacher creativity (96.4 percent), and had adequate learning materials, resources, and examples of activities (92.1 percent).

Table 4-2. Teachers liked their respective curricula for a variety of reasons

Reasons	Percent of teachers
Multiple domains addressed	99.1%
Easy to use/adapt	98.3%
Involves parents	96.6%
Room for teacher creativity	96.4%
Adequate learning materials	92.1%

4.3.2 Relationship Between Curricula and Family, Program, and Classroom Characteristics

4.3.2.1 Relationship Between Curricula and Family Characteristics

In the context of the study's conceptual framework, the relationship between the use of particular curricula and the characteristics of the families served by the programs were examined. The mean monthly family income of families from classrooms using the High/Scope curriculum was $1,641.18, from classrooms using Creative Curriculum was $1,559.60, and from classrooms using other curricula was $1,319.43. This indicates that classrooms using other curricula served the poorest families compared to classrooms using Creative Curriculum or High/Scope curriculum. The percentage of minority children served by teachers using other curricula was 75.8 percent, for teachers using High/Scope curriculum it was 69.6 percent, and for teachers using Creative Curriculum it was 48.8 percent. Thus, teachers using other curricula served the highest percentage of minority children.

4.3.2.2 Relationship Between Curricula and Program Characteristics

Table 4-3 shows that more teachers from the Northeast (78.7 percent), Midwest (59.4 percent), and West (69.4 percent) used Creative Curriculum or High/Scope curriculum than other curricula, while the majority of teachers from the South (54.4 percent) used other curricula.

Table 4-3. Most teachers from the northeast, midwest, and west used Creative Curriculum or High/Scope curriculum, while majority of teachers from the south used other curricula

Curricula	Percent of teachers			
	Northeast	Midwest	South	West
Creative Curriculum	61.4%	42.7%	24.5%	44.5%
High/Scope	17.3%	16.7%	21.1%	24.9%
Other	21.3%	40.6%	54.4%	30.5%

A greater percentage of rural teachers (77.5 percent) used either Creative Curriculum or High/Scope curriculum than urban teachers (51.4 percent). As shown in Table 4-4, although urban and rural teachers were equally likely to use the High/Scope curriculum (20 percent), urban teachers were more likely than rural teachers to use other curricula (48.6 percent vs. 22.4 percent).

Table 4-4. The majority of urban teachers used either Creative Curriculum or High/Scope curriculum while the majority of rural teachers used Creative Curriculum

	Creative Curriculum	High/Scope	Other
Urban	31.4%	20%	48.6%
Rural	57.5%	20%	22.4%

4.3.2.3 Relationship Between Curricula and Classroom Quality

The relationship between curricula and classroom quality was examined using independent samples t-tests on three different scores (weighted means) of classroom quality–ECERS-R Total score, ECERS-R Language subscale score, and a Quality Composite score.

The ECERS-R Total score is derived from the ECERS-R, which provides a global rating of classroom quality based on structural features of the classroom. Scores can range from 1 (inadequate) to 7 (excellent). As displayed in Table 4-5, the average ECERS-R Total scores for classrooms using Creative Curriculum (5.02, t =3.70) and High/Scope curriculum (5.04, t =3.83) were significantly higher (p < .05.), than those for classrooms using other curricula (4.55). However, the average ECERS-R Total scores for High/Scope curriculum and Creative Curriculum were not significantly different from each other.

The ECERS-R Language subscale score is a subscale of the ECERS-R and assesses classroom quality as it pertains to encouraging language-reasoning experiences (Table 4-5). Scores can range from 1 (inadequate) to 7 (excellent). Classrooms of teachers who used Creative Curriculum and High/Scope curriculum had significantly higher average ECERS-R Language subscale scores (5.03, t = 3.18 and 5.12, t = 3.64 respectively; $p < .05$) than classrooms of teachers using other curricula (4.58). However, the average ECERS-R Language subscale scores of classrooms that used Creative Curriculum (5.03) and High/Scope curriculum (5.12) were not significantly different from one another.

Table 4-5. The classrooms of teachers who used Creative Curriculum or High/Scope curriculum had significantly higher average ECERS-R total scores and ECERS-R Language subscale scores than classrooms of teachers who used other curricula

Curricula	Average ECERS-R total and Language subscale scores	
	Average ECERS-R total score	Average ECERS-R Language subscale scores
Creative Curriculum	5.02	5.03
High/Scope	5.04	5.12
Other	4.55	4.58

The Quality Composite score is derived from a principal components factor analysis of the ECERS-R Language subscale score, Assessment Profile Scheduling raw score, and Assessment Profile Learning Environment raw score. The standardized factor scores were calculated, so that the scores are in standard deviation units, with a mean of 0 and a standard deviation of 1 (see Appendix for further description of the Quality Composite Score). As displayed in Table 4-6, the classrooms of teachers who used Creative Curriculum and High/Scope curriculum had significantly higher Quality Composite scores (0.27, t = 3.63 and 0.26, t = 3.26 respectively; $p < .05$) than classrooms of teachers using other curricula (-0.18). The Quality Composite scores of classrooms that used Creative Curriculum (0.27) and High/Scope curriculum (0.26) did not differ significantly from one another.

Table 4-6. The classrooms of teachers who used Creative Curriculum or High/Scope curriculum had significantly higher quality composite scores than classrooms using other curricula

Curricula	Average quality composite score
Creative Curriculum	0.27
High/Scope	0.26
Other	-0.18

4.3.2.4 Other Curricula and Classroom Quality

In order to determine the relationship of curricula within the other category to classroom quality, other curricula have been further categorized as Widely Available Curricula and All Other Curricula. Widely Available Curricula include High Reach, Scholastic Curriculum, Newport Curriculum, and Montessori. These curricula appear to be established (for example High Reach has been commercially available for about 17 years, Scholastic for more than 82 years, and Montessori for 96 years), have information about the curricular goals and practices, printed materials, and in some cases have research available on their efficacy. All Other Curricula includes curricula not specified by teachers, as well as curricula termed "Head Start," Los Cantos Los Ninos, Theme Units, R.E.A.L, Global Curriculum, Creating Child Centered Classrooms—Step by Step, Building Bridges, Early Childhood Lesson Plan, Curriculum Workbook, Northern Kentucky Curriculum, Teacher Planning Wellbook, Therapeutic Intervention Program, and Kid College Curriculum. This group of curricula appears to be designed by programs locally or from sources not widely available.

The classroom quality scores for Widely Available Curricula, All Other Curricula, Creative Curriculum, and High/Scope curriculum are presented in Table 4-7. The average ECERS-R Total scores and average ECERS-R Language subscale scores for Creative Curriculum, High/Scope curriculum, and Widely Available Curricula are similar although the Quality Composite score for Widely Available Curricula is lower.

Table 4-7. Type of curricula and classroom quality scores

Type of curricula used	Weighted percentage of Head Start teachers	Average total ECERS-R score	SD	Average ECERS-R language subscale score	SD	Quality composite score	SD
Creative Curriculum	39.1	5.02	.92	5.03	1.32	0.27	.96
High/Scope Curriculum	20.0	5.04	.73	5.12	1.01	0.26	.84
Widely Available Curricula	9.8	4.82	.63	5.15	.96	0.07	.95
All Other Curricula	31.1	4.47	.86	4.4	1.19	-0.27	.95

Note:

- Widely Available Curricula includes: High Reach, Scholastic Curriculum, Newport Curriculum, and Montessori; and
- All Other Curricula includes: curriculum not specified, Head Start curriculum, Los Cantos Los Ninos, Theme Units, R.E.A.L, Global Curriculum, Creating Child Centered Classrooms—Step by Step, Building Bridges, Early Childhood Lesson Plan, Curriculum Workbook, Northern Kentucky Curriculum, Teacher Planning Wellbook, Therapeutic Intervention Program, and Kid College Curriculum.

Multivariate analyses of variance were used to examine if the classroom quality scores were significantly different between classrooms using Creative Curriculum versus High/Scope curriculum, Creative Curriculum versus Widely Available Curricula, Creative Curriculum versus All Other Curricula, High/Scope curriculum versus Widely Available Curricula, High/Scope curriculum versus All Other Curricula, and Widely Available Curricula versus All Other Curricula. Significant differences ($p < .05$) were found between classrooms using Creative Curriculum versus All Other Curricula on all three classroom quality scores, as well as between classrooms using High/Scope curriculum versus All Other Curricula on all three classroom quality scores. Classroom quality scores between classrooms using Widely Available Curricula versus All Other Curricula were not significantly different.

- Classrooms using the Creative Curriculum when compared to classrooms using All Other Curricula had significantly higher average ECERS-R Total scores (5.02 versus 4.47), average ECERS-R Language subscale scores (5.03 versus 4.4), and Quality Composite scores (0.27 versus - 0.27);

- Classrooms using the High/Scope curriculum when compared to classrooms using All Other Curricula had significantly higher average ECERS-R Total scores (5.04 versus 4.47), average ECERS-R Language subscale scores (5.12 versus 4.4), and Quality Composite scores (0.26 versus - 0.27); and

- Classrooms using Widely Available Curricula were not significantly different from classrooms using All Other Curricula. Classrooms using Widely Available Curricula were also not significantly different from those using Creative Curriculum or those using High/Scope curriculum. In addition, classrooms using Creative Curriculum were not significantly different from those using High/Scope curriculum.

Overall, these findings suggest that classrooms using curricula such as Creative Curriculum and High/Scope curriculum have relatively higher classroom quality than Head Start classrooms using other curricula.

4.4 Summary of Findings

Findings from FACES 2000 indicate that the great majority of Head Start programs use a curriculum, with the goal of providing a planned, developmentally appropriate early childhood program for children. Most teachers reported using a single curriculum. The most frequently used curricula are the Creative Curriculum (39.1 percent) and High/Scope curriculum (20 percent). Further, 93 percent of teachers reported receiving training in their curriculum, and 92 percent reported receiving ongoing support. Nearly all teachers (92.3 percent) reported liking their curriculum.

Data from FACES 2000 also indicate a relationship between program characteristics and the type of curriculum used. Specifically, classrooms using a curriculum other than the Creative Curriculum or High/Scope curriculum served the poorest Head Start families and the highest percentage of non-White children. There was also regional variation in curricula used by classrooms. More teachers from the Northeast (78.7 percent), Midwest (59.4 percent), and West (69.4 percent) used either the Creative Curriculum or High/Scope curriculum compared to any other type of curricula, while the majority of teachers from the South (54.4 percent) used a curriculum other than the Creative Curriculum or High/Scope curriculum. In addition, the majority of urban teachers (51.4 percent) used either the Creative Curriculum or High/Scope curriculum, while the majority of rural teachers (57.5 percent) used Creative Curriculum.

Finally, findings demonstrate a relationship between type of curricula and classroom quality. Classrooms using Creative Curriculum and High/Scope curriculum had significantly higher ECERS-R Total scores, ECERS-R Language subscale scores, and Quality Composite scores than classrooms using Other Curricula. However, the scores on these three measures for classrooms that used High/Scope curriculum and Creative Curriculum were not significantly different from each other. Analyses were also conducted separating the Other curricula category into Widely Available Curricula and All Other Curricula. The average ECERS-R Total scores and average ECERS-R Language subscale scores for Creative Curriculum, High/Scope curriculum, and Widely Available Curricula are similar although the Quality Composite score for Widely Available Curricula is lower. Multivariate analyses showed significant differences between classrooms using Creative Curriculum versus All Other Curricula on all three classroom quality scores, as well as between classrooms using High/Scope curriculum versus All Other Curricula on all three classroom quality scores. Classroom quality scores between classrooms using Widely Available Curricula versus All Other Curricula were not significantly different.

CHAPTER 5.
QUALITY IN HEAD START CLASSROOMS

This chapter examines the quality of Head Start classrooms in the FACES 2000 sample, in order to build a model that "unpacks" the critical ingredients of quality in Head Start by accounting for explainable variation at the levels of the classroom and the Head Start center. The nature of quality in Head Start programs will be described and key factors will be identified that help explain variations in quality across Head Start classrooms. The results of these analyses should provide information for improving quality in Head Start classrooms and general program improvement. Findings about the relationship between classroom quality and child outcomes are described in Chapters 6 and 7.

5.1 Conceptual Framework and Research Questions

The conceptual framework for this chapter is based on a multi-level, multi-causal model of the factors that explain variations in classroom quality. The model is presented in Figure 5-1 and describes a multi-level perspective on understanding or "unpacking" quality in Head Start classrooms.

There are a variety of indicators of quality in early childhood educational settings cited in the research literature (Phillips, Mekos, Scarr, McCartney and Abbott-Shim, 2000). In addition to "structural" indices, such as child:staff ratio (Cost, Quality and Outcome Studies Team, 1995; Howes, Phillips, and Whitebook, 1992) and group size (Ruopp, Travers, Glantz and Coelen, 1979), the backgrounds and experiences of early childhood teachers are important (Cost, Quality and Outcome Studies Team, 1995). Additionally, a variety of "process" aspects of the classroom environment have been considered important, including teacher-child interactions, child-child interactions, as well as curriculum, the schedule of activities and materials (Bredekamp, 1986; Hayes, Palmer and Zaslow, 1990; Phillips and Howes, 1987; Scarr and Eisenberg, 1993).

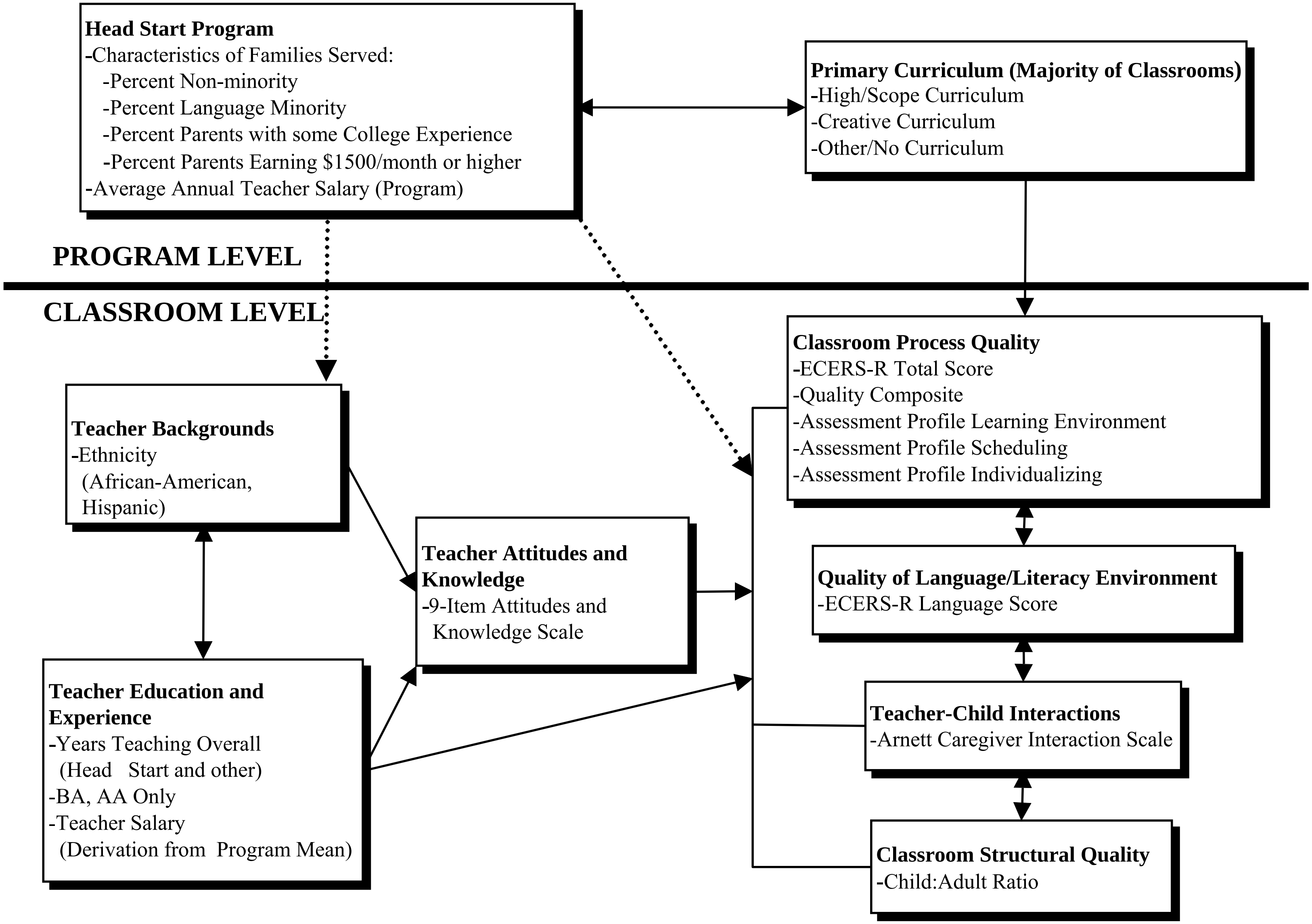


Figure 5-1. A conceptual model of program- and classroom-level factors explaining quality in Head Start

However, the search for understanding the mechanisms by which process, structural and teacher indicators influence quality has been relatively sparse and based more on logical assumptions than empirical evidence. For example, lower child:staff ratios (e.g., more staff per child) are assumed to create the conditions that support developmentally beneficial teacher-child interactions (Phillips, Mekos, Scarr, McCartney and Abbott-Shim, 2000). Although the empirical literature supports this connection (Phillips, Howes and Whitebook, 1992), only modest amounts of variance can be explained (Scarr, Eisenberg and Deater-Deckard, 1994). Little is known about how different indicators of quality in varying levels contribute to overall classroom quality. For example, the possible compensatory effect of more optimal levels on one indicator, such as child:staff ratios, on less favorable levels of other indicators, such as staff training is not known (Phillips, Mekos, Scarr, McCartney and Abbott-Shim, 2000).

Furthermore, factors "beyond the classroom door" may have potentially powerful influences on children's experiences of quality care in early childhood classrooms. Some of the factors suggested consist of the additional resources available to early childhood settings and the broader context within which these programs are situated (Blau, 1997; Hofferth and Chaplin, 1998 cited in Phillips, Mekos, Scarr, McCartney and Abbott-Shim, 2000). Program management styles, resources, and the demographics of the community may influence decisions about quality made by center directors (e.g., the quality of teachers that can be recruited) and, indirectly, by teachers in individual classrooms.

Although factors that exist beyond the classroom are typically difficult to measure in early childhood settings, the organizational structure of Head Start in which classrooms are located within centers, which in turn are located within a grantee or program, lends itself to the examination of this kind of factor. Head Start programs consist of the grantee and delegate agencies that administer the centers and classrooms within their purview. A program comprises the administrative entity primarily responsible for determining budgets, staffing, and the allocation of resources across centers and classrooms, as well as for choosing curricula for its classrooms and providing training and resources to support the curricula. Programs also define the geographic areas they serve and identify the needs of low-income families and children living there.

The FACES Third Performance Measures Report used FACES 1997 data to describe how classroom quality varies across classrooms and across programs suggesting that factors at the level of the Head Start program, in addition to those at the classroom level, may help to explain variations in quality (Administration for Children and Families, 2001). Thus, in developing a comprehensive model for understanding the many sources of influence on classroom quality, it is important to look at factors at both the classroom and the program levels. While some of the program-level factors were identified in earlier rounds of FACES, it was only with the FACES 2000 data that additional factors at the level of program were collected.

FACES 2000 employed similar measures as those used in the earlier cohort in order to make some comparisons in quality, but several key measures were added. Quality was considered to include not only the number of children and adults in each classroom, but process factors such as the availability of learning materials, the types of classroom activities, the scheduling and the variety of learning opportunities provided to all children. Many of the newer factors were collected through the lead teacher interview,[1] including teacher background information (experience and qualifications) as well as more detailed information about their curriculum, classroom activities, and attitudes and knowledge about early childhood education practices. The classroom quality measures are fully described in the Appendix.

The following research questions guided the analyses in this chapter:

- What is the quality of Head Start classrooms in 2000-2001, and how does it compare to quality reported in 1997-1998?
- What are the backgrounds, qualifications and experiences of Head Start teachers in 2000-2001 and were there any changes from the earlier cohort (1997-1998)?
- How are teaching qualifications and experience related to attitudes and knowledge of early childhood education and how are these factors linked to quality in Head Start classrooms? and
- Do factors beyond the classroom such as the types of curricula Head Start programs provide, the average teacher salaries, and characteristics of families served by the Head Start program explain variations in the quality of Head Start classrooms?

5.2 Analytic Methods

This chapter used a number of classroom quality measures, described more fully in the Appendix. The variables derived from these measures that were used in the analyses of classroom quality are as follows:

[1] The results of teacher data presented in this chapter were based on interviews and ratings of the lead or senior teacher in each classroom.

- **ECERS-R Total Score.** The mean of the 37 ECERS-R items[2] coded on a 7-point scale with a score of 1 representing "inadequate", a score of 3 representing "minimal quality," a score of 5 representing "good quality," and a score of 7 representing "excellent quality." The internal consistency of the ECERS-R mean score for all 37 items was .92 for both Fall 2000 and Spring 2001, and .89 for Spring 2002 (Head Start). A high score on the total ECERS-R indicates higher classroom quality, in terms of equipment, space and play materials, as well as the range of learning activities, supervision, and staff-child interactions;

- **ECERS-R Language Subscale Score.** The mean of 4 ECERS-R items related to language quality in classrooms, including informal use of language, books and pictures, encouraging children to communicate, and using language to develop reasoning skills. Scores range from 1 to 7, with high scores indicating a classroom with a rich language environment;

- **Assessment Profile Scheduling Scale Score.** The score consists of the sum of the 14 observation items, each with yes/no response formats. Raw scores used in these analyses ranged from 3 to 14. A high score is indicative of a teacher's "planfulness" and strategy for providing varied learning activities in whole groups, small groups, and one-on-one. The reliability of the Scheduling subscale was reported as .89 for Fall 2000, .87 for Spring 2001, and .82 for Spring 2002;

- **Assessment Profile Learning Environment Scale Score.** This score consists of the sum of the 18 observation items in this scale, with each item scored yes or no. Raw scores used in these analyses ranged from 4 to 18. A high score on this scale is indicative of a greater variety of materials accessible, that stimulate growth in a variety developmental domains in a "learning rich" environment. The reliability of the Learning Environment subscale was reported as .68 for Fall 2000, .77 for Spring 2001, and .65 for Spring 2002;

- **Assessment Profile Individualizing Scale Score.** This score is comprised of the sum of five observational items with raw scores ranging from 0 through 5. A high score indicates that teachers are able to adjust classroom activities to meet the learning needs of individual children. The reliability of the Individualizing subscale was reported as .50 for Fall 2000, .54 for Spring 2001, and .44 for Spring 2002;

- **Arnett Caregiver Interaction Scale for the Lead Teacher.** This score is the sum of 30 items each with four-point responses asking observers to rate the teacher according to a fixed descriptor either "not at all," "somewhat," "quite a bit," or "very much." Although there are five subscales measuring the teacher's sensitivity, punitiveness, detachment, permissiveness, and encouragement of child independence and self-help skills, in the foregoing analyses we primarily focus on the total score (since all subscales are highly correlated with this score). Scores ranged from 20 to 89 on this measure. A high score indicates greater teacher sensitivity, responsiveness and encouragement of children's independence and self-help skills, and lower levels of

[2] FACES did not use 6 items that comprise the Parents and Staff subscale.

punitiveness and detachment. The Cronbach Coefficient Alphas for all of the items were .94 for both Fall 2000 and Spring 2001, and .93 for Spring 2002;

- **Quality Composite.** Using a principal components factors analysis, scores from the ECERS-R Language subscale, and the Assessment Profile Scheduling and Learning Environment scales were combined to form a single quality factor score. The quality factor score was standardized with a mean of 0 and a standard deviation of 1, and scores ranged from -3.57 to 1.90. A higher score indicates higher levels of quality;

- **Child:Adult/Child:Staff Ratio.** This is a number obtained by dividing the number of children present in the class by the number of adult teachers, aides, or volunteers who were actively interacting with the children. Observers counted children and adults at two different times during their observation session and ratios were averaged over all classes visited in a given center. In addition, the observers took counts of the number of children present in the class and the number of adult teachers, aides, or volunteers who were actively engaged with the children. A higher ratio is indicative of lower quality, since this indicates there are more children with fewer adults;

- **Teacher Backgrounds and Qualifications.** A set of variables were derived from the teacher interview that assessed the teachers' backgrounds, qualifications and teaching experience. Continuous variables included: annual salary, years teaching Head Start, total years of teaching experience, and teacher age. A variety of variables were also binary coded 0 or 1 to allow for comparisons of proportions across time and between groups. Based on information in the teacher interview, variables were coded 1 for teachers who had:

    - CDA or state-awarded certificate;
    - AA degree or higher;
    - BA or AA degree;
    - BA degree or higher;
    - Teaching certificate or license;
    - Field of study in their highest degree obtained included Early Childhood Education;
    - Member of an Early Childhood Education association;
    - Teaches in Spanish;
    - Teaches in a language other than English or Spanish;
    - African American, Latino or Asian (three separate variables);

- Currently enrolled in teacher related training; or
- Primary curriculum is High/Scope curriculum or Creative Curriculum (two separate variables).

■ **Teacher Attitudes and Knowledge Score.** This score was derived from a single factor consisting of 9 items from the 24-item Teacher Beliefs Scale that explained most of the variation in scores for the entire scale. Each item used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) with negatively-worded items reverse scored. The 9 individual items were then combined to form the composite score with a minimum value of 1 and a maximum value of 10. A high score indicates higher positive attitudes and knowledge about early childhood education practices.

To answer the first three research questions, a number of descriptive and correlational analyses were conducted. To answer questions about change in quality and teacher backgrounds, qualifications or experiences from the first to the second cohorts, and to determine change in quality from Fall to Spring of the Head Start year, t-tests using the continuous (e.g., ECERS-R score, years of teaching) and the binary coded teacher variables (e.g., whether teacher had a BA or AA degree) were employed. Specifically, analyses of the binary coded variables used t-tests on independent samples to compare the proportions of "1" responses for each variable across the two time periods in question. To assess differences over time in categorical data, chi-square tests were used. However, since a significant chi-square statistic only indicates there was significant change over time but not where those changes occurred, that is, which of the categories was the source of the change, significant chi-square analyses were supplemented by binomial tests of the differences between proportions across time periods using unweighted data. This latter test allows for a comparison across time in the proportions for a specific response category.

Most descriptive statistics were derived from the weighted data wherever possible, with the exception of the Fall 1997–Spring 1998 data (first cohort) and Spring 2002 data,[3] where unweighted data were used. Analyses involving the first cohort data were all done using unweighted data because classroom weights were not developed for this cohort. Comparisons of Spring 2002 data with the earlier waves were done using only the unweighted data.

[3] The Spring 2002 data were based on a much smaller number of classrooms, because most of the children in the sample had graduated to kindergarten. The Spring 2002 classrooms consisted only of those in which FACES study children were still attending in their second year of Head Start. The Fall 2000 weights and the longitudinal weights were not considered appropriate for this small subsample and thus the unweighted data were analyzed.

To answer research question #4, a multi-level approach with two levels of factors: program and classroom[4] was used. This approach determines the joint influence of both program- and classroom-level factors in predicting classroom quality, by testing models for explaining quality in Head Start (see Figure 5-1). The first level of the model, the classroom (N=258), included the following predictors: Teacher having a Bachelor's or Associate's degree, teacher attitudes and knowledge, years of teaching experience, ethnicity and the teacher's salary as a deviation from the national average teacher salary of the Head Start program. The second level of the model, the program (N=43), included several variables describing the characteristics of families served by the Head Start program, the primary curriculum used by the program for all its classrooms (High/Scope curriculum, Creative Curriculum, or other/none), and the average of teacher salaries. Characteristics of the families enrolled in the program included: the percentage of parents with some college, the percentage of parents earning $1500/month or more, the percentage of non-minority children attending the program, and the percentage of language-minority children attending the program. The dependent variables consisted of the Fall 2000 quality scores, summarized earlier. Separate multi-level models were performed for each of these dependent variables.

During the Fall 2000 data collection period, which operated from September 2000 to December 2000, classroom quality and teacher data were obtained from 225 Head Start centers in 43 programs around the country. In Fall 2000, 278 classrooms were observed, from a possible 286 classrooms, for a completion rate of 97 percent.[5] In Spring 2001, 275 classrooms were observed out of 284 in the sample for a completion rate of 97 percent.[6] Finally, in Spring 2002, 204 classrooms were observed with sample children who were enrolled in Head Start for a second year (those who were three years of age in Fall 2000).

On the quality measures, agreement between two independent observers in a sample of classrooms during Fall 2000 averaged 93.5 percent for the Assessment Profile Scheduling Scale, 87.9 percent for the Assessment Profile Learning Environment Scale, and 86.7 percent for the Assessment Profile Individualizing Scale. Percent agreement across all ECERS scales (which includes direct hits and being off by one on a seven-point scale) averaged 79.5 percent, agreement on the ECERS-R Language subscale averaged 85.7 percent, and agreement on the Caregiver Interaction Scale averaged 93.9 percent.

---

[4] The PROC MIXED statistical procedure in SAS, as outlined by Singer (1998), was used.

[5] Due to some missing data among one or more of the quality indicators, the analytic sample was comprised of 258 classrooms.

[6] Due to missing data, the analytic sample was comprised of 256 classrooms.

5.3 Findings

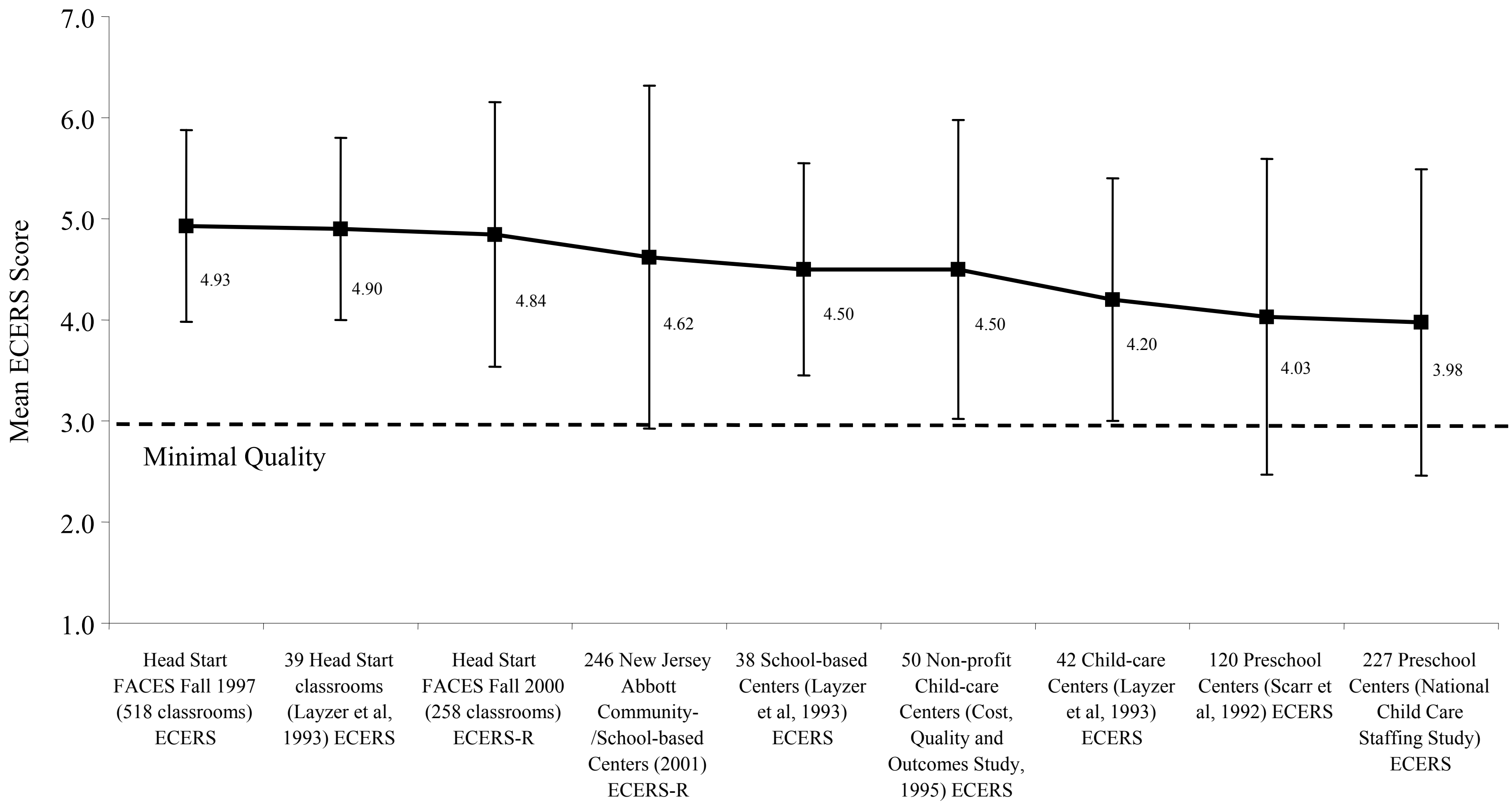
5.3.1 What is the Quality of Head Start Classrooms in 2000-2001, and How Does it Compare to Quality Reported in 1997-1998?

In FACES 2000, quality in Head Start classrooms continues to be good and above that found among other center-based preschools. Figure 5-2 compares the two FACES cohorts with other studies using the overall ECERS or ECERS-R score as the measure of quality. Not only is the average score higher, but the variation is less than that found in many other preschool settings.[7]

The Early Childhood Environment Rating Scale (ECERS-R). In FACES 2000, the overall average ECERS-R score for the 258 classrooms observed in Fall 2000 was 4.84 (with a standard deviation of .87). This compares with an overall average ECERS (the original scale) score of 4.93 (with a standard deviation of 0.63) for the 518 classrooms in the first FACES cohort (Fall 1997). While the FACES 2000 average score is comparable to that found earlier, the variability in the scores was higher, and approached that reported by the test developers for the revised measure. Since this comparison involved scores on the original ECERS and the revised ECERS, it is also hard to interpret any changes, because little is known of how the two measures are calibrated. In Spring 2001, the average ECERS-R total score remained virtually the same at 4.83 (standard deviation of 1.0). Thus, while the means remained the same from Fall to Spring, the variation in the quality scores increased, with a Spring 2001 standard deviation of 1.00 (from 0.84 in Fall 2000). In Spring 2002, with the reduced sample of classrooms, the ECERS-R total score averaged 4.84 with a standard deviation of 0.86, almost identical to the scores found for the total sample in Fall 2000. Overall, the average scores are consistent and indicate that Head Start classrooms continue to show good quality.

Table 5-1 compares the FACES 1997 and FACES 2000 cohorts on the ECERS/ECERS-R total score, with 1 indicating "inadequate" quality and 7 indicating "excellent" quality. The chi-square test for differences between the proportions of classrooms in the two cohorts was statistically significant (chi-square = 26.8234, df = 5, $p < .0001$) indicating significant differences from 1997 to 2000 in the proportion of classrooms rated from inadequate to excellent in quality. In FACES 1997, no classrooms were scored in the "inadequate" range (1 or 2) and only 4 of 518 classrooms were scored in the "minimal"

[7] It should be noted that, with the exception of the New Jersey study, all other studies used the original ECERS, and at this writing there are no published reports attesting to the appropriateness of comparing scores on the two versions.



Variability defined as +/- 1.5 SD's. It should be noted that studies published after 2001 used ECERS-R whereas all others used the original ECERS.

Figure 5-2. Classroom quality in Head Start compared with other preschool and child care settings range (score of 3). In FACES 2000, using the revised version of the ECERS which more strictly differentiated the highest and lowest ends of the scale, in Fall 2000, five classrooms (1.9 percent) were scored in the "inadequate" range and 15 of the 258 classrooms (5.7 percent) scored in the "minimal" range. A comparison of the proportions of classrooms that were rated "inadequate" from Fall 1997 and Fall 2000 reveals a non-significant difference ($z = -000003$), and the increase in classrooms rated "minimal" (from 0.8 percent to 5.7 percent) also did not reach statistical significance ($z = -1.50$). However, when combined, the number of classrooms rated "minimal" or lower (ECERS/ECERS-R scores of 3 or lower) increased significantly from 0.8 percent to 7.6 percent ($z = -2.12$, $p < .05$). Thus, compared to the earlier cohort, in FACES 2000 significantly more classrooms were rated lower in quality, but low-scoring classrooms still represented only 20 of 258 classrooms overall. On the other hand, the number of classrooms rated "excellent" (ECERS/ECERS-R scores of 6 or higher) increased significantly from FACES 1997 to FACES 2000 ($z = -3.31$, $p<.01$). In FACES 1997, there were 97 out of 518 (18.7 percent) classrooms rated "excellent" while in FACES 2000, 56 out of 258 classrooms (21.6 percent) were rated "excellent."

Table 5-1. Comparison of ECERS/ECERS-R categorical scores, Fall 1997, Fall 2000 and Spring 2001, percent of classrooms

ECERS Score category	Fall 1997* (N=518)	Fall 2000 (N=258)	Spring 2001 (N=264)	Spring 2002 (N=202)
1 "Inadequate"	0.0	0.0	0.0	0.0
2	0.0	1.9	2.3	1.5
3 "Minimal"	0.8	5.7	10.1	4.5
4	26.8	22.8	21.4	23.8
5 "Good"	53.7	48.1	36.7	50.5
6	18.1	20.9	28.0	18.3
7 "Excellent"	0.6	0.7	1.7	1.5
Total percentage	100	100	100	100

*Fall 1997 used the ECERS, whereas the ECERS-R was used in all other time periods.

A further comparison was done of the classrooms in the FACES 2000 sample, looking at any shifts in classroom quality from Fall to Spring (Table 5-1). The chi-square test for differences between the proportions of classrooms from Fall to Spring was statistically significant (chi-square = 11.27, df = 5, $p < .05$) indicating significant differences in proportion of classrooms rated from inadequate to excellent in quality from Fall to Spring of the same Head Start year. The proportion of classrooms in the lowest category (ECERS-R score of 2) increased slightly, from 1.9 percent in Fall 2000 and 2.3 percent in Spring 2001, but the differences did not reach statistical significance levels. However, there was a statistically significant increase in the proportion of classrooms rated a score of 3 for minimal quality, from 5.7 percent in Fall 2000 to 10.1 percent in Spring 2001 ($z = -2.15$, $p < .05$). When looking at the proportion of classrooms rated minimal or lower (ECERS-R scores of 3 or lower), there was a statistically significant increase from Fall to Spring, from 7.6 percent to 12.4 percent ($z = -2.66$, $p < .01$). There were also significantly fewer classrooms receiving a rating of 5 indicating good quality, from 48.0 percent in Fall 2000 to 36.4 percent in Spring 2001 ($z = 13.49$, $p < .001$). The Fall-to-Spring increases in the proportion of classrooms at the lower end of the quality scale were also mirrored by increases in the proportion of classes rated 6 or 7 for "excellent" quality. There was a statistically significant increase in the proportion of classrooms rated 6, from 20.9 percent in Fall to 28.4 percent in the Spring ($z = -6.66$, $p < .001$), although the increase in the percentage of classrooms rated 7 for "excellent" from 0.7 percent in the Fall to 1.7 percent in the Spring was not statistically significant. Overall, the proportion of classrooms rated "excellent" (ECERS-R scores of 6 or higher) increased significantly from 21.6 percent in Fall 2000 to 30.1 percent in Spring 2001 ($z = -7.70$, $p < .001$). Overall, it would appear that the slight shifts in categories reflects a movement towards both ends of the distribution in classroom quality, since the mean ECERS-R score remained almost the same from Fall to Spring.

Assessment Profile Scheduling, Learning Environment and Individualizing Scales. On the Scheduling scale, the raw scores were virtually identical from FACES 1997 to FACES 2000 (average raw scores of 11.17 in Fall 1997 and 11.12 in Fall 2000). In Spring 2001, the mean Scheduling score was 10.99, and in Spring 2002, with the reduced sample of classrooms, the mean Scheduling score was 11.39. The differences from Fall 2000 to Spring 2001 and Spring 2002 were not statistically significant. On the Learning Environment scale, the raw scores over the two cohorts were not significantly different (13.46 in Fall 1997 vs. 14.44 in Fall 2000). In Spring 2001, the Learning Environment mean score was 14.14 and in Spring 2002 the Learning Environment mean score was 14.63. The difference between the Fall 2000 and Spring 2001 Learning Environment scores was not statistically significant. However, the difference between the Spring 2001 and Spring 2002 scores was statistically significant, indicating an increase in the quality of learning materials in Head Start classrooms in the second year ($t=3.41$, $p < .001$). As well, from Fall 2000 to Spring 2002, Learning Environment raw scores increased significantly ($t=2.07$, $p=.04$). Finally, on the Individualizing Scale, a new measure in FACES 2000, the average raw score was 3.58 (out of a maximum raw score of 5), with a standard deviation of 1.2, indicating that overall Head Start classrooms provide an environment that takes into account the learning needs of individual students, but there is room for improvement. In particular, in Fall 2000 on the Individualizing Scale, 60 percent of classrooms maintained portfolios on individual children but only 37 percent of classrooms provided opportunities for children to evaluate their work, or to decide which products are included in their portfolios. However, with regard to the inclusion of children with disabilities, and making accommodations to allow these children to be included in classroom activities, 94 percent of classrooms were rated as having full inclusion and 90 percent of classrooms had provisions for accommodating children with special needs. In Spring 2001, the Individualizing raw score averaged 3.47, which was a statistically significant decrease from Fall 2000 ($t=-2.22$, $p=.03$). The Spring 2002 raw score averaged 3.52, which was not a statistically significant difference from either the Fall 2000 score or the Spring 2001 score.

Arnett Caregiver Interaction Scale. On this scale, measuring the sensitivity and responsiveness of teachers in Head Start classrooms, the average score for Fall 2000 was 71.5, almost identical to the score of 71.3 in FACES 1997. In Spring 2001, the average score was 72.24 and in Spring 2002, for the subsample of 201 classrooms, the average score was 73.65. Again, although the scores appear to increase from Fall 2000 to Spring 2001, the differences did not reach levels of statistical significance according to t-tests on the weighted means. Further, the differences between the Fall 2000 and the Spring 2002 means were also not statistically significant.

Summary of Changes in Process Quality Indicators Over Time. Comparisons across the two cohorts, FACES 1997 and 2000, revealed that, for all measures of classroom quality – the ECERS, the Assessment Profile scales and the Arnett Caregiver Interaction Scale – scores were consistent. Thus, it could be concluded that, in general, quality in Head Start classrooms was consistent over the two cohorts of classrooms, from 1997 to 2000.

When looking at changes in quality from Fall to Spring of FACES 2000 across a variety of indicators, very few of the differences reached levels of statistical significance. There was a statistically significant decrease in the Assessment Profile Individualizing raw score from Fall 2000 to Spring 2001, indicating a decrease in the focus on monitoring children's performance and adjusting the program to meet individual needs of each child. For the subset of classrooms observed in the second year of FACES 2000 (Spring 2002), most scores did not change significantly from the prior year. The only noteworthy difference was in the quality of the learning materials and literacy environment in the classroom, according to the Assessment Profile Learning Environment scale, in which the scores increased, indicating increased quality, from Spring 2001 to Spring 2002 and from Fall 2000 to Spring 2002.

Child:Adult Ratio. The average child:adult ratio for the FACES 2000 Head Start classrooms was 5.4 children per adult in Fall 2000, 5.7 in Spring 2001, and 6.1 in Spring 2002. These ratios compared with 6.3 children per adult at the Fall 1997 observation and 6.2 children per adult during the Spring 1998 observation.[8] Looking only at paid staff, the FACES 2000 cohort classrooms averaged 6.5 in Fall 2000, 6.4 in Spring 2001 and 6.8 in Spring 2002. Using weighted t-tests comparing the Fall 2000 and Spring 2001 means, the average child:staff ratios were not significantly different. However, the average child:adult ratio increased significantly from Fall 2000 to Spring 2001, indicating a decrease in quality ($t=2.66$, $p=.008$). The differences between Spring 2001 and Spring 2002 were not statistically significant and neither were the differences from Fall 2000 to Spring 2002 (likely due to the use of unweighted data in these comparisons with higher standard errors). The child:adult ratios were also lower than those found for Fall and Spring of the FACES 1997 cohort. The difference in these ratios suggests the important influence of volunteers on improving child:adult ratios in Head Start classrooms (lower ratios indicate higher quality). Despite a decline from Fall to Spring in FACES 2000, these ratios are far better than the NAEYC accreditation standard of eight or fewer three-year-olds or 10 or fewer four-year-olds for each adult and exceed the Head Start Program Performance Standards of 7.5 to 8.5 or fewer three-year-olds or 10 or fewer four-year-olds per adult.

Curricula. Teachers were asked at each interview what types of curricula they use and which curriculum is their primary curriculum. In Fall 2000, 91 percent of teachers reporting using a primary curriculum, with 23.5 percent of all teachers using the High/Scope curriculum and 41.3 percent using the Creative Curriculum. In Spring 2001, 93 percent of teachers reported that they used some type of curriculum, which represents a statistically significant increase from Fall 2000 in the use of any curriculum ($z = -3.48$, $p < .01$). In Spring 2001, 45.2 percent reported using the Creative Curriculum and 19.5 percent reporting the High/Scope curriculum. Compared to Fall 2000, in Spring 2001 more teachers reported using the Creative Curriculum ($z = -4.55$, $p < .001$) while fewer reported using the High/Scope curriculum ($z = 3.24$, $p < .01$).

---

[8] In Fall 1997, only the total number of adults in the classroom was measured, so that the discrepancy between 1997 and 2000 could be due to the slightly different measurement methods. In the remaining analyses in this chapter, the child-adult ratio rather than the child-staff ratio will be used, to remain consistent with the earlier cohort. In fact, the two measures are so highly correlated that results of statistical analyses using each will be almost identical.

In Spring 2002, 97.5 percent of teachers reported using a curriculum, significantly more than the 93 percent in Spring 2001 ($z = -5.88$, $p < .001$). In Spring 2002, 48.7 percent of teachers reported using the Creative Curriculum and 15.5 percent reported using the High/Scope curriculum. Using binomial tests comparing these proportions with those from Spring 2001, there was a significant decrease in the proportion of teachers reporting use of the High/Scope curriculum ($z = 4.66$. $p < .001$) and a significant increase in the proportion of teachers reporting use of the Creative Curriculum ($z = -7.29$, $p < .001$).

In general, from Fall to Spring of the same Head Start year there is an increase in the proportion of teachers reporting the use of one of the two most widely used integrated curriculum – Creative Curriculum and High/Scope curriculum – and the increase continues into the Spring of the following year when as many as 97.5 percent of teachers reported using one of the two most widely used integrated curriculum – Creative Curriculum and High/Scope curriculum. At the same time, when teachers were asked whether they use two of the most popular curricula, there was a steady decline in the proportion who reported using the High/Scope curriculum along with a steady increase in the proportion who reported using the Creative Curriculum, from Fall 2000 through to Spring 2001.

5.3.2 What are the Backgrounds, Qualifications and Experiences of Head Start Teachers in 2000-2001 and were there any Changes from the Earlier Cohort (1997-1998)?

In FACES 2000, Head Start teachers overall are experienced and qualified (see Tables 5-2 and 5-3) and these findings are consistent with the prior cohort. In Fall 2000, teachers in Head Start classrooms have been teaching in Head Start for an average of 7.9 years and they have been teaching for an average of 11.8 years in all educational settings. These data are almost identical to those reported for the earlier cohort, Fall 1997 (7.5 years teaching Head Start and 11.7 total years teaching, reported in the Third Progress Report, Administration for Children and Families, 2001).

In later waves of FACES 2000 (Spring 2001 and Spring 2002), the average years teaching were consistent (Table 5-3). In Spring 2001, the average number of years teaching Head Start was 7.95 and 11.7 total years teaching, whereas, one year later, teachers averaged 8.5 years of teaching Head Start and 12.1 total years teaching. Again, these differences did not reach levels of statistical significance.

Table 5-2. Comparison of lead teacher backgrounds, FACES Fall 1997 through Spring 2002

	Fall 1997 (N=437)	Fall 2000 (N=257)	Spring 2001 (N=264)	Spring 2002 (N=202)*
Years Teaching Head Start				
1-2 Years	14.2	21.1	23.4	19.8
3-4 Years	22.7	23.5	18.7	16.8
5-9 Years	34.1	27.6	29.3	33.5
10+ Years	29.0	27.8	28.6	30.0
Total	100.0	100.0	100.0	100.0
Highest Level of Education				
High School or Equivalent	10.8	10.4	12.1	11.5
Some College	31.4	32.2	32.5	31.5
Associate's Diploma	29.8	18.6	17.3	21.0
Bachelor's Degree or equivalent	24.9	27.8	28.1	29.0
Graduate or Professional Degree	3.2	10.9	10.0	7.0
Total	100	100	100	100.00
Teacher Age Category				
18-29	14.7	14.9	12.5	14.2
30-39	33.3	33.4	34.8	33.0
40-49	31.8	28.3	26.6	34.5
50-59	15.9	16.3	18.4	12.7
60 or Older	4.3	7.1	7.8	5.6
Total	100	100	100	100
Member in Early Childhood Education Association				
No	47.1	38.0	38.7	49.0
Yes	52.9	62.0	61.3	51.0
Total	100	100	100	100
Child Development Associates (DCA) Certificate or Equivalent				
No	23.9	26.0	28.4	34.0
Yes	76.1	74.0	71.6	66.0
Total	100	100	100	100
Teacher Ethnicity				
White, non-Hispanic	41.1	48.1	47.8	51.5
African American, non-Hispanic	34.2	33.7	33.7	28.5
Hispanic**	22.4	15.4	14.9	16.5
Asian	2.3	1.2	1.2	2.0
Multiple Race/Other	NA	1.7	2.4	1.5
Total	100	100	100	100

*Spring 2002 data are unweighted. No longitudinal weights for these classrooms were derived because they constitute only those classrooms where sampled children were three years of age in Fall 2000 and were attending a second year of Head Start.

**Puerto Rico was represented in FACES 1997, but not in the FACES 2000 sample, explaining the lower percentage of Hispanic teachers in this later cohort (2000).

Table 5-3. Comparison of teacher experiences and qualifications, Fall 2000 through Spring 2002

	Means or proportions (weighted)		
	Fall 2000 (N=257)	Spring 2001 (N=264)	Spring 2002 (N=202)*
Total Annual Salary (Mean Dollars)	20,750	20,823	21,727
Age in Years (Mean)	41.4	41.9	41.0
Total Years Teaching (Mean)	11.8	11.7	12.1
Years Teaching in Head Start (Mean)	7.9	8.0	8.5
Has a CDA (%)	58.0	55.8	58.3
Has a CDA or State-Awarded Certificate (%)	74.0	71.6	66.0
Has AA Degree or Higher (%)	57.0	54.9	57.0
Has BA or AA Degree (%)	41.0	39.2	43.0
Has BA Degree or Higher (%)	39.0	38.1	36.0
Has Teaching Certificate or License (%)	34.0	34.1	35.4
Field of Study Includes Early Childhood Education (%)	86.0	86.6	87.7
Currently enrolled in Teacher Related Training (%)	45.0	46.5	28.2

*Spring 2002 data are unweighted. No longitudinal weights for these classrooms were derived because they constitute only those classrooms where sampled children were three years of age in Fall 2000 and were attending a second year of Head Start.

The average amount of time teachers have spent in Head Start relative to their total years in teaching reflects that teachers spent most of their teaching careers in Head Start classrooms. However, there was a relatively wide range of experience in teaching Head Start that the average scores obscure. In Fall 2000, approximately 21 percent of the Head Start teachers were relatively new, having been teaching in Head Start for less than two years, and 28 percent had taught in Head Start for 10 years or more. The Spring 2001 and Spring 2002 data for the same cohort are not substantially different and reflect slight effect of the cohort as its members age across time, that is, fewer teachers were "new" in the following year and more teachers entered the older experienced group.

When teacher backgrounds in FACES 2000 were compared with the backgrounds of teachers in the earlier Fall 1997 cohort, several statistically significant differences emerged.[9] There was a statistically significant different in the number of years teaching in Head Start in Fall 2000 compared with Fall 1997 (chi-square = 8.38, df = 3, $p < .05$). Looking at the proportions in the two cohorts, there was an increase in the proportion of teachers who had taught for less than 2 years from 14 percent in Fall 1997 to 21 percent in Fall 2000 ($z = -8.26$, $p < .001$) and a decline in the percentage of teachers who taught from five to nine years from 34 percent in Fall 1997 to 28 percent in Fall 2000 ($z = 9.15$, $p < .001$).

[9] The chi-square statistic was calculated for these comparisons, using unweighted data on the independent samples because no classroom weights were created for the Fall 1997 sample.

Approximately the same number of teachers in both cohorts had been teaching in Head Start for 10 years or more (28 percent in Fall 2000 and 29 percent in Fall 1997).

Most Head Start teachers have good teaching qualifications, but lower than those of teachers in public elementary schools. In a survey of pre-kindergarten classrooms in the U. S. public schools in 2000-2001, 86 percent of pre-kindergarten teachers had a Bachelor's or higher degree (Smith, Kleiner, Parsad and Farris, 2002). Overall in Fall 2000, 46.4 percent of Head Start teachers had either a Bachelor's or an Associate's degree with 38.7 percent having a Bachelor's degree or higher and 57 percent having an Associate's degree or higher. In the Spring of that year (Spring 2001), 38 percent reported having a Bachelor's degree or higher and 55 percent reported having an Associate's degree or higher. In the following year, Spring 2002, these percentages remained very similar, 36 percent and 57 percent respectively.

Looking at the specific types of educational qualifications, in Fall 2000, 27.8 percent had a Bachelor's degree, 18.6 percent had an Associate's degree, and another 32.2 percent had some college but no degree in Fall 2000. In Spring 2001 and Spring 2002, the percentage of teachers reporting a Bachelor's degree was 28.1 percent and 29 percent respectively, and the percentage of teachers reporting an Associate's degree was 17.3 percent and 21 percent for Spring 2001 and Spring 2002. Thus, over the course of two years, the percentage of teachers with Bachelor's degree increased slightly from 27.8 percent to 29 percent while the proportion of teachers with Associate's degree increased from 18.6 percent in Fall 2000 to 21 percent in Spring 2002. However, caution must be exercised in interpreting these changes because the Spring 2002 sample consisted of only those classrooms with children in their second year of Head Start, and did not include all classrooms in the FACES sample (necessitating the use of unweighted data).

In Fall 2000, 74 percent of all teachers reported having the Child Development Associate (CDA) credential or a state-awarded preschool certificate, with 58 percent having the CDA only. In Spring 2001, teachers in the same cohort of classrooms reported similar percentages; 71.6 percent reported having a CDA or state-awarded certificate and 55 percent reported having a CDA diploma only. In Spring 2002, 66 percent reported having a CDA or state certificate, and 58 percent reported having just a CDA.

When looking at the change across cohorts, from Fall 1997 to Fall 2000, the proportion of teachers with a Bachelor's degree or higher increased significantly from 28.1 percent in Fall 1997 to

38.7 percent in Fall 2000, primarily due to an increase in the proportion of teachers with graduate level degrees, defined as a Master's degree, its equivalent or higher. In Fall 2000, 10.9 percent of teachers reported having a Master's degree or higher compared with only 3.2 percent in Fall 1997, with another 16 percent having received some graduate level training in Fall 2000.[10] The increase in the proportion of teachers with graduate level degrees was significantly correlated with the increase in the proportion of teachers with two years or less of teaching experience in Head Start, from 14 percent to 21 percent over the same time period. In Fall 2000, 32.9 percent of teachers with graduate level education or higher were new teachers, compared with only 13.3 percent in Fall 1997.[11] These results suggest that an increased number of new teachers with advanced degrees entered Head Start from 1997 to 2000.

In Spring 2001 the proportion of teachers with a Bachelor's degree or higher remained at the high level shown in the Fall, with 38 percent of teachers having these qualifications. In Spring 2002, 36 percent of teachers reported having a Bachelor's degree or higher, which is not a statistically significant change from the year earlier (especially given the reduced sample of teachers at this time period). In Spring 2001, the proportion of teachers with graduate level degrees also remained similar to the Fall 2000, with 10 percent of teachers, and in the following Spring (2002) the proportion dropped slightly (but not to a statistically significant degree) to 7 percent of teachers (with graduate level qualifications).

In addition to an increase from the prior FACES cohort in the proportion of teachers with advanced degrees, more teachers in FACES 2000 are reporting having studied Early Childhood Education or Child Development as part of their highest degree, whether for an Associate's, Bachelor's, or higher degree. In Fall 2000, 78 percent indicated that their field of study included Early Childhood Education or Child Development, compared with approximately 62 percent in Fall 1997. However, the proportion of teachers having a CDA or state preschool certificate has stayed approximately the same (74 percent in Fall 2000 compared with 76 percent in Fall 1997). In the same cohort of FACES 2000, from Fall 2000 to Spring 2001 and Spring 2002, the percentage of teachers who studied Early Childhood Education or Child Development in their degree work increased from 78 percent in Fall 2000 to 86.7 percent in Spring 2001 and 87.7 percent in Spring 2002, but these differences were not statistically significant. However, the percentage of teachers enrolled in teacher-related training was 45 percent in Fall 2000,

[10] The chi-square comparison using unweighted data on the independent samples was statistically significant at $p<.01$ (chi-square = 12.5371, 1 df).

[11] In Fall 2000 the chi-square test using weighted data was statistically significant at $p<.001$ (chi-square = 344.9624, 1 df). Further, a comparison of the two proportions from Fall 1997 to Fall 2000 using unweighted data also revealed a statistically significant increase at $p<.01$ ($z = 4.441$).

46.6 percent in Spring 2001 and declined to 28.2 percent in Spring 2002. The decrease in Spring 2002 was significant compared with both Fall 2000 (t=-2.01, p=.05) and Spring 2001 (t=-2.11, p=.04) and indicates that fewer teachers in Spring 2002 were engaged in continuing education activities.

In Spring 2001, the proportion of teachers reporting a CDA or state preschool certificate remained similar to that reported in the Fall, with 72 percent of teachers (compared with 74% in Fall 2000) having this qualification. In the following year (Spring 2001), the proportion of teachers having a CDA or state preschool certificate dropped slightly to 66 percent, which was not a statistically significant decrease.

Head Start teachers were, on average, 41.4 years old with a range from 23 to 73 years of age (standard deviation of 11.1 years). One-third of teachers belonged to the 30 to 39 year age group and another 28 percent were between the ages of 40 and 49 years. There were no differences in the proportions of teachers in each of the ten-year age groupings between the earlier cohort (Fall 1997) and this cohort (Fall 2000), with the exception that this cohort reflects an aging of teachers at the upper end of the distribution. Whereas only 4 percent of teachers were 60 years of age or older in Fall 1997, in this newest cohort (Fall 2000), 7 percent of teachers were 60 years of age or older.

There was also an increase in the membership of teachers in a national professional association for early childhood educators (e.g., NAEYC, NHSA, and NEA), from 53 percent in Fall 1997 to 62 percent in Fall 2000 but this increase did not reach levels of statistical significance. In Spring 2001, the proportion of teachers with membership in a national professional early childhood education association remained similar, with 61 percent of teachers reporting membership. In the following year (Spring 2002), the proportion dropped from 61 percent to 51 percent but this was not a statistically significant decline (likely due to the use of unweighted data for the comparisons).

In terms of racial and ethnic background, in Fall 2000, 33.7 percent of the teachers were African American, 15 percent were Hispanic, 1 percent were Asian and 48 percent were White. Compared with the Fall 1997 data, teachers in Fall 2000 were slightly less likely to be African American (the percentage of African American teachers declined to 33.7 percent from 34.2 percent in Fall 1997). Although it may appear that there was a strong decline in the percentage of Hispanic teachers, this decline is due entirely to the fact that Puerto Rico was included in the Fall 1997 but excluded from the Fall 2000 cohort. In FACES 2000, the proportion of teachers who were White increased slightly from 41 percent in Fall 1997 to 48 percent in Fall 2000 but this was not statistically significant and from Fall to Spring of

FACES 2000, the proportion of teachers remained very similar (see Table 5-2). In the following Spring (2002), the differences in the proportions of White (52 percent compared with 49 percent in Spring 2001) and African American teachers (29 percent compared with 34 percent in Spring 2001) over these two years were not statistically significant. The proportion of Hispanic teachers remained approximately the same in Spring 2002 (17 percent compared with 15 percent in Spring 2001).

In general, the data reveal that Head Start teachers are experienced and qualified to teach early childhood education. Compared with the first FACES cohort, Head Start teachers in FACES 2000 are more qualified. They are more likely to be White, younger, new to teaching Head Start and have higher educational levels including graduate degrees. They are also more likely to be trained in Early Childhood Education and to be members of a professional organization. There were few significant changes in the characteristics of the teachers across time periods within the FACES 2000 cohort, which likely reflects similarities in the teacher samples across each of the data collection periods.

Teacher Attitudes and Knowledge. In Fall 2000, on average, Head Start teachers showed generally positive attitudes and knowledge about instructional practice in early childhood education, with a mean score for all teachers of 7.9 out of a maximum score of 10 on the 9 items from the Teacher Beliefs Scale (see earlier description as well as Appendix). For example, they tended to agree with statements indicating positive attitudes and knowledge about early childhood education practices such as: "Head Start classroom activities should be responsive to individual differences in development" and "Children should be allowed to select many of their own activities from a variety of learning areas that the teacher has prepared (writing, science center, etc.)." Head Start teachers tended to disagree with statements indicating negative attitudes and a lack of knowledge about early childhood education practices such as: "Each curriculum area should be taught as a separate subject at separate times" and "Students should work silently and alone on seatwork." In Spring 2001, the average score was 7.9, indicating stable scores for this measure from Fall to Spring in the FACES 2000 cohort. In Spring 2002, the average score dropped to 7.2, which represents a statistically significant decline from Spring 2001 ($t = -3.27$, $p < .001$) and from Fall 2000 ($t = -3.73$, $p < .0003$) in teachers' knowledge and attitudes about early childhood education practices.

5.3.3 How are Teaching Qualifications and Experience Related to Attitudes and Knowledge of Early Childhood Education and How are these Factors Linked to Quality in Head Start Classrooms?

Teacher Background and Qualifications Correlated with Attitudes and Knowledge. The following analyses were comprised of bivariate correlations using either Spearman Rank-Order Correlations (for ordinal variables) or Pearson Product-Moment Correlations (for interval and binary-coded variables). The correlations show that teachers who had higher levels of knowledge about early childhood education practices tended to be more experienced and better educated (Table 5-4). Higher scores for teacher attitudes and knowledge of early childhood education practices were significantly correlated with higher levels of educational attainment (r=.16), some graduate school education or higher (r=0.13), more total years teaching (r=0.13), membership in an early childhood education association (r=0.19), teaching in a language other than Spanish or English (r=0.14), and teachers with White, Hispanic, or Asian racial backgrounds (r=-0.17). Teacher's attitudes and knowledge were not significantly correlated with teacher salary, years teaching Head Start, having a teaching certificate, having a CDA certificate, or having a course of study in Early Childhood Education.

Table 5-4. Teacher attitudes and knowledge correlated with teacher backgrounds and qualifications, Fall 2000 (N=257)

Teacher backgrounds and qualifications	Correlation coefficient	Significance (p level)
ECE Association Membership	0.19	0.01
Highest Level of Education	0.16	0.01
Teach in a language other than English or Spanish	0.14	0.01
Total Years Teaching	0.13	0.05
Some graduate school education or higher	0.13	0.05
Current Gross Annual Salary	0.09	Ns
Years Teaching Head Start	0.09	Ns
Field of Study includes ECE	0.09	Ns
Teaching Certificate	0.07	Ns
CDA Credential	0.06	Ns
Enrolled in teacher-related training	0.03	Ns
Hispanic Teacher	0.02	Ns
Teach in Spanish	0.00	Ns
Teach in English	-0.06	Ns
Teacher's Age	-0.06	Ns
African American Teacher	-0.17	0.01

*See Appendix for description of the variables. Highest level of education was ordinally-scaled. The rest were either interval or binary-coded.

This relationship between teacher backgrounds and qualifications and teacher's attitudes and knowledge about early childhood education practices is supported by findings from other studies.

Abbott-Shim, Lambert, and McCarty (2000) reported that teachers with higher levels of education also showed more positive attitudes and knowledge about early childhood education practices.

These results suggest that teachers who have higher levels of educational attainment especially at the graduate school level, more years of teaching experience overall, and who enroll in a related professional association are more likely to have knowledge and positive attitudes about early childhood education practices. These attitudes and knowledge should be expected to influence classroom quality, particularly since teacher credentials were related to classroom quality. In the next step toward building a model explaining the connection between teacher backgrounds, qualifications and experience and classroom quality, the relationship between teacher attitudes and knowledge and classroom quality is tested.

Teacher Backgrounds, Qualifications and Experience Correlated with Classroom Quality. In Fall 2000, teachers with more experience and higher levels of education tended to be in classrooms rated higher in classroom quality (Table 5-5). Specifically teachers with higher levels of education tend to be in classrooms rated higher on a number of quality indicators, including the ECERS-R Language subscale (r = 0.21), the Caregiver Interaction Scale (r=0.19), the ECERS-R total score (r=0.13), the quality factor score (r=0.15), and the classrooms had lower child:adult ratios, indicative of higher quality (r=-0.20). It is noteworthy that the correlation between education levels and the ECERS-R Language subscale was stronger than that for the total ECERS-R score.

Teachers with a BA or AA degree were in classrooms with lower child:adult ratios (r=-0.12), indicating higher quality, but also in classrooms rated lower on the ECERS-R Language subscale (r=-0.12), indicative of lower quality of language activities and materials. Teachers with a CDA tended to be in classrooms with higher ratios (r=0.16), indicative of lower quality, but were also in classrooms with higher Individualizing scores (r=0.19), indicating more individualized planning and attention. Teachers who had more years teaching overall (not just in Head Start) were rated higher in their sensitivity and responsiveness, as measured by the Caregiver Interaction Scale (r=0.17) and had higher ECERS-R Language subscale scores (r=0.13). Teachers with more years teaching Head Start were in classrooms rated higher on the Quality Composite score (r=.12). Teachers who were members of an early childhood education professional association were rated higher in their sensitivity and responsiveness, as measured by the Caregiver Interaction Scale (r=0.15), and were in classrooms rated higher on ECERS-R Language (r=0.16), the Assessment Profile Individualizing Scale (r=0.17), with higher Quality Composite scores (r=0.15). Teachers with at least some graduate school education (versus no graduate school education) were also in classrooms rated higher in quality as measured by the ECERS-R total score (r=0.16). Finally, teachers who had a teaching certificate (versus no certificate) were in classrooms rated higher on the overall ECERS-R (r=0.12) and the ECERS Language subscale (r=0.14), with lower child:adult ratios, indicative of higher quality (r=-0.19). Interestingly, teachers' total annual salaries were not significantly correlated with any of the quality indicators.

Table 5-5. Correlation of teacher backgrounds, qualifications and experience with classroom quality, Fall 2000 (N=257).

	ECERS-R Total	ECERS-R Language	Assessment profile individualizing	Caregiver interaction scale (Arnett)	Child-adult ratio	Quality factor score
	Correlation coefficient	Correlation coefficient	Correlation coefficient	Correlation coefficient	Correlation coefficient	Correlation coefficient
Highest Level of Education*	0.13*	0.21***	-0.06	0.19***	-0.20***	0.15**
Teacher has AA or BA	-0.07	-0.12*	-0.06	0.00	-0.12*	-0.10
Total Years Teaching	0.09	0.13*	0.05	0.17**	0.03	0.06
Years Teaching Head Start	0.05	0.07	0.04	0.08	0.05	0.12*
ECE Association Membership	0.10	0.16**	0.17**	0.15**	0.10	0.15**
CDA Credential	-0.10	-0.02	0.19***	-0.09	0.16**	0.05
Some Graduate School Education or higher	0.16**	0.21***	-0.03	0.14**	-0.11	0.10
Teaching certificate	0.12*	0.14**	-0.09	0.10	-0.19**	0.08

*Highest level of education was ordinally-scaled and the correlation coefficients were obtained from Spearman Rank-Order Correlations.

‡Significance levels as follows: *** indicates p < .001; ** indicates p < .01; * indicates p < .05. With the exception of highest level of education, all variables were either interval or binary-coded and coefficients were obtained from Pearson Product Moment Correlations.

These results indicate that teacher backgrounds and qualifications are related to higher levels of quality in Head Start classrooms. However, the above results are based only on simple correlations, and do not take into account the complex interplay between the various teacher-related factors. To understand the links between these factors, the relationships between teacher backgrounds and experience and teacher attitudes and knowledge about early childhood education practices were tested.

Teacher Attitudes and Knowledge Correlated with Classroom Quality. Teachers with more positive attitudes and knowledge about early childhood education practices tend to be in classrooms rated higher in quality (Table 5-6). This relationship was found for a wide variety of quality indicators, including the ECERS-R total score (r=0.32), the ECERS-R Language subscale (r=0.29), the Arnett Caregiver Interaction Scale (r=0.28), the Quality Composite score (r=0.27), and the Assessment Profile Individualizing subscale (r=0.15). It is noteworthy that the size of these correlations greatly exceed those found for the relationships between teacher attitudes and knowledge and teacher backgrounds, and they also exceed those found for the relationship between teacher backgrounds and classroom quality. Only child:adult ratio was not significantly related to teacher's attitudes and knowledge of early childhood education practices.

Table 5-6. Teacher attitudes and knowledge correlated with classroom quality indicators, Fall 2000 (N=257)

Classroom quality indicators	Correlation coefficient	Significance (p level)
ECERS-R Total Score	0.32	0.001
ECERS-R Language Subscale	0.29	0.001
Arnett Lead Teacher Total	0.28	0.001
Quality Factor Score	0.27	0.001
AP Individualizing Observed	0.15	0.01
Child:Adult Ratio	-0.06	Ns

Teachers holding more positive attitudes and knowledge of early childhood education practices were more sensitive and responsive to children, as measured by the Arnett Caregiver Interaction Scale. They were also more likely to adjust activities to meet the varying needs of individual children, as indicated by the Assessment Profile Individualizing scale. Teachers holding more positive attitudes and knowledge of practices in early childhood education were also in classrooms rated higher in overall quality on the ECERS-R and in classrooms with higher quality language activities, as rated by the ECERS-R Language subscale.

The above findings suggest that teacher backgrounds, experience and qualifications, notably their level of education, are significantly related to both teacher's attitudes and knowledge about early childhood education practice, and classroom quality. Teacher attitudes and knowledge alone was also significantly correlated with classroom quality. Thus, in variations in quality in Head Start classrooms, it would appear that teacher qualifications as well as their attitudes and knowledge about early childhood education practices are important.

These results are limited because they cannot specify, based on the simple correlations presented above, the complexities of the relationships between both teacher backgrounds and experiences and teacher attitudes and knowledge in explaining or predicting classroom quality. The relationships with classroom quality may be independent, suggesting that teacher education, for example, and teacher attitudes and knowledge contribute separately and uniquely to variations in quality, or there may be a more complicated, mediated pattern, in which teacher attitudes and knowledge may mediate the relationship between teacher backgrounds and experiences and classroom quality.

5.3.4 Do Factors Beyond the Classroom, Such as the Types of Curricula Head Start Programs Provide, the Average Teacher Salaries, and Characteristics of Families Served by the Head Start Program Explain Variations in the Quality of Head Start Classrooms?

As described earlier, a series of multi-level models were tested using the SAS PROC MIXED computer program (Singer, 1998; Bryk and Raudenbush, 1993). This method predicts average quality scores of a sample of classes, schools, or other educational units (in the present case, Head Start centers) from a set of characteristics of those units, such as measures of center demographics and program-level factors. Simultaneously, the models identify how the quality of Head Start in each center relate to a set of classroom-level teacher characteristics, such as teacher educational qualifications, experience, and attitudes and knowledge of early childhood education practice. The method provides a numerical estimate of how sizable the center-to-center variation in average quality scores is, relative to the class-to-class variation in scores within centers. In this approach, each level of the model is tested separately, with the results approximating those obtained from simple linear regression analyses, but then the final test involves both levels of the model entered simultaneously. This approach is possible only because of the hierarchical nature of the data, in which classrooms are nested within centers.

To determine how much Head Start programs differed in average quality, the total variability in classroom quality scores was divided into two components, the variation between programs in the average quality scores of the classrooms, and the variation among classrooms within the same program. If classrooms with differing levels of quality were randomly distributed across Head Start programs, one would expect that 90 percent or more of the variation in quality would Fall into the within-programs component, and only a small proportion would Fall into the between-programs component. On the other hand, if programs differed substantially in classroom quality, then, based on previous research, one might expect a significant amount of the variation to Fall into the between-program component (Bryk and Raudenbush, 1992). Such an approach was used effectively to study characteristics of school organization on teachers' sense of efficacy (Bryk and Driscoll, 1988, cited in Bryk and Raudenbush, 1992) and the effects of school characteristics on individual children's achievement (Lee and Bryk, 1989, cited in Bryk and Raudenbush, 1992).

In order to assess the contribution of the program- and classroom-level variables to the modeling of classroom quality, a sequence of regression models was run for each dependent variable (these are the indicators of quality listed earlier in the chapter, such as the ECERS-R total score). The first model, which Singer refers to as the "unconditional means model," a random effect parameter for the between-classroom variability in quality was added to the fixed effect for the overall mean (for all classrooms in the sample). This model had one fixed effect (the overall mean) and two variance components, one representing the variation between classroom means and the other representing the variation within classrooms. The second model added the classroom level teacher factors, including teacher education, years of teaching Head Start, membership in EC associations, teacher ethnicity, and teacher beliefs about developmentally appropriate practices to the unconditional model.[12] The third model tested the effect of the program-level factors including the effects of the proportion of families with incomes at the higher end of the poverty range, the proportion of non-minority students, the proportion of families with at least some college education, the use of Creative or High/Scope curriculum, and the center average teacher salaries to the unconditional model (independent of the classroom-level predictors). The fourth model added the classroom-level teacher factors to the program-level model and was the final model for predicting classroom quality scores.

As each successive model was run, the contribution of the additional variables was assessed by calculating the difference in the residual log likelihood statistic between the current model and the prior model in the sequence. This statistic multiplied by -2 is distributed approximately as chi-square, with the degrees of freedom equal to the difference between the degrees of freedom for the current model (with its additional parameters) and the prior model. Also examined was the extent to which the current model reduced the between-programs variance, the among-programs variance, and the total variance, relative to the sizes of these variances in the unconditional means model (Bryk and Raudenbush, 1992, p. 65; Singer, 1998, p. 332).

[12] The teacher variables were either interval level or binary-coded, allowing for their use and interpretation in these models as predictors. Teacher education was entered as a binary-coded variable coded 1 for teachers having an Associate's or a Bachelor's degree and 0 for teachers without these qualifications.

The results of the multi-level models are presented separately for each of the six key indicators of quality (ECERS-R total scores, ECERS-R Language subscale, Arnett Caregiver Interaction Scale, Assessment Profile Individualizing subscale, Child:Adult Ratio, and Quality Composite score), in Tables 5-7 through 5-12.[13] A summary of the final two-level models for all indicators of quality is presented in Table 5-13.

Predicting ECERS-R Total Score. Looking first at only the initial classroom-level model in which between-classroom variability in quality was added to the fixed effect for the overall ECERS-R total score, there were significant effects of teacher ethnicity and teacher knowledge and attitudes towards early childhood education. However, these classroom-level factors explained only 1.5 percent of the between-program variation and less than 1 percent of the variation overall. The second, program-level model revealed that programs with a higher proportion of non-minority students and with a higher proportion of language-minority students, tended to be rated higher in quality. In combining these two factors, one interpretation of these findings is that higher quality was found among programs serving white children and among programs serving children where English is not their primary language, such as Spanish-speaking children (representing the largest group of language-minority students). Program-level only effects accounted for 7.9 percent of overall variation in ECERS-R total scores and 32.2 percent of the between-program variation. However, in the final two-level model, when program- and classroom-level factors are combined, the initial effects of teacher ethnicity disappear and only teacher knowledge and attitudes at the classroom level remains a significant predictor. At the program-level the percentage of non-minority students and the percentage of language-minority students remained as significant predictors of quality. Overall, the two-level model accounted for 5.5 percent of the total variation and 28.3 percent of the between-program variation in ECERS-R total scores. These findings suggest that high quality classrooms are likely to be those where programs have a higher percentage of non-minority students, a higher percentage of language-minority students (e.g., Spanish speakers), and teachers who have greater knowledge about early childhood education practices (Table 5-7).

[13] The Assessment Profile Learning Environment and Scheduling subscales were not used in separate models because they were included in the Quality Composite.

Table 5-7. Multi-level model predicting ECERS-R mean score from teacher factors, controlling for program-level effects (n=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
(INTERCEPT = 4.8)	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	0.46				0.55	
Percent parents earning $1500/month or more	0.28				0.44	
Percent non-minority students	1.13	***			0.82	**
Percent language-minority students	1.04	*			1.04	*
High/Scope Curriculum	0.35				0.30	
Creative Curriculum	0.32				0.33	
Average Annual Teacher Salary	0.14				0.10	
Classroom-Level Teacher Factors						
Teacher having BA or AA			-0.05		-0.09	
Teacher Attitudes and Knowledge			0.12	***	0.11	***
Years Teaching Experience			0.01		0.01	
African-American Teacher			-0.33	**	-0.14	
Hispanic Teacher			-0.23		-0.18	
Teacher Salary (deviation from program mean)			0.02		0.03	

‡Significance levels as follows: *** indicates p < .001; ** indicates p < .01; * indicates p < .05.

Predicting ECERS-R Language Subscale. The initial classroom-level model revealed significant effects of teacher knowledge and attitudes and teacher ethnicity. These classroom-level factors predicted 55 percent of the between-program variation and 18.9 percent of the total variation in ECERS-R Language subscale scores. The second, program-level model revealed that programs with a higher proportion of non-minority students and with a higher proportion of language-minority students tended to be rated higher in quality. There was also a non-significant trend towards families whose parents had some college education or higher being in programs rated higher on ECERS-R Language subscale but this was only at p < .10. Program-level only effects accounted for 15.6 percent of overall variation in ECERS-R Language subscale scores and 56.2 percent of the between-program variation. However, in the final two-level model, when program- and classroom-level factors are combined, the initial effects of teacher ethnicity disappear and only teacher knowledge and attitudes at the classroom level remains a significant predictor. There was a non-statistically significant trend towards teacher education having an effect as well. At the program-level, none of the factors reached levels of statistical significance although percentage of non-minority and language-minority students in the program reached trend levels (p < .06

and .10 respectively). Overall, the two-level model accounted for 25.3 percent of the total variation and 73.4 percent of the between-program variation in ECERS-R total scores. These findings suggest that classrooms with higher quality of language activities and materials were likely to be those whose teachers had higher scores for attitudes and knowledge about early childhood education practice. Having a teacher with a BA or AA, or having a higher percentage of non-minority students or a higher percentage of language-minority students were associated with higher quality at the trend level, but did not reach statistical significance (Table 5-8).

Table 5-8. Multi-level model predicting ECERS-R language subscale scores from teacher factors, controlling for program-level effects (N=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
INTERCEPT = 4.80	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	0.53	†			0.46	
Percent parents earning $1500/month or more	0.24				0.36	
Percent non-minority students	1.17	***			0.68	†
Percent language-minority students	0.91	†			0.84	†
High/Scope Curriculum	0.29				0.24	
Creative Curriculum	0.25				0.27	
Average Annual Teacher Salary	0.11				0.08	
Classroom-Level Teacher Factors						
Teacher having BA or AA			-0.17		-0.23	†
Teacher Attitudes and Knowledge			0.17	***	0.16	***
Years Teaching Experience			0.01		0.01	
African-American Teacher			-0.50	***	-0.22	
Hispanic Teacher			-0.53	*	-0.34	
Teacher Salary (deviation from program mean)			0.11		0.11	

‡Significance levels as follows: *** indicates p < .001; ** indicates p < .01; * indicates p < .05, † indicates a non-significant trend at p < .10.

Predicting Assessment Profile Individualizing Score. The initial classroom-level model revealed no significant effects and classroom-level factors predicted only 1.5 percent of the between-program variation and 1.1 percent of the total variation in the Assessment Profile Individualizing scores. The second, program-level model revealed that programs with a higher proportion of parents earning $1,500 per month or more tended to be rated higher on this quality measure (Individualizing). Program-level only effects accounted for 3.3 percent of overall variation in Individualizing subscale scores and

8.5 percent of the between-program variation. In the final two-level model, when program- and classroom-level factors are combined, the initial effects of the program-level factor (parents earning $,1500 per month or more) remains a significant predictor. Overall, the two-level model accounted for 5.4 percent of the total variation and 12 percent of the between-program variation in Individualizing scores. These findings, while significant, are not as strong as many of the other models indicating that the program- and classroom-level factors in the models do not explain a large amount of variation in Assessment Profile Individualizing scores. Nevertheless, the significant findings do suggest that classrooms with a focus on individualizing the classroom activities for individual students (an indicator of greater quality) are likely to be located in Head Start programs with greater numbers of families with relatively higher incomes (Table 5-9).

Table 5-9. Multi-level model predicting assessment profile individualizing scores from teacher factors, controlling for program-level effects (n=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
INTERCEPT = 3.5	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	0.60				0.52	
Percent parents earning $1,500/month or more	3.67	**			3.82	**
Percent non-minority students	0.15				0.14	
Percent language-minority students	0.26				0.22	
High/Scope Curriculum	-0.34				-0.40	
Creative Curriculum	-0.38				-0.38	
Average Annual Teacher Salary	-0.33				-0.34	†
Classroom-Level Teacher Factors						
Teacher having BA or AA			0.09		0.10	
Teacher Attitudes and Knowledge			0.05		0.04	
Years Teaching Experience			0.01		0.01	
African-American Teacher			0.03		0.06	
Hispanic Teacher			-0.02		0.04	
Teacher Salary (deviation from program mean)			0.11		0.15	

‡Significance levels as follows: *** indicates $p < .001$; ** indicates $p < .01$; * indicates $p < .05$, † indicates a non-significant trend at $p < .10$.

Predicting Teacher Arnett Caregiver Interaction Scale Score. The initial classroom-level model in which between-classroom variability in quality was added to the fixed effect for the overall Arnett Caregiver Interaction Scale (CIS) score (lead teacher), there were significant effects of teacher ethnicity, teacher knowledge and attitudes towards early childhood education, and years of teaching experience. These classroom-level factors explained 13.6 percent of the total variation in Arnett CIS scores and 42.7 percent of the between-program variation. The second, program-level model revealed that programs with a higher proportion of non-minority students and that use the Creative Curriculum tended to be rated higher in quality. There was a non-significant trend towards programs using the High/Scope curriculum also being rated higher in quality, but this did not reach statistical significant levels ($p < .08$). Program-level only effects accounted for 12.6 percent of overall variation in Arnett CIS scores and 44.3 percent of the between-program variation. However, in the final two-level model, when program- and classroom-level factors are combined, the initial effects of teacher ethnicity disappear but teacher knowledge and attitudes and teacher years of experience remained significant predictors at the classroom level. At the program-level the percentage of non-minority students disappeared and only the use of Creative Curriculum and, to a much lesser extent, the use of the High/Scope curriculum remained as significant predictors of quality. Programs that used the High/Scope curriculum (19.5 percent in Spring 2001) had classrooms with somewhat higher scores for teacher sensitivity, although this did not reach statistically significant levels in the final model. Overall, the two-level model accounted for 20.4 percent of the total variation and 59.7 percent of the between-program variation in Arnett CIS scores. These findings suggest that curriculum, teacher experience, and teacher attitudes and knowledge were significant predictors, with the strongest being the teachers' attitudes and knowledge of early childhood education practice, followed by programs that used the Creative Curriculum. Classrooms with sensitive and responsive teachers are likely to be those that use the Creative Curriculum (45.2 percent in Spring 2001), where the teachers have more years of teaching experience and where the teachers hold more positive attitudes and knowledge about practices in early childhood education (Table 5-10).

Teacher education level, indicated by whether teachers had an Associate's or Bachelor's degree or not, was not significantly related to teacher sensitivity, even though it was related in earlier analyses without program level factors included. These results further suggest that the relationship between teacher education and classroom quality is not direct, but rather is mediated by their knowledge and attitudes towards early childhood education practice as well as by the type of curriculum used and the teacher's level of experience.

Table 5-10. Multi-level model predicting Arnett caregiver interaction scale scores from teacher factors, controlling for program-level effects (n=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
INTERCEPT = 70.96	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	9.60				10.59	
Percent parents earning $1500/month or more	-0.18				2.83	
Percent non-minority students	7.93	*			3.24	
Percent language-minority students	-3.28				-1.98	
High/Scope Curriculum	6.21	†			5.33	†
Creative Curriculum	6.22	*			6.41	*
Average Annual Teacher Salary	2.47				1.62	
Classroom-Level Teacher Factors						
Teacher having BA or AA			-0.15		-1.09	
Teacher Attitudes and Knowledge			1.62	***	1.55	***
Years Teaching Experience			0.17	*	0.17	*
African-American Teacher			-3.81	*	-1.75	
Hispanic Teacher			-6.67	***	-3.91	
Teacher Salary (deviation from program mean)			0.48		1.09	

‡Significance levels as follows: *** indicates p < .001; ** indicates p < .01; * indicates p < .05, † indicates a non-significant trend at p < .10.

Predicting Quality Composite Score. The initial classroom-level model in which between-classroom variability in quality was added to the fixed effect for Quality Composite score, there were significant effects only for teacher knowledge and attitudes towards early childhood education. This classroom-level factor explained 11.62 percent of the total variation in Arnett CIS scores and 28 percent of the between-program variation. The second, program-level model revealed that programs with a higher proportion of non-minority students and language-minority students tended to be rated higher in quality. There was a non-significant trend towards programs with a higher percentage of with at least some college level education being rated higher in quality, but this did not reach statistically significant levels ($p < .06$). Program-level only effects accounted for 12.2 percent of overall variation in Quality Composite scores and 52.4 percent of the between-program variation. In the final two-level model, when program- and classroom-level factors are combined, teacher knowledge and attitudes remained a significant predictor at the classroom level. At the program-level, the percentage of language-minority students remained as a significant predictor ($p < .01$) whereas the percentage of non-minority students was no longer statistically significant ($p < .06$) and the percentage of parents with some college education or higher remained as a non-significant trend ($p < .06$). Overall, the final two-level model accounted for 19.2 percent of the total variation and 55.9 percent of the between-program variation in Quality Composite scores. These findings suggest that classrooms with higher scores for quality on this indicator (comprising the ECERS-R Language subscale, and the Assessment Profile Scheduling and Learning Environment Scales) were likely to be those from Head Start programs with a higher percentage of language-minority students, and whose teachers had more positive attitudes and knowledge of early childhood education practice (Table 5-11).

Table 5-11. Multi-level model predicting quality factor score from teacher factors, controlling for program-level effects (n=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
INTERCEPT = 0.003066	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	0.48	†			0.46	†
Percent parents earning $1500/month or more	0.15				0.19	
Percent non-minority students	0.67	**			0.43	†
Percent language-minority students	1.20	**			1.27	**
High/Scope Curriculum	0.25				0.22	
Creative Curriculum	0.26				0.26	
Average Annual Teacher Salary	-0.07				-0.11	
Classroom-Level Teacher Factors						
Teacher having BA or AA			-0.06		-0.09	
Teacher Attitudes and Knowledge			0.15	***	0.14	***
Years Teaching Experience			0.01		0.01	
African-American Teacher			-0.23		-0.04	
Hispanic Teacher			-0.20		-0.21	
Teacher Salary (deviation from program mean)			-0.02		-0.03	

‡Significance levels as follows: *** indicates $p < .001$; ** indicates $p < .01$; * indicates $p < .05$, † indicates a non-significant trend at $p < .10$.

Predicting Child:Adult Ratio. The initial classroom-level model revealed no significant effects and classroom-level factors predicted only 1.5 percent of the between-program variation and less than 1 percent of the total variation in the child:adult ratio. The second, program-level model revealed that programs with higher average annual teacher salaries had lower child:adult ratios, indicative of higher quality. Program-level effects accounted for 7.9 percent of overall variation in child:adult ratios and 32.2 percent of the between-program variation in these ratios. In the final two-level model, when program- and classroom-level factors are combined, the initial effects of the program-level factor (average annual teacher salary) remained a significant predictor. Overall, the two-level model accounted for 5.5 percent of the total variation and 28.3 percent of the between-program variation in child:adult ratios. These findings, while significant, are not as strong as many of the other models indicating that the program- and classroom-level factors in the models do not explain a large amount of variation in child:adult ratios. Nevertheless, the significant findings do suggest that classrooms with lower child:adult ratios, indicating higher quality, are likely to be those where teachers are paid more (Table 5-12).

Table 5-12. Multi-level model predicting child-adult ratio from teacher factors, controlling for program-level effects (N=265), Fall 2000

Independent variables	Program-level effects		Class-level effects		Program + class effects	
(INTERCEPT = 5.4)	Estimate	Signif‡	Estimate	Signif	Estimate	Signif
Program-Level Factors						
Percent parents with some college or higher	-2.64				-2.57	
Percent parents earning $1500/month or more	0.49				0.50	
Percent non-minority students	-0.96				-0.87	
Percent language-minority students	-1.23				-1.31	
High/Scope Curriculum	-0.88				-0.88	
Creative Curriculum	-0.58				-0.56	
Average Annual Teacher Salary	-0.84	***			-0.81	**
Classroom-Level Teacher Factors						
Teacher having BA or AA			-0.16		-0.04	
Teacher Attitudes and Knowledge			0.03		0.04	
Years Teaching Experience			0.00		0.00	
African-American Teacher			0.35		0.14	
Hispanic Teacher			0.31		0.21	
Teacher Salary (deviation from program mean)			0.08		0.04	

‡Significance levels as follows: *** indicates $p < .001$; ** indicates $p < .01$; * indicates $p < .05$, † indicates a non-significant trend at $p < .10$.

Taken together, the results of these multi-level analytic models enhance our understanding of factors that explain quality in Head Start by "unpacking" some of the key elements that lead to higher quality (Table 5-13). The factors included in the model do a good job of explaining the more

Table 5-13. Summary of program- and classroom-level factors predicting classroom quality, Fall 2000

	Predicting ECERS-R total		Predicting AP individualizing		Predicting caregiver interaction scale		Predicting child-adult ratio		Predicting ECERS-R language		Predicting quality composite Score	
	(Intercept=4.8)		(Intercept=3.5)		(Intercept = 70.96)		(Intercept= 5.4)		(Intercept=4.80)		(Intercept=.003)	
	Estimate	Signif‡	Estimate	Signif	Estimate	Signif	Estimate	Signif	Estimate	Signif	Estimate	Signif
<u>Program-Level Factors</u>												
Percent parents with some college or higher	0.55		0.52		10.59		-2.58		0.47		0.463	†
Percent parents earning $1500/month or more	0.44		3.82	***	2.83		0.50		0.36		0.188	
Percent non-minority students	0.82	**	0.14		3.24		-0.87		0.68	†	0.425	†
Percent language-minority students	1.05	*	0.22		-1.98		-1.31		0.84	†	1.270	**
High/Scope Curriculum	0.30		-0.40		5.33	†	-0.88		0.24		0.217	
Creative Curriculum	0.32		-0.38		6.41	*	-0.56		0.27		0.264	
Average Annual Teacher Salary	0.10		-0.34	†	1.62		-0.81	**	0.08		-0.108	
<u>Classroom-Level Teacher Factors</u>												
Teacher having BA or AA	-0.09		0.10		-1.09		-0.04		-0.24	†	-0.095	
Teacher Attitudes and Knowledge	0.11	***	0.04		1.55	***	0.04		0.16	***	0.136	***
Years Teaching Experience	0.01		0.01		0.17	*	0.00		0.01		0.006	
African-American Teacher	-0.14		0.06		-1.75		0.14		-0.22		-0.045	
Hispanic Teacher	-0.18		0.04		-3.91		0.21		-0.34		-0.213	
Teacher Salary (deviation from program mean)	0.03		0.16		1.09		0.04		0.11		-0.035	

‡Significance levels as follows: *** indicates p < .001; ** indicates p < .01; * indicates p < .05, † indicates a non-significant trend at p < .10.

"process-oriented" aspects of quality, such as learning materials, quality of language activities and teacher-child interactions, but they do less well at predicting the more structural aspects of quality, such as child:adult ratios.

5.4 Summary and Conclusions

Head Start classrooms continue to show good levels of quality, based on the indicators of quality measured in FACES 2000 in the Fall (2000), Spring (2001) and the following Spring (2002). These levels of quality are consistent from the first cohort, and the consistency is evident across a wide variety of the indicators. Head Start teachers are qualified and experienced (although as a group they do not have the same level of credentials as public school teachers), and there appear to be substantially more teachers with higher educational attainment in this cohort compared with the first FACES cohort, in 1997. Head Start teachers in 2000 are younger, compared with those in 1997-1998, and more of them have been teaching in Head Start for two years or less. These newer teachers are also the ones most likely to have a graduate school degree. These results appeared to hold for the Spring 2001 classrooms (which consisted of many of the same teachers) and for the Spring 2002 classrooms (despite limitations of this sample).

The role of teacher attitudes and knowledge, experience and education has been illuminated by these analyses. Classrooms with higher levels of quality are likely to be those whose teachers have higher levels of education, experience, and knowledge and attitudes of early childhood education practices. However, the relationship between teacher education and classroom quality is explained by teacher's attitudes and knowledge of early childhood education practices, so that teachers who are more educated have more positive attitudes and knowledge, which translates into higher levels of classroom quality.

In general, when all three factors are included, the direct relationships occur most consistently and strongly for attitudes and knowledge about early childhood education practice, and to a lesser extent teacher experience, rather than with teacher education. When these other factors were not included there was a significant relationship between teacher education and classroom quality, suggesting that the role of teacher education in influencing classroom quality is an indirect one. Teachers with higher levels of education have more positive attitudes and knowledge about early childhood education practice, and they are more likely to be in classrooms rated higher in quality. Thus, teacher attitudes and knowledge mediates the role of teacher education in explaining classroom quality.

Interestingly, programs using one of the two most widely used integrated curriculum – High/Scope curriculum and Creative Curriculum – also have teachers with positive attitudes and knowledge about early childhood education practice. Both of these factors appear to have the strongest effect on teacher sensitivity and responsiveness compared with other indicators of quality.

The multi-level results suggest that variations in the quality of Head Start classrooms may be explained by characteristics of the families and children they serve, by the curriculum used in the program, and by teacher attitudes and knowledge about early childhood education practice. The results suggest that Head Start classroom quality may be affected by factors beyond the classroom door, that are characteristics of the program and the families who participate.

CHAPTER 6.
RELATIONSHIP OF PROGRAM, CLASSROOM, AND CHILD AND FAMILY CHARACTERISTICS TO CHILDREN'S COGNITIVE GAINS IN HEAD START

This chapter explores variations in child achievement and behavior across local Head Start programs and classes. It uses multilevel modeling to test hypotheses about early education program and classroom characteristics and child and family educational activities that many child development scholars believe to be associated with enhanced cognitive growth in preschool children (Phillips, Mekos, Scarr, McCartney, and Abbott-Shim, 2000; Whitebook, Howes, and Phillips, 1989). The characteristics that are thought to make a difference for program effectiveness include the following:

- Using an integrated and comprehensive preschool curriculum;
- Having more ample program resources;
- Providing classrooms that are of higher quality as early learning environments;
- Employing a better prepared teaching staff;
- Providing preschool services for a longer period each day;
- Conducting educational activities in smaller groups with more personal adult attention to the needs and preferences of individual children;
- Having parents who engage in more educational activities with their children at home; and
- Having children who attend Head Start for more than one year.

This chapter examines relationships between these characteristics and several measures of children's cognitive development while accounting for the influence of other variables. Dependent variables consisted primarily of direct assessment measures of children's letter recognition and pre-reading skills, vocabulary knowledge, early writing and early math skills, book knowledge, and knowledge of color names.[1] Covariates included measures of the socioeconomic and ethnic composition

[1] Analyses with norm-referenced measures (i.e., PPVT-III and the Letter-Word Identification, Applied Problems, and Dictation subtests from WJ-R), used W ability scores, or IRT scale scores based on item response theory. Analyses with criterion-referenced measures (e.g., book knowledge, color names) used raw scores. For description of the measures and their respective scores, please refer to the Appendix.

of the families and children participating in each program and classroom, characteristics of the child like age, sex, and disability status, and measures of parents' literacy skills.[2]

The multilevel analyses of FACES 2000 data did indeed show that some of the aforementioned characteristics listed above were significantly related to variations in the size of the gains children made in Head Start. This chapter enumerates the program, classroom, and child and family characteristics that seemed to make a difference, describes the nature and size of the relationships involved, and also lists factors that were *not* significantly related to gains in achievement or behavior and discusses possible reasons why hypothesized relationships failed to materialize.

6.1 Conceptual Framework

The conceptual framework that guided our analyses was a multi-level, multi-causal model of the influences that shape children's cognitive development and the factors that help determine the nature of the experience children have in Head Start. (See Figure 6-1.) This view posits that children's development in the early years is primarily a function of the experiences they have in their families. Children from low-income families, whose parents tend to have lower educational attainments than other parents, often do not experience the same extent or quality of intellectual stimulation at home as children from middle-class families (Phillips et al., 1998). Furthermore, their parents are less able to purchase high-quality supplementary or substitute care in the marketplace. In some cases, children from low-income families may also not receive as much emotional support from parents as they need for optimal development. A center-based early childhood learning environment such as Head Start may help provide experiences that would be beneficial for the development of all children, but especially for those from higher-risk family environments (NICHD Early Child Care Research Network, 2000). Also, parents' involvement in their children's educational experiences may be an important factor, and one that programs can foster.

[2] FACES 2000 included a wider range of program and classroom characteristics that could be related to differences in achievement than did FACES 1997. FACES 2000 added interview questions and observational procedures that looked more systematically at areas like the curriculum used in each center and classroom, training and support for that curriculum, teacher salary levels, teacher knowledge and beliefs, and the use of child portfolios and other procedures aimed at individualizing instruction. In addition, the sample design was modified to yield a larger number of sample children in each sample classroom. This produced more stable estimates of class means and more variation in child characteristics in each classroom subsample. The modified design made it possible to carry out multilevel regression analysis at three levels: the program, classroom, and child level. In FACES 1997, multilevel regression analyses could be carried out at only the center and child levels.

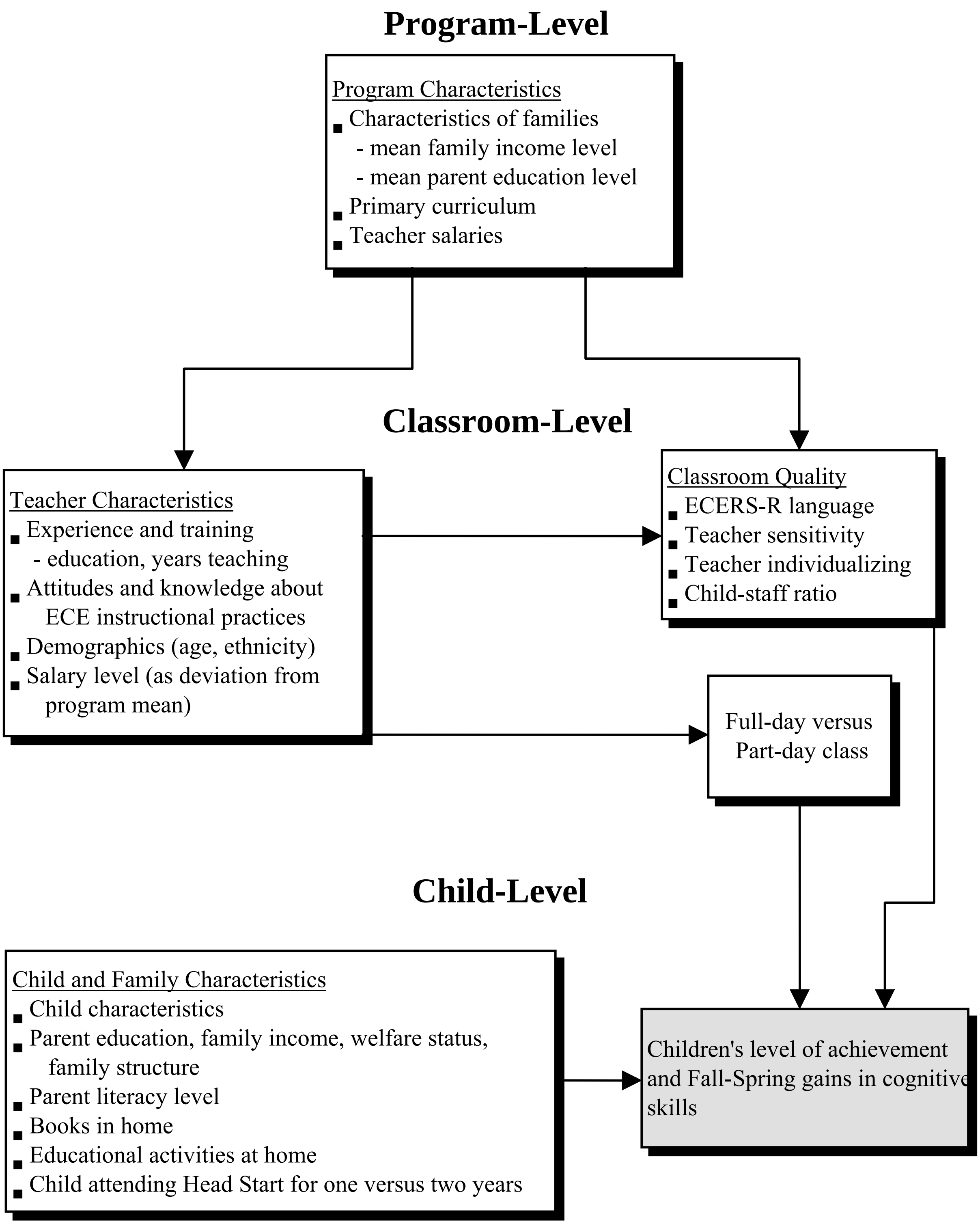


Figure 6-1. Analytical model of multi-level factors predicting classroom quality and children's achievement and gains in the Head Start year

The nature of the learning environment that a given child experiences in Head Start depends on the training and teaching experience of teachers in the program, and the resources available to them in terms of facilities, materials and teaching assistants. Programs with more resources are likely to be better able to provide adequate facilities and materials and recruit and retain talented and well-prepared teachers (Whitebook, Howes, and Phillips, 1989).

The character of the classroom environment that a given Head Start program is able to provide for a child is not just a question of program resources. It also depends on the educational philosophy to which the program adheres, and the kind of curriculum teachers use. Other things equal, children would be expected to do better in programs that employ well-thought-out curricula that are comprehensive and integrated in terms of educational activities and assessment methods. This is especially the case if the program is able to provide teachers with adequate training and support in the curriculum. At the same time, children's progress in a given cognitive or social-emotional area depends on whether the program's basic philosophy and curriculum of choice are supportive of efforts to bolster that area of child development.

6.2 Research Questions

The analyses reported in this chapter used multilevel regression models to address the following research questions:

1. Do Head Start programs and classes differ in the cognitive gains that children have attained when they leave the program?

2. Do children in Head Start programs that employ one of the two most widely used integrated early childhood curricula–Creative Curriculum or High/Scope curriculum–show larger cognitive gains than children in Head Start programs that employ other curricula?

3. Do children in Head Start programs that pay higher salaries to their lead teachers show larger cognitive gains than children in Head Start programs that have less ample resources?

4. Do children in Head Start classes that are of higher quality, as indicated by their receiving higher scores on the Language subscale of the Early Childhood Environment Ratings Scale–Revised (ECERS-R) or the Caregiver Interaction Scale (CIS), show larger cognitive gains than children in classes that are of lower quality on these measures?

5. Do children in Head Start classes led by better prepared teachers show larger cognitive gains than children in classes led by teachers who are less well prepared? Indicators of teacher preparation that were examined included whether the teacher had a Bachelor's degree or Associate's degree or higher, her years of teaching experience, her annual salary as a deviation from the program mean salary, and her score on a scale that measured positive attitudes and knowledge about early childhood educational practices.

6. Do children who participate in Head Start classes for a longer period each day–who attend "full-day" classes–show larger cognitive gains than children who participate for a shorter period of time–who attend "part-day" classes?

7. Do children in Head Start classes with lower child:staff ratios, and that provide more attention to the needs and preferences of individual children, as indicated by a higher score on the Assessment Profile Individualizing scale, show larger cognitive gains than children who are in classes with higher child:staff ratios or lower individualizing scores?

8. Do children whose parents engage in more educational activities at home with their children show larger cognitive gains than children whose parents engage in fewer educational activities? The presented analyses examined whether children whose parents read to them on a daily basis at home showed greater gains than children whose parents read to them less often. Frequency of reading was reported by the parents themselves in parent interviews in the Fall of 2003; and

9. Do children who attend Head Start for 2 years show larger cognitive gains than children who attend for only 1 year?

6.3 Analysis Method

The analysis method used to examine associations between Head Start program and class characteristics, as well as children's educational activities in the home and the number of years in Head Start, and children's cognitive development was multilevel linear regression modeling, using the SAS PROC MIXED computer program (Singer, 1998; Bryk and Raudenbush, 1992). Multilevel modeling shows how the average achievement scores of a sample of classes, schools, or other educational units (in the present case, Head Start programs and classes) relate to a set of characteristics of those units, such as measures of program demographics and classroom quality. Simultaneously, this type of modeling can examine how the achievement scores of individual children in each program and class relate to a set of child-level characteristics, such as child demographics and home literacy activities. This method provides a numerical estimate of how sizable the program-to-program and class-to-class variation in average scores is, relative to the child-to-child variation in scores within classes.

The primary dependent variables were the cognitive assessment scores at graduation from Head Start and the gains each child made between entry and graduation. Models were also constructed of assessment scores[3] attained by Head Start children in the FACES national sample in the Fall of 2000 and the Spring of their final Head Start year.[4] Each analytic model had three levels. The first level involved variation in average assessment scores or average gain scores across the 43 programs in the FACES national sample, expressed as deviations of the program means from the overall mean score for the entire sample. The second level involved variations of class means from the overall program means. And the third level involved variation in individual children's scores or gain scores around the class means.

There were three levels of independent variables used to model or predict the assessment scores and gain scores. At the program level, the independent variables consisted of measures representing the curriculum employed by the program, average teacher salary levels in the program, and average demographic and socioeconomic characteristics of the children who attended each program and their families. At the classroom level, independent variables consisted of measures of teacher preparation, teacher background characteristics, whether the class was of full-day or part-day duration, and indicators of classroom quality such as the ECERS-R Language subscale and Caregiver Interaction Scale. Class-level variables also included measures of the demographic composition of the class, expressed as deviations from the average demographic characteristics of the program. At the child level, the independent variables were measures that represented demographic characteristics of the child, socioeconomic, cultural, and structural characteristics of the family, parent literacy levels, disability status of the child, the frequency of parental reading to the child, and whether the child attended Head Start for one year or two years.

Statistical tests were conducted to assess whether a given set of independent variables (program-level, class-level, and child-level) improved the model's fit to the data, over and above simpler models that did not include that set. Tests were also done as to whether the regression coefficient for a given independent variable was reliably greater than zero. Details about variable definitions, means and

[3] In the multilevel regression models, assessment scores for the norm-referenced measures were converted to "W-ability scores," based on IRT scaling of item difficulties carried out by the test developers. These scale scores are purported to have equal-interval properties that are desirable in regression modeling, particularly of gain scores. In other analyses, standard score versions of the assessment scores were used. These scores show how Head Start children performed compared to national norms. But they do not have as strong equal-interval properties as the W-ability scores. Analyses with criterion-referenced measures used raw scores. For more information on the measures and scores used in FACES 2000, please refer to the Appendix.

[4] Children participating in FACES 2000 who entered Head Start in Fall 2000 were all new to Head Start. However, some children graduated after that first year of Head Start (Spring 2001) while others graduated after 2 years (Spring 2002). The entire sample of one- and two-year graduates are included in these analyses. The number of years of participation in Head Start was included in the models in the form of a dummy variable indicating if the child was a one-year graduate (being a two-year graduate is the reference group).

ranges, reliability of measures, and statistical tests used to ascertain the reliability of findings are described in the Appendix of the report.

6.4 Findings

Multilevel regression analyses of the assessment measures showed that the addition of the program- and classroom-level variables provided unique and significant contributions above and beyond the child-level variables to the prediction of children's assessment scores at Head Start entry and graduation, as well as gains from entry to graduation (see Table 6-1). The addition of the program- and classroom-level variables significantly contributed to the prediction of all nine assessment scores at entry into Head Start, increasing the variance accounted for by the model by approximately 1 to 7 percent. The addition of the program- and classroom-level variables significantly contributed to the prediction of eight of the nine graduation scores, increasing the variance accounted for in these models by approximately 2 to 10 percent. Finally, the addition of the program- and classroom-level variables significantly contributed to the prediction of five of the nine gain scores, increasing the variance accounted for in these models by approximately 2 to 4 percent.

The multilevel regression analyses showed that there were significant relationships between some of the program and class characteristics identified above and variations in children's cognitive gains. Tables 6-2 through 6-10 contain the details of the results from the multilevel regression analyses. The following factors seemed to make a difference for children's progress:

1. Programs using the High/Scope curriculum;
2. Programs having higher teacher salaries;
3. Teachers having Bachelor's or Associate's degrees or higher;
4. Children attending full-day rather than part-day classes;
5. Parents reporting that they read to their children more frequently; and
6. Children attending Head Start for 2 years.

Table 6-1. Percentage of variance accounted for by the child-level model and the combined model with program-, classroom-, and child-level factors, and the difference in the proportions

	Proportion of variance in dependent variable accounted for								
	Entry			Graduation			Gain		
	Child-level model	Combined model	Additional variance (change in R^2)	Child-level model	Combined model	Additional variance (change in R^2)	Child-level model	Combined model	Additional variance (change in R^2)
PPVT-III	51.2	54.1	2.9***	47.7	51.0	3.3***	35.5	37.1	1.6*
WJR Letter-Word ID	10.0	17.4	7.4***	15.2	24.0	8.8***	15.4	19.6	4.2**
WJR Applied Problems	26.1	31.7	5.6***	20.2	28.0	7.8***	15.8	18.7	2.9**
WJR Dictation	20.3	26.3	6.0***	14.9	24.4	9.5***	10.1	13.8	3.7
Book Knowledge	19.6	21.9	2.3***	15.4	20.9	5.5***	7.4	11.1	3.7**
Color Naming	29.0	31.4	2.4***	15.7	18.0	2.3***	25.9	28.2	2.3***
Design Copying	20.9	22.1	1.2**	13.3	15.0	1.7	11.1	12.5	1.4
One-to-One Counting	19.1	21.5	2.4***	12.7	15.4	2.7***	8.4	10.2	1.8
Social Awareness	18.3	20.8	2.5***	25.2	27.6	2.4***	13.6	15.9	2.3

Note: The test of model significance was a X^2 (df = 17) test. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 6-2. Three-level regression models of PPVT-III Vocabulary scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	PPVT-III score upon Head Start entry	PPVT-III score upon Head Start graduation	Gains in PPVT-III score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	2.57*	3.20**	0.66
Program Mean Family Income Level	1.47	-1.49	-2.99*
High/Scope Curriculum	0.49	1.91†	1.43
Creative Curriculum	1.22	1.21	0.00
Mean Teacher Salary Level	0.54	0.75	0.23
Proportion Language-Minority Children	-0.41	3.35	3.68†
Proportion Non-Minority Children	5.22**	6.76***	1.49
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	0.24	-0.19	-0.47†
Average Child-Adult Ratio	0.25	0.30	0.06
AP Individualizing Score	-0.17	-0.01	0.20
Teacher BA or AA or higher	0.42	-0.27	-0.67
Years Teaching Experience	0.00	-0.04	-0.05
African American Teacher	-0.38	0.43	0.69
Hispanic Teacher	-0.39	0.76	1.21
Teacher Salary Deviation Score	0.02	0.45	0.46
Teacher DAP Beliefs Score	0.06	0.01	-0.05
Average Lead Teacher Arnett Score	-0.01	0.03	0.04
Class Parent Education Level (deviation)	1.16†	0.94†	-0.23
Class Family Income Level (deviation)	-1.06	-0.60	0.41
Proportion Language-Minority (deviation)	0.11	-1.84	-2.26
Class Proportion Non-minority (deviation)	6.68***	4.50**	-2.19
Full -Day Class	-0.78	-0.34	0.48
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.50*	0.56**	0.08
Family Income Level (deviation)	0.13	0.27	0.13
Welfare Status	-0.48	-0.61	-0.14
Age of Child in Months	0.91***	0.70***	-0.21***
Sex of Child	0.08	-0.12	-0.18
Disability Status	-3.03***	-2.24***	0.86
African American Child	-2.57**	-3.61***	-1.00
Hispanic Child	-1.47	-2.18**	-0.70
Language-Minority Family	-8.82***	-6.58***	2.29*
Mother-Father Family	-0.79	-0.41	0.37
Neither Birth Parent In Home	1.83	-0.55	-2.33*
Parent Literacy Standard Score (KFAST)	0.14***	0.12***	-0.02
Books In Home	1.47*	1.24*	-0.21

See note at end of table.

Table 6-2. Three-level regression models of PPVT-III Vocabulary scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	PPVT-III score upon Head Start entry	PPVT-III score upon Head Start graduation	Gains in PPVT-III score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	-0.75	-0.26	0.52
One or twice	-1.45*	-1.55**	-0.11
Every day	1.04†	1.28*	0.27
One-year Head Start graduate	-0.64	-10.73***	-10.11***
Intercept	-8.40	21.13***	29.29***
Proportion of Variance Accounted For			
Between-Programs Variance	100	83	64
Between-Classes Variance	13	5	16
Within-Classes Variance	1	1	0
Total Variance	6	6	2

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** $p < .001$

** $p < .01$.

* $p < .05$.

† $p < .10$.

N = 1,600.

Table 6-3. Three-level regression models of WJR Letter-Word ID scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	WJR LW ID score upon Head Start entry	WJR LW ID score upon Head Start graduation	Gains in WJR LW ID score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	4.16*	4.90*	0.61
Program Mean Family Income Level	2.54	0.01	-2.30
High/Scope Curriculum	-1.17	3.85†	4.62*
Creative Curriculum	-1.35	0.94	1.97
Mean Teacher Salary Level	1.52	3.88**	2.03†
Proportion Language-Minority Children	5.34	2.94	-2.46
Proportion Non-Minority Children	-2.03	1.83	3.44
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	-0.63	-0.90	-0.11
Average Child-Adult Ratio	0.62*	1.15***	0.46
AP Individualizing Score	0.37	-0.17	-0.50
Teacher BA or AA or higher	2.30*	1.70	-0.54
Years Teaching Experience	0.00	0.00	0.00
African American Teacher	-3.10*	-0.59	2.43
Hispanic Teacher	-0.57	-0.13	0.53
Teacher Salary Deviation Score	-1.47	-1.46	-0.11
Teacher DAP Beliefs Score	-0.16	-0.88*	-0.77†
Average Lead Teacher Arnett Score	0.06	0.03	-0.03
Class Parent Education Level (deviation)	1.67†	3.23*	1.56
Class Family Income Level (deviation)	-0.10	0.39	0.57
Proportion Language-Minority (deviation)	-6.23	0.52	7.25
Class Proportion Non-minority (deviation)	1.59	4.18	2.73
Full -Day Class	0.42	2.26	1.67
Child-Level Predictor Variables			
Parent Education Level (deviation)	1.24**	1.99***	0.75†
Family Income Level (deviation)	0.95†	1.03	0.13
Welfare Status	-0.93	-2.94	-2.07†
Age of Child in Months	0.51***	0.64***	0.15
Sex of Child	0.50	1.54	1.14
Disability Status	-1.54	-4.03*	-2.40†
African American Child	0.95	3.00	2.22
Hispanic Child	-1.60	-0.61	1.23
Language-Minority Family	2.99	-0.89	-3.92†
Mother-Father Family	-1.12	-1.22	0.02
Neither Birth Parent In Home	4.02†	-0.71	-4.87*
Parent Literacy Standard Score (KFAST)	0.11**	0.06	-0.04
Books In Home	0.62	-0.10	-0.75

See note at end of table.

Table 6-3. Three-level regression models of WJR Letter-Word ID scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

	WJR LW ID score upon Head Start entry	WJR LW ID score upon Head Start graduation	Gains in WJR LW ID score from entry to graduation
Program-level predictor variables	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	-1.39	-2.86	-1.73
One or twice	-1.10	-3.63*	-2.78*
Every day	0.10	-1.25	-1.53
One-year Head Start graduate	-0.46	-14.80***	-14.05***
Intercept	291.38***	315.15***	22.84†
Proportion of Variance Accounted For			
Between-Programs Variance	75	96	100
Between-Classes Variance	100	58	34
Within-Classes Variance	3	--	0
Total Variance	8	10	5

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** *p* < .001

** *p* < .01

* *p* < .05

† *p* < .10

N = 797

Table 6-4. Three-level regression models of WJR Applied Problems scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	WJR AP score upon Head Start entry	WJR AP score upon Head Start graduation	Gains in WJR AP score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	9.35**	7.70*	-1.82
Program Mean Family Income Level	-3.44	-2.91	1.08
High/Scope Curriculum	7.74*	5.43†	-2.54
Creative Curriculum	1.22	3.22	1.57
Mean Teacher Salary Level	2.28	2.99†	0.79
Proportion Language-Minority Children	-2.84	-17.46*	-15.26*
Proportion Non-Minority Children	13.95*	8.30	-5.22
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	0.12	-1.11	-1.32
Average Child-Adult Ratio	1.93***	1.40**	-0.54
AP Individualizing Score	-0.03	0.53	0.61
Teacher BA or AA or higher	2.27	1.04	-1.27
Years Teaching Experience	0.05	0.01	-0.04
African American Teacher	1.30	-0.83	-2.01
Hispanic Teacher	2.32	5.53	3.01
Teacher Salary Deviation Score	-2.26	0.41	2.81
Teacher DAP Beliefs Score	0.00	-1.09†	-1.10†
Average Lead Teacher Arnett Score	0.01	0.12	0.13
Class Parent Education Level (deviation)	0.60	-0.31	-1.08
Class Family Income Level (deviation)	0.94	2.13	1.53
Proportion Language-Minority (deviation)	-7.14	1.29	7.36
Class Proportion Non-minority (deviation)	11.54*	8.20†	-2.92
Full -Day Class	1.08	0.52	-0.64
Child-Level Predictor Variables			
Parent Education Level (deviation)	1.43†	1.82**	0.44
Family Income Level (deviation)	2.06*	1.29	-0.68
Welfare Status	-0.16	1.79	1.75
Age of Child in Months	1.86	1.27***	-0.57**
Sex of Child	0.64	0.41	-0.20
Disability Status	-12.29***	-9.68***	2.74
African American Child	-7.09*	-8.94**	-1.65
Hispanic Child	1.70	-1.93	-3.08
Language-Minority Family	-8.17	-4.26	5.20
Mother-Father Family	-3.91*	-2.72†	1.17
Neither Birth Parent In Home	-0.15	-1.75	-1.12
Parent Literacy Standard Score (KFAST)	0.19**	0.19**	0.00
Books In Home	5.49*	1.30	-4.40*

See note at end of table.

Table 6-4. Three-level regression models of WJR Applied Problems scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	WJR Applied problems score upon Head Start entry	WJR Applied problems score upon Head Start graduation	Gains in WJR Applied problems score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	2.24	0.80	-1.90
One or twice	-1.67	-1.92	-0.30
Every day	0.83	-1.15	-1.98
One-year Head Start graduate	0.37	-16.60***	-16.98***
Intercept	224.43***	308.95***	82.52***
Proportion of Variance Accounted For			
Between-Programs Variance	100	100	100
Between-Classes Variance	61	48	100
Within-Classes Variance	--	--	3
Total Variance	8	10	3

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** *p* < .001

** *p* < .01

* *p* < .05

† *p* < .10

N = 674

Table 6-5. Three-level regression models of WJR Dictation scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	WJR Dictation score upon Head Start entry	WJR Dictation score upon Head Start graduation	Gains in WJR Dictation score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	8.82*	13.75**	4.85
Program Mean Family Income Level	0.93	1.13	-0.24
High/Scope Curriculum	4.63	8.90†	4.40
Creative Curriculum	0.53	1.43	1.12
Mean Teacher Salary Level	7.32**	6.07*	-1.40
Proportion Language-Minority Children	11.81	1.11	-11.05
Proportion Non-Minority Children	16.06*	4.91	-11.32
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	-2.01	-4.32**	-2.22
Average Child-Adult Ratio	1.31†	1.92**	0.57
AP Individualizing Score	1.48	0.82	-0.66
Teacher BA or AA or higher	5.22†	6.94*	1.70
Years Teaching Experience	0.16	0.15	-0.01
African American Teacher	2.55	-4.45	-6.81†
Hispanic Teacher	-0.70	2.11	3.04
Teacher Salary Deviation Score	-2.68	-1.76	0.92
Teacher DAP Beliefs Score	-1.00	0.31	1.25
Average Lead Teacher Arnett Score	0.19	0.27†	0.08
Class Parent Education Level (deviation)	0.81	6.93*	6.13*
Class Family Income Level (deviation)	-4.47	-5.26	-0.69
Proportion Language-Minority (deviation)	-17.78	-10.13	8.31
Class Proportion Non-minority (deviation)	3.79	7.97	4.62
Full -Day Class	5.62†	8.36**	2.55
Child-Level Predictor Variables			
Parent Education Level (deviation)	4.44***	4.52***	0.08
Family Income Level (deviation)	2.35†	1.06	-1.24
Welfare Status	-2.49	-5.53†	-2.96
Age of Child in Months	2.60***	2.18***	-0.43
Sex of Child	8.38***	7.03**	-1.40
Disability Status	-9.73**	-12.83***	-3.09
African American Child	-1.52	2.93	4.19
Hispanic Child	4.87	5.91	0.98
Language-Minority Family	0.37	12.94†	12.47†
Mother-Father Family	-3.98	-1.76	2.23
Neither Birth Parent In Home	-6.14	1.34	7.74
Parent Literacy Standard Score (KFAST)	0.19†	0.17	-0.02
Books In Home	3.06	3.29	0.05

See note at end of table.

Table 6-5. Three-level regression models of WJR Dictation scale scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	WJR Dictation score upon Head Start entry	WJR Dictation score upon Head Start graduation	Gains in WJR Dictation score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	0.87	3.78	2.78
One or twice	-3.68	-8.63*	-4.91
Every day	-2.19	-2.29	-0.20
One-year Head Start graduate	-1.41	-24.87***	-23.27***
Intercept	111.25***	162.75***	53.79†
Proportion of Variance Accounted For			
Between-Programs Variance	100	100	100
Between-Classes Variance	100	100	71
Within-Classes Variance	0	--	0
Total Variance	8	11	4

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** *p < .001*

** *p < .01*

* *p < .05*

† *p < .10*

N = 674

Table 6-6. Three-level regression models of Book Knowledge scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

	Book Knowledge score upon Head Start entry	Book Knowledge score upon Head Start graduation	Gains in Book Knowledge score from entry to graduation
Program-level predictor variables	Unstandardized regression coefficients		
Program Mean Parent Education Level	0.16	0.57***	0.42*
Program Mean Family Income Level	0.12	-0.13	-0.26
High/Scope Curriculum	0.20	0.25†	0.05
Creative Curriculum	0.04	0.10	0.06
Mean Teacher Salary Level	0.06	0.16†	0.09
Proportion Language-Minority Children	0.12	-0.17	-0.21
Proportion Non-Minority Children	0.17	0.07	-0.09
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	-0.01	0.00	0.00
Average Child-Adult Ratio	0.03†	0.04†	0.01
AP Individualizing Score	-0.03	0.00	0.04
Teacher BA or AA or higher	0.00	-0.02	-0.01
Years Teaching Experience	0.00	0.00	0.01
African American Teacher	-0.06**	-0.03	0.00
Hispanic Teacher	-0.11	0.03	0.06
Teacher Salary Deviation Score	-0.03	0.21*	0.22*
Teacher DAP Beliefs Score	0.06	-0.02	-0.08*
Average Lead Teacher Arnett Score	0.00	0.00	0.01
Class Parent Education Level (deviation)	0.09	0.01	-0.07
Class Family Income Level (deviation)	0.08	0.13	0.05
Proportion Language-Minority (deviation)	0.21	-0.30	-0.42
Class Proportion Non-minority (deviation)	0.49*	-0.15	-0.63*
Full -Day Class	0.03	0.25*	0.24*
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.07**	0.06*	0.00
Family Income Level (deviation)	0.00	0.06†	0.06
Welfare Status	-0.07	0.08	0.15†
Age of Child in Months	0.07***	0.06***	0.00
Sex of Child	0.11	0.24***	0.11†
Disability Status	-0.14†	-0.28**	-0.14
African American Child	-0.11	-0.35**	-0.24†
Hispanic Child	0.05	-0.06	-0.09
Language-Minority Family	-0.12	-0.07	0.07
Mother-Father Family	-0.10	0.00	0.10
Neither Birth Parent In Home	-0.06	-0.14	-0.08
Parent Literacy Standard Score (KFAST)	0.01*	0.01**	0.00
Books In Home	0.06	0.03	-0.02

See note at end of table.

Table 6-6. Three-level regression models of Book Knowledge scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	Book Knowledge score upon Head Start entry	Book Knowledge score upon Head Start graduation	Gains in Book Knowledge score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	-0.07	0.09	0.17
One or twice	-0.26***	-0.14†	0.13
Every day	0.00	0.20**	0.20*
One-year Head Start graduate	0.01	-0.74***	-0.74***
Intercept	-3.45***	-3.52***	-0.20
Proportion of Variance Accounted For			
Between-Programs Variance	46	73	100
Between-Classes Variance	40	22	25
Within-Classes Variance	0	0	0
Total Variance	3	6	4

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** $p < .001$

** $p < .01$

* $p < .05$

† $p < .10$

N = 674

Table 6-7. Three-level regression models of Color Naming scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	Color Naming score upon Head Start entry	Color Naming score upon Head Start graduation	Gains in Color Naming score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	1.26†	0.80†	-0.49
Program Mean Family Income Level	0.96	0.67	-0.32
High/Scope Curriculum	-0.98	-0.36	0.62
Creative Curriculum	-1.11†	-0.32	0.78
Mean Teacher Salary Level	0.36	0.16	-0.21
Proportion Language-Minority Children	2.35	0.41	-1.96
Proportion Non-Minority Children	2.78*	0.22	-2.53*
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	0.10	-0.09	-0.18
Average Child-Adult Ratio	0.26*	0.16*	-0.10
AP Individualizing Score	0.04	-0.06	-0.09
Teacher BA or AA or higher	-0.50	0.12	0.59
Years Teaching Experience	-0.02	0.00	0.02
African American Teacher	-0.23	-0.77*	-0.56
Hispanic Teacher	0.52	-0.31	-0.83
Teacher Salary Deviation Score	0.04	0.24	0.22
Teacher DAP Beliefs Score	0.06	-0.05	-0.12
Average Lead Teacher Arnett Score	-0.01	0.01	0.02
Class Parent Education Level (deviation)	0.60	0.68**	0.09
Class Family Income Level (deviation)	0.13	-1.17**	-1.29**
Proportion Language-Minority (deviation)	-0.22	0.38	0.56
Class Proportion Non-minority (deviation)	2.32*	0.18	-2.08*
Full -Day Class	-0.56	0.35	0.90*
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.52***	0.22*	-0.30*
Family Income Level (deviation)	0.22	-0.12	-0.34*
Welfare Status	-0.70†	-0.74**	0.00
Age of Child in Months	0.42***	0.19***	-0.23***
Sex of Child	1.39***	1.10***	-0.30
Disability Status	-1.95***	-0.67*	1.28**
African American Child	-0.84	-0.45	0.41
Hispanic Child	0.14	-0.28	-0.47
Language-Minority Family	-0.76	-0.23	0.56
Mother-Father Family	-0.19	-0.07	0.12
Neither Birth Parent In Home	2.03**	0.92†	-1.09
Parent Literacy Standard Score (KFAST)	0.07***	0.03***	-0.04**
Books In Home	0.50	0.17	-0.33

See note at end of table.

Table 6-7. Three-level regression models of Color Naming scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	Color Naming score upon Head Start entry	Color Naming score upon Head Start graduation	Gains in Color Naming score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	0.69	-0.08	-0.75
One or twice	-0.76†	-0.13	0.62†
Every day	0.60	0.06	-0.52
One-year Head Start graduate	-0.28	-3.55***	-3.28***
Intercept	-24.49***	2.15	26.95***
Proportion of Variance Accounted For			
Between-Programs Variance	74	39	68
Between-Classes Variance	--	35	22
Within-Classes Variance	1	0	0
Total Variance	3	3	3

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** *p* < .001

** *p* < .01

* *p* < .05

† *p* < .10

N = 1,710

Table 6-8. Three-level regression models of Design Copying scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

	Design Copying score upon Head Start entry	Design Copying score upon Head Start graduation	Gains in Design Copying score from entry to graduation
Program-level predictor variables	Unstandardized regression coefficients		
Program Mean Parent Education Level	0.14	0.20	0.06
Program Mean Family Income Level	0.26	0.03	-0.25
High/Scope Curriculum	-0.10	0.09	0.20
Creative Curriculum	-0.24*	-0.13	0.10
Mean Teacher Salary Level	-0.03	0.14	0.16†
Proportion -Minority Children	0.51†	-0.02	-0.49
Proportion Non-Minority Children	0.10	-0.18	-0.28
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	-0.06	0.02	0.08
Average Child-Adult Ratio	0.01	0.01	0.01
AP Individualizing Score	-0.02	0.00	0.03
Teacher BA or AA or higher	0.12	0.16	0.06
Years Teaching Experience	0.00	0.01	0.00
African American Teacher	0.01	0.10	0.07
Hispanic Teacher	-0.04	0.16	0.18
Teacher Salary Deviation Score	-0.01	-0.02	-0.02
Teacher DAP Beliefs Score	0.02	-0.07*	-0.09**
Average Lead Teacher Arnett Score	0.00	0.00	0.00
Class Parent Education Level (deviation)	-0.02	0.00	0.00
Class Family Income Level (deviation)	0.01	0.02	0.01
Proportion Language-Minority (deviation)	-0.01	-0.56	-0.58
Class Proportion Non-minority (deviation)	0.28	-0.35	-0.59*
Full -Day Class	0.03	0.10	0.06
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.03	0.09*	0.06
Family Income Level (deviation)	-0.03	0.01	0.04
Welfare Status	-0.09	-0.16	-0.09
Age of Child in Months	0.09***	0.11***	0.02**
Sex of Child	0.34***	0.26**	-0.08
Disability Status	-0.25**	-0.54***	-0.28*
African American Child	0.05	-0.53**	-0.57**
Hispanic Child	0.22†	0.00	-0.21
Language-Minority Family	0.38**	0.25	-0.13
Mother-Father Family	-0.01	0.12	0.13
Neither Birth Parent In Home	-0.33*	-0.10	0.25
Parent Literacy Standard Score (KFAST)	0.00	0.00	0.00
Books In Home	0.24**	0.15	-0.08

See note at end of table.

Table 6-8. Three-level regression models of Design Copying scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	Design Copying score upon Head Start entry	Design Copying score upon Head Start graduation	Gains in Design Copying score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	-0.10	0.20	0.32
One or twice	-0.04	-0.23†	-0.19
Every day	-0.02	-0.16	-0.14
One-year Head Start graduate	0.00	-1.42***	-1.42***
Intercept	-2.84***	-1.21	1.54
Proportion of Variance Accounted For			
Between-Programs Variance	100	100	54
Between-Classes Variance	20	23	26
Within-Classes Variance	0	0	0
Total Variance	2	2	2

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** *p* < .001

** *p* < .01

* *p* < .05

† *p* < .10

N = 1,729

Table 6-9. Three-level regression models of One-to-One Counting scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	One-One Counting score upon Head Start entry	One-One Counting score upon Head Start graduation	Gains in One-One Counting score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	0.22†	0.31*	0.09
Program Mean Family Income Level	0.07	0.13	0.03
High/Scope Curriculum	0.05	0.00	-0.05
Creative Curriculum	0.06	-0.12	-0.19
Mean Teacher Salary Level	0.03	-0.07	-0.09
Proportion Language-Minority Children	0.35	0.40	0.05
Proportion Non-Minority Children	0.08	-0.24	-0.33
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	0.03	-0.02	-0.05
Average Child-Adult Ratio	0.03†	0.02	-0.01
AP Individualizing Score	-0.04	-0.04	0.00
Teacher BA or AA or higher	-0.03	0.03	0.04
Years Teaching Experience	0.01	0.00	0.00
African American Teacher	0.11	-0.06	-0.19
Hispanic Teacher	0.02	0.01	-0.01
Teacher Salary Deviation Score	-0.06	0.18*	0.24*
Teacher DAP Beliefs Score	0.01	-0.03	-0.05
Average Lead Teacher Arnett Score	0.00	0.00	0.00
Class Parent Education Level (deviation)	0.14*	0.17*	0.03
Class Family Income Level (deviation)	0.07	-0.08	-0.14
Proportion Language-Minority (deviation)	-0.19	-0.25	-0.06
Class Proportion Non-minority (deviation)	0.37†	-0.04	-0.39
Full -Day Class	0.15†	0.05	-0.10
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.06**	0.08**	0.02
Family Income Level (deviation)	0.05	0.07†	0.02
Welfare Status	-0.03	-0.05	-0.02
Age of Child in Months	0.07***	0.07***	0.01
Sex of Child	0.24***	0.28***	0.04
Disability Status	-0.18*	-0.37***	-0.20†
African American Child	-0.05	-0.32**	-0.27†
Hispanic Child	-0.05	-0.14	-0.10
Language Minority Family	0.04	0.06	0.02
Mother-Father Family	-0.08	0.03	0.11
Neither Birth Parent In Home	0.26†	0.01	-0.23
Parent Literacy Standard Score (KFAST)	0.01**	0.00†	0.00
Books In Home	0.08	0.07	-0.01

See note at end of table.

Table 6-9. Three-level regression models of One-to-One Counting scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	One-One Counting score upon Head Start entry	One-One Counting score upon Head Start graduation	Gains in One-One Counting score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not at all	-0.04	-0.06	-0.03
One or twice	-0.09	-0.11	-0.02
Every day	0.05	0.04	0.00
One-year Head Start graduate	0.08	-0.84***	-0.92***
Intercept	-3.31***	-0.81	2.48**
Proportion of Variance Accounted For			
Between-Programs Variance	20	100	--
Between-Classes Variance	39	47	31
Within-Classes Variance	0	0	--
Total Variance	3	3	2

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** $p < .001$

** $p < .01$

* $p < .05$

† $p < .10$

N = 1,671

Table 6-10. Three-level regression models of Social Awareness scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation

Program-level predictor variables	Social Awareness score upon Head Start entry	Social Awareness score upon Head Start graduation	Gains in Social Awareness score from entry to graduation
	Unstandardized regression coefficients		
Program Mean Parent Education Level	0.20	0.20	0.01
Program Mean Family Income Level	0.43*	-0.23	-0.67**
High/Scope Curriculum	-0.09	0.25†	0.35*
Creative Curriculum	0.01	0.20†	0.19
Mean Teacher Salary Level	-0.07	0.11	0.19*
Proportion Language-Minority Children	-0.73*	-0.54†	0.26
Proportion Non-Minority Children	-0.56*	-0.24	0.37
Class-Level Predictor Variables			
Average ECERS Language Subscale Score	-0.04	-0.02	0.01
Average Child-Adult Ratio	0.02	0.01	-0.01
AP Individualizing Score	0.00	0.01	0.01
Teacher BA or AA or higher	-0.12	0.03	0.15
Years Teaching Experience	-0.01†	-0.01	0.00
African American Teacher	0.04	-0.15*	-0.19
Hispanic Teacher	-0.15	-0.04	0.08
Teacher Salary Deviation Score	-0.01	0.15	0.16†
Teacher DAP Beliefs Score	-0.03	-0.04	-0.02
Average Lead Teacher Arnett Score	0.01	0.00	0.00
Class Parent Education Level (deviation)	0.08	0.17*	0.09
Class Family Income Level (deviation)	0.09	-0.06	-0.16
Proportion Language-Minority (deviation)	-0.02	0.04	0.09
Class Proportion Non-minority (deviation)	-0.28	-0.31	-0.03
Full -Day Class	-0.20*	-0.03	0.18†
Child-Level Predictor Variables			
Parent Education Level (deviation)	0.06†	0.08**	0.02
Family Income Level (deviation)	0.03	0.04	0.00
Welfare Status	-0.16†	-0.10	0.05
Age of Child in Months	0.07***	0.06***	-0.01†
Sex of Child	0.26***	0.22***	-0.03
Disability Status	-0.61***	-0.18†	0.44***
African American Child	0.29*	0.44***	0.19
Hispanic Child	0.21	-0.05	-0.29†
Language-Minority Family	-0.18	-0.48***	-0.30†
Mother-Father Family	-0.25**	-0.09	0.17†
Neither Birth Parent In Home	0.02	0.31†	0.28
Parent Literacy Standard Score (KFAST)	0.01***	0.01***	0.00
Books In Home	0.16†	0.04	-0.11

See note at end of table.

Table 6-10. Three-level regression models of Social Awareness scores of Head Start children upon Head Start entry, Head Start graduation, and scale score gains from entry to graduation (continued)

Program-level predictor variables	Social Awareness score upon Head Start entry	Social Awareness score upon Head Start graduation	Gains in Social Awareness score from entry to graduation
	Unstandardized regression coefficients		
Frequency of Reading to Child			
Not At All	0.24	-0.33*	-0.57**
One or twice	0.02	-0.09	-0.11
Every day	0.10	0.03	-0.07
One-year Head Start graduate	0.05	-0.97***	-1.02***
Intercept	-2.10**	0.63	2.73**
Proportion of Variance Accounted For			
Between-Programs Variance	100	56	100
Between-Classes Variance	29	51	41
Within-Classes Variance	0	0	0
Total Variance	3	3	3

Note: Estimates of proportion of variance accounted for represent the proportion of explainable variance, based on the child-level model that is explained by the combined model.

*** $p < .001$

** $p < .01$

* $p < .05$

† $p < .10$

N = 1,707

Of these factors, only number of years of Head Start attendance was related to increased gains or Head Start graduation scores in all cognitive indices examined. But most were related to gains and/or Head Start graduation scores on two or more outcome measures, and the relationships were in the hypothesized direction. The following sections describe the nature and size of these relationships.

6.4.1 Use of an Integrated Curriculum

The High/Scope curriculum is a comprehensive, integrated preschool curriculum that has a long history of research and development. In the present analyses, children in programs that employed the High/Scope curriculum were found to have made significantly greater gains while at Head Start than children in programs that employed other curricula.

Children in programs that employed the High/Scope curriculum made small but significantly greater gains in letter recognition skills than children in programs that employed other curricula (see Table 6-3). In W ability-score terms, the average Head Start child made a gain on the Woodcock-Johnson-Revised Letter-Word Identification (WJ-R LWI) task of 11.4 W ability scale points from entry to graduation from Head Start. Children in programs employing the High/Scope curriculum showed an average gain of 14.1 scale points on the WJ-R LWI from entry to graduation ($p < .001$), whereas children in programs employing other curricula made gains of about 9.4 scale points ($p < .001$). The regression coefficient for the High/Scope curriculum in the three-level regression analysis, which is an estimate of the difference in average gains between it and other curricula adjusted for the influence of related variables, was 4.6 ($p < .05$).

Children in programs using the High/Scope curriculum were found to make significantly greater gains on a criterion-referenced measure of oral communication skills as well (see Table 6-10). The "Social Awareness" measure assessed children's ability to tell an adult basic information about themselves such as their age, and month and year of birth. The average Head Start child made a gain on the Social Awareness measure of 0.94 points. Children in programs employing the High/Scope curriculum showed an average gain of 1.14 points, whereas children in programs employing other curricula made gains of about 0.79 points. The regression coefficient for the High/Scope curriculum showed a gain of 0.35 points greater than that for other curricula ($p < .05$).

6.4.2 Higher Teacher Salary Levels

The three-level regression analyses also explored whether children attending Head Start programs with higher average teacher salary levels would make greater progress in their cognitive and social-emotional development. The multivariate analyses showed this to be the case with respect to children's pre-reading and oral communication skills.

Average Annual Salary for Lead Teachers was associated with higher graduation scores in the letter recognition task (see Table 6-3). The regression coefficient for Mean Teacher Salary Level in the three-level regression analysis of scores at Head Start graduation, which is an estimate of the difference in LWI scale scores associated with each $10,000 increment in average teacher salaries, adjusted for the influence of related variables, was 3.88 ($p < .01$).

Children in programs with higher average teacher salaries also had higher graduation scores in the dictation task than those in programs with lower average teacher salaries (see Table 6-5). The regression coefficient in the three-level regression model was 6.07 ($p < .05$).

Children in programs with higher average teacher salaries made greater gains as well on the criterion-referenced "Social Awareness" measure (see Table 6-10). The regression coefficient for the Mean Teacher Salary Level showed an increased gain of 0.19 points for every $10,000 increment in mean salary ($p < .05$).

6.4.3 Teachers with Bachelor's or Associate's Degrees or Higher

The possession of a four-year college degree or an Associate's degree in education or a closely-related field is among the most widely accepted indicators of teacher preparation. One of the current performance goals of the national Head Start program is to have all local programs staffed by teachers of whom a majority have Bachelor's degrees or Associate's degrees. These analyses explored whether the lead teacher having at least a BA or AA degree made a difference in children's progress on the FACES cognitive measures.

Children in Head Start classes taught by lead teachers with at least a Bachelor's degree or Associate's degree showed greater graduation scores in the dictation task than children in programs with teachers who did not have such a degree (see Table 6-5). The multilevel regression coefficient showed that children whose teachers had at least a Bachelor's or Associate's degree had average scores 6.94 W ability scale points higher than those whose teachers did not have a degree ($p < .05$).

6.4.4 Full-Day Versus Part-Day Classes

As of the 2000-2001 school year, the majority of children who attended Head Start participated in part-day classes that were conducted in morning or afternoon sessions only. These analyses explored whether children benefited more from the program in terms of academic achievement if they attended full-day classes. Children in FACES 2000 who did attend full-day Head Start programs made greater gains in several areas than children who attended part-day.

In the three-level analysis of Dictation scores at graduation from Head Start (see Table 6-5), children in full-day classes had a significant regression coefficient of 8.36 W ability scale points ($p < .01$). This meant that children in full-day classes had a mean score upon graduation from Head Start that was much higher than the mean for children in part-day classes, with related factors controlled. However, in the multilevel analysis of gains on the Dictation task, the regression coefficient for full-day classes (2.55) was not significant.

In the three-level analysis of book knowledge scores at graduation from Head Start (see Table 6-6), children in full-day classes had a significant regression coefficient of 0.25 points ($p < .05$), indicating that children in full-day classes had a mean book knowledge score upon graduation from Head Start that much higher than the mean for children in part-day classes, with related factors controlled. Further, in the multilevel analysis of gains on the book knowledge task, the regression coefficient for full-day classes (0.24) was also significant ($p < .05$). Children in full-day classes also made greater gains from Head Start entry to graduation in their knowledge of the parts of a book.

Children in full-day Head Start classes also made larger gains in color naming skills than children in part-day classes (see Table 6-7). The regression coefficient for full-day classes in the three-level regression analysis was 0.90 ($p < .05$).

6.4.5 More Frequent Parental Reading to Children

Analyses with FACES 2000 data suggest that Head Start programs would contribute to the school readiness of its children by encouraging parents to engage in more frequent and more effective educational activities at home with their children. Children are in preschool programs for only a limited time, both in terms of hours of each day and months out of the child's life. However, preschool programs may extend their influence by encouraging parents to participate in educational activities such as reading at home with their children. The national Head Start program recognized the importance of this function by stating, in its performance measures framework, that one of the major objectives of the program is, to "strengthen parents as the primary nurturers of their children."[5]

[5] See *Head Start FACES: Longitudinal findings on program performance*. Third progress report (ACYF, 2001) for a list of the 24 Program Performance Measures.

The three-level regression analyses with FACES 2000 data showed that more frequent parental reading in the Fall was associated not only with higher initial achievement for children as they entered the program, but also with larger gains during the program year. Larger gains were observed for children whose parents read to them every day. Smaller gains were observed for children who were read to only once or twice a week or not at all.

Parents were asked whether they read to their children, “not at all,” “once or twice,” “three to six times,” or “every day” during the previous week. Parental responses to the question were entered into the three-level regression analysis as a set of dichotomous variables, with the most frequent response, “three to six times,” as the omitted reference category. The reading responses were entered as child-level independent variables.

In the regression analysis of graduation vocabulary test scores, children whose parents reported reading to their children only “once or twice” had significantly lower mean scores than children whose parents reported reading “three to six” times (see Table 6-2). The mean score for children whose parents said they read “every day” was significantly larger than that of the “three to six times” group. In terms of W ability scores on the Peabody Picture Vocabulary Test, Third Edition, the mean for the “once or twice” group was 1.55 points lower ($p < .01$); in contrast, the mean for the “every day” group was 1.28 points higher ($p < .05$), than the mean for the “three to six times” reference group.

In the regression analysis of the graduation letter-word identification scores (see Table 6-3), the “once or twice” groups again had mean scores that were significantly lower (by 3.63 scale points, $p < .05$) than that for the “three to six times” reference group. In the regression analysis of letter-word identification gains from entry to graduation, the “once or twice” group had a significantly smaller gain than the reference group (by 2.78 scale points, $p < .05$).

For the graduation book knowledge scores (see Table 6-6), the “every day” groups had mean scores that were significantly higher than that for the “three to six times” reference group. Children whose parents reported reading to them “every day” in the past week had scores .20 scale points higher than those whose parents read to them “three to six times.” ($p < .01$) In the regression analysis of book knowledge gains from entry to graduation, the children whose parents reported reading to them “every day” had significantly larger gains than the reference group (by .20 scale points, $p < .05$).

For the graduation social awareness scores (see Table 6-10), children whose parents reported not reading to their children at all had mean scores that were significantly lower than that for the "three to six times" reference group ($b = -.33$; $p < .05$). In the regression analysis of social awareness gains from entry to graduation, the children whose parents reported not reading to them at all in the past week had gain scores that were .57 scale points lower than those whose parents read to them "three to six times" ($p < .01$).

In the regression analysis of the graduation dictation scores (see Table 6-5), the "once or twice" group again had mean scores that were significantly lower (by 8.63 scale points, $p < .05$) than that for the "three to six times" reference group. However, parents' reported frequency of reading to their children in the past week were not significantly associated with the entry-to-graduation gain scores for this cognitive outcome.

These differences related to frequency of parental reading were obtained even after controlling for parent education level, the mother's score on a measure of adult literacy (the K-FAST), and an indicator of the presence of books in the home. While these analyses do not address whether Head Start programs have an impact on parental reading to children in the home, they do support the notion that programs that are able to effectively encourage parents to engage in such activities show greater cognitive gains among their children.

6.4.6 Attending Head Start for More than One Year

Analyses of the FACES 2000 data also suggest that children who benefit from 2 years of Head Start show greater cognitive gains across their years in Head Start compared with their peers who attend for 1 year. The three-level regression analyses with FACES 2000 data showed that, for every measure, graduating from Head Start after 2 years as opposed to only 1 year was associated not only with higher achievement for children when they graduated from the program, but also with larger gains during their program years.

In the regression analysis of graduation vocabulary test scores (see Table 6-2), children who graduated from Head Start after 1 year had a significant regression coefficient of -10.73 W ability scale points ($p < .001$). This meant that children who graduated after 2 years had a mean score at graduation from Head Start that much higher than the mean for children who graduated after 1 year, with related factors controlled. Further, in the multilevel analysis of gains in vocabulary test scores from entry to graduation, the regression coefficient for one-year graduates (-10.11) was also significant (*p* < *.001*). Children who graduated after 2 years in Head Start also made greater gains from Head Start entry to graduation in their vocabulary size than those who graduated after 1 year.

In the regression analysis of the graduation letter-word identification scores (see Table 6-3), one-year graduates had a significant regression coefficient of -14.80 W ability scale points (*p* < *.001*). This meant that two-year graduates had a mean score at graduation from Head Start that much higher than that of one-year graduates. In the regression analysis of letter-word identification gains from entry to graduation, the one-year graduates had a significant regression coefficient of -14.05 (*p* < *.001*), indicating that two-year graduates showed greater improvement in their ability to identify letters and words.

For the graduation early math scores (see Table 6-4), the regression coefficient for one-year graduates was a significant -16.60 (*p* < *.001*), indicating that two-year graduates had a mean score at graduation from Head Start that much higher than their peers who graduated after 1 year. In the analysis of the early math gain scores, the regression coefficient was also significant (*b* = -16.98, *p* < *.001*). This meant that entry-to-graduation gains in early math W ability scores for two-year graduates were that much higher than those for one-year graduates.

For the graduation dictation scores (see Table 6-5), children who attended 1 year of Head Start had a significant regression coefficient of -24.87 (*p* < .001), indicating that the mean scores for two-year graduates was that much greater than that of one-year graduates. Likewise, analysis of the mean dictation gain scores shows that the one-year graduates had a significant regression coefficient (*b* = -23.27, *p* < *.001*). This meant that the entry-to-graduation mean gain in dictation scores was that much greater for the two-year graduates compared to the one-year graduates.

Similar patterns of results were found in each of the three-level regression analyses with the criterion-referenced variables (see Tables 6-6 through 6-10). The regression coefficients for the one-year Head Start graduates were significant and negative for the mean graduation scores in book knowledge, color naming, design copying, one-to-one counting, and social awareness. This meant that the mean graduation scores for the two-year graduates were significantly greater than those of the one-year graduates in each of these measures. Further, the regression coefficients for the one-year Head Start graduates were significant and negative for the mean entry-to-graduation gain scores in these measures.

This meant that the average gain scores for the two-year graduates were also significantly greater than those of the one-year graduates in each of these measures.

6.5 Child:Staff Ratio, More Individual Attention, and Teacher's Positive Attitudes and Knowledge About Early Childhood Educational Practices

The program and class characteristics described earlier in this chapter showed significant relationships with children's gains in Head Start, but there were some sets of characteristics that did not show the relationships with children's gains that were hypothesized. These were the indicators of child:staff ratio and the indicators of individualized attention to the needs of each child, and the teacher's positive attitudes and knowledge about early childhood educational practices.

6.6 Child:Staff Ratio and More Individual Attention

It was hypothesized that lower child:staff ratios in Head Start classroom activities, and more attention to the needs and preferences of individual children, would result in greater gains for children. These hypotheses were not confirmed by the data. Indeed, on some outcome measures, children actually showed higher scores at graduation from classrooms with more children per staff member.

Child:Staff Ratios. The mean child:staff ratio was a figure derived from counting the number of children in the Head Start classroom and dividing that number by the number of teachers or other adult staff members actively interacting with children. These counts were taken at two separate occasions on the day that the classroom was observed, and the two resulting ratios were averaged. When the mean child:staff ratio was entered into the three-level regression analyses of children's cognitive gains as a class-level variable, it was not significantly associated with gains in any of the cognitive outcomes measures.

In the analyses of children's scores at graduation, results were opposite to what was predicted. Children in classes with higher child:staff ratios had significantly *higher* scores at graduation in letter recognition skills, early math, early writing, and color naming than those in classes with lower ratios.

In the three-level regression analyses of letter-word identification scores at Head Start graduation (see Table 6-3), the value of the regression coefficient for the mean child:staff ratio was 1.15 ($p < .001$). This may be interpreted as the change in LWI W ability scores at Head Start graduation that would be expected for every unit increase in the child:staff ratio, net of the effects of other related variables. If the original hypothesis had been confirmed, this coefficient would be negative. Instead, it was significantly greater than zero in the positive direction. Significant positive regression coefficients were also found in the regressions predicting graduation scores for the applied problems task (b = 1.40, $p < .01$); (see Table 6-4) the dictation task (b = 1.92, $p < .01$; (see Table 6-5)), and the color naming task (b = .16, $p < .05$; (see Table 6-7)).

Assessment Profile Individualizing Scale. The Individualizing scale of the Assessment Profile instrument uses both observational and interview methods to assess the degree to which preschool teachers track the accomplishments of children in their classes and provide activities suited to the capabilities and interests of individual pupils. Class-level scores on this instrument did not relate to gains in any of the cognitive development areas. Nor did they show associations with graduation scores.

Caregiver Interaction Scale (CIS). The sensitivity and emotional tone of teacher-child interaction was assessed with the Caregiver Interaction Scale (CIS), an observation-based rating scale. CIS scores for the classroom teacher were not associated with gains for any of the cognitive development areas. Nor did they show associations with graduation scores.

ECERS-R Language Subscale. The other indicator of classroom quality, the ECERS-R Language subscale, was not associated with higher graduation scores or greater entry-to-graduation gain scores on any of the cognitive outcome measures. On the contrary, scores on the ECERS-R Language subscale had counterintuitive associations with early writing skills at graduation from Head Start. In the three-level regression analyses predicting Dictation scores at graduation (see Table 6-5), the regression coefficient for the ECERS-R Language subscale signified a Dictation score that was 4.32 W ability scale points *lower* for each unit increase in the classroom's ECERS-R Language subscale score ($p < .01$). The ECERS-R Language subscale score was not associated with any other cognitive outcomes.

6.7 Teacher's Positive Attitudes and Knowledge About Early Childhood Educational Practices

Teachers' knowledge about early childhood educational (ECE) practices was assessed by asking teachers their level of endorsement of various statements representing beliefs about how children in Head Start should be taught or managed. Examples of these statements are "Children should work silently and alone on seatwork" or "Head Start activities should be responsive to individual differences in development." Items were coded with respect to their conformity with developmentally appropriate ECE practices. High scores on this scale indicate strong knowledge of ECE practices. It was expected that children taught by teachers with stronger knowledge of ECE practices would display higher graduation scores and greater gains in the cognitive outcomes.

However, in the three-level analyses of graduation scores for letter recognition (see Table 6-3), children taught by teachers with stronger knowledge of ECE practices showed *lower* scores at graduation from Head Start. The value of the regression coefficient for the mean ECE practices score was -.88 ($p < .05$). This may be interpreted as the change in LWI W ability gain scores that would be expected for every unit increase in the teachers score on the ECE practices measure, net of the effects of other related variables. If the original hypothesis had been confirmed, this coefficient would be positive. Instead, it was significantly less than zero in the negative direction. Counterintuitive findings were also found for graduation scores in the design copying task (see Table 6-8). The value of the regression coefficient for the mean ECE practices score was -.07 ($p < .05$).

In the analyses of children's gain scores from Head Start entry to graduation, results were also opposite to what was predicted. Children taught by teachers with higher scores on the ECE practices measure had significantly *smaller* gain scores in book knowledge (b = -.08, $p < .05$; (see Table 6-6), and design copying (b = -.09, $p < .01$; (see Table 6-8) than those taught by teachers with lower scores.

6.8 Summary and Discussion

This chapter explored variations in child achievement across local Head Start programs and classes. It used multilevel modeling to test hypotheses about early education program, class, and family characteristics that many child development scholars believe to be associated with enhanced cognitive growth in preschool children. The conceptual framework posited that the gains a child makes in Head

Start depends on the nature of the learning environment that he or she experiences in the local program and the learning activities in the home. The nature of the learning environment depends in turn on the training and experience of teachers in the program and the resources available to them in terms of facilities, materials, and teaching assistants. Programs with more resources are likely to be better able to provide adequate facilities and materials and recruit and retain well-prepared teachers. Another hypothesis was that children would make larger gains in programs employing curricula that are comprehensive and integrated in terms of educational activities and assessment methods. Other expectations were that children would make more sizable gains in higher quality classrooms, in full-day as opposed to part-day classes, in classes with better child:staff ratios and more individualized attention to pupils, in families where parents engaged in more educational activities with their children, and had the benefit of Head Start for more than one year.

The addition of the program- and classroom-level variables provided unique and significant contributions to the prediction of children's assessment scores at Head Start entry and graduation, as well as gains from entry to graduation. Analysis of longitudinal data from FACES 2000 showed that children's gains in Head Start were significantly related to several of the hypothesized characteristics of programs and classes. Specifically, use of the High/Scope curriculum, higher teacher salaries, teachers' educational credentials (Bachelor's degrees or Associate's degrees or higher), and provision of preschool services for a longer period each day all serve as a pathway to greater cognitive gains.

Analyses of child-level data show that Head Start children whose parents read to them more often show greater gains and graduation scores on some measures. These analyses suggest that Head Start programs who implement efforts to encourage parents to engage in more educational activities with their children at home may contribute to the school readiness of children.

Analyses of FACES 2000 data also suggest that children's readiness for elementary school is enhanced by an extra year of Head Start. For each measure, children who attended Head Start for 2 years showed greater graduation scores and greater entry-to-graduation gains compared to their peers who attended Head Start for only 1 year.

However, other analytic results were not in line with expectations. In particular, within the narrow range of child:staff ratios in Head Start, higher child:staff ratios were associated with greater entry-graduation achievement gains in some of the cognitive outcomes. Further, teachers' scores on a scale assessing their knowledge of ECE practices were negatively linked with children's cognitive outcomes.

Some provisos about the results should be noted. Differences in cognitive gains, while statistically significant, were relatively modest in magnitude. By itself, each of the differences was not large enough to close the gap between where Head Start children typically end up at the end of the program year and the average achievement levels of American children at the start of elementary school. However, if several of the positive characteristics could be implemented simultaneously in a program, they might jointly make a more sizable difference.

The failure to find significant links between children's cognitive gains in Head Start and class-level scores on the ECERS-R Language subscale may have to do with the generally good quality of Head Start classrooms and the limited range of variation in classroom quality that FACES found in its national samples of programs and classes (Peisner-Feinberg, Burchinal, Clifford, Culkin, Howes, Kagan and Yazejian, 2001). Studies encompassing broader ranges of quality of childcare and early education facilities have shown greater variations in classroom quality measures and significant relationships between quality measures and children's gains (Peisner-Feinberg and Burchinal, 1997; Bryant, Burchinal, Lau, and Sparling, 1994; NICHD Early Child Care Research Network, 2000; Phillips, McCartney, and Scarr, 1987; Whitebook, Howes and Phillips, 1989; but for another failure to find a relationship, see Kontos and Fiene, 1987).

At the same time, the FACES results should make us wary of claims that Head Start could produce dramatically larger achievement gains in children from low-income families simply by raising ECERS scores or other indicators of classroom quality. It may be that good classroom quality is a necessary but not sufficient condition for practically significant gains in specific cognitive or behavioral areas. It may be that further progress depends on discovering and applying instructional approaches that can bolster gains in specific areas.

CHAPTER 7.
RELATIONSHIP OF PROGRAM, CLASSROOM, AND CHILD AND FAMILY CHARACTERISTICS TO CHILDREN'S SOCIAL BEHAVIOR IN HEAD START

This chapter explores variations in children's social behavior at the beginning and at the end of Head Start participation as a function of different program, classroom, and child- and family-level predictors. It uses hierarchical linear models to test hypotheses about the links between various predictors in each of these levels and social behavior in Head Start as reported by Head Start teachers and parents. Program characteristics that may influence social behaviors include the kind of curriculum used by the program, the mean teacher salary level, and different types of family characteristics as a function of the program mean. Head Start classroom and center characteristics that may influence children's social behavior include the child:adult ratio in the classrooms, the teacher's behavior in the classroom, and the teacher's educational level. Child and family characteristics that may influence social behaviors include the child's gender, race, disability status, the parent's educational level, and the family income level. These and other child and family, classroom, and program characteristics are generally believed to be associated with enhanced social skills in preschool children (Phillips, Mekos, Scarr, McCartney, and Abbott-Shim, 2000; Whitebook, Howes, and Phillips, 1989).

7.1 Conceptual Framework and Research Questions

The conceptual framework that guided the analyses presented in this chapter was a multi-level, multi-causal model of the influences that shape children's social-emotional development and the factors that help determine the nature of the experience children have in Head Start. This view posits that children's social development in the early years is primarily a function of the experiences they have in their families. In some cases, children from low-income families may not receive as much emotional support from parents as they need for optimal social development. A center-based early childhood learning environment such as Head Start may help provide experiences that would be beneficial for the socio-emotional development of all children, but especially for those from higher-risk family environments (NICHD Early Child Care Research Network, 2000).

The nature of the environment that a given child experiences in Head Start depends on the training and experience of teachers in the program and the resources available to them in terms of facilities, materials, and teaching assistants. Programs with more resources are likely to be better able to provide adequate facilities and materials and recruit and retain talented and well-prepared teachers (Whitebook et al., 1989). According to these assumptions, the following research questions are examined in this chapter:

1. Are Head Start program characteristics predictive of children's social behaviors?
2. Are Head Start classroom characteristics predictive of children's social behaviors?
3. Are child and family characteristics predictive of children's social behaviors? and
4. Do Head Start program and classroom characteristics explain children's social development over and above child and family characteristics?

7.2 Analysis Method

The analytic model used here examines all four research questions simultaneously. It is a multilevel linear regression modeling, using the SAS PROC MIXED computer program (Bryk and Raudenbush, 1992; Singer, 1998) which was designed for the simultaneous examination of the effects of different sets of characteristics which are not independent (i.e., one set of characteristics is nested within the other) on children's social behavior outcomes. It provides an estimate of the amount of variance accounted for by each set of variables separately, as well as an estimate of the amount of additional explanatory power of one or more sets of characteristics (e.g., program and classroom characteristics) over and above another set of characteristics (e.g., child and family characteristics).

The primary dependent variable sets were:

- Head Start entry year (Fall 2000) social behavior scores;
- Head Start graduation year (Spring 2001/2002) social behavior scores; and
- The gains each child made between the Fall of their entry year and the Spring of their final year in Head Start in their behavior ratings scale scores (for the social skills variables, gains refer to an increase in scale scores; for the problem behavior variables—(i.e., aggressive, hyperactive, and withdrawn behaviors)—gains refer to a decrease in scale scores).

Each analytic model had three levels. The first level involved variation in average assessment scores or average gain scores across the programs in the FACES national sample, expressed as deviations of the programs means from the overall mean score for the entire sample. The second level involved variation in average assessment scores or average gain scores across the classrooms in the FACES national sample, expressed as deviations of the classrooms means from the overall mean score for the entire sample. The third level involved variation in individual children's scores or gain scores. The sets of independent and dependent variables used in the analyses are presented in Table 7-1.

7.3 Findings

Twenty-four separate multilevel linear regression modeling analyses were conducted. Sixteen analyses were used to predict social behavior at entry and graduation. Eight analyses were used to predict gains from entry to graduation. In each analysis, the three sets of independent variables were used to predict 1 of the 24 social behavior outcomes or gains from entry to graduation (i.e., eight parent and teacher social behaviors ratings multiplied by three types of scores: entry, graduation, and gain from entry to graduation). The results are presented in four different sections corresponding to the four research questions.

7.3.1 Are Head Start Program Characteristics Predictive of Children's Social Behaviors?

There were significant relationships between some of the program characteristics and variations in children's social behaviors. All of these links were in the expected directions (i.e., better scores in the program-level characteristics predicted better social behaviors). The data related to the specific links between program characteristics and social behavior outcomes or gains in social behavior are detailed in the upper portions of Tables 7-2 (each table represents the links with 1 of the 8 social behavior outcomes). As can be seen in these tables, the program factors that seemed to influence children's social behaviors were:

- The salary level of teachers in the programs, which was related to 6 of the 8 social behavior variables (All the teacher-reported scales, and parent-reported social skills and withdrawn behavior);

Table 7-1. Independent and dependent variables used in the analyses

	Independent variables list		
Dependent variables	Program level	Classroom level	Child/family level
Teacher report:	1. Curriculum (High/Scope curriculum or Creative Curriculum)	1. Full-day class	1. Parent literacy
1 Social skills	2. Average teacher salary in the program	2. Proportion of non-minority children (deviation from program)	2. Parental education level
2. Aggressive behavior	3. Proportion of non-minority children	3. Average child:adult ratio in classroom	3. Family income level
3. Hyperactive behavior	4. Program mean parent education level	4. Average ECERS-R Language subscale	4. Welfare status
4. Withdrawn behavior	5. Program mean family income level	5. Average Arnett score	5. Books at home? (yes or no)
Parent report:	6. Proportion of language-minority children	6. Assessment Profile score	6. Frequency of reading to child: 1. Not at all 2. Once or twice 3. Every day
1. Social skills		7. Teacher education	7. Child's age
2. Aggressive behavior		8. Teacher experience	8. Child's gender
3. Hyperactive behavior		9. Teacher DAP beliefs score	9. Child's African American racial origin
4. Withdrawn behavior		10. African American teacher	10. Child's Hispanic racial origin
		11. Hispanic teacher	11. Child from a language-minority home
		12. Teacher salary deviation score	12. Disability status
		13. Parent education level (deviation)	13. Two-parent (bio or adoptive) household
		14. Family income level (deviation)	14. No biological parent at home
		15. Proportion language-minority (deviation)	15. Years in Head Start

Note: For a detailed description of these variables, see Technical Appendix.

- The use of the High/Scope curriculum, which was also related to 6 of the 8 social behavior variables (all the teacher-reported scales and parent-reported aggressive and hyperactive behavior); and

- The mean parent education level in the program, which was related to 4 of the 8 social behavior variables (All of the teacher-reported scales).

The social behavior variable that seemed to be most affected by program characteristics was teacher-reported social skills. It was linked to 4 of the 7 program characteristics: the mean family income in the program, the use of High/Scope curriculum, the use of Creative Curriculum, and the mean teacher salary in the program (see Tables 7-2). The program-level models explained only 0.4 to 5.2 percent of the variance in social behavior variables. In fact, the explanatory power of the program-level models was rarely higher than 2 percent (see Table 7-3). A notable exception was found in the case of parent-reported hyperactive behavior, for which the program-level model explained 5.2 percent of the variance at entry and 5 percent of variance at graduation. Links between program-level characteristics and social behaviors were most evident in the graduation period of data collection (21 of the 34 significant links in the program level were in this period).

7.3.2 Are Head Start Classroom Characteristics Predictive of Children's Social Behaviors?

Data related to specific links between classroom characteristics and children's social behavior are detailed in the middle portions of Tables 7-2. In general, only a few significant relationships emerged between the classroom characteristics and children's social behavior outcomes, but all were in the expected direction. The only classroom-level characteristic related to more than two social behavior outcomes was the teachers' education level (the proportion of teachers with an AA degree or higher). This variable was significantly related to four social behavior variables: teacher-reported aggressive, hyperactive, and withdrawn behaviors, and parent-reported withdrawn behavior. The classroom level models explained only 0.8 to 3.4 percent of the variance in social behavior variables. Again, the explanatory power of the classroom-level models was rarely higher than 2 percent (see Table 7-3). Links between classroom-level characteristics and social behaviors were most evident at graduation from Head Start (8 of the 17 significant links in the classroom level were in this period).

Table 7-2. Three-level regression models of Parent-reported Withdrawn Behavior scores in Head Start entry, graduation, and entry-graduation gains

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Program mean parent education level	-0.01	0.01	0.02
Program mean family income level	0.10	0.00	-0.11
High/Scope curriculum	-0.06	-0.10	-0.02
Creative Curriculum	-0.08	-0.10	0.00
Mean teacher salary level	-0.12*	0.00	0.12*
Proportion language-minority children	-0.13	0.10	0.25
Proportion non-minority children	0.19	0.11	-0.06
Class-level predictor variables			
Average ECERS-R language score	0.00	0.05†	0.05
Average child:adult ratio	0.00	-0.01	-0.01
AP individualizing score	0.03	0.00	-0.02
Teacher BA or AA or higher	-0.07	0.11†	0.16*
Years teaching experience	0.01*	0.00	-0.01**
African American teacher	-0.04	-0.05	0.02
Hispanic teacher	-0.04	0.09	0.13
Teacher salary deviation score	0.04	0.04	0.00
Teacher DAP beliefs score	-0.01	-0.02	-0.01
Average lead teacher Arnett score	0.00	0.00	0.00
Class parent education level (deviation)	0.15**	-0.01	-0.16*
Class family income level (deviation)	0.03	0.09	0.05
Class proportion non-minority (deviation)	0.14	-0.09	-0.22
Proportion language-minority (deviation)	0.11	0.06	-0.04
Full-day class	-0.08	0.07	0.14†
Child-level predictor variables			
Parent education level (deviation)	-0.01	-0.02	-0.01
Family income level (deviation)	0.02	0.01	-0.01
Welfare status	0.17**	0.12*	-0.05
Age of child in months	0.01*	0.01	0.00
Sex of child	0.06	0.01	-0.05
Disability status	0.19*	0.19*	0.00
African American child	-0.01	-0.05	-0.05
Hispanic child	0.21*	-0.03	-0.23*
Language-minority family	0.09	0.13	0.04
Mother-father family	-0.16**	-0.13*	0.03
Neither birth parent in home	0.23*	0.28*	0.04
Parent literacy standard score (KFAST)	0.00*	0.00	0.01*
Books in home	-0.09	-0.21**	-0.12†
Frequency of reading to child			
Not at all	0.11	0.25*	0.14
Once or twice	-0.12*	-0.09	0.04
Every day	-0.08	-0.05	0.03
One year in Head Start	0.14*	-0.04	-0.19*

Table 7-2. Three-level regression models of Parent-reported Withdrawn Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coefffICents (β) Entry	Graduation	Gains
Intercept	0.48	0.27	-0.22
Proportion of variance accounted for	64%	100%	100%
Between-programs variance	66%	65%	43%
Between-classes variance	4%	3%	2%
Within-classes variance			
Total variance	8%	5%	5%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

$N = 1{,}684$.

Table 7-3. Three-level regression models of Teacher-reported Social Skills scores in Head Start entry, graduation, and entry-graduation gains

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Program mean parent education level	0.33	-0.32	-0.68
Program mean family income level	-1.33	-2.80***	-1.38
High/Scope curriculum	-0.10	2.14***	2.18**
Creative Curriculum	-0.59	0.85*	1.40*
Mean teacher salary level	-0.49	0.84***	1.33***
Proportion language-minority children	0.87	-0.20	-0.99
Proportion non-minority children	-0.57	0.05	0.62
Class-level predictor variables			
Average ECERS-R language subscale score	-0.02	0.03	0.01
Average child:adult ratio	0.05	0.07	0.02
AP individualizing score	0.03	-0.06	-0.08
Teacher BA or AA or higher	-0.63	-0.58†	0.08
Years teaching experience	-0.04	-0.01	0.02
African American teacher	-0.85	-0.73	0.08
Hispanic teacher	-0.19	-0.38	-0.22
Teacher salary deviation score	0.36	0.22	-0.15
Teacher DAP beliefs score	-0.04	0.04	0.09
Average lead teacher Arnett score	0.03	0.00	-0.02
Class parent education level (deviation)	-0.43	-0.11	0.33
Class family income level (deviation)	0.39	-0.03	-0.36
Class proportion non-minority (deviation)	0.54	-0.44	-0.94
Proportion language-minority (deviation)	1.12	1.46†	0.31
Full-day class	-0.42	0.14	0.57

See note at end of table.

Table 7-3. Three-level regression models of Teacher-reported Social Skills scores in Head Start entry, graduation, and entry-graduation gains (continued)

	Unstandardized regression coeffficents (β)		
Program-level predictor variables	Entry	Graduation	Gains
Child-level predictor variables			
Parent education level (deviation)	0.03	0.07	0.04
Family income level (deviation)	0.21	0.03	-0.18
Welfare status	-0.35	-0.50	-0.13
Age of child in months	0.22***	0.13***	-0.09***
Sex of child	1.49***	1.53***	0.04
Disability status	-1.40***	-1.15***	0.21
African American child	0.23	-0.18	-0.40
Hispanic child	0.09	-0.04	-0.18
Language-minority family	0.19	0.19	0.00
Mother-father family	-0.04	0.51*	0.55*
Neither birth parent in home	-0.70	-0.63	0.04
Parent literacy standard score (KFAST)	0.00	0.01	0.01
Books in home	0.67**	0.38	-0.33
Frequency of reading to child			
Not at all	0.22	0.53	0.34
Once or twice	0.02	-0.16	-0.17
Every day	0.57*	0.36	-0.14
One year in Head Start	0.11	-2.31***	-2.36***
Intercept	3.42	12.78***	9.28**
Proportion of variance accounted for	100%	100%	100%
Between-programs variance	14%	24%	16%
Between-classes variance	15%	11%	10%
Within-classes variance			
Total variance	17%	14%	14%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

N = 1,678.

7.3.3 Are Child and Family Characteristics Predictive of Children's Social Behavior?

Data related to specific links between child and family characteristics and social behavior scores are detailed in the lower portions of Tables 7-2. As expected, the child- and family-level characteristics yielded the most significant links with children's social outcomes. The child characteristics that were most predictive of the child's social development were the child's disability status (8 of 8–

children with disabilities reported to have the less desirable behaviors), the number of years the child spent in Head Start (8 of 8–children with two years of Head Start reported to have the more desirable behaviors), gender (significant links in 7 of the 8 social behavior variables with girls reported to have the more desirable behaviors), and age (7 of 8–older children reported to have the more desirable behaviors). Family characteristics most predictive of child's social development included the presence of both biological/adoptive parents in the household (6 of 8–children with both parents reported to have the more desirable behavior) and the existence of books in the home (5 of 8–children with books at home reported to have the more desirable behaviors). All other child and family characteristics were related to at least one of the social outcomes and most (15 of 17) were related to at least 2 of the 8 social behavior outcomes. The social behaviors most affected from variables in this level were the teacher- and parent-reported withdrawn behavior (each were related to 12 of the 17 child and family characteristics). Even the least affected social behavior variables (teacher-reported social skills and aggressive behavior) had significant links with 7 of the 17 characteristics. The child- and family-level models explained 1.4 to 14.5 percent of the variance in social behavior variables, but most explained more than 5 percent of the variance (see Tables 7-3 through 7-10). Links between child- and family-level characteristics and social behaviors were evident in both entry and graduation periods (45 and 48 significant links, respectively), but were also fairly evident in terms of gains from entry to graduation (20 significant links).

Table 7-4. Three-level regression models of Teacher-reported Aggressive Behavior scores in Head Start entry, graduation, and entry-graduation gains

	Unstandardized regression coeffficents (β)		
Program-level predictor variables	Entry	Graduation	Gains
Program mean parent education level	0.22	-0.12	-0.31
Program mean family income level	0.35	0.76*	0.40
High/Scope curriculum	0.14	-0.49	-0.62**
Creative Curriculum	-0.06	-0.31	-0.25
Mean teacher salary level	-0.06	-0.46***	-0.39**
Proportion language-minority children	0.64	0.75	0.15
Proportion non-minority children	-0.05	-0.08	-0.05
Class-level predictor variables			
Average ECERS-R language subscale score	-0.01	-0.01	-0.01
Average child:adult ratio	0.02	-0.03	-0.05
AP individualizing score	0.07	0.05	-0.02
Teacher BA or AA or higher	-0.01	0.30*	0.31†
Years teaching experience	0.01	0.01	0.00
African American teacher	0.09	-0.05	-0.15
Hispanic teacher	0.03	0.21	0.19
Teacher salary deviation score	0.01	-0.10	-0.10
Teacher DAP beliefs score	-0.03	0.00	0.03
Average lead teacher Arnett score	-0.01	0.00	0.01
Class parent education level (deviation)	0.26†	0.26	0.01
Class family income level (deviation)	0.09	0.17	0.07
Class proportion non-minority (deviation)	1.30*	0.42	-0.87
Proportion language-minority (deviation)	-0.44	-0.31	0.12
Full-day class	0.18	0.12	-0.06
Child-level predictor variables			
Parent education level (deviation)	-0.05	-0.06	-0.01
Family income level (deviation)	-0.08	0.05	0.13*
Welfare status	0.16	0.24*	0.08
Age of child in months	-0.05***	-0.04***	0.01
Sex of child	-0.80***	-0.93***	-0.14
Disability status	0.43**	0.44**	0.00
African American child	0.02	0.30	0.28
Hispanic child	-0.13	0.02	0.11
Language-minority family	0.02	-0.18	-0.21
Mother-father family	-0.30***	-0.41**	-0.11
Neither birth parent in home	0.06	0.17	0.08
Parent literacy standard score (KFAST)	0.01	0.00	-0.01
Books in home	-0.18	-0.04	0.12

Frequency of reading to child			
Not at all	0.07	-0.16	-0.22
Once or twice	0.06	0.22†	0.16
Every day	-0.15	-0.10	0.03
One year in Head Start	0.03	0.53***	0.50***

Table 7-4. Three-level regression models of Teacher-reported Aggressive Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	3.07**	3.59**	0.46
Proportion of variance accounted for	99%	93%	98%
Between-programs variance	19%	2%	5%
Between-classes variance	9%	11%	3%
Within-classes variance			
Total variance	11%	13%	6%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

$N = 1,591$.

Table 7-5. Three-level regression models of Teacher-reported Hyperactive Behavior scores in Head Start entry, graduation, and entry-graduation gains

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Program mean parent education level	0.13	-0.33*	-0.45*
Program mean family income level	0.12	0.68**	0.52†
High/Scope curriculum	0.06	-0.38*	-0.41*
Creative Curriculum	0.12	-0.13	-0.23
Mean teacher salary level	-0.02	-0.24**	-0.21*
Proportion language-minority children	0.14	0.29	0.13
Proportion non-minority children	-0.34	-0.14	0.23
Class-level predictor variables			
Average ECERS-R language subscale score	-0.02	0.02	0.04
Average child:adult ratio	-0.04	-0.07**	-0.03
AP individualizing score	0.07	0.00	-0.07
Teacher BA or AA or higher	-0.04	0.22*	0.26*
Years teaching experience	0.01	0.00	0.00
African American teacher	-0.28†	-0.13	0.15
Hispanic teacher	-0.12	0.07	0.21
Teacher salary deviation score	0.03	0.04	0.00
Teacher DAP beliefs score	-0.01	-0.01	0.01
Average lead teacher Arnett score	0.00	0.00	0.01
Class parent education level (deviation)	0.17	0.19*	0.01
Class family income level (deviation)	-0.06	-0.10	-0.04
Class proportion non-minority (deviation)	0.49	0.28	-0.25
Proportion language-minority (deviation)	-0.31	-0.22	0.12
Full-day class	0.15	0.10	-0.05
Child-level predictor variables			
Parent education level (deviation)	-0.03	-0.03	0.01
Family income level (deviation)	-0.05	0.03	0.09*
Welfare status	-0.01	0.08	0.10
Age of child in months	-0.05***	-0.03***	0.02**
Sex of child	-0.42***	-0.49***	-0.06
Disability status	0.60***	0.61***	0.01
African American child	0.02	0.10	0.08
Hispanic child	-0.02	0.09	0.13
Language-minority family	-0.07	-0.32*	-0.24
Mother-father family	-0.13	-0.27***	-0.15†
Neither birth parent in home	-0.04	0.41*	0.45*
Parent literacy standard score (KFAST)	0.00	0.00	0.00
Books in home	-0.21*	-0.11	0.11

Frequency of reading to child			
Not at all	-0.20	0.00	0.23
Once or twice	-0.05	0.13	0.18†
Every day	-0.15	-0.12	0.02
One year in Head Start	0.06	0.46***	0.39***

Table 7-5. Three-level regression models of Teacher-reported Hyperactive Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	4.58***	3.38***	-1.22
Proportion of variance accounted for	100%	87%	100%
Between-programs variance	17%	14%	16%
Between-classes variance	10%	10%	5%
Within-classes variance			
Total variance	11%	13%	8%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

N = 1,611.

Table 7-6. Three-level regression models of Teacher-reported Withdrawn Behavior scores in Head Start entry, graduation, and entry-graduation gains

	Unstandardized regression coeffficents (β)		
Program-level predictor variables	Entry	Graduation	Gains
Program mean parent education level	0.04	-0.37	-0.40
Program mean family income level	0.76	1.33*	0.42
High/Scope curriculum	-0.23	-0.97*	-0.46
Creative Curriculum	-0.21	-0.26	0.09
Mean teacher salary level	-0.36†	-0.47*	0.04
Proportion language-minority children	-0.20	0.03	-0.56
Proportion non-minority children	-0.07	0.15	-0.05
Class-level predictor variables			
Average ECERS-R language score	0.11	0.10	-0.05
Average child:adult ratio	-0.09†	-0.12*	0.02
AP individualizing score	-0.09	-0.02	0.08
Teacher BA or AA or higher	0.24	0.73***	0.36
Years teaching experience	0.02†	0.01	-0.02
African American teacher	-0.47	-0.18	0.08
Hispanic teacher	-0.34	0.05	0.19
Teacher salary deviation score	0.24	0.09	0.00
Teacher DAP beliefs score	-0.02	-0.07	-0.02
Average lead teacher Arnett score	-0.01	0.00	0.01
Class parent education level (deviation)	0.14	0.24	0.02
Class family income level (deviation)	-0.18	-0.36	-0.27
Class proportion non-minority (deviation)	1.40	1.14†	-1.07
Proportion language-minority (deviation)	-0.12†	0.15	0.08
Full-day class	-0.23	-0.22	0.03
Child-level predictor variables			
Parent education level (deviation)	0.00	-0.07	-0.06
Family income level (deviation)	-0.11	-0.03	0.09
Welfare status	0.02	0.33*	0.33†
Age of child in months	-0.07***	-0.03*	0.04**
Sex of child	-0.21†	-0.40***	-0.19
Disability status	1.18***	1.42***	0.23
African American child	0.31	0.55*	0.12
Hispanic child	0.42†	0.54*	0.13
Language-minority family	-0.13	-0.82***	-0.64*
Mother-father family	0.28*	-0.01	-0.28
Neither birth parent in home	0.16	0.13	-0.04
Parent literacy standard score (KFAST)	-0.01	-0.02**	-0.01
Books in home	-0.30*	-0.12	0.16
Frequency of reading to child			
Not at all	0.13	-0.07	-0.18
Once or twice	-0.18	0.16	0.38*
Every day	-0.21	-0.17	0.05
One year in Head Start	0.08	0.40**	0.31†

Table 7-6. Three-level regression models of Teacher-reported Withdrawn Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	7.45***	5.38**	-2.17
Proportion of variance accounted for			
Between-programs variance	55%	43%	25%
Between-classes variance	13%	22%	12%
Within-classes variance	8%	10%	4%
Total variance	10%	14%	6%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

$N = 1,591$.

Table 7-7. Three-level regression models of Parent-reported Social Skills scores in Head Start entry, graduation, and entry-graduation gains

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Program mean parent education level	0.04	0.50**	0.45*
Program mean family income level	-0.51†	-0.17	0.33
High/Scope curriculum	0.17	0.21	0.05
Creative Curriculum	0.24†	0.08	-0.16
Mean teacher salary level	0.27**	0.25*	-0.02
Proportion language-minority children	0.16	0.29	0.13
Proportion non-minority children	0.45	0.13	-0.29
Class-level predictor variables			
Average ECERS-R language score	-0.01	-0.05	-0.03
Average child:adult ratio	0.02	0.00	-0.01
AP individualizing score	0.09†	0.01	-0.08
Teacher BA or AA or higher	-0.16	-0.17	-0.01
Years teaching experience	0.00	0.00**	0.00
African American teacher	0.05	0.03	-0.02
Hispanic teacher	0.22	0.20	-0.03
Teacher salary deviation score	-0.07	-0.10	-0.03
Teacher DAP beliefs score	-0.06†	0.00	0.06
Average lead teacher Arnett score	0.00	0.00	0.00
Class parent education level (deviation)	-0.07	0.09	0.16
Class family income level (deviation)	-0.30†	-0.03	0.32*
Class proportion non-minority (deviation)	0.60	-0.09	-0.41
Proportion language-minority (deviation)	0.45	-0.22	-0.61†
Full-day class	0.02	-0.05**	-0.03
Child-level predictor variables			
Parent education level (deviation)	0.06	0.06†	0.00
Family income level (deviation)	0.01	0.00	-0.01
Welfare status	0.25*	0.12	-0.13
Age of child in months	0.03**	0.02*	-0.01
Sex of child	0.16*	0.03	-0.15
Disability status	-0.62***	-0.75***	-0.14
African American child	0.27†	0.09	-0.17
Hispanic child	-0.15	-0.21	-0.06
Language-minority family	0.03	-0.32†	-0.33
Mother-father family	-0.08	0.09	0.17†
Neither birth parent in home	-0.25	0.00	0.27
Parent literacy standard score (KFAST)	0.00	-0.01	-0.01*
Books in home	0.23*	0.02	-0.20†
Frequency of reading to child			
Not at all	-0.34†	-0.18	0.17
Once or twice	-0.20†	-0.28*	-0.08
Every day	0.32**	0.26*	-0.06
One year in Head Start	-0.01	-0.50	-0.48***

Table 7-7. Three-level regression models of Parent-reported Social Skills scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	9.73***	10.23***	0.59
Proportion of variance accounted for			
Between-programs variance	100%	100%	63%
Between-classes variance	15%	27%	36%
Within-classes variance	7%	6%	3%
Total variance	9%	8%	6%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

$N = 1,709$.

Table 7-8. Three-level regression models of Parent-reported Aggressive Behavior scores in Head Start entry, graduation, and entry-graduation gains

	Unstandardized regression coeffficents (β)		
Program-level predictor variables	Entry	Graduation	Gains
Program mean parent education level	-0.37†	-0.55**	-0.15
Program mean family income level	-0.05	0.20	0.28†
High/Scope curriculum	-0.02	-0.33*	-0.28
Creative Curriculum	-0.25	-0.27*	0.00
Mean teacher salary level	-0.14	-0.05	0.09
Proportion language-minority children	-0.10	0.01	0.09
Proportion non-minority children	0.07	-0.03	-0.05
Class-level predictor variables			
Average ECERS-R language score	0.05	0.00	-0.07
Average child:adult ratio	-0.03	-0.05†	-0.01
AP individualizing score	-0.04	-0.05	-0.01
Teacher BA or AA or higher	-0.04	-0.02	0.02
Years teaching experience	0.00	-0.01†	0.00
African American teacher	0.07	0.18	0.17
Hispanic teacher	0.02	0.16	0.22
Teacher salary deviation score	-0.10	-0.06	0.05
Teacher DAP beliefs score	0.04	-0.02	-0.07*
Average lead teacher Arnett score	0.00	0.00	0.00
Class parent education level (deviation)	-0.06	-0.06	0.00
Class family income level (deviation)	0.26†	0.11	-0.16
Class proportion non-minority (deviation)	0.74†	0.37	-0.37
Proportion language-minority (deviation)	-0.19	-0.23	-0.03
Full-day class	-0.02	0.02	0.07
Child-level predictor variables			
Parent education level (deviation)	-0.09*	-0.02	0.07†
Family income level (deviation)	-0.05	-0.08	-0.03
Welfare status	0.02	0.00	-0.03
Age of child in months	-0.02*	-0.01	0.00
Sex of child	-0.20*	-0.35***	-0.15†
Disability status	0.54***	0.45***	-0.10
African American child	-0.25	-0.46***	-0.23
Hispanic child	-0.10	-0.18	-0.06
Language-minority family	0.08	-0.10	-0.19
Mother-father family	-0.05	-0.17†	-0.11
Neither birth parent in home	0.07	-0.06	-0.13
Parent literacy standard score (KFAST)	0.00	0.00	0.00
Books in home	-0.11	0.00	0.12
Frequency of reading to child			
Not at all	0.32	0.40†	0.07
Once or twice	0.25*	0.06	-0.19†
Every day	-0.14	-0.19†	-0.06
One year in Head Start	-0.08	0.26*	0.36**

Table 7-8. Three-level regression models of Parent-reported Aggressive Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	5.67***	5.91***	0.20
Proportion of variance accounted for			
Between-programs variance	64%	100%	100%
Between-classes variance	29%	54%	18%
Within-classes variance	4%	4%	2%
Total variance	8%	7%	5%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

$N = 1{,}694$.

Table 7-9. Three-level regression models of Parent-reported Hyperactive Behavior scores in Head Start entry, graduation, and entry-graduation gains

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Program mean parent education level	-0.53**	-0.41**	0.13
Program mean family income level	0.01	0.17	0.19
High/Scope curriculum	-0.10	-0.29*	-0.19
Creative Curriculum	-0.19	-0.24*	-0.04
Mean teacher salary level	-0.07	-0.06	0.00
Proportion language-minority children	-0.22	-0.04	0.18
Proportion non-minority children	0.21	-0.09	-0.30
Class-level predictor variables			
Average ECERS-R language score	0.08†	0.03	-0.05
Average child:adult ratio	-0.02	-0.01	0.00
AP individualizing score	0.00	0.04	0.03
Teacher BA or AA or higher	0.16†	0.10	-0.04
Years teaching experience	-0.01	-0.01†	0.00
African American teacher	0.19	0.05	-0.15
Hispanic teacher	0.17	0.13	-0.03
Teacher salary deviation score	-0.15†	0.02	0.18†
Teacher DAP beliefs score	0.00	-0.04	-0.04
Average lead teacher Arnett score	0.00	0.00	0.00
Class parent education level (deviation)	-0.01	-0.05	-0.05
Class family income level (deviation)	0.11	0.03	-0.07
Class proportion non-minority (deviation)	-0.02	0.15	0.15
Proportion language-minority (deviation)	0.16	-0.01	-0.20
Full-day class	-0.01	0.02	0.04
Child-level predictor variables			
Parent education level (deviation)	-0.09**	-0.04	0.05
Family income level (deviation)	-0.07†	-0.04	0.03
Welfare status	-0.01	0.11	0.12
Age of child in months	0.00	0.00	0.00
Sex of child	-0.13†	-0.24***	-0.11
Disability status	0.81***	0.77***	-0.03
African American child	-0.10	-0.28**	-0.17
Hispanic child	-0.06	-0.08	-0.02
Language-minority family	0.52***	0.36*	-0.15
Mother-father family	-0.08	-0.12	-0.05
Neither birth parent in home	0.41*	0.72***	0.30
Parent literacy standard score (KFAST)	-0.01***	-0.01*	0.01*
Books in home	-0.01	-0.15	-0.14
Frequency of reading to child			
Not at all	0.65***	0.57***	-0.06
Once or twice	0.33***	0.09	-0.23*
Every day	0.07	-0.02	-0.08
One year in Head Start	0.06	0.18†	0.13

Table 7-9. Three-level regression models of Parent-reported Hyperactive Behavior scores in Head Start entry, graduation, and entry-graduation gains (continued)

Program-level predictor variables	Unstandardized regression coeffficents (β) Entry	Graduation	Gains
Intercept	4.34***	3.95***	-0.37
Proportion of variance accounted for			
Between-programs variance	90%	100%	100%
Between-classes variance	76%	55%	7%
Within-classes variance	10%	9%	3%
Total variance	17%	15%	3%

Note: Deviation scores in child/family level represent deviations from classroom level mean. Deviation scores in classroom level represent deviations from program mean.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

N = 1,696.

Table 7-10. Variance accounted by program-, classroom-, and child- and family-level models in social behavior outcomes

	Program-level models		
	Percentage of variance accounted for in:		
Social behavior	Entry	Graduation	Entry to end gain
Teacher report:			
Social skills	1.2	2.6**	2.9*
Aggressive behavior	0.9	1.9	1.3
Hyperactive behavior	0.4	1.7	1.6
Withdrawn behavior	0.6	1.8	0.7
Parent report:			
Social skills	1.3	1.9**	0.5
Aggressive behavior	1.6	2.4***	0.7
Hyperactive behavior	5.2***	5.0***	0.5
Withdrawn behavior	1.8*	0.8	0.8
	Classroom-level models		
	Percentage of variance accounted for in:		
Social behavior	Entry	Graduation	Entry to end gain
Teacher report:			
Social skills	2.1	1.5	1.4
Aggressive behavior	2.0	1.4	0.8
Hyperactive behavior	1.8	1.7	1.2
Withdrawn behavior	3.1†	2.1	1.1
Parent report:			
Social skills	1.2	1.0	1.4
Aggressive behavior	1.9	1.8	1.6
Hyperactive behavior	2.7*	3.4**	0.9
Withdrawn behavior	1.9	1.3	1.6
	Child and family-level models		
	Percentage of variance accounted for in:		
Social behavior	Entry	Graduation	Entry to end gain
Teacher report:			
Social skills	14.4***	9.1***	8.7***
Aggressive behavior	8.7***	8.8***	2.3***
Hyperactive behavior	9.5***	8.7***	4.7***
Withdrawn behavior	6.7***	7.5***	3.5***
Parent report:			
Social skills	7.1***	5.8***	3.5***
Aggressive behavior	5.7***	4.7***	2.4**
Hyperactive behavior	14.5***	12.5***	1.9**
Withdrawn behavior	5.6***	4.0***	1.4†

Note: Numbers represent percentage of variance accounted for by each model.

Tests of model significance–χ^2.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

7.3.4 Do Head Start Program and Classroom Characteristics Explain Children's Social Development Over and Above Child and Family Characteristics?

The variance accounted for by the combined program- and classroom-level models in social behavior scores is presented in Table 7-3. As can be seen in the table, the combined model typically explained 3 to 5 percent of the variance in the social behavior scores. The only two cases where the combined model explained more than 6 percent of the variance were in relation to the parent-reported hyperactive behavior at entry and graduation (6.8 percent and 6.4 percent, respectively). Nine of the 24 combined models were significant and three more were close to significance (see Table 7-3). Next, it was examined whether the inclusion of the program- and classroom-level sets of independent variables in the analyses improved models' fit to the data over and above models which included only the child- and family-level set of independent variables. In Tables 7-11 and 7-12, the comparison between the child-level models and the combined model (of all three levels) is presented. As can be seen in the table, these levels did add a considerable amount of explained variance to the combined models. Eleven of the 24 additions were significant, and 3 others were close to significance, but all of the combined models accounted for more variance in social behavior outcomes than the child and family models by themselves, typically adding 2 to 4 percent to the models' fit to the data.

Table 7-11. Variance accounted by the combined program- and classroom-level model in social behavior outcomes

	Program- and classroom-level models		
	Percentage of variance accounted for in:		
Social behavior	Entry	Graduation	Entry to end gain
Teacher report:			
Social skills	4.1	4.4***	4.0**
Aggressive behavior	2.5	3.1	2.7
Hyperactive behavior	2.2	3.8	3.1
Withdrawn behavior	3.9†	5.4***	1.9
Parent report:			
Social skills	2.6†	2.6*	1.9†
Aggressive behavior	3.6	3.7	3.5**
Hyperactive behavior	6.8*	6.4*	1.3
Withdrawn behavior	3.0**	1.9	2.8***

Note: Numbers represent percentage of variance accounted for by the model.

Tests of model significance–χ^2.

† *p* < .1; * *p* < .05; ** *p* < .01; *** *p* < .001.

Table 7-12. Variance accounted by child- and family-level models and the addition program- and class-level models to the combined models

Social behavior	Percentage of variance accounted for: Child-level models	Combined models	Addition
Teacher report:			
Social skills:			
Entry	14.4***	17.2***	2.8
Graduation	9.1***	14.1***	5**
Gain	8.7***	13.7***	5†
Aggressive behavior:			
Entry	8.7***	11.2***	2.5
Graduation	8.8***	12.7***	3.9
Gain	2.3***	5.7***	3.4
Hyperactive behavior:			
Entry	9.5***	11.3***	1.8
Graduation	8.7***	13.1***	4.4*
Gain	4.7***	8.0***	3.5†
Withdrawn behavior:			
Entry	6.7***	9.9***	3.2*
Graduation	7.5***	13.6***	6.1*
Gain	3.5***	5.6***	2.1
Parent report:			
Social skills:			
Entry	7.1***	9.1***	2.0†
Graduation	5.8***	7.9***	2.1*
Gain	3.5***	5.7***	2.2
Aggressive behavior:			
Entry	5.7***	7.9***	2.2**
Graduation	4.7***	7.4***	2.7***
Gain	2.4**	4.8***	2.4
Hyperactive behavior:			
Entry	14.5***	16.7***	2.2***
Graduation	12.5***	14.6***	2.1***
Gain	1.9**	3.2**	1.3
Withdrawn behavior:			
Entry	5.6***	8.4***	2.8*
Graduation	4.0***	5.0***	1.0
Gain	1.4†	4.2**	2.8*

Note: Numbers represent percentage of variance accounted for by each model.

Tests of model significance–χ^2.

† $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

7.4 Summary and Conclusions

The strongest predictors of children's social behaviors were predictors at the child and family level, mainly the child's age, gender, disability status, and number of years spent in Head Start, as well as the presence of both biological/adoptive parents in the house and the existence of books in the home. Most child and family characteristics have shown stronger links with the social behavior variables than characteristics in the program and classroom levels.

The program- and especially the classroom-level characteristics showed relatively low predictive power over children's social behaviors. Nonetheless, the addition of the program and classroom levels to the child and family level set of characteristics did result in higher explanatory power of the combined model over and above the explanatory power of the child- and family-level model by itself. This addition was significantly higher in almost half of the cases examined (see Table 7-4). Thus, the inclusion of program and classroom level sets of variables in the combined models increased the models' overall fit to the data. This finding supports our initial expectations that (at least in some cases) the Head Start experience may serve to compensate for disadvantages in the child- and family-levels.

The FACES 2000 data are not completely comparable to data from the FACES 1997 cohort because only two levels were examined in the 1997 modeling: child and family level and center level. Nonetheless, the similarities between the two cohorts are still noticeable. In both cohorts, the same child and family characteristics mentioned above had the strongest links to social behaviors and in both cohorts the Head Start-related levels added a significant amount of explained variance to increase the models' fit to the data. It should be noted, however, that these results do not supports claims that Head Start could produce dramatic social gains in children with disadvantaging circumstances. In particular, important program and classroom characteristics such as the kind of curriculum used, classroom quality variables (e.g., ECERS-R and Arnett scores), and child:adult ratio were not as strong predictors of children's social outcomes as expected.

CHAPTER 8.
HEAD START CHILDREN WITH DISABILITIES: SOCIAL-EMOTIONAL AND COGNITIVE DEVELOPMENTAL OUTCOMES

8.1 Introduction

This chapter examines two hypotheses regarding children with disabilities and their non-disabled counterparts. The first hypothesis is that Head Start children with disabilities exhibit lower social-emotional and cognitive developmental outcomes than their counterparts without disabilities upon entrance into Head Start programs. This hypothesis is consistent with expectations regarding the apparent linkages between types and severity of disabilities and negative outcomes for children with these disabilities. The second hypothesis is that children with disabilities will not progress as well as their non-disabled counterparts over the course of a Head Start year in most social-emotional and cognitive developmental outcomes even after controlling for differences in baseline starting points for these outcomes and socio-demographic characteristics. This hypothesis will be verified as correct if it can be shown that children with disabilities exhibited statistically significantly different gains, either in a positive direction or a negative direction, from their non-disabled counterparts, favoring the later group of children, in most social-emotional and cognitive developmental outcomes.

8.2 Methodology

This chapter of the report describes the socio-demographic and other family characteristics of Head Start children[1] with disabilities and those without disabilities and compares these two groups of children on their social-emotional and cognitive developmental outcomes.

Numerous social-emotional and cognitive developmental outcome measures were utilized for the analyses in this chapter. The social-emotional outcomes as reported by each child's primary caregiver (parent) and teacher include aggression, hyperactivity, withdrawn behavior, total problem behaviors (a compilation of previous three), positive approaches to learning (parent only), and cooperative classroom behavior (teacher only). The cognitive developmental outcomes as derived from

[1] There were 2,429 sampled cases with both completed parent interviews and child assessments in the Fall 2000 sample. Only 2,337 of these children and their families were utilized in the analyses for this chapter based on the availability of parent-reported and child assessment data for both the Fall 2000 and Spring 2001 data collection periods, and specifically, the parental reporting of disability status in the Spring 2001 parent interviews.

direct child assessments include receptive vocabulary knowledge, as measured by the Peabody Picture Vocabulary Test-III (PPVT-III); letter and word knowledge, early mathematical skills, and early writing skills, as measured by the Woodcock Johnson–Revised (WJ-R) subtests Letter Word Identification, Applied Problems, and Dictation; perceptual-motor skills, assessed by the McCarthy Draw-a-Design; and book knowledge. Detailed information about the measures is included in the Appendix.

Head Start requires 10 percent of enrolled children to meet the definition of disability (45 CFR - Part 1308.4). For these analyses, disability status was determined by the Spring 2001 parent interview. Children were categorized as disabled only if (a) the parent reported that the child possessed a disability as reported by a doctor or other health or education professional at some point prior to or during enrollment in Head Start, and (b) the parent reported that the child had an Individualized Education Plan (IEP) or an Individualized Family Service Plan (IFSP). Children's disability status as reported in the Spring of 2001 was used in lieu of their disability status as reported in the Fall of 2000. This was done to reflect the diagnoses of disabilities over the course of the Head Start year so as not to underreport the numbers of children with verifiable disabilities. Some of the children categorized as non-disabled may have actually had disabilities that had not yet been diagnosed as of Spring 2001. For these analyses, these children will be categorized as non-disabled.

The analyses conducted for this section of the report entailed the generation of results at the beginning of the Head Start year (Fall 2000) to delineate baseline results and across the Head Start year (Fall 2000 to Spring 2001) for both children with disabilities and children without disabilities. Data from the Fall 2000 and Spring 2001 parent interviews, teacher-child reports, and child assessments were utilized for these analyses. For measures of children's knowledge and cognitive skills, the sample size was limited to children who completed assessments in English during both interview periods. The sample size was limited in this way because it would have been impossible to gauge change over time with assessments that had been administered in different languages using non-comparable assessment tools.

Progress in the social-emotional and cognitive developmental domains over the course of the Head Start year was demonstrated by taking the difference between Spring 2001 scores and Fall 2000 scores. For example, a difference of 5.0 on the PPVT standard score (standardized with regard to the age of the child) represented a 5-point gain from Fall to Spring, and a score of -5.0 on the PPVT standard score represented a 5-point loss. Generally, positive difference scores indicated progress in that domain. The exception to this is with ratings of problem behaviors, where higher scores for these social-emotional variables indicated more problematic behaviors. In this case, negative numbers represented reductions in problematic behaviors, and thus, progress.

All of the differences to be discussed with regard to socio-demographic and other family characteristics, and all of the differences to be discussed with regard to children's social-emotional and cognitive developmental outcomes, were determined to be statistically significant through statistical tests and regression modeling. For the results pertaining to the social-emotional and cognitive developmental outcome measures, significance tests were run in SAS utilizing the GLM procedure with a least squares means option to control for the large disparities in sample sizes (unbalanced design) between the disabled subgroup and the non-disabled subgroup. This procedure essentially equates the sample sizes between subgroups, thus controlling for large disparities in sample sizes between subgroups in statistical tests to determine if averages are significantly different.

8.3 Findings

As reported by parents, 6 percent of the Head Start children had a disability (see Table 8-1). The majority of the children with disabilities had either speech or language impairments or cognitive impairments.

Table 8-1. Disability categories for children with disabilities (Spring 2001)

Disability categorizations	Weighted percentages diagnosed
Percent of Head Start Children with Disabilities	6%
Cognitive impairment*	23%
Speech or language impairment**	82%
Behavioral impairment***	20%
Physical impairment†	17%
Sensory impairment‡	13%

*Cognitive impairment includes the following: Specific learning disability, mental retardation, autism, and non-categorical/developmental delay.

**Speech or language impairment includes speech impairment and language impairment.

***Behavioral impairment includes emotional/behavioral disorder.

†Physical impairment includes the following: orthopedic impairment, health impairment lasting six months or more, and traumatic brain injury.

‡Sensory impairment includes the following: deafness, other hearing impairment, blindness, and other visual impairment.

Note: Individuals can have more than one impairment type within any impairment category and/or can have more than one impairment type across the impairment categories. Eleven cases could not be incorporated into any of the five disability categorization types, referenced above, and thus were excluded from this analysis.

White children were three times more likely than African American children (9 percent vs. 3 percent) to have disabilities (see Table 8-2). Boys were nearly three times more likely than girls to have disabilities (8 percent vs. 3 percent). Children of parents with more than a high school education were two times more likely than children of parents with a high school education or less to have disabilities (8 percent vs. 4 percent). Finally, the parents of children with disabilities were older, on average, than the parents of children without disabilities (31.19 vs. 29.59) (see Table 8-3). There were no significant differences between children with disabilities and those without disabilities on any of the other socio-demographic measures referenced in Tables 8-2 and 8-3.

Table 8-2. Characteristics of children with disabilities and their families

Demographic/socioeconomic variables – categorical	Weighted percentages
Percent of Head Start children with disabilities	6%
African American	3%***
White	9%***
Hispanic	5%
Other Race/Ethnicities	7%
Male	8%***
Female	3%
Child's age – 3 years of age	5%
Child's age – 4, 5, or 6 years of age	6%
Parent with high school education or less	4%**
Parent with more than high school education	8%
Married	6%
Not married	5%
Parent working	5%
Parent not working	6%
Urban	5%
Non-urban	6%
Welfare receipt status – Yes	4%
Welfare receipt status – No	6%
Food stamps receipt status – Yes	5%
Food stamps receipt status – No	6%
Language-minority household	4%
Non-language-minority household	6%

Difference of Means T-Tests were run in Excel using standard errors generated in WesVar v4.1.

*** p<.001; ** p<.01.

For race/ethnicity, the only significant difference is between African American children and White children.

Table 8-3. Characteristics of the families of children with disabilities and the families of children without disabilities (Fall 2000)

Demographic/socioeconomic variables – continuous	Weighted average	
	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Parent's age	31.19**	29.59
	(7.26)	(7.76)
Monthly family income	$1,668.13	$1,482.68
	($1,157.98)	($1,082.60)
Family size	4.74	4.66
	(1.77)	(1.80)

Standard deviations are in parentheses.

Difference of Means T-Tests were run in Excel using standard errors generated in WesVar v4.1.

** p<.01.

The results generated with disability status predicting to all beginning of the Head Start year social-emotional and cognitive developmental outcomes indicated that, on average, children with disabilities did not perform as well as other children on any of these measures with the exception of two (see Tables 8-4 and 8-5). The two exceptions were for receptive vocabulary knowledge and book knowledge, for which no significant differences were observed.

Table 8-4. Parent- and teacher-reported social-emotional outcomes for both children with disabilities and children without disabilities (Fall 2000) [bivariate results]

Social-emotional measures Outcome variables – continuous	Weighted averages	
	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Aggression scale (parent-reported)	3.59**	3.14
Hyperactivity scale (parent-reported)	2.46***	1.84
Withdrawn Behavior scale (parent-reported)	0.95***	0.59
Total Problem Behaviors scale (parent-reported)	8.01***	6.08
Positive Approaches to Learning scale (parent-reported)	11.38***	12.17
Aggression scale (teacher-reported)	3.02***	1.66
Hyperactivity scale (teacher-reported)	2.34***	1.28
Withdrawn Behavior scale (teacher-reported)	4.58***	2.35
Total Problem Behaviors scale (teacher-reported)	9.93***	5.30
Cooperative Classroom Behavior scale (teacher-reported)	12.26***	14.79

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status as the independent variable in each model.

*** p<.001; ** p<.01.

Table 8-5. Cognitive developmental outcomes for both children with disabilities and children without disabilities (Fall 2000) [bivariate results]

Cognitive developmental measures Outcome variables – continuous – E1E2	Weighted averages Disabled	 Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Receptive vocabulary knowledge – PPVT standardized score	83.48	85.23
Letter/word knowledge – WJ-R letter/word identification standardized score	89.57*	92.73
Early mathematical skills – WJ-R applied problems standardized score	82.09**	88.27
Early writing skills – WJ-R dictation standardized score	79.03***	85.58
Perceptual – motor skills – McCarthy draw-a-design score	2.47***	2.96
Book knowledge – book knowledge score	1.64	1.61

E1E2 – Assessment administered in English in both Fall 2000 and Spring 2001.

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status as the independent variable in each model.

*** $p<.001$; ** $p<.01$; * $p<.05$.

The results generated with disability status predicting beginning of the Head Start year social-emotional and cognitive developmental outcomes, controlling for numerous socio-demographic characteristics of the sampled children and their families, indicated that, on average, children with disabilities did not perform as well as other children on any of these outcomes with the exception of one (see Table 8-6 and 8-7). The one exception was for book knowledge, for which no significant difference was observed. The covariates included in the models were child's race/ethnicity, gender, and age; parent's (respondent's) age; highest education level of parent(s); monthly family income; parental marital status; parent's (respondent's) employment status; family size; urbanicity of school; household language-minority status; and parental reading to child on a daily basis.

The results generated with disability status predicting children's progress over the course of the Head Start year with regard to all social-emotional and cognitive developmental outcomes, controlling for numerous socio-demographic characteristics of the sampled children and their families, indicated that, on average, children with disabilities did not progress as well as other children on parent- and teacher-reported hyperactivity, withdrawn behavior, and total problem behaviors (see Table 8-8 and 8-9). The covariates included in the models were Fall 2000 baseline score (for each individual scale), child's race/ethnicity, gender, and age; parent's (respondent's) age; highest education level of parent(s); monthly family income; parental marital status; parent's (respondent's) employment status; family size; urbanicity of school; household language-minority status; and parental reading to child on a daily basis. On average, children with disabilities progressed similarly to children without disabilities on all remaining social-emotional outcomes. With regard to the cognitive developmental outcomes, children with disabilities did not progress as well as other children, on average, on early math skills, perceptual-motor skills, and book knowledge. No other significant differences were observed.

Table 8-6. Parent- and teacher-reported social-emotional outcomes for both children with disabilities and children without disabilities (Fall 2000) [multivariate results]

Social-emotional measures	Weighted averages	
Outcome variables – continuous	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Aggression scale (parent-reported)	3.59**	3.13
Hyperactivity scale (parent-reported)	2.40***	1.81
Withdrawn Behavior scale (parent-reported)	0.96***	0.57
Total Problem Behaviors scale (parent-reported)	7.85***	5.99
Positive Approaches to Learning scale (parent-reported)	11.54***	12.18
Aggression scale (teacher-reported)	2.91***	1.68
Hyperactivity scale (teacher-reported)	2.26***	1.28
Withdrawn Behavior scale (teacher-reported)	4.46***	2.36
Total Problem Behaviors scale (teacher-reported)	9.63***	5.32
Cooperative Classroom Behavior scale (teacher-reported)	12.60***	14.86

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status and other covariates as the independent variables in each model.

*** $p<.001$; ** $p<.01$.

Table 8-7. Cognitive developmental outcomes for both children with disabilities and children without disabilities (Fall 2000) [multivariate results]

Cognitive developmental measures	Weighted averages	
Outcome variables – continuous – E1E2	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Receptive vocabulary knowledge – PPVT standardized score	79.50***	85.67
Letter/word knowledge – WJ-R Letter/Word Identification standardized score	89.54*	92.52
Early mathematical skills – WJ-R Applied Problems standardized score	81.51***	88.54
Early writing skills – WJ-R Dictation standardized score	79.41**	85.53
Perceptual – motor skills – McCarthy Draw-a-Design score	2.40***	2.95
Book knowledge – Book Knowledge score	1.51	1.63

E1E2 – Assessment administered in English in both Fall 2000 and Spring 2001.

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status and other covariates as the independent variables in each model.

*** $p<.001$; ** $p<.01$; * $p<.05$.

Table 8-8. Parent- and teacher-reported social-emotional outcomes for both children with disabilities and children without disabilities (Fall 2000 – Spring 2001) [multivariate results]

Social-emotional measures	Weighted averages	
Outcome variables – continuous	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Aggression scale (parent-reported)	-0.07	-0.29
Hyperactivity scale (parent-reported)	0.12**	-0.23
Withdrawn Behavior scale (parent-reported)	0.17*	-0.00
Total Problem Behaviors scale (parent-reported)	0.18*	-0.53
Positive Approaches to Learning scale (parent-reported)	-0.10	0.05
Aggression scale (teacher-reported)	0.11	-0.08
Hyperactivity scale (teacher-reported)	0.17**	-0.18
Withdrawn Behavior scale (teacher-reported)	0.43**	-0.25
Total Problem Behaviors scale (teacher-reported)	0.63**	-0.51
Cooperative Classroom Behavior scale (teacher-reported)	1.36	2.01

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status and other covariates as the independent variables in each model.

** *p*<.01; * *p*<.05.

Table 8-9. Cognitive developmental outcomes for both children with disabilities and children without disabilities (Fall 2000 – Spring 2001) [multivariate results]

Cognitive developmental measures	Weighted averages	
Outcome variables – continuous – E1E2	Disabled	Nondisabled
Percents of Head Start children with and without disabilities	6%	94%
Receptive vocabulary knowledge – PPVT standardized score	4.76	4.10
Letter/word knowledge – WJ-R Letter/Word Identification standardized score	-0.54	0.76
Early mathematical skills – WJ-R Applied Problems standardized score	-4.62***	1.79
Early writing skills – WJ-R Dictation standardized score	-0.34	2.34
Perceptual – motor skills – McCarthy Draw-a-Design score	0.31*	0.63
Book knowledge – Book Knowledge score	0.48**	0.85

E1E2 - Assessment administered in English in both Fall 2000 and Spring 2001.

Significant differences were determined using the GLM Procedure (Least Squares Means option) in SAS v8.0.

A regression model was run for each dependent variable outcome with disability status and other covariates as the independent variables in each model.

*** *p*<.001; ** *p*<.01; * *p*<.05.

8.4 Conclusion

In summary, at the beginning of the Head Start year children's disability status was highly associated with both social-emotional and cognitive developmental outcomes. Results generated without covariates indicated that, on average, children with disabilities did not perform as well as other children on all baseline social-emotional and cognitive developmental outcomes with the exception of receptive vocabulary knowledge and book knowledge. Results generated with covariates indicated that, on average, children with disabilities did not perform as well as other children on all baseline social-emotional and cognitive developmental outcomes with the exception of book knowledge.

Results generated pertaining to children's progress over the course of the Head Start year with regard to both social-emotional and cognitive developmental outcomes, controlling for baseline outcomes and child and family socio-demographic characteristics, indicated that, on average, children with disabilities did not progress as well as other children on 9 out of 16 social-emotional and cognitive developmental outcomes. The ones for which they did not progress as well as their non-disabled counterparts included the following: parent- and teacher-reported hyperactivity, withdrawn behavior, total problem behaviors, early math skills, perceptual-motor skills, and book knowledge. The remaining seven outcomes were not statistically significantly different for the two groups.

8.5 Discussion

As noted earlier, White children were more likely than African American children to be disabled. Male children were more likely than female children to be disabled. These findings are consistent with findings from other studies indicating the greater prevalence of identifications of disabilities for White children as compared to others and male children as compared to female children (Coiro, Zill, & Bloom, 1994; Bloom & Tonthat, 2002; Child Trends Data Bank, 2002). Because young boys are more likely than young girls to have speech or language problems (Dale, Price, Bishop, & Plomin, 2003) that are classified as disabilities, it is not surprising that more male children than female children in the Head Start sample were classified as disabled. This is particularly the case given the fact that speech and language impairments constituted the largest category within the disabilities categorization.

As delineated earlier, children with disabilities exhibited more problematic behaviors (i.e., had higher scores on negative behavior measures and lower scores on positive behavior measures) than other children for all social-emotional outcomes in the Fall of 2000 as consistent with expectations regarding the apparent linkages between the types and severity of the disabilities and negative outcomes for children with these disabilities. Over the course of a Head Start year, children with disabilities were not able to achieve similar gains to their non-disabled counterparts in 6 out of 10 social-emotional outcomes, controlling for baseline starting point outcomes and other child and family socio-demographic characteristics. Interestingly enough, for these six outcomes, children with disabilities actually experienced deterioration in both parent- and teacher-reported behaviors while their counterparts experienced improvements in these behaviors. There were no significant differences in gains between children with disabilities and their non-disabled counterparts for any of the other social-emotional outcomes.

The results obtained for the cognitive developmental outcomes were quite similar to those obtained for the social-emotional outcomes in the Fall of 2000. Children with disabilities had lower baseline outcomes for most of the cognitive developmental outcomes. Over the course of a Head Start year, children with disabilities were not able to achieve similar gains to their non-disabled counterparts in three out of six cognitive developmental outcomes, controlling for baseline starting point outcomes and other child and family socio-demographic characteristics. For two of the outcomes, children with disabilities experienced significantly smaller gains than their non-disabled counterparts. Interestingly enough, for one of the outcomes (early math skills), the former group of children experienced a significantly large decline while the latter group experienced a sizable increase in this skill area. There were no significant differences in gains between children with disabilities and their non-disabled counterparts for any of the other cognitive developmental outcomes.

CHAPTER 9.
FAMILY STRUCTURE AND FAMILY CHANGES

9.1 Introduction

Past research shows that the nature of the family and other household socioeconomic characteristics has important implications for children's well being (Amato, 2000; Duncan and Brooks-Gunn, 1997). Attempts have also been made to better understand the mechanisms through which this relationship operates. Some argue that family dissolution resulting in single parenthood has a negative effect on children because of a decline in emotional support from the single parent (Astone and McLanahan, 1991) or that deterioration in child outcomes is a result of a possible decline in familial economic resources following marital dissolution (Duncan and Brooks-Gunn, 1997; Duncan, Brooks-Gunn, Yeung, and Smith 1998; White and Rogers, 2000 for example). Other research shows that parents' mental health and wellbeing could also affect child outcomes (Simons, Beaman, Cogner, and Chao, 1993). More recent research using longitudinal data on young children distinguishes between the effects of various household and family characteristics based on the type of child outcomes discussed. Ram and Hou (2003) argue that while household poverty mediates the relationship between family structure and cognitive outcomes, a deterioration of familial resources including mothers' marital dissolution has a negative impact on emotional and behavioral outcomes of children.

This chapter uses four waves of data from the FACES 2000 data collection effort to gain a better understanding of the parental and other household characteristics of children who enroll in the Head Start program. The main interest in this chapter lies in examining these household characteristics both at the time of initial enrollment into the Head Start program as well as observing annual changes as the children move from Head Start to kindergarten. The differences in these characteristics are observed between children who spend additional years in the Head Start program and those who move on to kindergarten by Spring 2002. This chapter also examines changes in family structure and composition including mothers' presence in the home and marital status, her education and employment, as well as her health and well-being as children move through the Head Start program. The relationships among mothers' marital status, family structure and selected children's cognitive outcomes are also examined after controlling for mothers' race and other household socioeconomic characteristics such as parents' education, household poverty status, and mothers' health and well-being. Chapter 10 of this report further examines the extent of risk factors that are experienced by Head Start families and their relationship to children's outcomes after taking into account parent involvement, and Head Start's protective role for children.

9.2 The FACES 2000-2003 Longitudinal Sample

The FACES 2000 longitudinal data come from interviews of parents of Head Start children at four points of data collection (the base year, Fall 2000, and follow up surveys in Spring 2001, Spring 2002, and Spring 2003). Children between the ages of 3 and 5 years with no previous Head Start experience who enrolled into the program were selected for this study in Fall 2000 and were subsequently followed through Spring 2003.[1] The analysis presented in this chapter is based on information for the longitudinal sample of 1,645 Head Start children interviewed in Fall 2000 for whom data from all three time points, Fall 2000, Spring 2001 and Spring 2002 were available. All these children continued in the Head Start program in Spring 2001. However, while 678 children remained in Head Start in Spring 2002, 967 children had moved on to kindergarten. In Spring 2003, the 678 children who were in Head Start in Spring 2002 had also moved on to kindergarten.[2] Therefore, the results presented in this chapter are not representative of the entire Fall 2000 sample as they exclude households which had dropped out of the sample over time or for whom insufficient information was available at later time points. Furthermore, while Table 9-1 reports information obtained on mothers of all Head Start children, the results presented in Tables 9-2 to 9-5 are restricted to households where the mother was the primary caregiver.[3]

[1] The results presented later in this chapter are based on the use of appropriate weights. The Spring 2003 longitudinal weight is used for longitudinal analysis of the data following Head Start children from Fall 2000 through Spring 2003.

[2] A small number of kindergarten children interviewed in Spring 2003 were those repeating kindergarten. The numbers presented in all tables do not include information on these "repeater" kindergarten children.

[3] As Table 9-1 shows, slightly less than 7 percent of the mothers do not reside in the households of their children and approximately 11 percent are not primary caregivers. However, the data show that the characteristics of the respondent of the parent interview (i.e., the primary caregiver) on the topics of interest to the study are very similar to that of the mother.

Table 9-1. Changes in mother's and other family characteristics of children in the Head Start program at various time points: FACES 2000-2003

	Fall 2000	Spring 2001	Spring 2002	Spring 2003
	HS	HS	HS and KG	KG
Mother's Characteristics				
Mother's education				
Some high school or lower	34.3	-	-	-
High school graduate/GED	36.1	-	-	-
High school plus voc/tech	5.6	-	-	-
Some college/Associate's degree	21.5	-	-	-
Bachelor's degree or more	2.5	-	-	-
Mother's education in last 12 months				
No classes	61.0	87.6	-	-
High school diploma	-	0.9	-	-
Classes but not towards a degree/certificate	14.2	-	-	-
Certificate/license/Associate's degree etc.	17.2	9.6	-	-
Bachelor's degree or more	4.1	0.1	-	-
Other	3.6	1.7	-	-
Mother's employment status				
Working full-time	37.6	40.4	42.9	43.2
Working part-time	14.7	12.9	14.4	11.1
Looking for work	6.3	7.1	2.6	3.1
Not in labor force	34.9	34.5	38.1	40.4
No mother in household	6.5	5.1	2.0	2.2
Mother's marital status				
Married	44.3	45.7	49.8	43.7
Single, never married	37.0	35.1	34.0	38.2
Divorced/widowed	11.4	11.9	10.0	9.5
Married living separately	7.3	7.3	6.2	8.6
Family Characteristics				
% With Mother in Household	93.5	94.9	92.2	91.9
% With Mother as Primary Caregiver	89.2	90.0	88.7	87.8
Family size				
2-3	25.5	24.9	23.9	26.3
4-5	49.9	51.8	52.2	52.9
6 +	24.6	23.3	23.9	20.8
Family structure				
Mother-father	48.2	48.6	47.8	43.2
Mother only	45.3	46.0	44.4	48.7
Father only	2.1	2.1	2.5	2.1
Neither mother nor father	4.5	3.4	5.3	5.9

See note at end of table.

Table 9-1. Changes in mother's and other family characteristics of children in the Head Start program at various time points: FACES 2000-2003 (continued)

	Fall 2000	Spring 2001	Spring 2002	Spring 2003
	HS	HS	HS and KG	KG
Household SES				
Monthly Family Income				
Less than $500	11.3	8.0	8.6	7.1
$500 to $999	22.4	19.7	16.9	18.6
$1000 to $1499	26.3	25.2	24.6	24.6
$1500 to $1999	14.6	17.0	17.2	18.1
$2000 or more	25.3	30.2	32.7	31.7
Poverty Status				
% Poor	65.4	59.2	57.3	59.3
% Receiving Welfare	22.0	19.6	18.5	17.0
N	1,645	1,645	1,645	678

Notes: HS refers to children in the Head Start program, KG refers to children who have moved on to Kindergarten but were interviewed as part of FACES:2000-2003. Information on mother's education is available only for Fall 2000. For mothers in Spring 2001, information on recent education refers to education since Thanksgiving, therefore comparisons between mother's recent education for Fall 2000 and Spring 2001 should be made with caution. Information on mother's recent education between data collection points is not presented for Spring 2002 and Spring 2003 due to missing information in the data. In Spring 2002, 678 children were enrolled in their second year of Head Start while 967 children had moved on to Kindergarten.

9.2.1 Mothers' Characteristics

In order to explore the changes in family characteristics and their implications for child outcomes over time, the first step is to get a good picture of the nature of mothers' and family characteristics of Head Start children between Fall 2000 and Spring 2003. Table 9-1 shows that 65 percent of all mothers had at least completed high school or had obtained a GED at the time their child entered Head Start. Although about 21 percent of mothers had attended some college or had obtained an Associate's degree, only 2.5 percent had a Bachelor's degree. Much of this education was completed well before Fall 2000. About 61 percent of mothers did not attend any classes in the last 12 months before the Fall 2000 survey was conducted. About 21 percent had attended classes towards a certificate, license, an Associate's degree, or a Bachelor's degree. Furthermore, as many as 81 percent of mothers in Spring 2001 did not attend any classes since the last Thanksgiving.

With regard to employment, 35 percent of mothers were not in the labor force in Fall 2000, implying that they were not employed as well as not looking for a job. In later years, this percentage increased to 38 percent of mothers in Spring 2002 being out of the labor force. This increase is matched by a corresponding decline in the percentage of mothers looking for work. Most others worked full time (38 percent in Fall 2000), again with percentages increasing over time.[4]

Over 44 percent of all mothers of Head Start children were currently married in Fall 2000. This percentage increases steadily over the years, and was as high as 50 percent among mothers of children in Spring 2002. Over the years, the increase in marriage rates is matched by a decline in the percentage of mothers who were divorced or widowed or those married but living separately.

9.2.2 Family and Household Characteristics

Table 9-1 also shows that approximately half of the children enrolled in Head Start in Fall 2000 and later years lived in families comprised of 4 to 5 members. About a quarter were in larger families, and a quarter in smaller families with 2 to 3 members. A majority of the children were also either the youngest or the only child in the household.

About 94 percent of the families of these children had a mother present in the household. Furthermore, about 89 percent of mothers in the household were the primary caregivers for their children. In Fall 2000, slightly less than half the families had the mother and father of the children present in the household while approximately 45 percent of families had only the mother in the household and a very marginal number (about 2 percent) had only the father residing in the household. The same general patterns persist across the years with some small differences. These results generally fit the pattern of family structure reported in earlier analysis on Head Start children in the FACES 1997 data (FACES 1997 Technical Report II, forthcoming).

Since the Head Start program is targeted towards economically-deprived children, it is not surprising that household income levels of the Head Start children are low. In Fall 2000, 11 percent of households had a monthly income of less than $500, about half the households had incomes ranging between $500 and $1,500 a month, and about 25 percent had a monthly income over $2,000. Over the years however, there is a slight increase in the economic status of these households. A lower incidence of

[4] Full time work is defined as working 35 or more hours a week.

poverty (57 percent) is also evident by Spring 2002 as compared to 65 percent of households in Fall 2000. Similarly, a decline in the percentage of households receiving welfare is also evident over the years.

9.2.3 Mothers' Health and Well-being

Research has shown that mothers' mental and physical health and general well-being is especially important in the context of its implications for children's outcomes. An examination of the rates of depression among mothers in Table 9-2 indicates a general betterment in mothers' mental health over time. The percentage of mothers who were not depressed increased from 47 percent in Fall 2000 to close to 60 percent in Spring 2002. This corresponds with the steady decline in the percentage of mothers with various levels of depression ranging from mild depression to severe depression. Similarly, on a locus of control scale of 0 to 21 that includes information on feelings and behavior, mothers reported an average score of 14.8 in Fall 2000 with a marginal but statistically significant increase in the score in subsequent years.

Table 9-2. Changes in health and wellbeing of mother's of children in the Head Start program at various time points: FACES 2000-2003

	Fall 2000	Spring 2001	Spring 2002	Spring 2003
	HS	HS	HS and KG	KG
Depression				
Not depressed	47.1	45.5	59.2	61.2
Mildly depressed	26.5	30.0	20.4	18.6
Moderately depressed	14.2	13.5	11.1	11.1
Severely depressed	12.3	11.0	9.3	9.1
Locus of Control	14.8	15.3	15.3	15.3
Health				
Excellent	22.3	26.4	21.2	20.5
Very good	33.0	30.1	33.0	33.0
Good	31.0	28.7	32.4	35.0
Fair	11.3	12.8	11.3	9.8
Poor	2.5	2.0	2.2	1.7
Mother is primary caregiver	1,461	1,472	1,262	590

Notes: Analysis is restricted to mothers who were respondents in each of the three time points, Fall 2000, Spring 2001 and Spring 2002. HS refers to children in the Head Start program, KG refers to children who have moved on to Kindergarten but were interviewed as part of FACES:2000-2003 In Spring 2002, 678 children were enrolled in their second year of Head Start while 967 children had moved on to Kindergarten.

9.3 Mothers' Characteristics and Children's Cognitive Outcomes

Regression analyses are used to examine the relationship between mothers' marital status and family structure of children enrolled in Head Start (their mothers' marital status, whether both parents reside in the household) between Fall 2000 and Spring 2002 and children's cognitive outcomes assessed using the standardized scores for the Peabody Picture Vocabulary Test and the Woodcock Johnson-Revised Letter Word Identification and Applied Problem tests in Spring 2002.[5] These analyses are restricted to those children whose mothers were respondents and had complete information on household characteristics at all time periods (Fall 2000, Spring 2001, and Spring 2002) as well as cognitive assessments in Spring 2002 thus bringing the sample size down to 1,180 children. The regression models control for mothers' and household socio-economic characteristics such as mothers' ethnicity, highest parental education, and household poverty status in Fall 2000; mothers' health status between Fall 2000 and Spring 2002; and whether the child was enrolled in the Head Start program or kindergarten in Spring 2002.[6]

Table 9-3 presents means of children's cognitive outcomes in relation to their mothers' characteristics. The table shows that there are some differences in cognitive scores of children of mothers who stayed married between Fall 2000 and Spring 2002 and those who were single and never married. Children of married mothers performed better than children of never married mothers on the assessment of math skills. Interestingly, children of mothers who had been married only for some of the time during the period Fall 2000-Spring 2002 also performed better than children of single, never married mothers with respect to their vocabulary and math skills.[7] However, no differences in cognitive outcomes were observed for children based on their co-residence with both parents between Fall 2000 and Spring 2002. The table also shows that children of mothers who reported that that they were always in good health performed better on two of the measures than children of mothers with changing health status.

The means presented in Table 9-3 do not control for mothers' and other household socio-economic characteristics such as race, parents' education, poverty and mothers' health status. Regression results that take these characteristics into account are presented in Tables 9-4 and 9-5. Table 9-4 presents

---

[5] The Woodcock Johnson-Revised tests assess the child's writing and math skills while the PPVT assesses the child's vocabulary skills. The WJ-R LWI test was not administered to children in kindergarten in Spring 2002.

[6] Since the mother could be either the biological or adoptive mother, there is a possibility that the respondent could vary between the survey time points. However, this variation is expected to be small.

[7] It is important to note that this category of persons includes mothers who were married in Fall 2000 but may have been divorced/separated subsequently or others who were divorced or separated in Fall 2000 and got married by Spring 2002.

results for the relationship between mothers' marital status and the three cognitive outcomes while Table 9-5 shows results of the relationship between family structure and the three cognitive outcomes. In each table, for each of the cognitive outcomes, results from two models are presented. The first column examines the relationship between mothers' marital status (Table 9-4) and family structure (Table 9-5) and cognitive outcomes of Head Start children without controlling for any household socio-economic characteristics. The second column takes into account the role of additional characteristics such as race, parents' education, poverty, and mothers' health status.

Table 9-3. Means and standard deviations of children's cognitive outcomes in Spring 2002 in relation to selected mother's characteristics (Fall 2000-Spring 2002)

	Writing skills		Math skills		Vocabulary		N
Marital Status							
Always Single, Never Married	93.52	(12.49)	88.51 $_{ab}$	(17.90)	89.70 $_{a}$	(14.26)	335
Always Married	91.36	(12.72)	92.74 $_{a}$	(19.60)	90.77	(16.33)	486
Any Divorce/Separation/Widowhood	92.28	(12.23)	92.08 $_{b}$	(18.34)	92.43 $_{a}$	(14.86)	352
Family Structure							
Both Parents Always in Household	91.36	(12.57)	91.94	(19.25)	90.17	(16.64)	531
Sometimes Less than Two Parents in Household	92.96	(12.44)	90.88	(18.44)	91.67	(14.17)	648
Depression							
Never Depressed	91.95	(12.87)	92.06	(19.03)	90.87	(15.58)	679
Always Depressed	93.39	(10.94)	92.94	(17.40)	92.68	(14.15)	83
Sometimes Depressed	92.52	(12.31)	89.84	(18.66)	90.83	(15.20)	418
Health							
Always Good Health	93.08 $_{a}$	(12.60)	91.66	(19.11)	91.59 $_{a}$	(14.96)	876
Always Poor Health	88.92	(12.27)	90.68	(16.88)	92.86	(14.68)	51
Changing Health Status	89.96 $_{a}$	(11.89)	90.38	(18.14)	88.52 $_{a}$	(16.50)	253

Note: Standard deviations are given in parentheses. Analysis is restricted to mothers who were respondents at all three time points, Fall 2000, Spring 2001 and Spring 2002. Subscripts indicate that means are significantly different from each other ($p<0.05$). Differences in math skills between children of mothers who were always single and always married are significant at $p < 0.01$. Differences in vocabulary between children of mothers who were always in good health and those with changing health are significant at $p < 0.01$. Writing and Math skills were assessed using the Woodcock Johnson tests, vocabulary skills were assessed using PPVT. Assessment of child's writing skills using the WJ Word Test was administered only to children in the Head Start program in Spring 2002 while other tests were administered to children in Kindergarten as well. Family structure is defined as "sometimes less than two parents in the household" if at any of the time points between Fall 2000 and Spring 2002, one of the child's parents was not present in the household. A mother is reported to have changing health status if she was in poor health in at least one but not all of the time points between Fall 2000 and Spring 2002.

Table 9-4. Relationship between longitudinal change in mothers' marital status and various cognitive outcomes in Spring 2002: FACES 2000 (Fall 2000-Spring 2002)

	Writing skills		Math skills		Vocabulary	
Mothers' Marital Status						
(Always Married: reference group)						
Always Single, Never Married	3.35*	1.25	-4.53**	-1.24	-1.44	0.62
Any Divorce/Separation/Widowhood	1.90	1.61	-0.97	-0.62	1.56	1.12
Enrolled in KG (Spring 2002)		-		7.03**		3.23**
Mothers' Race (White: reference group)						
African American		5.54**		-7.70**		-10.19**
Hispanic		1.12		-6.81**		-15.78**
Other race		3.57		-9.74**		-15.68**
Parents' Education						
(Less than High School: reference group)						
High School		5.94**		6.88**		4.84**
Some College or higher		8.13**		11.35**		8.46**
Poverty Status		-2.36+		-1.89		-1.56+
Mothers' Health Status						
(Always Good Health: reference group)						
Always Poor Health		-3.02		-0.55		0.91
Changing Health Status		-1.08		-0.55		-0.99
Constant	90.16**	85.03**	92.70**	87.15**	90.59**	92.72**
Observations	539	525	1148	1117	1137	1106
R-squared	0.01	0.11	0.01	0.15	0.01	0.30

Note: + significant at 10%; * significant at 5%; ** significant at 1%. Analysis is restricted to mothers who were respondents at all three time points, Fall 2000, Spring 2001 and Spring 2002. Writing and Math skills were assessed using the Woodcock Johnson tests, vocabulary skills were assessed using PPVT. Assessment of child's writing skills using the WJ Word Test was administered only to children in Spring 2002 while other tests were administered to children in Kindergarten as well. A mother is reported to have changing health status if she was in poor health in at least one but not all of the time points between Fall 2000 and Spring 2002.

Consistent with results presented in Table 9-3, the regression results in Table 9-4 indicate that mothers' marital status bears a statistically significant relationship to children's cognitive outcomes when examined in isolation without controlling for other family characteristics. Surprisingly, children of single, never married mothers perform better than children of married mothers with regard to children's writing skills. But the opposite is true with regard to math skills. However, this relationship disappears after controlling for other family characteristics. Coresidence with both parents is also not related to children’s cognitive outcomes in Table 9-5 after controlling for other family and background characteristics.

Table 9-5. Relationship between longitudinal change in family structure and various cognitive outcomes in Spring 2002: FACES 2000 (Fall 2000-Spring 2002)

	Writing skills		Math skills		Vocabulary	
Family Structure (Both parents in household: reference group)						
Less than Two Parents in Household	2.01	0.63	-1.86	-0.32	0.98	1.22
Enrolled in KG (Spring 2002)		-		7.02**		3.06**
Mothers' Race (White: reference group)						
African American		5.71**		-7.97**		-10.57**
Hispanic		0.83		-6.79**		-15.57**
Other race		3.39		-9.85**		-15.88**
Parents' Education (Less than High School: reference group)						
High School		5.88**		6.84**		4.93**
Some College or higher		8.21**		11.33**		8.45**
Poverty Status		-2.00		-2.04		-1.73+
Mothers' Health Status (Always Good Health: reference group)						
Always Poor Health		-3.47		-0.56		0.76
Changing Health Status		-0.99		-0.59		-1.13
Constant	90.56**	85.25**	92.16**	87.03**	90.14**	92.88**
Observations	544	530	1154	1122	1143	1111
R-squared	0.01	0.11	0.00	0.15	0.00	0.30

Note: + significant at 10%; * significant at 5%; ** significant at 1%. Analysis is restricted to mothers who were respondents at all three time points, Fall 2000, Spring 2001 and Spring 2002. Writing and math skills were assessed using the Woodcock Johnson tests, vocabulary skills were assessed using PPVT. Assessment of child's writing skills using the WJ Word Test was administered only to children in Spring 2002 while other tests were administered to children in Kindergarten as well. A mother is reported to have changing health status if she was in poor health in at least one but not all of the time points between Fall 2000 and Spring 2002.

Children in kindergarten in Spring 2002 have better cognitive outcomes as compared to children still enrolled in Head Start (Tables 9-4 and 9-5). With regard to the role of household socio-economic characteristics, higher education of parents, particularly at the college level, is associated with better cognitive outcomes while the poverty status of the household is associated with poorer outcomes.

Moreover, children of mothers who identified themselves as White also performed better than others. An exception is the case of children's writing skills where children whose mothers were African American performed better than children with White mothers.

9.4 Summary and Conclusions

This chapter uses data from the FACES 2000-2003 longitudinal surveys to present an overall picture of the family characteristics of children enrolled in Head Start. In particular, the emphasis is on the family structure and characteristics of these families including the marital status and general health and well-being of the mother, the predominant primary caregiver in the household, in relation to child outcomes. The longitudinal data collection effort of FACES provides the ability to examine changes in these characteristics and relationships over time.

Overall, the results indicate that the condition of Head Start children's families and especially their mothers has generally remained the same or marginally improved as children move from enrollment in Head Start to kindergarten between Fall 2000 and Spring 2002. Although not all mothers reside in the household of their children, the percentage remains consistently over 90 percent. It is also interesting to note that a high percentage of mothers are also single, never married mothers. Over 40 percent of households have only the mother (and no father) residing in the child's household. The longitudinal analysis shows evidence of an improvement in the socioeconomic status of households over the years. The data show that a greater percentage of households report higher levels of monthly income in Spring 2002 and Spring 2003. Moreover, a smaller percentage of households remains in poverty or depends on welfare as children move through the Head Start program and possibly to kindergarten.

Examining the role of these mothers' and household characteristics on children's cognitive outcomes using the FACES 2000 data shows that a change in the family situation represented in terms of a change in mothers' marital status or coresidence with parents bears little relationship with children's cognitive outcomes. Instead, household characteristics such as parents' education and household poverty status are more important predictors. While greater parents' education is related to better cognitive outcomes, a poor household economic situation is associated with lower assessment scores.

CHAPTER 10.
FAMILY RISK FACTORS, PARENTAL INVOLVEMENT, AND HEAD START'S PROTECTIVE ROLE

Head Start has a strong interest in supporting parents as the primary educators of their children. This chapter explores relationships between family and parental characteristics and child outcomes (including teacher ratings of social skills and behavior problems, and direct cognitive assessments) that may be related to the development of school readiness in the FACES 2000 sample of Head Start children.

These data provide further insight into the development of cognitive and social skills, particularly with regard to family challenges and strengths. The family risk factors investigated in this chapter include parental depression, exposure to violence, domestic violence, involvement in the criminal justice system, substance use, and cumulative social and economic risks. The chapter concludes with a presentation of protective family factors and investigates how experiences in Head Start may help some families cope with the multiple life challenges they encounter.

10.1 Research Questions

The following research questions will be addressed:

1. How are family and parental characteristics related to children's cognitive and social development?
2. How are family interactions with children related to children's cognitive and social development? and
3. How do parents interact with Head Start and how is their interaction related to children's cognitive and social development?

10.2 Findings

The findings presented in this chapter are primarily based on partial correlations, tests of group differences (t-test, ANOVA), and multivariate logistic regression models used to estimate levels of risk while controlling for demographic variables. Unless noted in the text, all findings tested significant

($p \leq .05$). Additional information on specific parent interview scales and child assessment measures is found in the Appendix. Means and standard deviations for scales from the parent interview and from the teachers' reports of children's behavior are also included in the Appendix.

10.2.1 Prevalence of Family and Parental Risk Factors and their Relationship to Children's Outcomes

Depression.[1] Because depression is a frequent challenge facing low-income families with young children (Hall, Williams and Greenberg, 1985; Liaw and Brooks-Gunn, 1994), depression among the Head Start parents was measured using the CES-D Depression Scale (Radloff, 1977). Overall, parents had a mean score of 6.8 in the Fall of 2000, which was in the mildly depressed range. As seen in Figure 10-1, most parents were classified as not depressed (47.7 percent) or only mildly depressed (27.0 percent), while one-fourth of the parents were classified as moderately depressed (13.6 percent) or severely depressed (11.7 percent). From Fall to Spring, there was a small decline in the overall mean depression scores (Spring 2001 mean score of 6.6), but the difference was not statistically significant. Over 40 percent of parents reported moderate or severe depression at two or more time points.

The level of depression did not vary significantly by race or ethnicity. Similar proportions of African American parents (27.7 percent), Hispanic parents (27.4 percent), and White parents (25.0 percent) were classified as moderately or severely depressed.

Group differences, evaluated using t-tests, were found between parents who were moderately or severely depressed and those who were not at all or only mildly depressed (Table 10-1). Parents who were moderately or severely depressed reported a lower household income, had a more external locus of control, had a more authoritarian parenting style (i.e., more directive and harsh), and spanked their children more frequently. When asked about activities with the children, parents who were moderately or severely depressed were more likely to report participating in fewer activities with their children and were also less likely to be involved with their children's Head Start program. A higher proportion of mothers living without a father in the home were classified as moderately or severely depressed (30.4 percent) than those who had a father present in the home (20.3 percent).

---

[1] The literature cited on depression focuses largely on maternal depression. All parent figures who served as primary caregivers, including fathers and grandfathers, were included in this analysis. However, 87.9 percent of the FACES primary caregivers were mothers.

Figure 10-1. The prevalence of depression among Head Start parents in Fall 2000

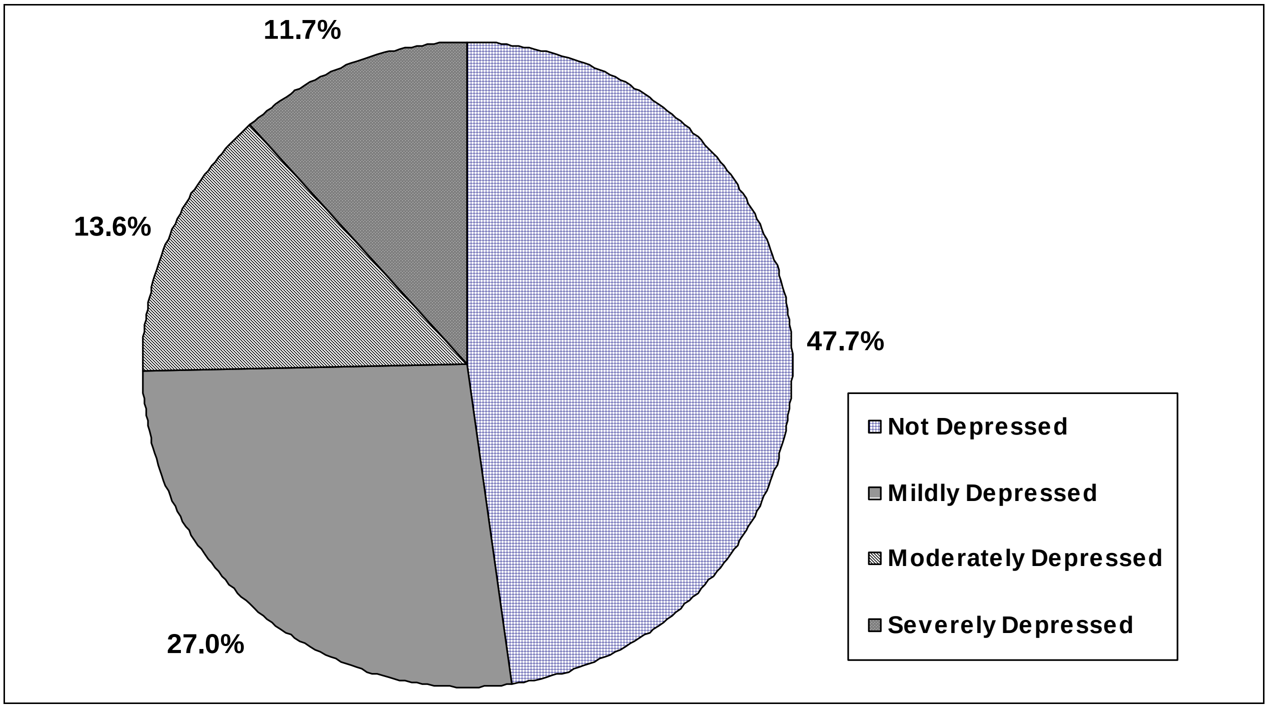


Table 10-1. Significant group differences between moderately or severely depressed parents and mildly or not at all depressed parents on selected family factors

Family factors	Moderately or severely depressed		Mildly or not depressed		T
	Mean	SD	Mean	SD	
Household Income	1.54	0.66	1.64	0.50	3.38***
Locus of Control	13.0	3.27	15.4	3.06	16.19***
Authoritarian Parenting Style	2.22	0.78	2.17	0.67	1.74*
Spanking	1.59	0.50	1.51	0.50	2.85**
Involvement at Head Start	21.3	5.86	22.00	7.59	1.96*

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

Results of partial correlations (Table 10-2) controlling for household income, education, employment, and child gender, age, and ethnicity also revealed that the overall parental depression scale score was significantly correlated with parent and teacher reports of the children's behavior. Parents who were more depressed reported that children had more problem behavior, including being more aggressive, hyperactive, and withdrawn. Similarly, teachers also reported more aggressive, hyperactive, and overall problem behavior for children of parents who were more depressed. Negative correlations were also found between parental depression and children's positive social behavior ratings and emergent literacy, indicating that less depressed parents reported having children with better social and academic skills.

Table 10-2. Significant partial correlations between parental depression and child behavior and cognitive outcomes

Problem Behavior–Parent Report $r = .23^{***}$	Aggressive Behavior–Parent Report $r = .18^{***}$	Withdrawn Behavior–Parent Report $r = .20^{***}$	Hyperactive Behavior–Parent Report $r = .15^{***}$
Positive Social Behavior $r = -.06^{**}$	Emergent Literacy–Parent Report $r = -.05^{**}$	Problem Behavior–Teacher Report $r = .08^{***}$	Aggressive Behavior–Teacher Report $r = .06^{**}$
Hyperactive Behavior–Teacher Report $r = .05^{**}$	One-to-One Counting $r = -.07^{**}$	Early Math Skills $r = -.09^{**}$	Creativity–Teacher Report $r = -.06^{**}$

$^{*}p \leq .05$; $^{**}p \leq .01$, $^{***}p \leq .001$.

Significant correlations were also found between parental depression and selected child cognitive outcomes. Children of parents who were depressed did worse on one-to-one counting and early math tasks, as well as on the teacher reports of creativity (descriptions of the child outcome measures are found in the Appendix).

Additional risks identified in the literature on maternal depression were explored more fully through multivariate logistic regression models, controlling for parent's education and employment; household income; and child's race, age, and gender. Odds ratios and 95 percent confidence intervals for each model are presented in Table 10-3. These risk estimates indicate that parents who were moderately or severely depressed, compared to parents who were not at all or only mildly depressed, were 1.45 times more likely to be single parents, and almost twice as likely to report that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children. In addition, depressed parents were 1.36 times more likely than non-depressed parents to drink or live with someone who drank alcohol, and 1.69 times more likely to have been exposed to violence in their neighborhoods or homes. The risk for screening positive for domestic violence greatly increased for those parents who were moderately or severely depressed. These parents were almost 3 times more likely to have been victims of domestic violence compared to parents who were only mildly depressed or not depressed.

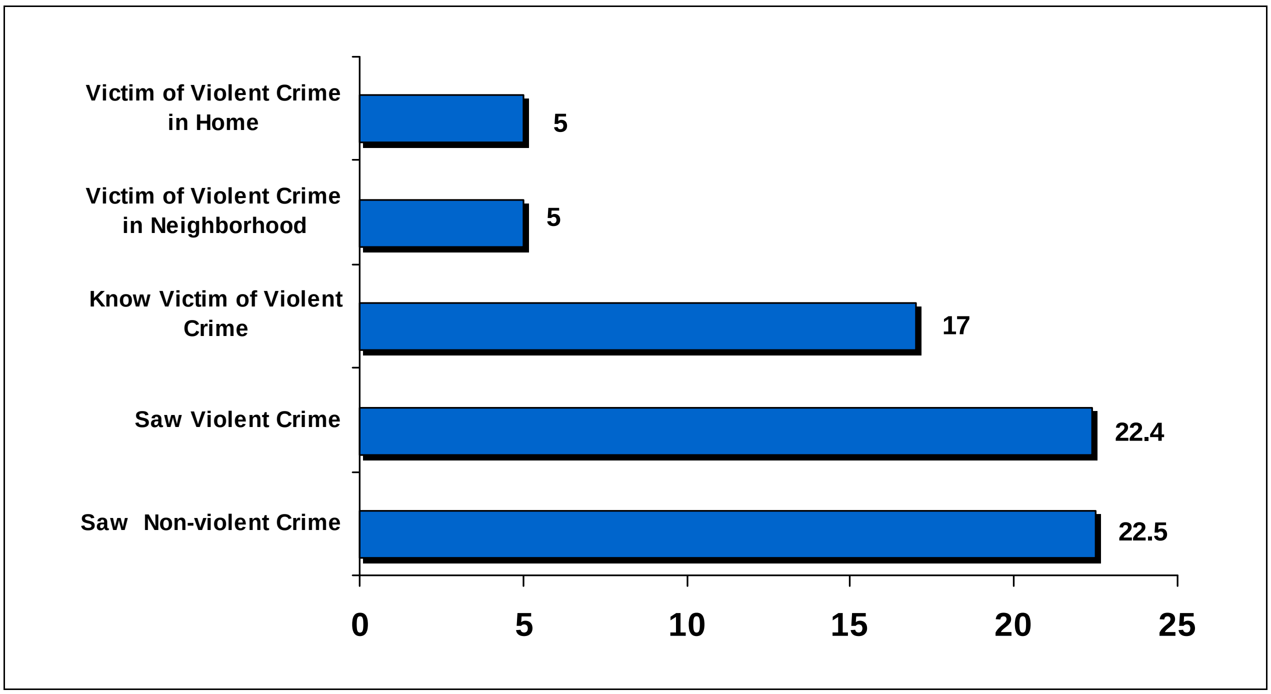
Table 10-3. Significant adjusted odds ratios and 95 percent confidence intervals for parental depression, exposure to violence, domestic violence, alcohol use, and involvement with the criminal justice system

	Parental depression	
Risk factors	Adjusted odds ratio	95 Percent confidence interval
Exposure to Violence	1.69	(1.38–2.06)
Domestic Violence	2.99	(2.33–3.84)
Alcohol Use	1.36	(1.12–1.65)
Criminal Justice System Involvement	1.97	(1.57–2.47)
Single Parent	1.45	(1.17–1.79)
	Exposure to violence	
Risk factors	Adjusted odds ratio	95 Percent confidence interval
Domestic Violence	2.68	(2.09–3.44)
Criminal Justice System Involvement	1.73	(1.38–2.16)
Alcohol Use	1.46	(1.21–1.75)
Single Parent	1.43	(1.18–1.74)
	Domestic violence	
Risk factors	Adjusted odds ratio	95 Percent confidence interval
Criminal Justice System Involvement	3.47	(2.68–4.50)
Single Parent	2.61	(1.97–3.46)
Alcohol Use	1.50	(1.18–1.91)
	Alcohol use	
Risk factors	Adjusted odds ratio	95 Percent confidence interval
Criminal Justice System Involvement	1.48	(1.19–1.83)
Exposure to Violence	1.45	(1.21–1.74)
Domestic Violence	1.48	(1.17–1.89)
	Criminal justice system involvement	
Risk factors	Adjusted odds ratio	95 Percent confidence interval
Parental Depression	1.97	(1.57–2.47)
Single Parent	2.64	(2.06–3.39)
Domestic Violence	3.47	(2.68–4.50)
Child Witness to Violence	3.23	(2.04–5.12)
Child Witness to Domestic Violence	3.60	(2.63–2.90)
Child Victim of Violence	3.06	(1.35–6.97)
Child Victim of Domestic Violence	3.14	(1.57–6.30)

Exposure to Violence. Neighborhoods have long been recognized in theory and research as important contexts for child development (Leventhal and Brooks-Gunn, 2000). Parents were asked about the violence they knew to occur in their neighborhoods, and were asked additional questions about their own personal exposure to violence and domestic violence, as well as their children's exposure to violence.

As depicted in Figure 10-2, more than one-fifth of all parents (22.5 percent) reported seeing nonviolent crime such as selling drugs or stealing in their neighborhoods in Fall 2000 (15.3 percent more than once), as well as having been a witness to violent crime (22.4 percent; 17.2 percent more than once). Approximately 17.0 percent of the parents knew someone who was the victim of a violent crime in their neighborhood, bringing the reality of violence very close to many of the Head Start families. Five percent of parents reported being a victim of violent crime in their neighborhood, and similarly, 5.0 percent of the parents reported being victims of violence in their homes.

Figure 10-2. The frequency of parents' exposure to crime in their neighborhoods and their homes



Exposure to violence varied across ethnic groups. Among parents of African American children, 34.5 percent reported seeing nonviolent crimes in their neighborhoods, a figure that was over twice the rate reported by parents of White children (15.0 percent) and approximately one-third more than parents of Hispanic children (24.2 percent). This pattern held for each type of exposure to violent crime. Over twice as many parents of African American children (36.3 percent) reported witnessing violent crimes in their neighborhoods compared to parents of White children (15.5 percent) and Hispanic children

(17.5 percent). For reports of victimization, parents of African American children were again highest, with 7.4 percent indicating they were victims of crime in their neighborhoods, and 6.2 percent reporting they were victims of violent crime in their homes. These victimization rates were closer to reports for parents of Hispanic children (5.8 percent in the neighborhoods; 5.4 percent in their homes) than for parents of White children (4.4 percent in the neighborhoods; 4.5 percent in their homes).

As for the Head Start children, 3.8 percent were reported by parents to have witnessed a violent crime and 8.6 percent were reported to have witnessed domestic violence during the previous year. Slightly more than 1 percent of the children were reported by their parents to have been victims of violent crime (1.1 percent), while almost 2 percent were victims of domestic violence (1.7 percent) during the previous year.

Exposure to violence had direct and indirect associations with child outcomes. Partial correlations controlling for household income, education, employment, and child gender, age, and ethnicity revealed small but significant positive correlations between parents' reports of exposure to violence and parents' reports of child problem behavior. In these analyses, exposure to violence is represented as a summary score of how often each of the five types of exposure to violence (as noted above) was reported by parents. Scores ranged from 5 (no exposure) to 15 (more than one exposure to every type). The mean score for the sample was 6.1. Parents who reported greater exposure reported their children engaged in fewer positive social behaviors and more overall problem behavior; teachers of these children also reported they were more aggressive. Exposure to violence did not have a direct relationship with child cognitive outcomes. Parents who reported more exposure to violence were significantly more depressed, but interestingly, they also were more likely to have an authoritative parenting style (e.g., less harsh, more use of rationales).

Controlling for mother's education and employment; household income; and child's race, age, and gender; multivariate logistic regression models were used to further explore the risks associated with exposure to violence. Odds ratios and 95 percent confidence intervals for risk factors (alcohol use, single parent, domestic violence, and criminal justice system involvement) included in the models are presented in Table 10-3. Parents who reported exposure to violence in their neighborhoods or homes, compared to parents who were not exposed to violence, were 1.45 times more likely to be single parents as well as to live in households where either they or someone else drank alcohol. Parents who were exposed to violence in their neighborhoods or homes, compared to parents not exposed to violence, were also 1.73 times more likely to have reported that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children, and over 2.68 times more likely to have screened positive for domestic violence.

Domestic Violence. A three-item screening measure for domestic violence was administered to the parents (Feldhaus, Koziol-McLain and Amsbury, 1997). They were asked if they had ever been hit, kicked, punched, or otherwise hurt by anyone within the past year; if they felt safe in their current relationship; and if they currently felt unsafe because of a partner from a previous relationship. Almost 13 percent of the parents answered 'yes' to one of these questions, thereby screening positively for experiencing domestic violence. Differences on family and child outcomes were found between families with a parent who screened positively for domestic violence versus those who did not. For example, parents experiencing domestic violence were significantly more depressed, and reported their children to be more aggressive, more hyperactive, more withdrawn, and to have more overall problem behavior. Teachers also reported these children to be more withdrawn and have more overall problem behavior than children whose parents were not experiencing domestic violence (Table 10-4). However, no direct relationship was found between parent-reported domestic violence and child cognitive outcomes.

Table 10-4. Significant group differences between parents who did and did not screen positive for domestic violence on parental depression and child behavior

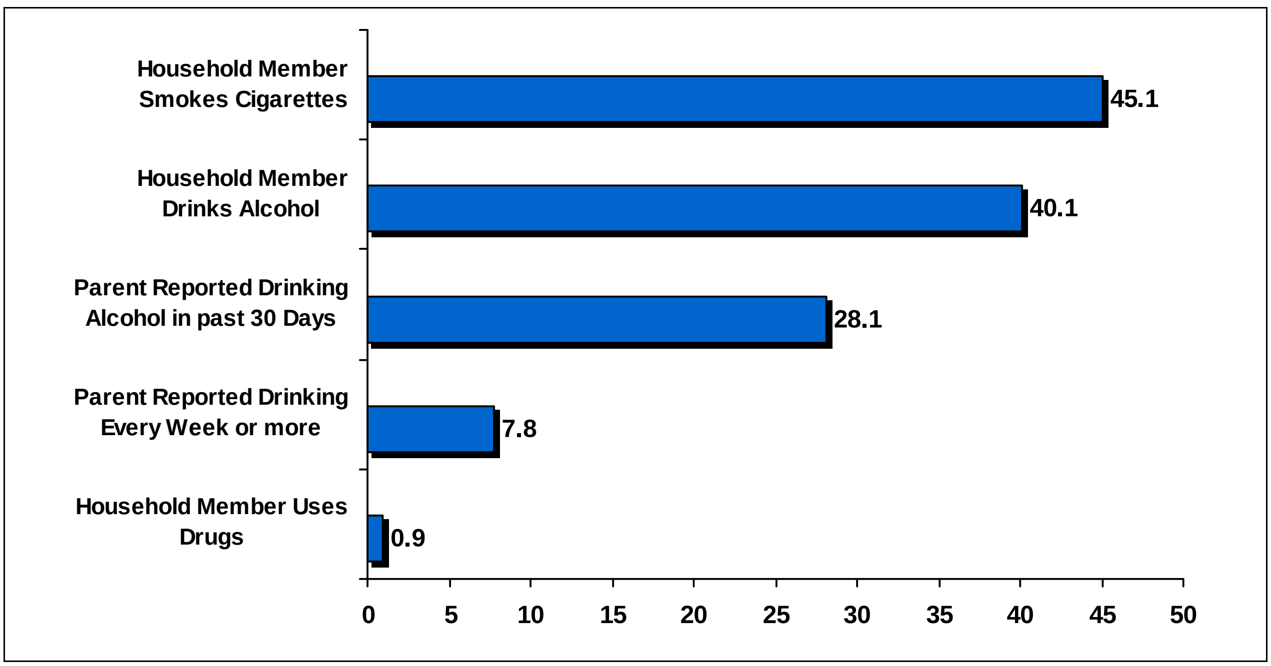
	Domestic violence		No domestic violence		
	Mean	SD	Mean	SD	T
Parental Depression	10.8	8.04	6.17	6.26	-12.19***
Child Behavior (Parent-Reported)					
Aggressive Behavior	3.59	1.77	3.09	1.73	-4.87***
Hyperactive Behavior	2.08	1.58	1.81	1.49	-2.92**
Withdrawn Behavior	0.85	1.08	0.58	0.90	-4.95***
Overall Problem Behavior	7.08	3.87	6.02	3.48	-5.17***
Child Behavior (Teacher-Reported)					
Withdrawn Behavior	2.98	2.71	2.53	2.64	-2.76**
Overall Problem Behavior	6.39	5.16	5.67	5.04	-2.34*

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

Multivariate logistic regression models, controlling for mother's education and employment; household income; and child's race, age, and gender; estimated that parents who screened positive for domestic violence, compared to those who did not, were 2.61 times more likely to be single parents, 1.5 times more likely to drink alcohol or live with a drinker, and almost 3.5 times more likely to report that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children. (See Table 10-3.)

Substance Use in the Home. The occurrence of substance use in homes, especially cigarette smoking and alcohol use, is another challenge that faced a number of Head Start families (Figure 10-3). Almost one-half of the children (45.1 percent) lived in households with at least one individual who smoked cigarettes. Smoking varied by ethnicity. Cigarette smoking was reported less frequently in households where African American children lived (38.8 percent) than in households of Hispanic children (52.2 percent) or White children (54.3 percent). More than one-fourth of the parents (28.1 percent) reported drinking alcohol such as beer, wine, or liquor in the past 30 days; 20.0 percent drank less than once a week, and 7.8 percent reported drinking between 1-2 times per week to every day. Slightly over 40 percent of all households reported having at least one individual who drank alcohol. Among families who lived in households where someone drank alcohol, 7.9 percent reported alcohol-related problems with family members, 5.9 percent experienced trouble with the police because of alcohol, and 4.2 percent missed work or school due to alcohol-related illness. Less than 1 percent of the families reported having anyone in the household who used drugs.

Figure 10-3. The frequency of substance use in the home



Parents who drank or who lived with a drinker were significantly more likely to be depressed and to report their children to have more overall problem behavior than parents who did not drink or live in a household with a drinker. Children who lived with someone who drank alcohol scored lower on vocabulary, color naming, and social awareness tasks.

Further exploration of the risks associated with drinking or having a drinker in the household was done using multivariate logistic regression models, again controlling for mother's education and employment; household income; and child's race, age, and gender. Odds ratios and 95 percent confidence intervals for the associated risk factors included in the models are presented in Table 10-3. Parents who drank alcohol or lived with a drinker were approximately 1.5 times more likely to report involvement in the criminal justice system, exposure to violence, or domestic violence in their lives compared to parents who did not drink or live with a drinker. Parents who drank or lived with a drinker were no more likely to be single parents than those who lived in alcohol-free households.

Involvement with the Criminal Justice System. In order to assess how many families had involvement with the criminal justice system, parents were asked if they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children. Almost one-fifth (19.2 percent) of the parents reported that someone had been arrested and charged with a crime and 16.7 percent reported someone had spent time in jail. Involvement with the criminal justice system did not vary much by race or ethnicity. About one-fifth of parents of White children, parents of Hispanic children, and parents of African American children reported having someone in their family who was arrested (21.8 percent, 20.1 percent, and 18.9 percent, respectively).

T-tests were conducted to examine if child behavior and cognitive outcomes differed between children who were members of families with involvement in the criminal justice system and children from families without such involvement (Table 10-5). Children from families who had someone arrested scored lower on assessed vocabulary and were reported by their parents to be more aggressive, more hyperactive, more withdrawn, and to have more overall problem behavior. They were also reported by their teachers to be more aggressive and have more overall problem behavior than children from families who did not have someone arrested.

Multivariate logistic regression models, controlling for mother's education and employment; household income; and child's race, age, and gender were used to determine estimates of risk among those families who had someone close to them who was involved in the criminal justice system (odds ratios and 95 percent confidence intervals are reported in Table 10-3). These risk estimates indicated that parents who reported that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children were twice as likely to be depressed, over 2.5 times more likely to be single parents, and over 3.5 times more likely to have been a victim of domestic violence than parents who did not have someone in their families involved in the criminal justice system.

Table 10-5. Significant group differences between families with involvement in the criminal justice system and families with no involvement in the criminal justice system on child behavior and cognitive outcomes

	Criminal justice involvement		No criminal justice involvement		
	Mean	SD	Mean	SD	T
Vocabulary	80.72	17.53	80.72	17.53	-3.91***
Child Behavior (Parent-Reported)					
Aggressive Behavior	3.42	1.73	3.09	1.74	-3.70***
Hyperactive Behavior	2.01	1.51	1.81	1.50	-2.67**
Withdrawn Behavior	0.77	0.99	0.57	0.91	-4.04***
Overall Problem Behavior	6.75	3.59	6.01	3.53	-4.03***
Child Behavior (Teacher-Reported)					
Aggressive Behavior	2.09	2.08	1.73	2.03	-3.34***
Overall Problem Behavior	6.18	4.86	5.67	5.10	-2.01*

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

Children in families from which someone had been arrested were at great risk for witnessing or being a victim of violent crime and domestic violence compared to children in families where no one had been arrested or charged with a crime. These children were more than 3 times more likely to have been a witness to violent crime and almost 3.5 times more likely to have witnessed domestic violence in the past year. Their risk of victimization also increased greatly. These children were 3 times more likely to have been both a victim of domestic violence and a victim of violent crime in the past year than children whose families did not report involvement with the criminal justice system. While there is a strong literature to suggest that children of incarcerated parents are at increased risk for negative outcomes (Lange, 2000), these findings suggest that risk greatly increases when there is any family involvement with the criminal justice system.

Cumulative Social and Economic Risk. Previous work has noted that social and economic risk factors may be associated with negative child outcomes, particularly when multiple risks are present (Annie E. Casey Foundation, 1999; ACYF, 2002). To explore this further, a standard list of family risk factors (Annie E. Casey Foundation, 1999) was adapted to investigate relationships with child outcomes. These risk factors include:

- The child lived with a single parent;
- The mother was a high school dropout;
- The household income was below the poverty line;
- The child was living with a parent(s) who did not have steady, full-time employment;
- The family was receiving welfare benefits; and
- The child did not have private health insurance.

While a number of these risks are already included in broader analytic models looking at child outcomes (see earlier chapters), the interest here is in how the cumulative effects of these risks may be associated with reduced school readiness. At least one of the six risks was evident in 89 percent of the Head Start families. Across families, the most prevalent risks were having a mother without a high school education (38.2 percent), being in a single-parent household (48.2 percent), and children not having private health insurance (64.6 percent).[2] The three other risks were reported for less than 30 percent of the families.

Increases in the number of risk factors, particularly counts of four or more risks, increased the likelihood of negative child outcomes. About one-fifth of the families (20.3 percent) were found to have four or more risk factors, with almost 10 percent having five or six risks. Less than one-fifth of the families of White children (18.6 percent) and Hispanic children (18.6 percent) had four or more of the risk factors, while one quarter of the families of African American children (24.9 percent) were found to have this high level of risk.

As shown in Table 10-6, parents reporting four or more risk factors also had higher depression scores, a more external locus of control, and, interestingly, scored higher than parents with fewer cumulative risks on both the authoritarian and authoritative parenting style scales. These same parents also reported more problem behavior for their children, including aggressive and withdrawn behavior, and gave lower ratings on emergent literacy. In addition, the Head Start teachers also rated the children from families with four or more risks as having more withdrawn and hyperactive behavior.

[2] The Fall 2000 FACES parent interview identified children without private health insurance, but was not clear on whether children received Medicaid or SCHIP.

In terms of the child cognitive assessments, the cumulative effect of four or more risk factors was clearly associated with lower child outcomes. This high level of risk meant significantly poorer performance on social awareness, design copying, color naming, one-to-one counting, book knowledge, comprehension, print concepts, vocabulary, letter identification, early math, and early writing.

Table 10-6. Significant group differences between families with fewer than four risks and families with four or more risks on child behavior, child cognitive outcomes, and parent variables

	Less than four risk factors N = 1,706		Four or more risk factors N = 505		
	Mean	SD	Mean	SD	T
Child Behavior (Parent-Reported)					
Aggressive Behavior	3.14	1.75	3.31	1.74	-1.94*
Withdrawn Behavior	0.6	0.94	0.69	0.97	-1.84*
Child Behavior (Teacher-Reported)					
Hyperactive Behavior	1.33	1.52	1.5	1.56	-2.22*
Withdrawn Behavior	2.48	2.69	2.74	2.56	-1.93*
Overall Problem Behavior	5.57	5.08	6.12	5.10	-2.13*
Child Cognitive Status (Parent-Reported)					
Emerging Literacy	2.02	1.45	1.54	1.69	4.95***
Child Cognitive Assessments					
Vocabulary	81.19	17.56	78.23	15.71	3.29***
Early writing	85.52	14.21	82.66	13.01	2.53**
Early math	88.31	15.10	83.54	16.61	3.75***
Letter identification	92.06	9.34	90.02	7.22	2.94**
Social awareness	3.36	1.63	3.09	1.67	3.28***
Design copying	3.02	1.38	2.87	1.25	2.20*
Color naming	11.17	7.30	8.95	7.22	6.05***
Counting	0.35	0.66	0.27	0.06	2.31*
Book knowledge	1.57	1.25	1.14	1.24	2.48*
Comprehension	0.54	0.71	0.46	0.66	2.27*
Print concepts	0.24	0.57	0.18	0.48	1.97*
Parent self-reports					
Parent Depression	6.18	6.64	7.96	7.21	-6.17***
Parent Locus of Control	14.71	3.20	13.85	3.39	4.42
Authoritative Parenting	4.20	0.62	4.28	0.63	-3.45***
Authoritarian Parenting	2.18	0.69	2.30	0.73	-2.40*

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

10.2.2 Family Involvement and its Relationship to School Readiness

Family Activities. While the findings presented above show that the aggregated effect of selected family characteristics has negative associations with children's school readiness, the frequency of families' activities that include the Head Start children reflects potential family strengths that could improve school readiness. For example, do families that are more active with their children have children with better behavior and better cognitive skills?

Parents were asked how often family members engaged in selected activities over the previous week as well as the previous month with their Head Start children. The weekly activities included telling the child a story; teaching letters, words, or numbers; teaching songs or music; doing arts and crafts; playing toys or games; doing errands; or doing household chores. Parents were also asked about how often they read to their children over the past week, and findings related to this variable are addressed in Chapter 5. The monthly activities included visiting the library, shows, museums, and zoos; attending community or sporting events; and discussing family history. Total activity scores are based on the sum of the weekly and monthly activity scores.

The total activity score for combined weekly and monthly activities in Fall 2000 indicated that families engaged in a mean of 7.6 activities with the children, out of a possible 14 activities (National Center for Education Statistics, 1996). Weekly activities made up most of that total, with a reported mean of 5.7 activities of a possible seven, while a mean of 1.9 monthly activities was reported, also out of a possible seven. Total reported activity with the children increased significantly from Fall 2000 to Spring 2001 and from Spring 2001 to Spring 2002 (means = 7.6, 8.3, and 8.5, respectively).

Racial or ethnic differences were noted in the number of activities families engaged in with their children. For the activities in the previous week, African American children had higher family activity than both White or Hispanic children, and White children had higher activity scores than Hispanic children. For the activities covered for the previous month, there was a similar effect for ethnicity, as African American children were engaged in more activity than White or Hispanic children.

There were several positive relationships between family activities with children and child behavioral and cognitive outcomes (Table 10-7). Partial correlations were run controlling for household income, education, and employment; child gender, age, and ethnicity; and how often the child was read to in the previous week. The frequency of weekly, monthly, and combined family activities was positively correlated with parent reports of positive child behaviors and emergent literacy skills. All three types of activity scores were negatively correlated with parent reports of aggressive behavior, while weekly and total behavior scores were negatively associated with overall problem behavior and the hyperactive behavior subscale. Teacher ratings of hyperactive and overall problem behavior were negatively correlated with reported monthly activities. Among parents, increased engagement in activities with their children was correlated positively with both locus of control and authoritative parenting (Table 10-8).

Table 10-7. Significant partial correlations between family activities with the child and child behavior and cognitive outcomes

Weekly Activities with Child–Positive Child Behavior *r* =.22***	Monthly Activities with Child–Positive Child Behavior *r* =.11***	Overall Activities with Child–Positive Child Behavior *r* =.20***	Weekly Activities with Child–Emerging Literacy *r* =.12***
Monthly Activities with Child–Emerging Literacy *r* =.12***	Overall Activities with Child–Emerging Literacy *r* =.15***	Weekly Activities with Child–Aggressive Child Behavior *r* =-.10***	Monthly Activities with Child–Aggressive Child Behavior *r* =-.04*
Overall Activities with Child–Aggressive Child Behavior *r* =-.09***	Weekly Activities with Child–Hyperactive Child Behavior *r* =-.09***	Overall Activities with Child–Hyperactive Child Behavior *r* =-.05*	Weekly Activities with Child–Overall Problem Behavior *r* =-.09***
Overall Activities with Child– Overall Problem Behavior *r* =-.06*	Monthly Activities with Child–Teacher-reported Hyperactive Child Behavior *r* =-.05*		
Weekly Activities with Child–Color Naming *r* =-.05*	Overall Activities with Child–Color Naming *r* =-.05*	Weekly Activities with Child–Vocabulary *r* =-.08***	

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

Table 10-8. Significant partial correlations between family activities with parent outcomes

Weekly Activities with Child–Internal Locus of Control r =.08***	Overall Activities with Child–Internal Locus of Control r =.07***	Weekly Activities with Child–Authoritative Parenting r =.15***
Monthly Activities with Child–Authoritative Parenting r =.10***	Overall Activities with Child–Authoritative Parenting r =.15***	

With regard to child cognitive scores, partial correlations identified a number of positive relationships, with weekly activities having the strongest association with cognitive skills. For example, engaging in more activities with their children during the previous week was positively correlated with higher scores for the children on the color naming and vocabulary assessments. Participating in the monthly activities also had small positive correlations with the social awareness, color naming, one-to-one counting, book knowledge, and print concepts assessments.

Parenting Style. FACES 2000 provided the opportunity to look at the correlations of parenting style with selected child cognitive and behavioral outcomes. The parents were scored on two different parenting styles: authoritarian and authoritative. The first of these styles is generally considered a stricter, directive style, while the authoritative style reflects a less harsh style with greater use of rationales or explanations. The differences in these styles were apparent in their relationships with the parent and child outcomes in Table 10-9. As expected, the authoritative and authoritarian scales were negatively correlated with each other.

Table 10-9. Significant partial correlations between parenting style and child and parent outcomes

Authoritative Parenting–Internal Locus of Control r =.09***	Authoritative Parenting–Social Awareness r =.06**	Authoritative Parenting–Prosocial Child Behavior r =.17***	Authoritative Parenting–Emerging Literacy r =.08***
Authoritarian Parenting–Problem Child Behavior r =.06**	Authoritarian Parenting–Aggressive Child Behavior r =.04*	Authoritarian Parenting–Hyperactive Child Behavior r =.04*	Authoritarian Parenting–Withdrawn Child Behavior r =.05*
Authoritarian Parenting–Emerging Literacy r =-.05*	Authoritarian Parenting–Color Naming r =-.05*	Authoritarian Parenting–Teacher-Reported Hyperactive Child Behavior r =.05*	Authoritarian Parenting–Internal Locus of Control r =-.10***

*$p \le .05$; **$p \le .01$, ***$p \le .001$.

Partial correlations, controlling for household income, education, and employment; child gender, age, and ethnicity; and how often the child was read to in the previous week, were used to assess the relationship between parenting style and child behavioral and cognitive outcomes. Parents who scored higher on the authoritative subscale also had higher internal locus of control (i.e., they felt more in control of their lives rather than at the mercy of external factors) and gave their children higher ratings of positive social behavior and emergent literacy. In contrast, the authoritarian subscale scores were positively correlated with the overall ratings of children's problem behavior, as well as with the three problem behavior subscales of aggression, hyperactivity, and withdrawn behavior, and negatively correlated with parents' locus of control.

There were generally weak relationships between parenting style and the cognitive outcomes. The authoritative parenting style scores were positively correlated with children's social awareness and the authoritarian parenting style was negatively correlated with color naming, but neither parenting style was correlated with any of the other cognitive measures. Hyperactive behavior as rated by teachers was positively correlated with an authoritarian parenting style.

Family Support from Head Start. In the Spring of 2001, parents were asked about the ways that they were involved in the Head Start program throughout the past school year. As presented in Figure 10-4, parents most frequently reported attending parent-teacher conferences (79.0 percent), observing in their children's classrooms for at least 30 minutes (74.9 percent), and meeting with a Head Start staff member in their homes (69.4 percent). More than one-half of the parents volunteered in their children's classrooms (60.0 percent) or prepared food or materials for special events (58.1 percent) and slightly less than one-half helped with field trips (42.5 percent), or attended bazaars (42.3 percent) and workshops (42.5 percent). Less than one-fourth of the parents participated in Policy Council (22.5 percent).

A summary score[3] measuring total involvement was created for each parent who responded to the interview. A series of partial correlations, controlling for household income, education, and employment, and child gender, age, and ethnicity, were conducted to examine the relationship between involvement at Head Start and other family and child factors (Table 10-10). Higher levels of Head Start

[3] Summary parent involvement score is based on respondents' reports of how frequently (not yet, 1-2 times, 3 or more times) they participated in each of 10 activities over the past school year. Summary score ranges from 10 to 30, with higher scores representing more involvement. Mean score = 21.4; SD = 5.9.

involvement were significantly related to positive outcomes for the children. Parents who were more involved at Head Start were likely to report more positive social behavior for their children and less aggressive and overall problem behavior, as well as higher emergent literacy skills than parents who were less involved at Head Start. There was also a relationship between involvement and teachers' reports of the children's skills and with the children's scores on cognitive assessments. For example, teachers rated the children of more involved parents as more creative and more socially aware. On the cognitive measures, children whose parents were more involved at Head Start scored higher on vocabulary, book knowledge, early writing, early math, and letter-identification tasks.

Figure 10-4. The frequency of family participation in Head Start activities

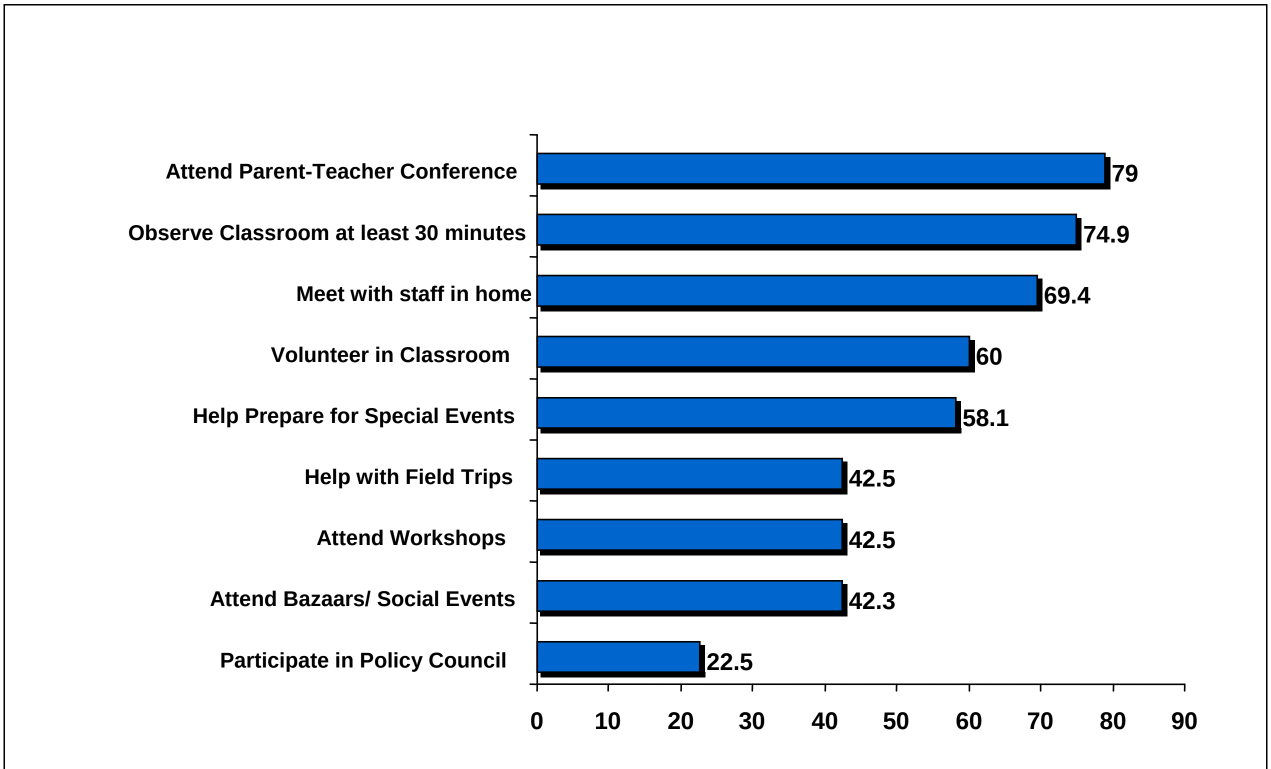


Table 10-10. Significant partial correlations between parental involvement at Head Start and child behavior and child cognitive outcomes

Positive Social Behavior–Parent Report *r* = .09***	Aggressive Behavior–Parent Report *r* = -.08***	Overall Problem Behavior–Parent Report *r* = -.04*	Emergent Literacy–Parent Report *r* = .10***
Creativity–Teacher Report *r* = .04*	Social Awareness–Teacher Report *r* = .06*	Vocabulary–Direct Assessment *r* = .09***	Book Knowledge–Direct Assessment *r* = .07***
Early Writing–Direct Assessment *r* = .06*	Word-Letter ID–Direct Assessment *r* =.11***	Early Math Skills–Direct Assessment *r* =.08***	

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

Consistent with previous findings (ACYF, 2002), most parents reported high levels of satisfaction with the Head Start program. For example, while more than one-half were very satisfied with Head Start in every area, more than 82 percent of the parents reported they were very satisfied with how the program helped their children grow and develop, respected the family's culture, identified child services, maintained a safe program, and prepared the children for kindergarten. There were very few unsatisfied parents, with reports of dissatisfaction generally under 2.5 percent. The area with the greatest degree of reported dissatisfaction was how Head Start helped parents become involved in their community, but even this concern was cited by less than 7 percent of the parents.

Parents also reported that they and their children had positive experiences at Head Start. Almost 85 percent of the parents reported that their children "often" or "always" received individual attention, while more than 90 percent reported their children "often" or "always" felt safe and secure at Head Start, were happy in the program, were treated with respect, and felt accepted by their teachers. With regard to the teachers, more than 90 percent of the parents reported that their children's teachers "often" or "always" were open to new learning, were warm towards the children, were interested in the children, were supportive to parents, were welcoming, did not use harsh discipline, and were happy.

10.2.3 Head Start's Protective Role for Families and Children

Findings from the FACES 1997 cohort of families suggest that Head Start may play an important role in protecting families and children from the challenges that many families face. To validate these findings in the FACES 2000 cohort, a series of linear regression models, controlling for household income, education, and employment; child gender, age, and ethnicity; and parent level of activity with their children, tested whether the effect of parental depression, exposure to violence, domestic violence, alcohol use, or involvement in the criminal justice system on child behavior and child cognitive outcomes varied as a function of (or was moderated by) Head Start satisfaction, experience, or involvement. Summary scores were created for parents' involvement at Head Start, for parents' reports of having a positive experience at Head Start, and parents' reports of satisfaction with the program. Means and standard deviations for these three variables are found in the Appendix.

Table 10-11 presents the interaction terms found to be significant moderators of risk factors in the regression equations. These findings support the earlier results and suggest that Head Start may play an important role in protecting families from the negative outcomes associated with these challenges. For example, there was a direct positive relationship between domestic violence and parent and teacher reports of children's problem behavior. This relationship remained consistent for children of parents who had fewer positive experiences with the program; however, when parents had positive experiences at Head Start, the effects of domestic violence on parent and teacher reports of increased aggressive, hyperactive, withdrawn, and overall problem behavior for children were no longer evident. These findings indicate that the negative effects of being exposed to domestic violence on children were buffered by parents' positive experiences at Head Start.

Table 10-11. Significant Head Start moderators of negative outcomes associated with risk factors

Interaction term (independent variable x moderator)	Outcome variable	Sig
Head Start satisfaction, experience, and involvement as moderators of the negative outcomes for children and families related to parental depression		
Depression x Head Start Satisfaction	Withdrawn Behavior (parent report)	**
Depression x Head Start Involvement	Book Knowledge	*
Depression x Head Start Experience	Early Math	**
Depression x Head Start Experience	Creativity (teacher report)	*
Head Start satisfaction, experience, and involvement as moderators of the negative outcomes for children and families related to exposure to violence		
Interaction term (independent variable x moderator)	Outcome variable	Sig
Violence x Head Start Involvement	Hyperactive Behavior (parent report)	*
Violence x Head Start Involvement	Emergent Literacy (parent report)	**
Violence x Head Start Involvement	Social Awareness	**
Violence x Head Start Involvement	Color Naming	**
Violence x Head Start Experience	One-to-One Counting	*
Head Start satisfaction, experience, and involvement as moderators of the negative outcomes for children and families related to domestic violence		
Interaction term (independent variable x moderator)	Outcome variable	Sig
Domestic Viol x Head Start Experience	Social Skills (parent report)	*
Domestic Viol x Head Start Experience	Total Problem Behavior (parent report)	***
Domestic Viol x Head Start Experience	Aggressive Behavior (parent report)	*
Domestic Viol x Head Start Experience	Withdrawn Behavior (parent report)	**
Domestic Viol x Head Start Satisfaction	Hyperactive Behavior (parent report)	*
Domestic Viol x Head Start Experience	Total Problem Behavior (teacher report)	*
Domestic Viol x Head Start Experience	Social Score (teacher report)	*

* See note at end of table.

Table 10-11. Significant Head Start moderators of negative outcomes associated with risk factors (continued)

Head Start satisfaction, experience, and involvement as moderators of the negative outcomes for children and families related to alcohol use by family member		
Interaction term (independent variable x moderator)	Outcome variable	Sig
Alcohol Use x Head Start Experience	Social Skills (parent report)	**
Alcohol Use x Head Start Satisfaction	Total Problem Behavior (parent report)	*
Alcohol Use x Head Start Experience	Withdrawn Behavior (parent report)	**
Alcohol Use x Head Start Satisfaction	Hyperactive Behavior (parent report)	**
Alcohol Use x Head Start Experience	Total Problem Behavior (teacher report)	**
Alcohol Use x Head Start Experience	Aggressive Behavior (teacher report)	**
Alcohol Use x Head Start Involvement	Emergent Literacy (parent report)	*
Alcohol Use x Head Start Experience	Social Score (teacher report)	*
Head Start satisfaction, experience, and involvement as moderators of the negative outcomes for children and families related to involvement with the criminal justice system		
Interaction term (independent variable x moderator)	Outcome variable	Sig
Crime x Head Start Experience	Social Skills (parent report)	*
Crime x Head Start Experience	Withdrawn Behavior (parent report)	**
Crime x Head Start Involvement	Vocabulary	*

*$p \leq .05$; **$p \leq .01$, ***$p \leq .001$.

The moderating effects of Head Start on negative outcomes for children were consistent across all risk factors presented in Table 10-11, and included moderation of the relationship between risk factors and lower cognitive outcomes, as well as negative child behavior. For example, parents' depression, exposure to violence, or involvement in the criminal justice system were related, in some cases, to poorer cognitive outcomes for children, such as early math skills, book knowledge, color naming, one-to-one counting, and vocabulary. Having a parent who reported positive Head Start experiences or who was more involved with the program moderated this relationship and appeared to protect the children from these negative outcomes. All significant modifying relationships are presented below.

The difficult challenges or risks that parents face—depression, exposure to violence, domestic violence, alcohol use, and involvement with the criminal justice system—are often associated with negative behavioral and cognitive outcomes for their children. The findings cited above provide some evidence that Head Start may play an important role in protecting children and their families from the consequences of these risk factors. Parent involvement at Head Start, parent reports that they and their children had positive experiences at Head Start, and parent satisfaction with the program significantly moderated relationships between these risk factors and many negative child behavior and cognitive outcomes. These findings provide support for the theory that children's school readiness is enhanced when programs work with families as well as with children. It is suggested that studies of subsequent cohorts of FACES families ask teachers to record and report on parents' activities as a means of improving the overall measures of parent involvement in the program.

10.3 Summary

This chapter examined family and parent characteristics that are related to school readiness in early childhood, and parents' important role in preparing children for school. The challenges families face and the strengths they possess are important considerations in understanding how best to prepare children for school. The main findings are summarized below.

- One-quarter of the parents were classified as moderately or severely depressed. Parents who were more depressed reported that their children had more problem behaviors and fewer positive social behaviors, a finding supported by the teachers' reports of children's behavior. Their children also had lower scores on one-to-one counting, creativity, design copying, early writing, letter identification, and early math assessments;

- More than one-fifth of the parents reported they had witnessed violent crime. Five percent reported being victims of violent crime in the neighborhood, while a similar percentage reported being victims of violence in their homes. Almost 10 percent of the children were reported to have witnessed domestic violence during the previous year. Less than 2 percent of the children were reported to have been victims of violent crime or victims of domestic violence. Positive correlations were found between increased exposure to neighborhood violence and reports of child problem behavior, while children in more violent neighborhoods had lower assessment scores on the color naming and book knowledge assessments;

- Almost 13 percent of the parents indicated that they have been victims of domestic violence. Teachers and parents reported children in these families had more overall problem behaviors. However, no direct relationship was found between parent reports of domestic violence and child cognitive outcomes;

- Almost one-half of the Head Start children lived in households with at least one individual who smoked cigarettes and about two-fifths of the households reported having at least one individual who drank alcohol. Living in a household with someone who drinks increased the risk of parental depression, while the children in these homes were reported to have more overall problem behavior and scored lower on vocabulary, color naming, and social awareness assessments;

- Almost one-fifth of the parents reported that someone in their household had been arrested and charged with a crime. Children in these families were more than three times more likely to have been a witness to violent crime or domestic violence in the past year. These children were also three times more likely to have been a victim of domestic violence or violent crime. These children had lower vocabulary scores, and were reported by both parents and teachers to be more aggressive and have more overall problem behaviors;

- At least one of a set of selected risk factors was evident in over 90 percent of the families. Almost one-quarter of the families had four or more risk factors. Children in these families had lower parent ratings on emergent literacy and higher teacher and parent ratings of problem behavior. In the assessments, these children scored lower on design copying, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification;

- Families engaged their children in a number of weekly and monthly activities. The number of activities was positively correlated with positive child behaviors and emergent literacy and negatively correlated with problem behaviors. In particular, the weekly activities had positive correlations with scores on the social awareness, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification tasks;

- Higher authoritative parenting style scores were significantly positively correlated with children's social awareness, but not with any of the other cognitive measures. On the other hand, higher authoritarian scores were significantly negatively correlated with comprehension, color naming, vocabulary, and early math assessments;

- More than two-thirds of parents had attended parent-teacher conferences, observed in their children's classrooms for at least 30 minutes, or met with a Head Start staff member in their homes. Parent involvement in Head Start was positively correlated with parental reports of positive social behavior and higher emergent literacy skills and negatively correlated with aggressive and overall problem behavior. Children with more involved parents scored higher on vocabulary, book knowledge, early writing, early math, and letter identification tasks; and

- Head Start may play an important role in protecting families and children from the challenges that low-income families face. Parent involvement at Head Start, parent reports that they and their children had positive experiences at Head Start, or parent satisfaction with the program significantly moderated relationships between risk factors (e.g., parental depression, exposure to violence and domestic violence, substance use, and involvement with the criminal justice system) and many negative child behavior and lower cognitive outcomes.

CHAPTER 11.
PREDICTIVE VALIDITY OF COGNITIVE AND BEHAVIORAL MEASURES: RELATIONSHIPS WITHIN AND ACROSS COGNITIVE AND SOCIAL DEVELOPMENTAL DOMAINS

The child assessment measures for the Family and Child Experiences Survey (FACES) was designed to provide a comprehensive description of Head Start children's development in both the cognitive and social-emotional domains. As an indicator of the validity of the FACES instruments, criterion-related validity, that is the degree to which the test or questionnaire correlates with one or more outcome criteria, was assessed. Systematic evidence for criterion validity is often described in terms of *predictive* validity. Predictive validity is assessed when the outcome criterion is measured at a later time point after the evaluated test (e.g., later school year). This chapter evaluates how well the FACES measures predict children's achievement and school adjustment at the end of their kindergarten year. The analyses presented in this chapter are conducted with data from FACES measures administered in Spring 2001 of the Head Start year and on kindergarten data from Spring 2002. All of the described analyses were conducted on children who were assessed in English at all timepoints.

11.1 Research Questions

This chapter will address the following research questions:

- How well do the FACES measures predict children's early reading skills and general knowledge at the end of kindergarten?
- How well do the FACES parent and teacher ratings of children's social competence predict children's school adjustment at the end of kindergarten?
- Do the FACES parent and teacher ratings of children's social competence contribute to the FACES measures' prediction of children's reading skills and general knowledge at the end of kindergarten?
- How well do cognitive gains during Head Start predict reading and general knowledge at the end of kindergarten?
- How well do changes in behavior ratings during Head Start predict social competence at the end of kindergarten? and
- Do FACES cognitive and behavior measures predict promotion to first grade?

11.2 Findings

11.2.1 Children's Scores on the FACES Instruments at the End of Head Start Predict Kindergarten Outcomes

The FACES cognitive measures were designed to assess preliteracy skills, as well as general school readiness. In order to determine the predictive validity of these measures, relationships between the FACES assessment scale scores obtained during the Head Start year and the Reading and General Knowledge scale scores obtained at the end of kindergarten were examined by two approaches. In the first approach, Reading and General Knowledge scale scores from the end of the kindergarten year were correlated with the scores from each of the FACES subtests from the end of the Head Start year. The second approach assessed the ability of the FACES scale scores from the end of the Head Start year to predict Reading and General Knowledge scale scores at the end of the kindergarten year, controlling for child's age at time of assessment (i.e., the end of their Head Start year), in two multiple regression analyses.

Children's scores on each of the component tasks in the FACES measures at the end of Head Start were significantly positively correlated with their Reading scale scores at the end of kindergarten (Table 11-1). Bivariate correlations with the Reading scale ranged from .51 (for the Woodcock-Johnson-Revised Letter-Word Identification or WJ-R LWI) to .26 (for the WJ-R Draw-a-Design). These moderate-to-high correlations indicate that the FACES measures may have predictive power on outcome criteria at later time points. Similarly, children's scores on each of the component tasks in the FACES battery at the end of Head Start also correlated significantly with their scores on the General Knowledge scale at the end of kindergarten. Bivariate correlations with the General Knowledge scale ranged from .61 (for the Peabody Picture Vocabulary Test-III or PPVT-III) to .18 (for the WJ-R LWI).

Table 11-1. Pearson's product moment correlation coefficients between FACES assessment scores at end of Head Start and Reading and General Knowledge scale scores at end of kindergarten

	Correlation coefficients	
Assessment score	Reading	General Knowledge
PPVT-III	.36	.61
WJR Applied Problems	.39	.48
WJR Dictation	.43	.26
WJR Letter-Word ID	.51	.18
Social Awareness	.28	.20
Draw-A-Design	.26	.19
Color Naming	.35	.27
One-to-One Counting	.31	.23
Book Knowledge	.29	.37

Note: Ns for bivariate correlations with Kindergarten Reading scale scores range from 1,394-1,511. Ns for bivariate correlations with Kindergarten General Knowledge scale score range from 1,386-1,501. All correlation coefficients are significant at $p < .0001$.

When the subtest scores were combined in a multiple regression model, the model did quite well at predicting children's early reading skills at the end of kindergarten, accounting for 34 percent of the variance in Reading scale scores. The best predictor of Reading scale scores was the WJ-R LWI task (beta = .30), which also showed the highest bivariate correlation with the Reading scale (Table 11-2). The model predicting General Knowledge scale scores accounted for 41 percent of the variance in General Knowledge scale scores. The best predictor in this multiple regression was the PPVT-III (beta = .50), which also showed the highest bivariate correlation with the General Knowledge scale.

Table 11-2. Summary of multiple linear regression analysis for FACES assessment scores at end of Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($N_{Reading}$ = 1,282; $N_{General\ Knowledge}$ = 1,275)

	Reading scale			General Knowledge scale		
Assessment score	B	SE B	Beta	B	SE B	Beta
PPVT-III	.13	.03	.13****	.41	.02	.50****
WJR Applied Problems	.03	.02	.06	.05	.01	.13****
WJR Dictation	.04	.01	.13***	.00	.01	.02
WJR Letter-Word ID	.16	.02	.30****	-.04	.01	-.09**
Social Awareness	.33	.19	.04	.30	.15	.05*
Draw-A-Design	.27	.14	.05	.03	.11	.01
Color Naming	.31	.07	.12	.07	.05	.03
One-to-One Counting	.23	.21	.03	-.01	.16	-.00
Book Knowledge	.05	.21	.01	.68	.16	.11****
Child's Age at Head Start Assessment	-.18	.05	-.09***	-.12	.04	-.07**

Note: $R^2 = .34$ for Reading Scale; $R^2 = .41$ for General Knowledge; ($ps < .0001$).

*$p < .05$; **$p < .01$; ***$p < .001$; ****$p < .0001$.

The Book Knowledge score was not a significant predictor of the Reading scale score in its multiple regression model, though scores on this task correlated significantly with the Reading score ($\underline{r}$ = .29). It is noteworthy to also mention that the best predictor of the Reading score, namely WJ-R LWI, had a very small, and counterintuitive, relationship with General Knowledge in the multiple regression analysis.

Although these findings are not as strong as what has been found in similar analyses with FACES 1997 data (ACF, 2003), they do support the notion that the Reading and General Knowledge assessments may be tapping two distinct clusters of skills, both of which have been shown to be important for children's future reading proficiency and academic achievement. The skills tapped by the Reading scale during kindergarten were primarily what Whitehurst and Lonigan (1998) have called "inside-out" skills such as letter recognition, letter-sound association, and word decoding. The skills tapped by the General Knowledge assessment were "outside-in" skills, such as general information, word knowledge, and conceptual understanding that children need to help them comprehend and evaluate what they read and relate it to facts and concepts they have previously acquired. The FACES battery showed its validity by predicting well children's later learning in both skill domains. In addition, these results suggest that development in both skill domains should receive attention in preschool curricula and practice in order to foster both types of school-age abilities.

Although both domains were predicted well, the combination of subtests that produced the best forecasts differed across the two skill clusters. The LWI test was the best predictor of inside-out skills, with Dictation, Color Naming, and PPVT-III contributing additional predictive power. The PPVT-III was by far the best predictor of outside-in skills, with Book Knowledge, Applied Problems, Social Awareness, and LWI showing much smaller but significant regression coefficients as well.

Predictive Validity of Abbreviated FACES Battery. In order to further explore the validity of the FACES battery, multivariate regression analyses were carried out with the set of norm-referenced tests (i.e., PPVT-III and the WJ-R subtests) at the end of Head Start predicting the Reading and General Knowledge scale at the end of kindergarten, controlling for the child's age at the end of the Head Start assessment. These analyses were repeated with the set of criterion-referenced measures (i.e., Social Awareness, McCarthy Draw-a-Design, Color Names, One-to-One Counting, and Book Knowledge) as the predictors.

The four norm-referenced tests at the end of Head Start did almost as well as the full battery at predicting the Reading scores at the end of kindergarten (Table 11-3), predicting 32 percent of the variance. The LWI was most closely associated with the Reading scores. Applied Problems, Dictation, and PPVT-III had significant coefficients as well. The four norm-referenced tests at the end of Head Start also did as well as the full battery at predicting the General Knowledge scores at the end of kindergarten, predicting 40 percent of the variance. The PPVT-III was most closely associated with the General Knowledge scores. Applied Problems also had a significant regression coefficient, but Dictation and LWI did not.

Table 11-3. Summary of multiple linear regression analysis for norm reference FACES assessment scores at end of Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($\underline{N}_{Reading}$ = 1,315; $\underline{N}_{General\ Knowledge}$ = 1,308)

	Reading scale			General Knowledge scale		
Assessment score	B	SE B	Beta	B	SE B	Beta
PPVT-III	.15	.03	.15****	.43	.02	.53****
WJR Applied Problems	.05	.01	.11***	.07	.01	.18****
WJR Letter-Word ID	.19	.02	.34****	-.02	.01	-.05
WJR Dictation	.05	.01	.16****	.01	.01	.03
Child's Age at Head Start Assessment	-.17	.05	-.08**	-.15	.04	-.09***

Note: R^2 = .32 for Reading Scale; R^2 = .40 for General Knowledge; ($\underline{p}s$ < .0001).

*$\underline{p}$ < .05; **$\underline{p}$ < .01; ***$\underline{p}$ < .001; ****$\underline{p}$ < .0001.

The five criterion-referenced FACES tasks significantly predicted the Reading and General Knowledge assessments at the end of kindergarten, though notably less well than either the full battery or the abbreviated battery composed of norm-referenced tests, predicting 20 percent and 18 percent of the respective variances (Table 11-4). All five subtests were significant predictors of the Reading score. As with the full battery, tasks that tapped perceptual-motor skills and knowledge of print conventions were more closely related to kindergarten Reading scores than were tasks that tapped vocabulary or general information. In the model predicting General Knowledge, four of the five subtests had significant regression coefficients (One-to-One Counting was the only measure with a non-significant regression coefficient). The pattern of relative magnitudes of the significant regression coefficients was different than that for Reading. Book Knowledge was most closely related with General Knowledge.

Table 11-4. Summary of multiple linear regression analysis for criterion reference FACES assessment scores at end of Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($N_{Reading}$ = 1,469; $N_{General\ Knowledge}$ = 1,459)

Assessment score	Reading scale			General Knowledge scale		
	B	SE B	Beta	B	SE B	Beta
Social Awareness	.97	.18	.13****	.41	.15	.07**
Draw-A-Design	.69	.13	.13****	.26	.11	.06*
Color Naming	.41	.06	.19****	.19	.05	.11***
One-to-One Counting	.93	.20	.12****	.31	.17	.05
Book Knowledge	1.00	.19	.13	1.71	.16	.28****
Child's Age at Head Start Assessment	-.09	.04	-.05	.04	.04	.03

Note: R^2 = .20 for Reading Scale; R^2 = .18 for General Knowledge; (ps < .0001).

*p < .05; **p < .01; ***p < .001; ****p < .0001.

The combination of the four norm-referenced tests at the end of Head Start that were more extended (and hence more reliable in terms of internal consistency) did a better job of predicting children's kindergarten achievement than the combination of the five FACES criterion-referenced tasks. However, further analyses revealed that the five criterion-referenced measures provide significant unique contributions to the prediction of the Reading scale score, over and above the contributions of the norm-referenced tests, increasing the predictive power of the assessment battery by 1 percent (Table 11-5). The criterion-referenced measures also provided significant unique contributions to the prediction of the General Knowledge scale score increasing the predictive power by 2 percent. These results indicate that the criterion-referenced measures significantly contribute to the assessment battery by picking up variance not accounted for by the norm-referenced tests.

11.2.2 FACES Behavioral Ratings at the End of Head Start Predict Children's Social Competence in Kindergarten

Social competence is an important developmental domain measured by the FACES battery. Two sources of information are tapped for assessing children's social competencies by collecting ratings of the children's behavior from their teachers and parents. These analyses examine the ability of the teacher and parent ratings of children's social competencies during Head Start to predict children's school adjustment at the end of kindergarten. Analyses of the predictive validity of the behavior ratings mirror those for the cognitive measures. First, teacher ratings of cooperative classroom behavior and total problem behaviors from the end of the kindergarten year were correlated with the behavior ratings from teachers and parents at the end of the Head Start year. Then the teacher and parent ratings were combined in multiple regressions predicting teacher ratings of cooperative classroom behavior and total problem behaviors at the end of the kindergarten year, controlling for child's age at the end of the Head Start assessment, in two multiple regression analyses.[1]

Table 11-5. Summary of hierarchical regression analysis for FACES assessment scores at end of Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($N_{Reading}$ = 1,282; $N_{General\ Knowledge}$ = 1,275)

	Reading			General Knowledge		
Assessment score	B	SE B	Beta	B	SE B	Beta
Step 1						
PPVT-III	.15	.03	.15****	.43	.02	.53****
WJR Applied Problems	.05	.01	.11***	.07	.01	.18****
WJR Letter-Word ID	.19	.02	.34****	-.02	.01	-.05
WJR Dictation	.05	.01	.16****	.01	.01	.03
Child's Age at Head Start Assessment	-.17	.05	-.08**	-.15	.04	-.09***
Step 2						
PPVT-III	.13	.03	.13****	.41	.02	.50****
WJR Applied Problems	.03	.02	.06	.05	.01	.13****
WJR Letter-Word ID	.04	.01	.13****	.00	.01	.02
WJR Dictation	.16	.02	.30****	-.04	.01	-.09**
Social Awareness	.33	.19	.04	.30	.15	.05*
Draw-A-Design	.27	.14	.05	.03	.11	.01
Color Naming	.31	.07	.12****	.07	.05	.03
One-to-One Counting	.23	.21	.03	-.01	.16	-.00
Book Knowledge	.05	.21	.01	.68	.16	.11****
Child's Age at Head Start Assessment	-.18	.05	-.09***	-.12	.04	-.07**

Note: Reading: R^2 = .32 for Step 1 (p < .0001); ΔR^2 = .01 for Step 2 (p < .001). General Knowledge: R^2 = .40 for Step 1 (p < .0001); ΔR^2 = .02 for Step 2 (p < .001).

*p < .05; **p < .01; ***p < .001; ****p < .0001.

All 10 parent and teacher ratings of behavior at the end of Head Start were significantly correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten (Table 11-6). Correlations were significant in the expected directions. Problem behaviors as rated by both parents and teachers at the end of Head Start had significant negative correlations with teacher ratings of cooperative classroom behavior at the end of kindergarten, ranging from -.33 (for teacher reported ratings of total behavior problems) to -.06 (for parent reported ratings of withdrawn behavior). Kindergarten teacher

[1] In all multiple regression analyses involving the behavior ratings, teacher and parent ratings of total problem behavior were excluded as predictor variables from the models. Because these ratings are summative scores of the ratings for aggression, withdrawn behavior, and hyperactivity, including them would introduce multicollinearity among these predictor variables to the model. They are included in the bivariate correlation analyses.

ratings of cooperative classroom behavior were positively correlated with teacher ratings of cooperative classroom behavior (r = .33) and parent ratings of positive approaches to learning (r = .11) at the end of Head Start. These moderate-sized correlation coefficients indicate that the FACES measures may have predictive power on kindergarten outcomes. In general, teacher ratings showed stronger relationships than the parent ratings did with the kindergarten outcomes.

Table 11-6. Pearson's product moment correlation coefficients between FACES behavior ratings at end of Head Start and teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten

	Correlation coefficients	
Behavior rating	Cooperative classroom behavior	Total problem behaviors
Teacher Reported Aggression	-.31	.39
Teacher Reported Hyperactivity	-.32	.41
Teacher Reported Withdrawal	-.23	.27
Teacher Reported Total Problem Behaviors	-.33	.41
Teacher Reported Cooperative Classroom Behavior	.33	-.34
Parent Reported Aggression	-.19	.22
Parent Reported Hyperactivity	-.23	.26
Parent Reported Withdrawal	-.06*	.10**
Parent Reported Total Problem Behaviors	-.22	.26
Parent Reported Positive Approaches to Learning	.11***	-.15

Note: Ns for correlations with Kindergarten Teacher Ratings of Cooperative Classroom Behavior range from 1,146-1,191. Ns for correlations with Kindergarten Teacher Ratings of Total Problem Behavior range from 1,147-1,193.

All p-values < .0001 except for *p < .05; **p < .01; ***p < .001.

Similarly, all 10 parent and teacher ratings of behavior at the end of Head Start were moderately correlated with teacher ratings of total problem behaviors at the end of kindergarten (Table 11-6). Correlations were significant in the expected directions. Problem behaviors as rated by both parents and teachers at the end of Head Start had significant positive correlations with teacher ratings of total problem behaviors at the end of kindergarten, ranging from .41 (for teacher reported ratings of hyperactive behavior and total problem behaviors) to .10 (for parent reported ratings of withdrawn behavior). Teacher ratings of cooperative classroom behavior (r = -.34) and parent ratings of positive approaches to learning (r = -.15) were both negatively correlated with teacher ratings of total problem behavior at the end of kindergarten. In general, teacher ratings showed stronger relationships than the parent ratings did with the kindergarten outcomes.

When the ratings were combined in a multiple regression, the model accounted for 14 percent of the variance in kindergarten teacher ratings of cooperative classroom behavior (Table 11-7). The best predictor in the multiple regression was teacher-reported ratings of cooperative classroom behavior (beta = .15). The model predicting kindergarten teacher ratings of total problem behavior accounted for 21 percent of the variance in kindergarten teacher ratings of total problem behavior. The best predictor in the multiple regression was teacher-reported ratings of aggressive behavior (beta = .19).

Table 11-7. Summary of multiple linear regression analysis for FACES behavior ratings at end of Head Start predicting teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten ($N_{Cooperative\ Classroom\ Behavior}$ = 1,060; $N_{Total\ Problem\ Behaviors}$ = 1,061)

	Cooperative classroom behavior			Total problem behaviors		
Behavior rating	B	SE B	Beta	B	SE B	Beta
Teacher Reported Aggression	-.21	.10	-.09*	.49	.10	.19****
Teacher Reported Hyperactivity	-.40	.14	-.13**	.61	.15	.17****
Teacher Reported Withdrawal	.01	.07	.00	.06	.08	.03
Teacher Reported Cooperative Classroom Behavior	.16	.04	.15***	-.07	.05	-.06
Parent Reported Aggression	-.15	.09	-.05	.15	.10	.05
Parent Reported Hyperactivity	-.33	.11	-.10**	.37	.12	.10**
Parent Reported Withdrawal	.02	.16	.00	.13	.17	.02
Parent Reported Positive Approaches to Learning	-.05	.09	-.02	-.06	.09	-.02
Child's Age at Head Start Assessment	-.01	.02	-.01	-.01	.03	-.02

Note: R^2 = .14 for Teacher Ratings of Cooperative Classroom Behavior; R^2 = .21 for Teacher Ratings of Total Problem Behaviors; (ps < .0001).
*p < .05; **p < .01; ***p < .001; ****p < .0001.

Predictive Validity of Abbreviated Sets of Behavior Ratings. In order to further explore the validity of the behavior ratings, multivariate regression analyses were carried out with the set of teacher-reported ratings at the end of Head Start predicting the teacher-reported ratings of cooperative classroom behavior and total problem behavior at the end of kindergarten.

The teacher-reported ratings at the end of Head Start did almost as well as the full set of behavior ratings at predicting the teacher-reported ratings of cooperative classroom behavior at the end of kindergarten (Table 11-8). The model explained 14 percent of the variance in kindergarten cooperative classroom behavior ratings, and the rating of cooperative classroom behavior at the end of Head Start was most closely associated with the teacher ratings at the end of kindergarten (beta = .18). The ratings of hyperactive (beta = -.15) and aggressive behavior (beta = -.11) both had significant regression coefficients as well, but the ratings of withdrawn behavior did not.

Table 11-8. Summary of multiple linear regression analysis for FACES teacher ratings at end of Head Start predicting teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten
($N_{Cooperative\ Classroom\ Behavior}$ = 1,125; $N_{Total\ Problem\ Behaviors}$ = 1,126)

	Cooperative classroom behavior			Total problem behaviors		
Behavior rating	B	SE B	Beta	B	SE B	Beta
Teacher Reported Aggression	-.25	.09	-.11**	.50	.10	.19****
Teacher Reported Hyperactivity	-.47	.14	-.15***	.74	.15	.20****
Teacher Reported Withdrawal	.00	.07	.00	.04	.08	.02
Teacher Reported Cooperative Classroom Behavior	.18	.04	.18****	-.12	.04	-.10**
Child's Age at Head Start Assessment	-.01	.02	-.02	-.01	.03	-.01

Note: R^2 = .14 for Teacher Ratings of Cooperative Classroom Behavior; R^2 = .20 for Teacher Ratings of Total Problem Behaviors; ($ps < .0001$).
*$p < .05$; **$p < .01$; ***$p < .001$; ****$p < .0001$.

Similarly, the teacher-reported ratings at the end of Head Start did almost as well as the full set of behavior ratings at predicting the teacher-reported ratings of total problem behavior at the end of kindergarten. The model explained 20 percent of the variance in kindergarten total problem behavior ratings, and the rating of hyperactive behavior was most closely associated with the total problem behavior ratings (beta = .20). The ratings of cooperative classroom behavior (beta = -.10) and aggressive behavior (beta = .19) both had significant regression coefficients as well, but the ratings of withdrawn behavior did not.

The four parent-reported behavior ratings significantly predicted the teacher ratings of cooperative classroom behavior at the end of kindergarten, though notably less well than either the full set of behavior ratings or the set of teacher-reported ratings (Table 11-9). The model explained 6 percent of the variance in teacher-reported cooperative classroom behavior and the strongest predictor was ratings of hyperactive behavior (beta = -.18), followed by ratings of aggressive behavior (beta = -.10). Parent ratings of positive approaches to learning and withdrawn behavior did not have significant regression coefficients.

Table 11-9. Summary of multiple linear regression analysis for FACES parent ratings at end of Head Start predicting teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten ($\underline{N}_{\text{Cooperative Classroom Behavior}}$ = 1,166; $\underline{N}_{\text{Total Problem Behaviors}}$ = 1,168)

Behavior rating	Cooperative classroom behavior			Total problem behaviors		
	B	SE B	Beta	B	SE B	Beta
Parent Reported Aggression	-.27	.09	-.10**	.32	.10	.11**
Parent Reported Hyperactivity	-.59	.11	-.18****	.74	.12	.20***
Parent Reported Withdrawal	.19	.15	.04	-.09	.17	-.02
Parent Reported Positive Approaches to Learning	.07	.08	.03	-.19	.09	-.06*
Child's Age at Head Start Assessment	.04	.02	.05	-.06	.03	-.06*

Note: R^2 = .06 for Teacher Ratings of Cooperative Classroom Behavior; R^2 = .09 for Teacher Ratings of Total Problem Behaviors; ($\underline{p}$s < .0001).

*$\underline{p}$ < .05; **$\underline{p}$ < .01; ***$\underline{p}$ < .001; ****$\underline{p}$ < .0001.

The four parent-reported behavior ratings also significantly predicted the teacher ratings of total problem behavior at the end of kindergarten, though notably less well than either the full set of behavior ratings or the set of teacher-reported ratings. The model explained 9 percent of the variance in kindergarten teacher-reported total problem behavior and parent-reported ratings of hyperactive behavior had the largest regression coefficient (beta = .20), followed by ratings of aggressive behavior (beta = .11) and ratings of positive approaches to learning (beta = -.06). Parent ratings of withdrawn behavior did not have a significant regression coefficient.

The combination of the four teacher-reported behavior ratings at the end of Head Start did a better job of predicting children's kindergarten behavior than the combination of the four parent-reported behavior ratings (Table 11-10). When added to the regression model, the four parent ratings provided unique contributions at the trend level (*p* < .10) to the prediction of the cooperative classroom behavior ratings, over and above the contributions of the teacher ratings, increasing the predictive power of the assessment battery by 1 percent. These results indicate that the parent ratings provide some unique contributes to the assessment battery.

The four parent ratings also provided significant unique contributions to the prediction of the total problem behavior ratings, over and above the contributions of the teacher ratings, increasing the predictive power of the assessment battery by almost 2 percent. These results further indicate that the parent ratings significantly contribute to the assessment battery, by picking up significant variance not accounted for by the teacher-reported ratings.

Table 11-10. Summary of hierarchical regression analysis for FACES behavior ratings at end of Head Start predicting teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten ($\underline{N}_{\text{Cooperative Classroom Behavior}}$ = 1,060; $\underline{N}_{\text{Total Problem Behaviors}}$ = 1,061)

	Cooperative classroom behavior			Total problem behaviors		
Behavior rating	B	SE B	Beta	B	SE B	Beta
Step 1						
Teacher Reported Aggression	-.25	.09	-.11**	.50	.10	.19****
Teacher Reported Hyperactivity	-.47	.14	-.15***	.74	.15	.20****
Teacher Reported Withdrawal	.00	.07	.00	.04	.08	.02
Teacher Reported Cooperative Classroom Behavior	.18	.04	.18****	-.12	.04	-.10**
Child's Age at Head Start Assessment	-.01	.02	-.02	-.01	.03	-.01
Step 2						
Teacher Reported Aggression	-.21	.10	-.09*	.49	.10	.19****
Teacher Reported Hyperactivity	-.40	.14	-.13**	.61	.15	.17****
Teacher Reported Withdrawal	.01	.07	.00	.06	.08	.03
Teacher Reported Cooperative Classroom Behavior	.16	.04	.15***	-.07	.05	-.06
Parent Reported Aggression	-.15	.09	-.05	.15	.10	.05
Parent Reported Hyperactivity	-.33	.11	-.10**	.37	.12	.10**
Parent Reported Withdrawal	.02	.16	.00	.13	.17	.02
Parent Reported Positive Approaches to Learning	-.05	.09	-.02	-.06	.09	-.02
Child's Age at Head Start Assessment	-.01	.02	-.01	-.01	.03	-.02

Note: Teacher Ratings of Cooperative Classroom Behavior: R^2 = .14 for Step 1 ($\underline{p}$ < .0001); ΔR^2 = .01 for Step 2 ($\underline{p}$ < .10). Teacher Ratings of Total Problem Behaviors: R^2 = .20 for Step 1 ($\underline{p}$ < .0001); ΔR^2 = .02 for Step 2 ($\underline{p}$ < .05).

*$\underline{p}$ < .05; **$\underline{p}$ < .01; ***$\underline{p}$ < .001; ****$\underline{p}$ < .0001.

11.2.3 Behavior Ratings at the End of Head Start Predict Reading Skills and General Knowledge at the End of Kindergarten

Given that certain positive behaviors may foster learning, while other negative behaviors may impede learning, the ability of these behavior ratings obtained during the Head Start year to predict Reading and General Knowledge scale scores obtained at the end of kindergarten was also examined.

Eight of the ten parent and teacher ratings of behavior at the end of Head Start correlated significantly with Reading scale scores at the end of kindergarten, indicating that the FACES measures may have some predictive power on outcome criteria at later time points. The parent-reported ratings of positive approaches to learning and withdrawn behavior were correlated with kindergarten Reading scores at the trend level ($p < .10$). Significant correlations were in the expected directions. Ratings of problem behaviors were negatively correlated with kindergarten Reading scale scores, suggesting that behaviors that may impede learning are associated with lower reading skills in kindergarten (Table 11-11).

Table 11-11. Pearson's product moment correlation coefficients between FACES behavior ratings at end of Head Start and Reading and General Knowledge scale scores at end of kindergarten

	Correlation coefficients	
Behavior rating	Reading	General Knowledge
Teacher Reported Aggression	-.14***	-.12***
Teacher Reported Hyperactivity	-.21***	-.16***
Teacher Reported Withdrawal	-.27***	-.25***
Teacher Reported Total Problem Behaviors	-.25***	-.22***
Teacher Reported Cooperative Classroom Behavior	.20***	.19***
Parent Reported Aggression	-.08**	-.06*
Parent Reported Hyperactivity	-.12***	-.10***
Parent Reported Withdrawal	NS	NS
Parent Reported Total Problem Behaviors	-.12***	-.09**
Parent Reported Positive Approaches to Learning	NS	.06*

Note: Ns for correlations with Reading Scale Scores range from 1,395-1,453. Ns for correlations with General Knowledge Scale Scores range from 1,387-1,443.

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

Significant correlations were also found between children's scores on 9 of the 10 parent- and teacher-reported behavior ratings at the end of Head Start and General Knowledge scale scores at the end of kindergarten (the association with parent ratings of withdrawn behavior was not significant). Significant bivariate correlations between parent teacher ratings and General Knowledge scale scores were in expected directions and ranged in absolute value from .25 (for teacher ratings of withdrawn behavior) to .06 (for parent ratings of positive approaches to learning behavior), again indicating that the FACES measures may have some predictive power on outcome criteria at later time points. Ratings of problem behaviors were negatively correlated with kindergarten General Knowledge scale scores, again suggesting that behaviors that may impede learning are associated with lower skills in the natural sciences and social studies in kindergarten.

When the behavior ratings were combined in a multiple regression model, the model accounted for 9 percent of the variance in Reading scale scores (Table 11-12). However, the only significant predictor in the multiple regression was teacher-reported ratings of withdrawn behavior (beta = -.22) which also showed the strongest bivariate correlation with reading scores (*r* = -.27).

Table 11-12. Summary of multiple linear regression analysis for FACES behavior ratings at end of Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($\underline{N}_{Reading}$ = 1,283; $\underline{N}_{General\ Knowledge}$ = 1,277)

	Reading			General knowledge		
Behavior rating	B	SE B	Beta	B	SE B	Beta
Teacher Reported Aggression	.09	.19	.02	-.07	.15	-.02
Teacher Reported Hyperactivity	-.36	.28	-.05	.25	.22	.05
Teacher Reported Withdrawal	-.88	.14	-.22*	-.72	.11	-.22*
Teacher Reported Cooperative Classroom Behavior	.04	.09	.02	.08	.07	.04
Parent Reported Aggression	.03	.18	.01	.05	.15	.01
Parent Reported Hyperactivity	-.42	.22	-.06	-.38	.18	-.07*
Parent Reported Withdrawal	.06	.30	.01	.66	.24	.08**
Parent Reported Positive Approaches to Learning	-.10	.17	-.02	.06	.14	.01
Child's Age at Head Start Assessment	.21	.05	.12*	.21	.04	.15*

Note: R^2 = .09 for Reading; R^2 = .09 for General Knowledge; (*ps* < .0001).

*p < .0001.

When the behavior ratings were combined in a multiple linear regression model predicting General Knowledge, the model accounted for 9 percent of the variance in General Knowledge scale scores. The strongest predictor in the multiple regression was teacher-reported ratings of withdrawn behavior (beta = -.22), which also had the strongest bivariate correlation.

FACES Behavior Ratings Contribute to the Prediction of General Knowledge at the End of Kindergarten. In order to further examine the usefulness of the behavior ratings, the unique contribution of the set of behavior rating measures to the prediction of kindergarten outcomes was assessed (Table 11-13). The eight behavior ratings did significantly provide unique contributions to the prediction of the General Knowledge scale score, over and above the cognitive assessments, increasing the predictive power of the assessment battery by 2 percent. Further, the eight behavior ratings did provide unique contributions at the trend level (*p* < .10) to the prediction of the Reading scale score, over and above the contributions of the cognitive tests, increasing the predictive power of the assessment battery by almost 1 percent. This indicates that the behavior ratings provide some unique contributions to the prediction of kindergarten outcomes.

Table 11-13. Summary of hierarchical regression analysis for FACES behavior ratings at end of Head Start predicting Reading and General Knowledge scale scores at the end of kindergarten ($N_{Reading}$ = 1,105; $N_{General\ Knowledge}$ = 1,100)

Assessment score/behavior rating	Reading			General Knowledge		
	B	SE B	Beta	B	SE B	Beta
Step 1						
PPVT-III	.13	.03	.13****	.41	.02	.50****
WJR Applied Problems	.03	.02	.06	.05	.01	.13****
WJR Letter-Word ID	.04	.01	.13****	.00	.01	.02
WJR Dictation	.16	.02	.30****	-.04	.01	-.09**
Social Awareness	.33	.19	.04	.30	.15	.05*
Draw-A-Design	.27	.14	.05	.03	.11	.01
Color Naming	.31	.07	.12****	.07	.05	.03
One-to-One Counting	.23	.21	.03	-.01	.16	-.00
Book Knowledge	.05	.21	.01	.68	.16	.11****
Child's Age at Head Start Assessment	-.18	.05	-.09***	-.12	.04	-.07**
Step 2						
PPVT-III	.13	.03	.13****	.41	.02	.51****
WJR Applied Problems	.02	.02	.05	.05	.01	.13***
WJR Dictation	.03	.01	.11**	.00	.01	.00
WJR Letter-Word ID	.16	.02	.30****	-.04	.01	-.10**
Social Awareness	.13	.21	.02	.32	.16	.05*
Draw-A-Design	.30	.15	.06*	.05	.11	.01
Color Naming	.27	.08	.10***	.07	.06	.03
One-to-One Counting	.36	.22	.05	-.13	.17	-.02
Book Knowledge	.07	.22	.01	.68	.17	.11****
Teacher Reported Aggression	-.08	.17	-.02	-.05	.13	-.01
Teacher Reported Hyperactivity	-.26	.26	-.04	.12	.19	.02
Teacher Reported Withdrawal	-.42	.13	-.10**	-.40	.10	-.12****
Teacher Reported Cooperative Classroom Behavior	-.14	.08	-.06	-.06	.06	-.03
Parent Reported Aggression	-.07	.17	-.01	-.11	.13	-.02
Parent Reported Hyperactivity	.10	.20	.01	-.11	.15	-.02
Parent Reported Withdrawal	-.22	.28	-.02	.27	.21	.03
Parent Reported Positive Approaches to Learning	-.32	.16	-.05	-.01	.12	-.11
Child's Age at Head Start Assessment	-.16	.06	-.08**	-.12	.04	-.07**

Note: Reading: R^2 = .34 for Step 1 (p < .0001); ΔR^2 = .01 for Step 2 (p < .10). General Knowledge: R^2 = .41 for Step 1 (p < .0001); ΔR^2 = .02 for Step 2 (p < .05).

*p < .05; **p < .01; ***p < .001; ****p < .0001.

11.2.4 Head Start Fall-to-Spring Gain Scores from the FACES Battery Predict Kindergarten Outcomes

As a further assessment of the predictive validity of the cognitive measures from the FACES battery, the predictive validity of their Fall-to-Spring gain scores was assessed. To determine the children's gains during their Head Start year, differences between the Fall 2000 and Spring 2001 mean scores were calculated for each of the measures by subtracting the Fall score from the Spring score for each child in the study.

The ability of the Fall-to-Spring gain scores from the cognitive measures of the FACES battery to predict later school readiness was assessed by two approaches. In the first approach, Reading and General Knowledge scale scores from the end of the kindergarten year were correlated with the gain scores from each of the FACES subtests from the end of the Head Start year. In these analyses, partial correlations were calculated, controlling for individual difference in Fall 2000, baseline scores.[2] The second approach assessed the ability of the Fall-to-Spring gain scores from the FACES instruments to predict Reading and General Knowledge scale scores at the end of the kindergarten year, controlling for child's age at time of assessment (i.e., the end of their Head Start year), in two multiple linear regression analyses.[3]

Children's Fall-to-Spring gain scores on each of the component tasks in the FACES battery correlated significantly with their Reading scale scores at the end of kindergarten (Table 11-14). Bivariate correlations with the Reading scale ranged from .33 (for the WJ-R LWI gain score) to .18 (for the Social Awareness gain score). Similarly, children's Fall-to-Spring gain scores on all of the component tasks in the FACES battery correlated significantly with their General Knowledge scale scores at the end of kindergarten. Significant bivariate correlations with the General Knowledge scale ranged from .33 (for the PPVT-III gain score) to .10 (for the gain scores for WJ-R LWI and Color Naming).

[2] In analyses of gains scores, baseline scores are controlled for, effectively examining the effect of the gain scores if all students had the same baseline score.

[3] Control of the Fall baseline score for each gain score in the multiple linear regression models was accomplished through the use of residual scores. A residual score was created for each gain score with the effects of its respective baseline score partialled out. These residual scores were then entered as independent variables in the multiple linear regression, predicting the kindergarten outcome variables.

Table 11-14. Pearson's product moment correlation coefficients between FACES assessment gain scores during Head Start and Reading and General Knowledge scale scores at end of kindergarten

Assessment gain score	Correlation coefficients	
	Reading	General Knowledge
PPVT-III	.23	.33
WJR Applied Problems	.23	.23
WJR Letter Word	.33	.10*
WJR Dictation	.30	.11**
Social Awareness	.18	.11
Draw-A-Design	.20	.14
Color Naming	.19	.10
One-to-One Counting	.22	.17
Book Knowledge	.22	.29

Note: Ns for bivariate correlations with Kindergarten Reading scale scores range from 637 -1,475. Ns for bivariate correlations with Kindergarten General Knowledge scale score range from 629-1,468. All correlation coefficients are significant at $p < .0001$ except *$p < .05$ and **$p < .01$.

When the gain scores were combined in a multiple regression model, the model did well at predicting children's early reading skills at the end of kindergarten, accounting for 20 percent of the variance in Reading scale scores (Table 11-15). The best predictor of Reading scale scores was the gain score for WJ-R Dictation task (beta = .16), followed by the LWI gain score (beta = .15) (which showed the highest bivariate correlation with the Reading scale). The model also did well at predicting children's general knowledge at the end of kindergarten, accounting for 23 percent of the variance in General Knowledge scale scores. The best predictor in the model was the gain score for the PPVT-III gain score (beta = .31), which had the highest bivariate correlation with the General Knowledge scores.

11.2.5 Head Start Fall-to-Spring Change Scores from the Behavior Ratings Predict Social Competence in Kindergarten

The Fall-to-Spring change scores from all but one of the parent and teacher ratings of behavior were moderately correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten. Parent ratings of changes in withdrawn behavior were not significantly correlated with the kindergarten teacher ratings (Table 11-16). Correlations were significant in the expected directions. Declines in problem behaviors as rated by both parents and teachers at the end of Head Start were significantly associated with higher teacher ratings of cooperative classroom behavior at the end of kindergarten, with correlation coefficients ranging from -.24 (for the change score for parent-reported ratings of hyperactivity) to -.13 (for the change score for parent-reported ratings of aggression). Increases in teacher ratings of cooperative classroom behavior at the end of Head Start were significantly associated with higher teacher ratings of cooperative classroom behavior at the end of kindergarten (r = .24), as were increases in parent ratings of positive approaches to learning (r = .09). In general, change scores for the teacher ratings showed stronger relationships with the kindergarten outcomes than those for the parent ratings. However, this could be due to the fact that parents did not report significant Fall-to-Spring change in their children's behavior.

Table 11-15. Summary of multiple linear regression analysis for FACES gain scores during Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($N_{Reading}$ = 510; $N_{General\ Knowledge}$ = 503)

	Reading scale			General Knowledge scale		
Assessment gain scale	B	SE B	Beta	B	SE B	Beta
PPVT-III	.17	.07	.11*	.39	.05	.31****
WJR Letter-Word ID	.11	.03	.15***	-.02	.03	-.04
WJR Dictation	.06	.02	.16****	.01	.01	.02
WJR Applied Problems	.01	.03	.02	.02	.02	.04
Draw-A-Design	.29	.21	.06	-.05	.17	-.01
Book Knowledge	.23	.36	.03	1.44	.29	.21****
Social Awareness	.76	.37	.08*	.51	.30	.07
Color Naming	.26	.14	.08	-.15	.11	-.06
One-to-One Counting	.77	.37	.09*	.69	.29	.10*
Child's Age at Head Start Assessment	.11	.10	.05	.11	.08	.06

Note: R^2 = .20 for Reading Scale; R^2 = .23 for General Knowledge; (ps < .0001).

*p < .05; **p < .01; ***p < .001; ****p < .0001.

Similarly, the Fall-to-Spring change scores from all of the parent and teacher ratings of behavior were moderately correlated with teacher ratings of total problem behavior at the end of kindergarten. Correlations were significant in the expected directions. Declines in problem behaviors as rated by both parents and teachers at the end of Head Start were significantly associated with lower teacher ratings of total problem behavior at the end of kindergarten, with correlation coefficients ranging from .29 (for the change score for teacher-reported ratings of hyperactive behavior) to .07 (for the change score for parent-reported ratings of withdrawn behavior). Increases in teacher ratings of cooperative classroom behavior were significantly associated with lower teacher ratings of total problem behavior at the end of kindergarten (r = -.24), as were increases in parent ratings of positive approaches to learning (r = -.12). In general, change scores for the teacher ratings showed stronger relationships with the kindergarten outcomes than those for the parent ratings.

Table 11-16. Pearson's product moment correlation coefficients between FACES behavior ratings gain scores at end of Head Start and teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten

	Correlation coefficients	
Behavior rating gain score	Cooperative classroom behavior	Total problem behaviors
Teacher Reported Aggression	-.18	.24
Teacher Reported Hyperactivity	-.24	.29
Teacher Reported Withdrawal	-.18	.19
Teacher Reported Total Problem Behaviors	-.23	.28
Teacher Reported Cooperative Classroom Behavior	.24	-.24
Parent Reported Aggression	-.13	.17
Parent Reported Hyperactivity	-.17	.17
Parent Reported Withdrawal	Ns	.07*
Parent Reported Total Problem Behaviors	-.15	.17
Parent Reported Positive Approaches to Learning	.09**	-.12

Note: Ns for correlations with Kindergarten Teacher Ratings of Cooperative Classroom Behavior range from 1,146-1,191. Ns for correlations with Kindergarten Teacher Ratings of Total Problem Behavior range from 1,147-1,193.

All p-values < .0001 except for *$p < .05$; **$p < .01$.

When the rating change scores were combined in a multiple regression predicting kindergarten teacher ratings of behavior, the model accounted for 8 percent of the variance in kindergarten teacher ratings of cooperative classroom behavior. The best predictor in the multiple regression was the change score for teacher-reported ratings of cooperative classroom behavior (beta = .14). The model predicting kindergarten teacher ratings of total problem behaviors accounted for 10 percent of the variance in teacher ratings of total problem behaviors. The best predictor in the multiple regression was the change score for teacher-reported ratings of hyperactive behavior, (beta = .14) (Table 11-17).

11.2.6 Head Start Fall-to-Spring Change Scores from the Behavior Ratings Predict Reading Skills and General Knowledge at the End of Kindergarten

All of the Fall-to-Spring change scores for teacher ratings of behavior correlated significantly with Reading scale scores at the end of kindergarten, however, only one of the parent ratings change scores of behavior had a significant correlation coefficient with kindergarten Reading (parent ratings of changes in hyperactive behavior: $r = -.06$; $p < .05$ - see Table 11-18). Correlations for teacher ratings were significant in the expected directions. Declines in teacher ratings of problem behaviors were significantly associated with higher kindergarten Reading scale scores, adding more evidence that behaviors that may impede learning are associated with lower reading skills in kindergarten. All of the

Fall-to-Spring change scores for teacher ratings of behavior also correlated in the expected directions with General Knowledge scale scores at the end of kindergarten. None of the parent rating change scores were significantly associated with kindergarten General Knowledge scores.

Table 11-17. Summary of multiple linear regression analysis for FACES behavior rating gain scores at end of Head Start predicting teacher ratings of cooperative classroom behavior and total problem behaviors at end of kindergarten ($N_{Cooperative\ Classroom\ Behavior}$ = 946; $N_{Total\ Problem\ Behaviors}$ = 947)

	Cooperative classroom behavior			Total problem behaviors		
Behavior rating gain score	B	SE B	Beta	B	SE B	Beta
Teacher Reported Aggression	.01	.11	.01	.20	.12	.07
Teacher Reported Hyperactivity	-.48	.16	-.13**	.60	.18	.14***
Teacher Reported Withdrawal	.01	.09	.01	-.02	.09	-.01
Teacher Reported Cooperative Classroom Behavior	.16	.05	.14***	-.13	.05	-.10*
Parent Reported Aggression	-.12	.11	-.04	.26	.12	.07*
Parent Reported Hyperactivity	-.37	.13	-.10**	.31	.15	.07*
Parent Reported Withdrawal	-.05	.18	-.01	.22	.20	.04
Parent Reported Positive Approaches to Learning	.01	.10	.00	-.02	.11	-.01
Child's Age at Head Start Assessment	.01	.03	.01	-.05	.03	-.05

Note: R^2 = .08 for Teacher Ratings of Cooperative Classroom Behavior; R^2 = .10 for Teacher Ratings of Total Problem Behaviors; (ps < .0001).
*p < .05; **p < .01; ***p < .001; ****p < .0001.

Table 11-18. Pearson's product moment correlation coefficients between FACES behavior ratings gain scores during Head Start and Reading and General Knowledge scale scores at end of kindergarten

	Correlation coefficients	
Behavior rating gain score	Reading	General knowledge
Teacher Reported Aggression	-.09**	-.09**
Teacher Reported Hyperactivity	-.13****	-.10***
Teacher Reported Withdrawal	-.21****	-.20****
Teacher Reported Total Problem Behaviors	-.18****	-.16****
Teacher Reported Cooperative Classroom Behavior	.13****	.10***
Parent Reported Aggression	NS	NS
Parent Reported Hyperactivity	-.06*	NS
Parent Reported Withdrawal	NS	NS
Parent Reported Total Problem Behaviors	NS	NS
Parent Reported Positive Approaches to Learning	NS	NS

Note: Ns for correlations with Reading Scale Scores range from 1,395-1,453. Ns for correlations with General Knowledge Scale Scores range from 1,387-1,443.
*p < .05; **p < .01; ***p < .001; ****p < .0001.

When the teacher ratings change scores were combined in a multiple regression model, the model accounted for 6 percent of the variance in Reading scale scores (Table 11-19).[4] The only significant predictor in the multiple regression was the change score for teacher-reported ratings of withdrawal behavior (beta = -.19). The multiple regression model of teacher rating change scores predicting kindergarten General Knowledge scale scores accounted for 7 percent of the variance in General Knowledge scale scores. Similar to the model predicting kindergarten Reading, the only significant predictor in this multiple regression was the change score for teacher-reported ratings of withdrawal behavior (beta = -.22).

Table 11-19. Summary of multiple linear regression analysis for FACES behavior rating gain scores during Head Start predicting Reading and General Knowledge scale scores at end of kindergarten ($N_{Reading}$ = 1,262; $N_{General\ Knowledge}$ = 1,338)

	Reading			General knowledge		
Behavior rating gain score	B	SE B	Beta	B	SE B	Beta
Teacher Reported Aggression	.07	.20	.01	-.10	.16	-.02
Teacher Reported Hyperactivity	.05	.30	.01	.40	.24	.06
Teacher Reported Withdrawal	-.88	.16	-.19*	-.80	.13	-.22*
Teacher Reported Cooperative Classroom Behavior	.06	.09	.02	-.00	.07	-.00
Child's Age at Head Start Assessment	.26	.05	.15*	.27	.04	.19*

Note: R^2 = .06 for Reading; R^2 = .07 for General Knowledge; (ps < .0001).

*p < .0001.

11.2.7 Children's Scores on the FACES Instruments and Behavior Ratings at the End of Head Start Predict Promotion to First Grade

Another measure of the predictive validity of the FACES battery is to examine how well scores on the FACES instruments and behavior ratings at the end of Head Start are related to practical decision-making at the end of kindergarten, namely, the teacher's decision of whether the child gets promoted to first grade or repeats the kindergarten year. In these analyses, the ability of the FACES scale scores from the end of the Head Start year to predict the teacher's decision at the end of the kindergarten year to have the child repeat a year of kindergarten (versus promote the child to first grade) was examined

[4] Because of their low correlations with kindergarten Reading and General Knowledge scale scores, parent rating gain scores were not included in the multiple regression models.

in a multiple logistic regression analysis. This approach was then repeated with the parent- and teacher-reported behavior ratings.

Children's Scores on the FACES Instruments at the End of Head Start Predict Repeating Kindergarten. When the subtest scores were combined in a multiple logistic regression model, the model did quite well at predicting whether children repeated kindergarten, accounting for 20 percent of the variance in the prediction of repeating kindergarten. Information on scores from the FACES instruments led to an 81 percent accuracy rate in predicting whether or not a child was assigned by his or her teacher to repeat kindergarten. The strongest predictor of whether or not children were assigned by their teachers to repeat kindergarten was the Draw-A-Design task, in which for every unit increase in Draw-A-Design scores, children were 22 percent less likely to repeat kindergarten (Table 11-20).

Table 11-20. Summary of multiple logistic regression analysis for FACES assessment scores at end of Head Start predicting teacher decision to have child repeat kindergarten (N = 906)

Assessment score	B	SE B	Odds ratio	95 Percent confidence limit
PPVT-III	-.02	.00	.981	.979 - .983
WJR Applied Problems	-.03	.00	1.004	1.003-1.005
WJR Letter Word	.00	.00	.975	.973-.976
WJR Dictation	-.01	.00	.989	.989-.990
Social Awareness	.16	.01	.864	.853-.875
Draw-A-Design	-.15	.01	.778	.766-.790
Color Naming	-.33	.00	.937	.933-.940
One-to-One Counting	.02	.01	.900	.887-.914
Book Knowledge	-.05	.01	.872	.858-.886
Child's Age at Head Start Assessment	-.01	.00	.976	.973-.979

Note: R^2_{Logit} = .20 (p < .0001). Percent Concordant = 81 percent. All regression coefficients significant at p < .0001.

Children's Behavior Ratings at the End of Head Start Predict Repeating Kindergarten. When the teacher ratings were combined in a multiple logistic regression model, the model did well at predicting whether children repeated kindergarten, although not as well as the FACES assessment tasks, accounting for 6 percent of the variance in the prediction of repeating kindergarten. Information from the FACES teacher ratings led to a 66 percent accuracy rate in predicting whether or not a child was assigned by his or her teacher to repeat kindergarten. The strongest predictor of whether or not children were assigned by their teachers to repeat kindergarten was parent ratings of withdrawn behavior, in which for every unit increase in these parent ratings, children were 13 percent more likely to repeat kindergarten (Table 11-21).

Table 11-21. Summary of multiple logistic regression analysis for teacher behavior ratings and parent behavior ratings at end of Head Start predicting teacher decision to have child repeat kindergarten (N = 926)

Behavior rating	B	SE B	Odds ratio	95 Percent confidence limit
Teacher Reported Aggression	-.01*	.01	.987	.977-.997
Teacher Reported Hyperactivity	.10**	.01	1.110	1.094-1.126
Teacher Reported Withdrawal	.12**	.00	1.126	1.118-1.134
Teacher Reported Cooperative Classroom Behavior	-.04**	.00	.964	.960-.969
Parent Reported Aggression	.10**	.01	1.106	1.094-1.117
Parent Reported Hyperactivity	.02*	.01	1.016	1.004-1.029
Parent Reported Withdrawal	.12**	.01	1.128	1.109-1.147
Parent Reported Positive Approaches to Learning	.12**	.01	1.122	1.111-1.134
Child's Age at Head Start Assessment	-.04**	.00	.966	.963-.968

Note: R^2_{Logit} = .06 (p < .0001). Percent Concordant = 66 percent. *p < .05; **p < .0001.

The Combination of Children's Scores on the FACES Instruments and Behavior Ratings at the End of Head Start Predict Repeating Kindergarten. When the teacher ratings were combined with the subtest scores from the FACES instruments in a multiple logistic regression model, the model did quite well at predicting whether children repeated kindergarten, accounting for 30 percent of the variance in the prediction of repeating kindergarten (Table 11-22). This is an increase of 10 percent in the variance that was explained by the subtest scores alone. Information from the combination of the FACES behavior ratings and the subtest scores led to an 82 percent accuracy rate in predicting whether or not a child was assigned by his or her teacher to repeat kindergarten. This suggests that the additional information provided by the behavior ratings adds to the predictive validity of the FACES instruments in predicting kindergarten repetition.

11.3 Conclusions

The FACES measures have strong predictive validity with outcomes at the end of kindergarten. As an indicator of preliteracy skills, the cognitive measures show strong associations with reading ability at the end of the kindergarten year. As an indicator of school adjustment and social competence, the behavior ratings demonstrate ability to predict kindergarten behaviors that promote learning. These analyses show that:

- The instruments used in FACES predict later behavior and performance in kindergarten;

- The instruments used in FACES also predict whether a child gets promoted to first grade;

- The instruments used in FACES may tap different types of abilities (“inside-out” vs. “outside-in”) that are important for children’s future reading proficiency and academic achievement; and

- The multi-measure and multi-method approach to the measurement of children’s abilities provides a variety of information sources that significantly contribute to the prediction of kindergarten outcomes.

Table 11-22. Summary of multiple logistic regression analysis for FACES assessment scores, teacher behavior ratings, and parent behavior ratings at end of Head Start predicting teacher decision to have child repeat kindergarten (N = 794)

Assessment score/behavior rating	B	SE B	Odds ratio	95 Percent confidence limit
PPVT-III	-.03	.00	.968	.965-.970
WJR Applied Problems	.01	.00	1.007	1.006-1.008
WJR Letter Word	-.02	.00	.976	.974-.977
WJR Dictation	-.01	.00	.992	.991-.993
Social Awareness	-.17	.01	.847	.835-.860
McCarthy Drawing	-.26	.01	.768	.754-.781
Color Naming	-.07	.00	.932	.928-.936
One-to-One Counting	-.13	.01	.880	.866-.894
Book Knowledge	-.15	.01	.860	.845-.876
Teacher Reported Aggression	.01(ns)	.01	1.005	.993-1.017
Teacher Reported Hyperactivity	.10	.01	1.110	1.091-1.131
Teacher Reported Withdrawal	.04	.00	1.038	1.028-1.047
Teacher Reported Cooperative Classroom Behavior	.00(ns)	.00	1.003	.997-1.010
Parent Reported Aggression	.10	.01	1.108	1.095-1.122
Parent Reported Hyperactivity	-.02*	.01	.983	.969-.998
Parent Reported Withdrawal	.19	.01	1.210	1.186-1.235
Parent Reported Positive Approaches to Learning	.22	.01	1.247	1.232-1.263
Child’s Age at Head Start Assessment	-.02	.00	.980	.977-.983

Note: R^2_{Logit} = .30 (p < .0001). Percent Concordant = 82 percent. All regression coefficients significant at p < .0001 except where indicated (*p < .05).

Children who made greater gains during the Head Start year on the LWI, Dictation, One-to-One Counting, Social Awareness, and the PPVT-III tasks tended to have greater early reading skills at the end of kindergarten. Children’s improved scores on the PPVT-III, Book Knowledge, and the One-to-One Counting tasks at the end of Head Start were associated with greater General Knowledge scores at the end of kindergarten. These results suggest that children who show greater improvement in preschool show greater school readiness and school adjustment when they are preparing to enter first grade.

In the assessment of children's social competencies, the use of parent and teacher ratings provides data on children's behavior skills in different situations and provides a more comprehensive picture of their behavior. Equally important, both parent and teacher ratings significantly contribute to the prediction of social skills at the end of kindergarten. The parent and teacher ratings also significantly predict reading skills and general knowledge at the end of kindergarten. Ratings of problem behaviors were negatively correlated with kindergarten Reading and General Knowledge scale scores, suggesting that behaviors that may impede learning are associated with lower reading skills in kindergarten. High ratings of behaviors that enhance learning, positive approaches to learning and cooperative classroom behavior, were positively correlated with kindergarten outcomes.

The multi-measure and multi-method approach to the measurement of children's development and school readiness provides a comprehensive assessment of children's abilities. The addition of the criterion-referenced measures to the norm-referenced measures improves the cognitive assessment battery in many ways. First, they are short tasks that cover more specific topic areas that are typically taught in preschool curricula and are also fun for the children to do in the assessment. And second, they significantly (although moderately) increase the measures' ability to predict kindergarten outcomes, improving its predictive validity. The addition of the parent and teacher behavior ratings adds another source of information that predicts kindergarten outcomes, namely an assessment of behaviors that can either foster or impede learning. They also provide some unique contributions to the battery's ability to predict kindergarten outcomes, particularly in the practical decision of whether a child is promoted to first grade or repeats kindergarten.

REFERENCES

Abbott-Shim, M., Lambert, R., and McCarty, F. (2000). Structural Model of Head Start Classroom Quality. *Early Childhood Research Quarterly*, 15, 1, 115-134.

Abbott-Shim, M., and Sibley, A. (1998). *Assessment profile for early childhood programs*: Research version. Atlanta, GA: Quality Assist.

Abbott-Shim, M., and Sibley, A. (1987). *Assessment profile for early childhood programs*. Atlanta, Georgia: Quality Assist Inc.

Achenbach, T.M., Edelbrock, C., and Howell, C.T. (1987). Empirically based assessment of the behavioral/emotional problems of 2-3 year-old children. *Journal of Abnormal Child Psychology*, 15, 629-650.

Administration for Children and Families. (2001). Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report. (Washington, DC: Child Outcomes Research and Evaluation Branch and the Head Start Bureau, Administration for Children and Families, U.S. Department of Health and Human Services, DHHS). http://www.acf.dhhs.gov/programs/core/pubs_reports/faces/meas_99_intro.html

Administration for Children and Families. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start.* Washington, DC: U.S. Department of Health and Human Services.

Administration for Children and Families. (2003). *Head Start FACES 2000: A Whole-Child Perspective on Program Performance, Fourth Progress Report*. (Washington, DC: Child Outcomes Research and Evaluation Branch and the Head Start Bureau, Administration for Children and Families, U.S. Department of Health and Human Services, DHHS). http://www.acf.dhhs.gov/programs/core/ongoing_research/faces/faces2000_final/faces02_title.html

Administration for Children and Families. (Forthcoming). FACES 1997 Technical Report II. Chapter on Marital Status, Family Composition and Childbearing: Changes during Head Start and beyond. Washington, DC: Administration for Children and Families, Department of Health and Human Services.

Administration on Children, Youth, and Families. (1995). *Charting our progress: Development of the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth, and Families. (1997). *First progress report on the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth, and Families. (1998). *Head Start Program Performance Measures: Second progress report.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth, and Families. (2002). *A descriptive study of Head Start families: FACES technical report I.* Washington, DC: U.S. Department of Health and Human Services.

Alborough, J. (1992). *Where's My Teddy?.* London, England: Walker Books Ltd.

Alborough, J., and Castro, M. (1995). ?*Donde Esta Mi Osito?.* Miami, Fl: Santillana USA Publishing Company, Inc.

Alexander, K.L., and Entwisle, D.R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53 (2),* Serial No. 218.

Amato, P.R. (2000). The consequences of divorce for adults and children. *Journal of Marriage and the Family,* 62, 1269-1287.

Annie E. Casey Foundation. (1999). *1999 Kids count data book: State profiles of child well-being*. Baltimore, MD: Annie E. Casey Foundation.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-552.

Astone, N.M., and McLanahan, S.S. (1991). Family structure, parenting practices, and high school completion. *American Sociological Review,*56, 309-320.

Barnett, W.S. (1998). Long-term effects on cognitive development and school success. Pp. 11-44 in *Early Care and Education for Children in Poverty*, edited by W.S. Barnett and S.S. Boocock. Albany, NY: State University of New York Press.

Blau, D. (1997). The production of quality in child care centers. *Journal of Human Resources*. 32. 354-387.

Bloom, B., and Tonthat, L. (2002). Summary health statistics for U.S. children: National Health Interview Survey. 1997 Vital Health Stat 10(203).

Bredekamp, S. (1986). The reliability and validity of the early childhood classroom observation scale for accrediting early childhood programs. *Early Childhood Research Quarterly*, 1, 103-118.

Bryant, D.M., Burchinal, M., Lau, L.B., and Sparling, J.J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9,* 289-309.

Bryk, A.S., and Driscoll, M.E. (1988). *An empirical investigation of school as a community.* Madison: University of Wisconsin Research Center on Effective Secondary Schools.

Bryk, A.S., and Raudenbush, S.W., (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Burts, D.C., Hart, C.H., Charlesworth, R., and Kirk. L. (1990) "A Comparison of Frequencies of Stress Behaviors Observed in Kindergarten Children in Classrooms with Developmentally Appropriate versus Developmentally Inappropriate Practices." *Early Childhood Research Quarterly*, vol. 5, pp. 407-423.

Child Trends Data Bank. (2002). Learning disabilities. Retrieved from the World Wide Web on 10/10/2003 - http://revised.childtrendsdatabank.org/indicators/65LearningDisabilities.cfm.

Clay, M.M. (1979). *What did I write?* Auckland, N.Z.; Exeter, NH: Heinemann.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Lawrence Earlbaum Associates.

Coiro, M.J., Zill, N., and Bloom, B. (1994). Health of our nation's children. National Center for Health Statistics. Vital Health Stat 10(191).

Consortium for Longitudinal Studies. (1983). *As the twig is bent…Lasting effects of preschool programs.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Cost, Quality and Child Outcomes Study Team. (1999). *The children of the cost, quality, and outcomes study go to school. Executive summary.* Chapel Hill: Frank Porter Graham Child Development Center.

Cost, Quality and Child Outcomes Study Team. (1995). *Cost, quality and child outcomes in child care centers, public report, second edition.* Denver: Economics Department, University of Colorado at Denver.

Currie, J., and Thomas, D. (1999). Does Head Start help Hispanic children? *Journal of Public Economics, 74*(2), 235-262.

Currie, J., and Thomas, D. (1998). *School quality and the longer-term effects of Head Start.* Washington, DC: National Bureau of Economic Research.

Currie, J., and Thomas, D. (1995). Does Head Start make a difference? *American Economic Review, 85*, 341-364.

Dale, P. S., Price, T. S., Bishop, D. V. M., and Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research,* 46, 544-560.

Duncan, G.J., Brooks-Gunn, J., Yeung W.J., and Smith J.R. (1998). How much does child poverty affect the life chances of children? *American Sociological Review,* 63, 406-423.

Duncan, G.J., and Brooks-Gunn, J. (eds.). (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.

Dunn, L.M., and Dunn. L.M. (1997). *Peabody Picture Vocabulary Test, Third Edition. Examiner's manual and norms booklet.* Circle Pines, MN: American Guidance Service.

Dunn, L.M., Padilla, E.R., Lugo, D.E., and Dunn, Leota M. (1986). *Test de Vocabulario en Imagenes Peabody*. Circle Pines, MN: American Guidance Service.

Elliot, S.N., Gresham, F.M., Freeman, R., and McCloskey, G. (1988). Teacher and observer ratings of children's social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment, 6,* 152-161.

Entwistle, D.R., Alexander, K.L., Cadigan, D., and Pallis, P.M. (1987). The emergent academic self-image of first graders: Its response to social structure. *Child Development*, 58, 1190-1206.

Feldhaus, K.M., Koziol-McLain, J., and Amsbury, H.L. (1997). Accuracy of 3 brief screening questions for detecting partner violence in the emergency department. *Journal of the American Medical Association, 227*(17), 1357.

Frede, E., Austin, A., and Lindauer, S. (1993). The relationship of specific developmentally appropriate teaching practices in preschool to children's skills in first grade. *Advances in Early Education and Day Care, 5*, 95-111.

Goal One Technical Planning Group. (1993). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Draft report to the National Education Goals Panel.* Washington, DC: National Education Goals Panel.

Goal One Technical Planning Group. (1991). The Goal One Technical Planning Subgroup report on school readiness. In National Education Goals Panel (Ed.), *Potential strategies for long-term indicator development: Reports of the technical planning subgroups* (Report No. 91-0, pp. 1-18). Washington, DC: National Education Goals Panel.

Hall, L., Williams, C., and Greenberg, R. (1985). Supports, stressors, and depressive symptoms in low-income mothers of young children. *American Journal of Public Health, 75*, 518-522.

Harms, T., and Clifford, R.M. (1980). *Early childhood environment rating scale*. New York: Teachers College Press.

Harms, T., Clifford, R.M., and Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised Edition*. New York: Teachers College Press.

Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist,* 44(2), 274-282.

Hayes, C., Palmer, J., and Zaslow, M. (1990*). Who cares for America's children? Child care policy for the 1990s.* Report of the Panel on Child Care Policy, Committee on Child Development Research and Public Policy, National Research Council. Washington, DC: National Academy Press.

High/Scope Educational Research Foundation. (1992). *Child Observation Record-Manual*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Hofferth, S.I., and Chaplin, D.D. (1998). State regulations and child care choice. *Population Research and Policy Review*. 17. 111-140.

Howes, C., Phillips, D., and Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based care. *Child Development, 63*, 449-460.

Infant Health Development Project (IHDP). (1990). Enhancing the outcomes of low-birth weight, premature infants: A multisite, randomized trial. *Journal of the American Medical Association, 262*, 3035-3042.

Kontos, S., and Fiene, R. (1987). Child care quality, compliance with regulations, and children's development: The Pennyslvania Study. In D. Phillips (ed.), *Quality in child care: What does the* *research tell us?* (pp. 57-79). Washington, DC: National Association for the Education of Young Children.

Lange, S.M. (2000). The challenges confronting children of incarcerated parents. *Journal of Family Psychotherapy, 11*(4), 61-68.

Layzer, J., Goodson., B., and Moss, M. (1993). *Final report volume 1: Life in preschool*. Cambridge, MA: Abt Associates, Inc.

Lazar, I., and Darlington, R. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development. 47*(2-3), Serial No. 195.

Lee, V., and Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Educatio*n. *62*, 172-192.

Leventhal, T., and Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337.

Liaw, F., and Brooks-Gunn, J. (1994). Cumulative familial risks and low birth weight children's cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*, 360-372.

Love, J., Schochet, P.Z., and Meckstroth, A.L. (1996). *Are they in any real danger? What research does – and doesn't – tell us about child care quality and children's well being.* Princeton, NJ: Mathematica Policy Research, Inc.

Mason, J., and Stewart, J. (1989). *Early childhood diagnostic instrument*. Iowa City, IA: American Testronics.

McCarthy, D. (1972). *Manual for the McCarthy Scales of Children's Abilities*. San Antonio, TX: The Psychological Corporation.

McKey, R.H., Condelli, L., Ganson, H., Barrett, B.J., McConkey, C,. and Plantz, M.C. (1985). *The impact of Head Start on children, families and communities. Final report of the Head Start evaluation, synthesis and utilization project*. (Contract No. 105-81-C-026). Washington, DC: Administration for Children, Youth and Families.

National Center for Education Statistics. (1996). National Household Education Survey Parent and Family Involvement Questionnaire. Washington, DC: U.S. Department of Education.

NICHD Early Child Care Research Network (ECCRN). (2000). The relation of child care to cognitive and language development. *Child Development, 71,* 823-839.

Peisner-Feinberg, E.S., and Burchinal, M.R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly*. 43, 3. 451-477.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L, Howes, C., Kagan, S.L., and Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development, 72(5),* 1534-1553.

Phillips, D. and Howes, C. (1987). *Indicators of quality in child care: Review of the research*. In D. A. Phillips (Ed.), Quality in child care: What does research tell us? (pp. 1-19). Research Monograph of the National Association for the Education of Young Children, Vol. 1. Washington, DC: NAEYC.

Phillips, D., Howes, C. and Whitebook, M. (1992). The social policy context of childcare: Effects on quality. *American Journal of Community Psychology*. 20. 25-51.

Phillips, D., McCartney, K., and Scarr, S. (1987). Child care quality and children's social development. *Developmental Psychology, 23,* 537-543.

Phillips, D., Mekos, D., Scarr, S., McCartney, K., and Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly*. 15, 4. 475-496.

Phillips, D.A., Voran, M., Kisker, E., Howes, C., and Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? *Child Development*, 65, 472-492.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Klebanov, P., and Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. Pp. 103-145 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips, Washington DC: Brookings Institution Press.

Pianta, R.C., and McCoy, S.J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology*, *18*, 1-22.

Quint, J. C., Bos, J.M., and Polit, D.F. (1997). *New Chance: Final report on a comprehensive program for young mothers in poverty and their children*. New York: Manpower Demonstration Research Corporation.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.

Ram, B., and Hou, F. (2003). Changes in family structure and child outcomes: Roles of economic and familial resources. *Policy Studies Journal*, 31(3), 309-330.

Resnick, G., McKey, R.H., and Klayman, D. (2001). *The Evaluation of Early Childhood Education Programming in the 30 Abbott School Districts: First-Year Report on Program Implementation and Descriptions of Children and Families*. (Contract No. A43078). Trenton, NJ: New Jersey Department of Human Services and New Jersey Department of Education.

Roid, G.H., and Miller, L.J. (1997). Examiners Manual: Leiter International Performance Scale - Revised. Wood Dale, Il: Stoelting Company.

Rosenthal, R., and Rosnow, R.L. (1984*). Essentials of behavioral analysis: Methods and data analysis.* New York: McGraw-Hill.

Ruopp, R., Travers, J., Glantz, F., and Coelen, C. (1979*). Children at the center: Final results of the National Day Care Study*. Boston: Abt Associates.

Salvia, J., and Ysseldyke, J.E. (1991). *Assessment in Special and Remedial Education*, 5th ed. Boston: Houghton Mifflin Co.

Scarr, S., and Eisenberg, M. (1993). Child care research: Issues, perspectives and results. Annual Review of Psychology. 44. 613-644.

Scarr, S., Eisenberg, M., and Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131-151.

Seitz, V. (1990). Intervention programs for impoverished children: A comparison of educational and family support models. *Annals of Child Development, 7*, 73-103.

Simons, R.L., Beaman, J., Cogner, R.D., and Chao, W. (1993). Stress, support and antisocial behavior trait as determinants of emotional well-being and parenting practices among single mothers. *Journal of Marriage and the Family,* 55, 385-398.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4), 323-355.

Smith, T., Kleiner, A., Parsad, B., and Farris, E. (2002). *Prekindergarten in U.S. Public Schools: 2000–2001* (NCES 2003-019). U.S. Department of Education, National Center for Education Statistics: Washington, DC.

Snow, C.E., Tabors, P.O., Nicholson, P.A., and Kurland, B.F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Early Childhood Education, 10*, 37-48.

St. Pierre, R.G., Layzer, J.I., Goodson, B.D., and Bernstein, L.S. (1997). *National Impact Evaluation of the Comprehensive Child Development Program: Final Report.* Washington, DC: U.S. Department of Health and Human Services, Administration on Children, Youth and Families.

Teale, W.H. (1990). Issues in early childhood assessment. In P.Afflerbach (Ed.). *Issues in statewide reading assessment*, 35-53. Washington, DC: ERIC/American Institutes for Research.

Teale, W.H. (1988). Developmentally appropriate assessment of reading and writing in early childhood classroom. *Elementary School Journal*, 89, 173-183.

Vaden-Kiernan, M., D'Elio, M.A., and Sprague, K. (1999). *The FACES embedded case study: Documenting the methodology and early findings.* Paper presented at the biennial meeting of the Society for Research in Child Development, Albuquerque, NM.

Weikart, D.P., and Schweinhart, L.J. (1991). Disadvantaged children and curriculum effects. *New Directions for Child Development, 53,* 57-64.

White, L., and Rogers, S.J. (2000). Economic circumstances and family outcomes: A review of the 1990s. *Journal of Marriage and the Family,* 62, 1035-1051.

Whitebook, M., Howes, C., and Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America* (Final report of the National Child Care Staffing Study). Oakland, CA: Child Care Employee Project.

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

Willet, J.B. (1988). Questions and answers in the measurement of change. In E. Rothkepf (Ed.), Review of research in education (1988-1989), (pp. 345-422). Washington, DC: American Educational Research Association.

Woodcock, R.W., and Johnson, M.B. (1989). *Woodcock-Johnson Revised Tests of Achievement*. Itasca, IL: Riverside Publishing.

Woodcock, R.W., and Mather, N. (1989, 1990). WJ-R Tests of Achievement: Examiner's manual. In: R.W. Woodcock and M.B. Johnson, *Woodcock-Johnson Psycho-Educational Battery—Revised.* Chicago: Riverside.

Zill, N. (1990). Behavior Problems Index based on parent report [memorandum]. Washington, DC: Child Trends.

Zill, N., and West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

APPENDIX A

METHODOLOGY

A.1 Sample Design Overview

The Head Start children for FACES 2000 were selected as a two-stage sample. The first stage sampling units were Head Start programs; the second stage units were classes within sampled programs. In each sampled classroom, all children in their first year of Head Start were included in the sample.

A.1.1 Programs

The sampling frame of eligible Head Start programs was constructed from the 1998-1999 Program Information Report (PIR). Migrant and Seasonal Head Start programs, American Indian/Alaska Native Head Start programs, Early Head Start programs, programs in the territories, and programs that do not serve children directly were excluded, resulting in a frame of 1,675 programs. The programs were stratified by Census region (North East, Midwest, South, and West), percent minority (above/below 50 percent), and metro or urban/rural status (Metropolitan Statistical Area (MSA)/non-MSA). These are the same stratification variables used in sampling programs for the FACES 1997 cohorts.

A sample of 45 programs was selected for FACES 2000. The sample size in each stratum was proportional to the stratum first year Head Start enrollment. The programs were selected with probability proportional to the program's first year enrollment using systematic sampling. The first year enrollment was calculated from the PIR by subtracting the reported second and third year enrollment from the total enrollment. A Keyfitz procedure was used to minimize the overlap with the 40 programs sampled for the FACES 1997 cohorts. As a result, there was no overlap with the previous program sample. Of the 45 programs selected, two were later discovered to be ineligible because they had been defunded, resulting in a total of 43 sampled programs for FACES 2000.

A.1.2 Classrooms

In the 43 remaining Head Start programs, lists of the anticipated classes for Fall 2000 were obtained in late Summer 2000. The programs also provided the expected number of first-year Head Start children in each class.

These lists formed the basis for the classroom sampling frame, after excluding classes with no first-year children. Classes with fewer than five first-year children expected were combined with another class in the same center to form a “class group.” The class groups were treated as a single unit for sampling purposes and sample size calculations. The total target sample size of first-year children was 2,825 or 66 per program. In general, the desired sample size of classes in each program was determined as 66/(average class size for the program), where the average class size was in terms of the number of first-year children. The actual initial sample size was increased by 2 classes to allow for a reserve sample in each program. In programs where the total first year enrollment, as obtained from the class rosters, was more than twice the measure of size used to sample the program, the initial class sample size was increased to prevent large variation in class weights. In small programs where the initial sample size exceeded the number of classes available, all classes were taken with certainty.

Classes were sorted by center within program and were sampled with equal probabilities. A subsample of the initial sample was selected with equal probabilities to obtain a main sample of the desired sample size and a reserve sample of two classes in each program. A total of 367 classes was selected: 279 classes for the main sample and 88 for the reserve sample. The number of main sample classes in each program varied from 3 to 15, with an average of 6 classes. (In terms of collapsed classrooms, a total of 252 classroom groups was sampled for the main sample and 82 for the reserve sample, for an average of 6 per program and a range of 3 to 10.)

In Fall 2000, the eligibility status of the main sample classes was determined. One or two reserve classes were added in some programs to prevent a shortfall in the target number of first-year children for the study. The final sample for weighting purposes included all sampled classes where an attempt was made to collect data from the classroom, including those discovered to be ineligible. The rationale for this is because ineligible classes in the sample represent ineligible classes on the Head Start program frame. A total of 307 main and reserve classes were in the sample in Fall 2000. When the program was contacted by field staff in Fall 2000, 20 of the 307 classes were discovered to be ineligible because they no longer existed, they did not receive Head Start funding, or they had no first-year Head Start children. In 286 of the remaining 287 eligible classes (one teacher refused to allow the children in her class to be sampled), all first-year Head Start children were taken into the sample.

A.2 Response Rates

Fall 2000

- 2,508 Child assessments were completed out of 2,790 for a completion rate of 90 percent;
- 2,488 Parent interviews were completed out of 2,790 families selected for the sample (89 percent);
- Teacher Report forms were obtained on 2,532 of the 2,790 sample children (91 percent);
- Assessment, parent, and teacher data were obtained on 2,396 of the 2,790 sample children (86 percent); and
- A total of 278 classrooms were observed out of 286 in the sample for a completion rate of 97 percent.

Spring 2001

- 2,232 Child assessments were completed out of 2,288, representing 98 percent of the children who remained in the program, and 80 percent of the original sample of 2,790 children;
- 2,166 Parent interviews were completed out of 2,288, representing 95 percent of the children who remained in the program, and 78 percent of the original sample;
- Teacher report forms were obtained on 2,236 of the sample children, representing 98 percent of the children who remained in the program and 80 percent of the original sample;
- Assessment, parent, and teacher data were obtained on 2,115 of the 2,288 sample children who remained in the program (92 percent); and
- A total of 275 classrooms were observed out of 284 in the sample for a completion rate of 97 percent.

A.3 Sampling Weights

There are two main reasons why weights are desirable and are calculated for FACES data. The most important reason for the use of weights is that all units generally do not have the same probability of selection. In FACES, probabilities of selection vary among sample programs, among sample centers, among sample classes, and among sample children. The first function of weights is to reflect the probability of selection. For example, large programs were selected with higher probabilities than small programs, and large centers within a given program were selected with higher probabilities than small centers. Thus, probabilities of selection among centers vary considerably. If there tend to be differences in center characteristics according to the size of the center or the size of the program the center is in, then weighted estimates may be different from unweighted estimates.

A second important reason for using weights is to reduce bias and sampling error. Inevitably, there is some level of nonresponse in all surveys. Use of nonresponse adjustment factors in the weighting process can reduce the bias caused by nonresponse. For example, there are differences in response rates in FACES by gender of child. To the extent that there are also differences in test scores by gender, then forming nonresponse cells based on gender reduces the bias caused by differential gender response rates. Also, known population totals can frequently be used as poststratification controls in the weighting, which reduces both bias and sampling error. Poststratification, however, has not been deemed feasible for FACES.

A.3.1 Program Weights

The program weight was calculated as the inverse of the program's probability of selection. As mentioned earlier, a Keyfitz procedure was used to minimize the overlap with the program sample drawn for FACES 1997. This procedure involved calculating conditional probabilities of selection which are based on whether the program was sampled previously or not, and whether its probability of selection increased compared with the previous sample. Prior to sampling for the FACES 2000 cohort 3, the unconditional probability of selection for each program on the cohort 3 frame was calculated as

$$NEWPSEL_i = \frac{FIRSTYR_i}{\left(\sum_{i=1}^{Nh} FIRSTYR_i\right) / NEWSMPSZ_h}$$

where N_h is the number of programs on the frame in stratum h, $NEWSMPSZ_h$ is the sample size for stratum h for the cohort 3 design, and $FIRSTYR_i$ is the first year enrollment for program i from the PIR. The probability of selection for each program under the Abt sample design for Cohorts 1 and 2 was also calculated as

$$ORIGPSEL_i = \frac{ENRTOT_i}{\left(\sum_{i=1}^{Nh} ENRTOT_i\right)/SAMPSIZE_h}$$

where N_h was the number of programs on the Abt frame in stratum h, $SAMPSIZE_h$ was the sample size for stratum h under the Abt design, and $ENRTOT_i$ was the total enrollment for the *i*-th program from an earlier PIR.

The conditional probability of selection was calculated for each program on the cohort 3 frame according to the Keyfitz procedure as:

Case 1: $NEWPSEL \geq 1 - ORIGPSEL$

$CONDPROB$ = $\frac{NEWPSEL - (1 - ORIGPSEL)}{ORIGPSEL}$ if the program was sampled for cohorts 1 and 2,

= 1 if the program was not sampled for cohorts 1 and 2.

Case 2: $NEWPSEL < 1 - ORIGPSEL$

$CONDPROB$ = 0 if the program was sampled for cohorts 1 and 2,

= $\frac{NEWPSEL}{1 - ORIGPSEL}$ if the program was not sampled for cohorts 1 and 2.

These conditional probabilities of selection were the measures of size used to select the cohort 3 program sample. It can be shown that the Keyfitz procedure preserves the unconditional cohort 3 program probabilities of selection, while at the same time minimizing the overlap. Thus the cohort 3 program weight is the inverse of NEWPSEL, the unconditional probability of selection under the cohort 3 design.

All 43 eligible programs cooperated with the study, so that nonresponse adjustments at the program level were unnecessary.

For each program, a set of 43 jackknife replicate weights was created for calculating standard errors. The replicate weights were created using a standard stratified jackknife procedure. One program at a time was dropped (i.e., given a zero replicate weight) and the weights of the remaining programs in the same stratum were adjusted by a factor of $n_h/(n_h-1)$, where n_h is the number of sampled programs in stratum h. The program weights in the other strata were left unchanged. By repeating it 43 times, 43 replicate weights were obtained for each program. For estimates involving child or classroom data from all 43 programs, the degrees of freedom for the variance of the estimate is #PSUs - #varstrat = 43 – 12 = 31. (One of the 13 original sampling strata was collapsed with an adjacent stratum for variance estimation purposes because it contained only one eligible sampled program.)

A.3.2 Classroom Weighting

Two sets of class weights were produced for classroom level estimation: one set for Fall 2000 cross-sectional estimates and a second set for Fall 2000–Spring 2001 longitudinal classroom analysis. Class base weights were first created that reflected the overall probability of selection for the class, including the program probability of selection. These base weights were adjusted for classroom level nonresponse, using the following criteria for a complete classroom:

Fall 2000 Cross-Sectional Estimates. Classroom must have complete Fall 2000 observation data. Classroom observation data includes counts of children and adults, Assessment Profile (Scheduling, Learning Environment, and Individualizing), ECERS-R, Arnett Caregiver Interaction Scale, Teacher-Directed Activities Checklist and Wrap-Up measures; and

Fall 2000–Spring 2001 Longitudinal Analysis. Classroom must have complete observation data for either Fall 2000 or Spring 2001 and child assessment data for both Fall 2000 and Spring 2001.

A.3.2.1 Class Base Weights

A class base weight was created for each of the 367 initially sampled classes in Fall 2000. Fifty-four reserve classes that were never used were given base weights of zero. Six main sample classes were sampled out on an ad hoc basis by field staff to reduce burden and to have independence between classes. They were assigned base weights of zero, since they were not part of the final sample. In this situation, a teacher had both a morning and an afternoon class in the sample. One class out of the morning/afternoon pair was subsampled.

The remaining 307 classes considered to constitute the sample were each assigned a class base weight equal to the inverse of their overall probability of selection. The overall probability of selection is the product of the program probability of selection and the probability of selecting the class within the program. The inverse of the overall probability of selection can also be written as the product of the program weight and the within-program class weight:

Class Base Weight = Program Weight * (Total # Classes in Program / # sampled classes fielded).

Collapsed classrooms were counted as one classroom in the base weight calculations, since they were treated as a single unit in sampling. The ad hoc subsampling was reflected by multiplying the base weight of the retained class in the am/pm pair by a factor of 2 and the dropped class by zero. One class that had merged with another was given a zero base weight, and the newly merged class had its base weight multiplied by a factor of .5 to reflect its increased probability of selection.

Forty-three jackknife class replicate base weights were created from the program replicate weights:

> Class Replicate Base Weight j = Program Replicate Weight j * (Total # Classes in Program / # sampled classes fielded); $j = 1, 2, \ldots 43$.

A.3.2.2 Cross-Sectional Fall 2000 Class Weights

Of the 307 sampled classes that were fielded in Fall 2000, 279 were eligible and had complete classroom data, 8 were eligible but didn't complete data collection, and 20 were discovered to be ineligible. A class nonresponse adjustment factor was applied to the class base weights of the 279. The nonresponse adjustment factor was computed separately by program. Both the 8 incomplete and the 20 ineligible classes were given a zero final class weight. The classroom replicate base weights were also adjusted for nonresponse by program, so that the sampling variability in the nonresponse adjustments were reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000 classroom weights is 34,638. The unweighted and weighted completion rates are both 97 percent, excluding ineligibles from both numerator and denominator. The unweighted and weighted eligibility rates are both 94 percent. The class base weight was used in calculating the weighted rates.

A.3.2.3 Longitudinal Fall 2000–Spring 2001 Class Weights

Of the 286 eligible classes in Fall 2000, 280 completed data collection in Spring 2001. Note that the 279 Fall 2000 classroom completes are not a subset of the 280 Spring 2001 completes. Five classes that completed Fall 2000 data collection did not complete the Spring 2001, and six classes that completed Spring 2001 data collection did not complete the Fall 2000. There were 79 new classes added in Spring 2001 because children who switched classes after the Fall 2000 data collection were followed to the new class. However, no classroom observations were done at these new classes, so they were not considered to be part of the classroom sample and were assigned a zero base weight.

A class nonresponse adjustment factor was applied to the class base weights of the 280 eligible completes. The nonresponse adjustment factor was computed separately by program. The classroom replicate base weights were also adjusted for nonresponse by program, so that the sampling variability in the nonresponse adjustments were reflected in the standard error estimates. The incomplete and ineligible classes, along with the 79 new classes, were given a final class weight of zero.

The sum of the nonresponse-adjusted Fall 2000 – Spring 2001 classroom weights is 34,768. The unweighted and weighted completion rates are both 98 percent, excluding ineligibles from both numerator and denominator. Both unweighted and weighted eligibility rates are 94 percent. The class base weight was used in calculating the weighted rates.

A.3.3 Child Weights

Three sets of child weights were produced: a cross-sectional set for Fall 2000 estimates, a Fall 2000 – Spring 2001 set for base year longitudinal analyses, and a Fall 2000 – Spring 2003 set for longitudinal analysis including the kindergarten school year. Child base weights were first created that reflected the overall probability of selection for the child, including the program and classroom stages of sampling. These base weights were adjusted for child nonresponse, using the following criteria for a complete child case:

> **Fall 2000 Cross-Sectional Analysis.** A child is considered a complete case if the child has a parent interview from either a Fall 2000 or Spring 2001 and a Fall 2000 child assessment or teacher rating;

Fall 2000 – Spring 2001 Longitudinal Analysis. A child is considered a complete case if the child has either a Fall 2000 or Spring 2001 parent interview and one of the following data pairs: a child assessment for both Fall 2000 and Spring 2001, or a teacher rating for both Fall 2000 and Spring 2001; and

Fall 2000 – Spring 2003 Longitudinal Analysis. A child is considered a complete case if there was at least one parent interview, an assessment while the child was in Head Start, either in Fall 2000 or Spring 2001, and an assessment while the child was in kindergarten, either Spring 2002 or Spring 2003.

A.3.3.1 Child Base Weights

In 286 eligible Fall 2000 classes, all eligible children in their first year of Head Start were included in the sample with certainty. A base weight was created for each child as the product of their program weight and nonresponse-adjusted classroom weight. Note that these nonresponse adjusted class weights are not the same as those described earlier, which were designed for use in classroom level analyses. The creation of special classroom weights for the child weights was necessary because there were eligible classrooms that did not have complete classroom observations, but did allow their children to be sampled, and vice versa. To create this special classroom weight, the classroom base weight was adjusted for classes which had eligible children but where "sampling" of children did not take place. This nonresponse-adjusted classroom weight was then used in calculating the child base weight. Since there was no subsampling of children within classrooms, the within-classroom child weight is equal to one and the overall child weight can be written as:

Child Base Weight = Program Weight * Nonresponse-adjusted Classroom Weight.

A set of 43 jackknife (JKn) replicate base weights was also created for each child using the program replicate weights and the special full-sample nonresponse-adjusted classroom weight:

Child Replicate Base Weight j = Program Replicate Weight j * Nonresponse-adjusted Classroom Weight; $j = 1, 2, \ldots 43$.

A.3.3.2 Child Fall 2000 Cross-Sectional Weights

Of the 3,100 children in the Fall 2000 sample, 2,535 were considered complete for the Fall 2000 data collection, 251 were eligible but incomplete (30 of these had assessments but no parent interview), and 314 were ineligible. Children could be ineligible if either they came from classrooms that were ineligible or they were discovered to be in their second year of Head Start or were otherwise ineligible when Fall 2000 data collection began.

The child base weights of the eligible, complete children in each classroom were adjusted for nonresponse separately by classroom. The ineligible and incomplete children were given a zero final child weight and were dropped from the sample for the Spring 2001 data collection. The replicate child base weights were also adjusted for nonresponse by classroom, so that the sampling variability in the nonresponse adjustments were reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000 child weights is 337,247. The unweighted and weighted completion rates are both 91 percent, excluding ineligibles from both the numerator and denominator. The unweighted and weighted eligibility rates are 90 percent and 91 percent, respectively. The child base weight was used in calculating the weighted rates.

A.3.3.3 Child Fall 2000 – Spring 2001 Longitudinal Weights

In Spring 2001, the eligible first-year children were again given assessments, a teacher rating, and an attempt was made to interview the child's parent(s). Of the 2,535 eligible children who had completed Fall 2000 data collection, 2,359 were eligible, complete cases for the Fall 2000 – Spring 2001 data collection; 171 were eligible, incompletes; and five became ineligible because they moved out of the area.

Children who had switched to new classes in the Spring 2001 were followed up, but classroom observations were not done at the new classes. There were 91 children from the Fall 2000 sample who were followed to 79 new classrooms in Spring 2001. In calculating their base weights, these children were given the classroom probability of selection associated with the classroom from which they were originally sampled in Fall 2000.

The child base weights of the eligible, complete children in each classroom were adjusted for nonresponse separately by classroom. The ineligible and incomplete children were given a zero final child weight. The replicate child base weights were also adjusted for nonresponse by classroom, so that the sampling variability in the nonresponse adjustments will be reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000 – Spring 2001 child weights is 338,047. The unweighted and weighted conditional Spring 2001 completion rates are both 93 percent. The conditional rate is the percent of Fall 2000 eligible completes who also completed the Spring 2001 data collection. The overall (unconditional) completion rate is the product of the completion rates for the Fall 2000 and Spring 2001 data collections: 91 percent * 93 percent = 85 percent. This rate is the percent of eligible, sampled children in Fall 2000 that completed the Spring 2001 data collection.

A.3.3.4 Child Fall 2000 – Spring 2003 Longitudinal Weights

For the year in which the children were in kindergarten, either Spring 2002 or Spring 2003, the children were again given assessments, a teacher rating, and an attempt was made to interview the child's parent(s). No classroom observations were done for these children. Of the 2,535 eligible children who had completed Fall 2000 data collection, 1,895 were respondents again for kindergarten data collection, either Spring 2002 or Spring 2003, and 640 were nonrespondents.

The child base weights of the eligible, responding children in each classroom were adjusted for nonresponse separately by classroom. Classrooms were collapsed within centers and within programs when necessary to prevent excessively large nonresponse adjustment factors. The ineligible and nonresponding children were given a zero final child weight. The replicate child base weights were also adjusted for nonresponse by classroom.

The sum of the nonresponse-adjusted Spring 2002-2003 child weights is 337,247. The unweighted completion rate is 68 percent, i.e., 68 percent of eligible sampled children were Spring 2002-2003 (kindergarten) respondents. The weighted completion rate is 69 percent.

A.4 Variance Estimation

Estimates obtained from the FACES sampled children will differ from the true population parameters because they are based on a randomly chosen subset of the population, rather than on a complete census of all Head Start children. This type of error is known as sampling error or variance. The precision of an estimate is measured by the standard error (defined as the square root of the variance). The calculation of the standard error must reflect not only the sample size on which the estimate is based, but the manner in which the sample was drawn. Otherwise, the standard errors can be misleading and result in incorrect confidence intervals and p-values in hypothesis testing. The FACES sampling involved stratification, clustering, and unequal probabilities of selection, all of which must be reflected in the standard error calculations.

Jackknife replication is a commonly used variance estimation method for complex surveys involving multi-stage sampling (Wolter, 1985). Replication methods work by dividing the sample into subsample replicates that mirror the design of the sample. A weight is calculated for each replicate using the same procedures as for the full-sample weight. This produces a set of replicate weights for each sampled child. To calculate the standard error of a survey estimate, the estimate is first calculated for each replicate using the replicate weight and the same form of estimator as for the full sample. The variation among the replicates is then used to estimate the variance for the full sample estimate. Replication has the advantage that it can reflect the different features of the weighting and estimation by simply repeating all steps separately for each replicate.

For each child, a set of 40 jackknife replicate weights was created for calculating standard errors. The replicate weights were created using a standard stratified jackknife ("JKn") procedure. One Head Start program at a time was dropped (i.e., given a zero replicate weight) and the weights of children in the remaining programs in the same stratum were adjusted by an inflation factor to account for the reduction of the sample. The weights in the other strata were left unchanged. By repeating this for each of the 40 sampled programs, 40 replicate weights were obtained for each child. The replicate weights are used in the formula below to calculate the variance:

$$v(\hat{\theta}) = \sum_{g=1}^{G} \frac{n_h - 1}{n_h} (\hat{\theta}_{(g)} - \hat{\theta})^2$$

where $\hat{\theta}_{(g)}$ is the estimate of θ based on the observations included in the g-th replicate (i.e., using the g-th replicate weight), $\hat{\theta}$ is the estimate of θ based on the full sample, G is the total number of replicates formed, n_h is the number of programs in stratum h, and $(n_h-1)/n_h$ is the "JKn" factor. When creating a var file in WesVar, the JKn factors must be imported using the "Attach Factors" option in the Data menu. For estimates involving data from all 40 programs, the degrees of freedom for the variance of the estimate is # programs - # strata = 40 – 10 = 30. (Six of the original 14 sampling strata were collapsed for variance estimation purposes because they contained only one eligible sampled program, resulting in 10 pseudo strata.)

A.5 Data Collection Instruments

A.5.1 Direct Child Assessment

The child assessment was an essential component of FACES. It provided direct measures of how well Head Start programs were achieving the goals of assisting children to be physically, socially, and educationally ready for success in kindergarten. The assessment battery was composed of a short series of tasks that were feasible and interesting for preschoolers and kindergartners to carry out, and that have been shown to be predictive of later school achievement or learning difficulties. The areas of the FACES assessment battery included vocabulary development, emerging literacy (recognizing letters of the alphabet, hearing similarities in the sounds at the beginnings or ends of different words, showing familiarity with printed words and story books), emerging numeracy (counting, adding, or taking away blocks to show a given number), perceptual-motor development (drawing copies of simple geometric figures), and social and communicative competence (telling basic facts about self and family to another person). These tasks were drawn or adapted from well-established and widely used instruments.

A Spanish version of the assessment battery was also developed for assessing children whose primary language was Spanish. Spanish versions of the measures, when available, (e.g., *Test de Vocabulario en Imagenes Peabody*, *Woodcock-Muñox Pruebas de Aprovechamiento-Revisada*) were employed in the Spanish battery. Otherwise, the English versions of the measures (e.g., one-to-one counting, social awareness, color names, etc.) were directly translated.

A screener was used to determine whether English-language learners were to be administered the direct child assessment battery in English or not. The screener involved information provided by teachers and assessors which was used to determine the language of administration. In Fall 2000, English-language learners who were determined to be primarily Spanish-speaking received the entire direct child assessment battery in Spanish, e.g., TVIP, Woodcock Munoz Letter-Word Identification, Applied Problems, Dictation, etc. They also were administered the PPVT and Woodcock Johnson Letter-Word Identification in English, as well. In Spring 2001 and Spring 2002 (for children who were in Head Start for 2 years), these same children received the entire direct child assessment battery in English. They were also administered the TVIP and Woodcock Munoz Letter Word Identification in Spanish for the purpose of comparison. The children who had been administered assessments in Spanish and English in Fall 2000, with some Spanish sections, in Spring 2001 and Spring 2002 (Head Start only) were administered the entire assessment in English during the Spring of their kindergarten year (either Spring 2002 or Spring 2003).

In Fall 2000, English-language learners who were determined to primarily speak a language other than Spanish did not receive any portion of the direct child assessment battery in their native languages and were assessed in English, if possible. In Spring 2001, Spring 2002, and Spring 2003 (if applicable), these same children received the entire direct child assessment battery in English, if possible.

Norm Referenced Cognitive Tests

In these assessment tasks, norms are available for a nationally representative sample of U.S. children of the same age (including children from all family income groups). Average scores, factoring in the age of the children, a.k.a. standard or scale scores, are reported throughout this report. However, most of the results detailed in Appendix A. Tables A-2 through A-18 are for raw scores only. These standard scores are adjusted for the child’s age and constructed to have an overall mean and standard deviation of specific values, e.g., a mean of 100 and a standard deviation of 15. This is useful for comparing children from Head Start programs to other children on a nationwide basis on specific school readiness measures to determine the growth and progress of these Head Start children in these domains relative to all others in the nation.

A.5.1.1 Peabody Picture Vocabulary Test – Third Edition – Revised

The Peabody Picture Vocabulary Test (PPVT-III) (Dunn and Dunn, 1997) is designed to assess children's knowledge of the meaning of words by asking them to say or indicate by pointing which of four pictures best shows the meaning of a word that is said aloud by the assessor. A series of words is presented, ranging from easy to difficult for children of a given age, each accompanied by a picture plate consisting of four line drawings. The test takes about 10 minutes to administer. It is suitable for a wide range of ages from 2½ through adulthood and has established age norms based on a national sample of 2,725 children and adults tested at 240 sites across the U.S.

The PPVT-III has been extensively revised from earlier versions of the test. These improvements were undertaken to promote easier testing and more accurate scoring. Also, new drawings have been added and dated illustrations dropped so as to achieve better gender and ethnic balance. Individual test items that showed statistical bias by race or ethnicity, gender, or region were deleted from the item pool for the scale prior to standardization.

PPVT-III scores have high reliability, with the test publisher reporting internal-consistency reliability (alpha) coefficients ranging from .92 to .98, with a median of .95, and test-retest reliability ranging from .91 to .94. The alpha coefficients for the PPVT-III results from FACES were .97 for Fall 2000, Spring 2001, Spring 2002/2003 (Kindergarten), and .96 for Spring 2002 (Head Start).

A Spanish-language test, the *Test de Vocabulario en Imagenes Peabody* (TVIP), is also available, but has not been updated to be directly comparable to the PPVT-III. For FACES, the TVIP was used with children whose primary language was Spanish. The TVIP was reported to be highly reliable utilizing FACES data with internal-consistency alpha coefficients of .92 for both Fall 2000 and Spring 2001, and .94 for Spring 2002 (Head Start).

A.5.1.2 Woodcock-Johnson Psycho-Educational Battery – Revised

The updated edition of the Woodcock-Johnson Battery (WJ-R) is a carefully constructed and widely used test battery. The set of individually administered tests is designed to assess the intellectual and academic development of individuals from preschool through adulthood (Woodcock and Johnson, 1989; Salvia and Ysseldyke, 1991). FACES used three subtests from the Achievement Battery that together constitute an "Early Development -- Skills" cluster, according to the test developers. The cluster is comprised of the Letter-Word Identification, Applied Problems, and Dictation tests. The same three subtests of the Spanish version (*Woodcock-Muñox Pruebas de Aprovechamiento-Revisada*) were used in the Spanish version of the FACES assessment battery.

Letter-Word Identification. The first five Letter-Word Identification items involve symbolic learning, or the ability to match a rebus (pictographic representation of a word) with an actual picture of the object. The remaining items measure children's reading identification skills in identifying isolated letters and words that appear in large type on the pages of the test book. As well as being part of the Early Development cluster, this subtest is also part of the Basic Reading Skills cluster. The internal reliability of the Letter-Word Identification subtest with preschool age children averages .92 (Woodcock and Johnson, 1989). The internal reliability of this subtest with FACES children averaged .84 for Fall 2000, and .86 for Spring 2001 and Spring 2002 (Head Start). The internal reliability of the Spanish version of this subtest (Woodcock Munoz) was .75 for Fall 2000, .78 for Spring 2001, and .83 for Spring 2002 (Head Start).

Applied Problems. This subtest measures children's skill in analyzing and solving practical problems in mathematics. In order to solve the problems, the child must recognize the procedure to be followed and then perform relatively simple counting, addition or subtraction operations. Because many of the problems include extraneous stimuli or information, the child must also decide which data to include in the count or calculation. As well as being part of the Early Development cluster, the subtest is also part of a Broad Mathematics cluster. The internal reliability of the Applied Problems subtest with preschool age children averages .91 (Woodcock and Johnson, 1989). The internal reliability of this subtest with FACES children averaged .90 for Fall 2000, .91 for Spring 2001, .89 for Spring 2002 (Head Start), and .88 for Spring 2002/2003 (Kindergarten). The internal reliability of the Spanish version of this subtest (Woodcock Munoz) was .85 for Fall 2000.

Dictation. The first six items in this subtest measure prewriting skills, such as drawing lines and copying letters. The remaining items measure the child's skill in providing written responses when asked to write specific upper- or lower-case letters of the alphabet. Later parts of the test ask the child to write specific words and phrases, punctuation, and capitalization. The internal reliability of the Dictation subtest with preschool age children averages .90 (Woodcock and Johnson, 1989). The internal reliability of this subtest with FACES children averaged .77 for Fall 2000, Spring 2001, and Spring 2002/2003 (Kindergarten), and .71 for Spring 2002 (Head Start). The internal reliability of the Spanish version of this subtest (Woodcock Munoz) was .77 for Fall 2000.

A.5.1.3 Leiter International Performance Scale – Revised (Leiter-R) – Attention Sustained

The Leiter-R by Roid and Miller (1997) assesses cognitive function in children and adolescents. The battery includes measures of nonverbal intelligence in fluid reasoning and visualization, as well as appraisals of visuospatial memory and attention. In Spring 2001, the Leiter-R Attention Sustained (AS) Subtest was added to the FACES direct child assessment battery to permit assessments of children’s visuospatial memory and attention. The subtest is primarily nonverbal and is administered in two subsections – the first being for those 2-3 years of age and the second being for those 4-5 years of age. Assessors provide minimal instructions throughout the administration of the Leiter-R AS. Children are presented with a series of pages containing pictures and are instructed to mark off all pictures that resemble a reference picture. The assessor times the child, with times ranging from 30 seconds to 120 seconds allotted for completion of the tasks. The internal reliability of the Leiter-R Attention Sustained subtest for 2-3 year olds and 4-5 year olds is .83 (Roid and Miller, 1997). The internal reliability of this subtest with FACES children, by age groupings, averaged .71 for 3 year olds and .81 for 4-5 year olds in Spring 2001 and .80 for 4-6 year olds in Spring 2002 (Head Start).

A.5.1.4 Test of Language Development (TOLD) – Primary – 3rd Edition – Phonemic Analysis (Kindergarten Only)

The Phonemic Analysis test from the Test of Language Development (TOLD; Newcomer and Hamill, 1997) is a supplemental subtest designed to assess children's awareness of phonemes, that is, the significant speech sounds that comprise words. For this test, the child is presented with a compound word, and then asked to repeat part of the word's component phonemes back to the assessor (e.g., "Say 'popcorn.' Now say it again without 'pop'.") The internal reliability of this supplemental subtest with FACES children averaged .96 for Spring 2002/2003 (Kindergarten).

A.5.1.5 ECLS-K Reading and General Knowledge Scales (Kindergarten Only)

In the Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K), the Reading scale taps a variety of skills that indicate reading ability (including familiarity with print), recognition of letters and phonemes, vocabulary, and reading comprehension skills (e.g., children's understanding of the text), as well as their personal reflection and critical evaluation of the text. The General Knowledge scale taps skills in the natural sciences (e.g., their conceptual understanding of why things occur as they do, and their ability to pose questions and investigate answers in the natural sciences) and social studies (e.g., their basic knowledge of History, Government, and Culture). Both scales follow the guidelines of the 1996 National Assessment of Educational Progress, have been reviewed by curriculum experts, as well as elementary school teachers, and have been found to be both reliable and valid measures of reading achievement and basic knowledge acquisition.[1]

The Reading assessment was administered in two stages. First, a routing test was administered to estimate the child's reading ability. Based on his/her performance on the routing test (either "high," "medium," or "low"), an appropriate "second stage" test was administered. The Reading assessment had three levels of second stage tests: low (red), medium (yellow), and high (blue). For the

[1] For more information on the Reading and General Knowledge measures, please refer to the Early Childhood Longitudinal Study website at http://nces.ed.gov/ecls/.

General Knowledge assessment, each child was administered only the routing test. Estimates of reliability with FACES data, as measured by Cronbach's coefficient alpha, were as follows: 1.) for Reading – Spring 2002/2003 (Kindergarten) – Routing = .87, Red (Low Form) = .95, Yellow (Middle Form) and Blue (High Form) = .94, and 2.) for General Knowledge – Spring 2002/2003 (Kindergarten) = .77.

Criterion Referenced Cognitive Tasks

These tasks cover areas of basic knowledge, verbal, mathematical, and perceptual-motor skills that children typically learn in the preschool and kindergarten years and are often included in assessments of children's school readiness and progress. These tasks do not have national norms. All results referenced in Table A-2a through A-3c and throughout the entire Report for these Tasks pertain to raw scores.

A.5.1.6 McCarthy Scales of Children's Abilities

The McCarthy Scales of Children's Abilities is a widely used and well-documented test battery. FACES employed one subtest from the battery, the Draw-A-Design Task. The Draw-A-Design Task was used to assess children's perceptual-motor skills. This task asks the child to draw copies of a series of increasingly complex geometric figures. For FACES, this task was directly translated as part of the Spanish version of the assessment. The FACES reliabilities for the McCarthy Draw-A-Design measure were .58 for Fall 2000, .70 for Spring 2001, and .72 for Spring 2002 (Head Start). The FACES reliability for the Spanish version of this measure for Fall 2000 was .57.

A.5.1.7 Story and Print Concepts

The Story and Print Concepts task was an adaptation of earlier prereading assessment procedures developed by Marie Clay (1979), William Teale (1988, 1990), and Mason and Stewart (1989). In these procedures, a child is handed a children's storybook (FACES Battery - *Where's My Teddy?* (Alborough, 1992) or *¿Dónde Está Mi Osito?* (Alborough, Castro, Trans. 1992)) upside down and backwards. The assessor asks a series of questions designed to test the child's knowledge of books. These include questions regarding the location of the front of the book, the point at which one should begin reading, and information relating to the title and author of the book. The assessor reads the story to the child and asks basic questions about both the mechanics (print conventions) of reading and the content (comprehension) of the story. The print conventions questions pertain to children's knowledge of the left-to-right and up-and-down conventions of reading, while the comprehension questions pertain to children's recall of key facts from the story. Additionally, for FACES, questions were added tapping rhyming awareness (e.g., "I'll say some words from the story and you tell me whether they rhyme, OK - bawl and small, etc.") and phonological awareness (e.g., "What word would be left if I took "teh" away from Ted?"). These additions were only included in the Fall 2000 direct child assessment battery. The FACES reliabilities for Fall 2000, Spring 2001, and Spring 2002 (Head Start) were as follows: Book Knowledge (.57, .59, and .61); Print Conventions (.73, .75, and .84); and Comprehension (.43, .42, and .40). The reliabilities for the Spanish version of these measures for Fall 2000 were .43 – Book Knowledge, .59 - Print Conventions, .39 – Comprehension.

A.5.1.8 Social Awareness

This measure was adapted from a subtest of the Comprehensive Assessment Program (CAP) Early Childhood Diagnostic Instrument used by Snow et al. (1995) among others to test children's general knowledge and awareness of the social environment. The child is asked to give his/her "full name," which includes both first and last name, his/her age (either verbally, which is given full credit or by holding up the correct number of fingers, which is given partial credit) and month/day of birth. The FACES reliabilities for the Social Awareness measure were .63 for Fall 2000, .62 for Spring 2001, .62 for Spring 2002 (Head Start), and .65 for Spring 2002/2003 (Kindergarten). The FACES reliability for the Spanish version of the Social Awareness measure for Fall 2000 was .36.

A.5.1.9 Color Names and One-to-One Counting

This was also a subtest of the CAP Early Childhood Diagnostic Instrument used by Snow et al. (1995) and developed by Marie Clay (1979), William Teale (1988, 1990) and Mason and Stewart (1989) as a battery of emergent literacy and school readiness measures. For the FACES battery, 10 teddy bears of different colors are presented randomly arranged on a page and the child is asked to point to each in turn and name the color. Following the Color Names task, the child is asked to count the bears and the assessor marks the final number the child arrives at when finished counting (correct answer is “10”). After this, the child is asked to report the total number of bears. The verbatim response is then recorded. Following these questions, the assessor must rate the child’s one-to-one counting performance using a 5-point scale. At the extremes, a score of 5 indicated that the child made no mistakes and score of 1 indicated that the child could not count or did not try to count. The FACES reliabilities for the Color Naming task were .95 for Fall 2000, .94 for Spring 2001, and .90 for Spring 2002 (Head Start). The FACES reliability for the Spanish version of the Color Naming task for Fall 2000 was .92.

A.5.1.10 Writing Name (Kindergarten Only)

The Writing Name task was designed to assess the child’s ability to write his or her first or last name correctly.

A.5.1.11 Interviewer Ratings

At the end of the one-on-one testing sessions with the children, the assessor completes a set of rating scales evaluating the child’s behavior in the test situation, including the child’s approaches to learning and problem behaviors. There are two sections to these ratings. The first consists of eight scales rating the child’s response during the assessment on eight different domains: task persistence, attention span, body movement, attention to directions, comprehension of directions, verbalization, ease of relationship, and the child’s level of confidence. Ratings use 4-point scales with descriptive anchors at each point. For example, the “task persistence” scale consists of the following anchor points: persists with task (4), attempts task briefly (3), attempts task after much encouragement (2), refuses (1). The FACES reliabilities for the Interviewer Ratings were .82 for Fall 2000, .80 for Spring 2001, .70 for Spring 2002 (Head Start), and .75 for Spring 2002/2003 (Kindergarten). The FACES reliability for the Spanish version of the Interviewer Ratings for Fall 2000 was .77.

The second section asks the assessor to indicate any special concerns regarding the child’s ability to complete the assessment: responding nonverbally, using nonstandard English such as dialect, speaking English as a second language, having limited English proficiency, experiencing difficulty hearing or seeing the assessor/test materials, or reporting the child’s speech was difficult to understand. These items use 3-point ratings to indicate the degree to which the child displayed any of these characteristics (i.e., “not at all,” “somewhat,” and “very much”).

Table A-1 summarizes the modifications to the assessment battery from the Fall 2000 through the Spring 2002-2003 kindergarten followup.

A.5.2 Classroom Observation Instruments

In FACES 2000, quality was considered to include not only the number of children and adults in each classroom, but process factors such as the availability of learning materials, the types of classroom activities, scheduling and the variety of learning opportunities provided to all children. Lead teachers[2] in Head Start classrooms were also interviewed to collect teacher background information (experience and qualifications) as well as more detailed information about their curriculum, classroom activities, and attitudes and knowledge about early childhood education practices.

A.5.2.1 Counts of Children and Adults, Child-Adult Ratio

The Counts of Children and Adults provide information needed to calculate child-adult ratios and for other calculations to be used in assessing specific measures of classroom quality. Classroom observers counted the number of children, the number of adults and the number of paid staff at two separate time periods during the classroom day. The two occasions were separated by at least one hour and involved one structured (teacher-directed) and one unstructured activity. The child-adult ratio is calculated as the average number of children per adult (both paid and volunteer) across the two observations. A related measure, the child-staff ratio, was calculated using only the number of paid staff across the two observations. Higher child-adult or child-staff ratios are indicative of lower quality.

A.5.2.2 Assessment Profile for Early Childhood Programs: Research Edition I

The Assessment Profile for Early Childhood Programs: Research Edition I (Abbott-Shim and Sibley, 1987) is a structured observation guide designed to provide a quantitative assessment of classrooms and teaching practices that facilitate the learning and development of children. Three subscales were used in FACES: Scheduling, Learning Environment, and Individualizing.

---

[2] The results of teacher data presented in this chapter were based on interviews and ratings of the lead or senior teacher in each classroom.

Table A-1. Summary of measures administered from Fall 2000 through Spring 2003

Fall 2000 (Head Start)	Spring 2001 and Spring 2002 (Head Start)	Spring 2002 and Spring 2003 (kindergarten)
Social Awareness	Social Awareness	Social Awareness
PPVT-III / TVIP	PPVT-III / TVIP	PPVT-III
McCarthy Draw-A-Design	McCarthy Draw-A-Design	TOLD – Primary -3rd Phonemic Analysis
Color Names and Counting	Leiter-R AS (Attention Sustained) Subset	Reading ECLS-K Routing and Second Stage Sections
Woodcock Johnson R (Munoz): Letter-Word Identification	Color Names and Counting	Woodcock Johnson R: Applied Problems
Woodcock Johnson R (Munoz): Applied Problems	Woodcock Johnson R (Munoz): Letter-Word Identification	Woodcock Johnson R: Dictation
Woodcock Johnson R (Munoz): Dictation	Woodcock Johnson R: Applied Problems	Writing Name
Story and Print Concepts	Woodcock Johnson R: Dictation	General Knowledge ECLS-K Routing
Interviewer Rating: Assessment Behavior	Story and Print Concepts	Interviewer Rating: Assessment Behavior
	Interviewer Rating: Assessment Behavior	

The **Scheduling** subscale assesses the written plans for classroom scheduling and how classroom activities are implemented. The appropriateness and completeness of the classroom activity plan are also noted. The subscale also assesses the balance and variety of learning contexts (e.g., individual, small group, and large group) and learning opportunities (i.e., child- vs. teacher-directed and active vs. quiet activities). The 14 observation items are scored in a yes/no format. High scores on this measure are indicative of a teacher that uses a "planful" approach to classroom activities. The reliability of the Scheduling subscale was reported as .89 for Fall 2000, .87 for Spring 2001, and .82 for Spring 2002 (Head Start).

The **Learning Environment** subscale focuses on the accessibility of a variety of learning materials to children in the classroom. Variety is assessed across various conceptual areas, such as science, math, language, fine motor, etc. and also within each conceptual area. The subscale also assesses how classroom space is arranged to determine whether the classroom encourages independence (e.g., whether the learning materials are located on low shelves and clearly labeled) and reflects the child as an individual. When materials are both available and accessible, and in sufficient numbers (typically a minimum of three in each group), the item is given a positive score. High scores on this 7-item measure indicate a "learning rich" environment filled with toys and learning materials that address a variety of developmental domains. The reliability of the Learning Environment subscale was reported as .68 for Fall 2000, .77 for Spring 2001, and .65 for Spring 2002 (Head Start).

The **Individualizing** subscale is based on a scale from the Assessment Profile for Early Childhood Programs. For FACES 2000 it was shortened to five observational items measuring whether the teacher plans classroom activities to meet the varying learning needs of each child, how the teacher keeps track of the children's work during the year through the use of individual child portfolios and whether the teacher accommodates children with disabilities through an inclusionary approach. A high score indicates that teachers are able to adjust classroom activities to meet the learning needs of individual children. The reliability of the Individualizing subscale was reported as .50 for Fall 2000, .54 for Spring 2001, and .44 for Spring 2002 (Head Start).

A.5.2.3 Early Childhood Environment Rating Scale – Revised (ECERS-R)

The Early Childhood Environment Rating Scale (ECERS) is a global rating of classroom quality based on structural features of the classroom (Harms and Clifford, 1980). It has been widely used in child development research and has predicted optimal child outcomes in a number of studies (e.g., Phillips, Voran, Kisker, Howes, and Whitebook, 1994). The revised version of the ECERS (ECERS-R) provides improvements to the items and allows for a more standardized approach to assigning scores. In addition, the ECERS-R is easier to train on and gain inter-rater reliability. The ECERS-R contains 37 items representative of classroom quality. Each item is coded on a 7-point scale with a score of 1 representing "inadequate", a score of 3 representing "minimal quality," a score of 5 representing "good quality," and a score of 7 representing "excellent quality." The internal consistency of the ECERS-R mean score for all combined items was .92 for both Fall 2000 and Spring 2001, and .89 for Spring 2002 (Head Start).

The ECERS-R items were grouped into seven subscales for usage in analyses of FACES classroom quality, each pertaining to different elements of classroom quality.[3] These are as follows:

- Personal Care Routines are measured using six items: greeting/departing, meals/snacks, nap/rest, toileting/diapering, health practices, and safety practices;
- Furnishings is measured using four items: indoor space, furniture for routine care, play, and learning, furniture for relaxation and comfort, and room arrangement for play;
- Language Skills are measured using four items: books and pictures, encouraging children to communicate, using language to develop reasoning skills, and informal use of language;
- Motor Skills are measured using four items: space for gross motor play, gross motor equipment, fine motor activities, and supervision of gross motor activities;
- Creativity is measured using six items: child-related display, art, music/movement, blocks, sand/water, and dramatic play;
- Social Skills are measured using four items: supervision other than gross motor activity, discipline, staff-child interactions, and interactions among children; and
- Program Structure is measured using four items: space for privacy, schedule, free play, and group time.

[3] It should be noted that these subscales do not correspond to the subscales and item groupings developed by the instrument developer.

Five items were not incorporated into any of the subscales which are as follows: nature/science, math/numbers, use of TV, video, and/or computers, promoting acceptance of diversity, and provisions for children with disabilities. Thus there were only 32 of the 37 available items included in the subscales.

The Language Skills subscale was used as a key classroom quality measure in many of the analyses, and also contributed to the Quality Factor score (described in Chapter 4). Devised to assess the quality of the language environment in Head Start classrooms, a high score indicates a classroom with a rich language environment, in terms of the availability and use of books and printed materials, receptive and expressive language activities, language to engage logical and reasoning skills, and the informal use of language throughout the classroom day.

A.5.2.4 Classroom Observation of Teacher – Directed Activities

The Classroom Observation of Teacher-Directed Activities is a checklist completed by classroom observers of observed teacher-directed activities in 21 specific areas, e.g., reading stories, singing songs, etc. The classroom observer indicates whether observed activities were directed toward individual children (Individual Attention), a small group of children (Small Group = 3 to 8 children), or a whole group of children (Whole Group = entire classroom). Observers were instructed to mark down, only once for any item, any teacher-directed activities observed throughout the course of the classroom observation and whether these observed activities were directed toward individuals, a small group of children, or the entire classroom. This checklist was introduced in Spring 2001.

A.5.2.5 Arnett Caregiver Interaction Scale

The Arnett Caregiver Interaction Scale (Arnett, 1989) is a rating scale of teacher behavior towards the children in the classroom. The Arnett Caregiver Interaction Scale consists of 30 items and five subscales labeled Sensitivity, Harshness, Detachment, Permissiveness, and Independence. At the end of the observational period, the observer completes the scale for an individual teacher, typically the lead teacher in the classroom. For example, in evaluating whether the teacher "speaks warmly to the children," the observer will assign a rating indicating the extent to which the statement is characteristic of the teacher, from 1 "never seen" to 4 "always or almost always." A high score indicates greater teacher sensitivity, responsiveness and encouragement of children's independence and self-help skills, and lower levels of punitiveness and detachment. The Cronbach Coefficient Alphas for all of the items were .94 for both Fall 2000 and Spring 2001, and .93 for Spring 2002 (Head Start).

A.5.2.6 Teacher Backgrounds, Qualifications, and Attitudes

The **Lead Teacher Background Information** is based on individual interviews with the lead teacher of the classrooms that were being observed. The interviews collected extensive information about the teachers' backgrounds (e.g., age, ethnicity), experience (e.g., total years teaching, years teaching Head Start), and qualifications (e.g., whether the teacher has a BA or AA, whether the teacher had a graduate degree or some graduate school education, whether the teacher has a Child Development Associate certificate, teaching certificate, or whether the teacher took any early childhood education or development courses). Ethnicity was included in these analyses because it may be related to differences in teacher qualifications and experience and because the types of teachers in the classrooms may be influenced by the backgrounds of the families and children attending the Head Start program as well as the larger community served by the program.[4] There were also questions about the nature of the curriculum used, the training and resources provided to support the curricula used, as well as questions about how teachers monitor the progress of individual children, and what accommodations the teacher makes to meet the learning needs of each student, including those with special needs.

The 24-item Teacher Beliefs Scale (Burts, Hart, Charlesworth, and Kirk, 1990) was included in the teacher interview, and consists of statements worded to reflect positive attitudes and knowledge of generally accepted practices in preschool settings, or to reflect a lack of these attitudes and knowledge. In FACES 2000, one factor comprising 9 items that explained most of the variation in scores for the entire scale was used. A high score indicates higher positive attitudes and knowledge about early childhood education practices.

[4] While age was also included in the teacher interview, it was so highly correlated with the teacher's years of experience that it was not included in analyses beyond the descriptive level. However, since ethnicity did prove to be related to other factors in classroom quality, we continued to include it in our analyses.

A.5.2.7 Quality Composite

As a result of principal components analyses of the quality measures, several of the measures–the ECERS-R Language subscale, the Assessment Profile Scheduling subscale and the Assessment Profile Learning Environment subscale–were found to be highly correlated with each other, suggesting that a greater amount of variation in quality can be explained by reducing these three quality indicators to one measure. Additional principal components analyses which included other quality measures (e.g., the Arnett Caregiver Interaction Scale, the child:adult ratio and even the ECERS-R total score) found that these measures did not add significantly to the explained variation and in some cases may have detracted from it. Thus, these other quality measures were used independently in the analyses of classroom quality. Scores from the three measures that formed the quality composite were combined to form a single factor score for quality. A higher score indicates higher levels of quality.

A.5.3 Teacher's Child Reporting Form

Teacher ratings of children were important sources of information about children's learning and behavior because teachers see children over extended periods of time and in a variety of settings. Using a rating form known as the Teacher's Child Report (TCR), teachers were first asked to rate each child on a set of behaviors that assessed the child's basic social skills and classroom behavior. In these two sections, the teacher is asked to indicate the extent to which a given statement (e.g., "follows the teacher's directions") is characteristic of the child, from 1 "never" to 3 "very often." The items making up these ratings form two scales.

A.5.3.1 Cooperative Classroom Behavior

There are 12 ratings items for the teacher to indicate how often the child engages in cooperative classroom behaviors such as following teacher's directions, helping put things away, complimenting classmates, and following rules when playing games. The ratings include items drawn from the Personal Maturity Scale (Alexander and Entwisle, 1988) and the Social Skills Rating System (Elliott, Gresham, Freeman, and McCloskey, 1988) to assess positive behavior such as cooperation, sharing, and expression of feelings. A summary score is created from the 3-point scale items with a range from 0 to 24, with high scores indicating more frequent cooperative behavior. The internal consistency for this measure was .88 for Fall 2000, Spring 2001, and Spring 2002/2003 (Kindergarten), and .87 for Spring 2002 (Head Start).

A.5.3.2 Total Behavior Problems

The Behavior Problems scale is based on measures of negative child behaviors that are associated with learning problems and later grade retention. Items come from an abbreviated adaptation of the Personal Maturity Scale (Alexander and Entwisle, 1988), the Child Behavior Checklist for Preschool-Aged Children, Teacher Report (Achenbach, Edelbrock, and Howell, 1987) and The Behavior Problems Index (Zill, 1990). The items ask about the frequency of aggressive behavior (e.g., hits/fights with others), hyperactive behavior (e.g., is very restless), and anxious or depressed and withdrawn behavior (e.g., is unhappy). The summary score from the scale's 14 behavior items ranges from 0 to 28, with higher scores representing more frequent or severe negative behavior.

The Total Problem Behavior items were grouped into three subscales for usage in analyses of children's behavior, each pertaining to different types of problem behavior: Aggressive, Hyperactive, and Withdrawn behavior. The Aggressive behavior subscale, comprised of four items, assesses the frequency of disobeying rules or requests, disrupting ongoing activities, hitting or fighting with others, and having temper tantrums. The Aggressive behavior scale score ranges from 0 to 8, with higher scores representing more frequent or severe aggressive behavior. The Hyperactive behavior subscale, comprised of three items, assesses the frequency of being unable to concentrate or pay attention for long; being nervous, high strung, or tense; and being very restless. The Hyperactive behavior scale score ranges from 0 to 6, with higher scores representing more frequent or severe hyperactive behavior. The Withdrawn behavior subscale, comprised of seven items, assesses the frequency of acting too young for age; being hard to understand; keeping to self; lacking confidence in learning new things or trying new activities; often seeming sleepy or tired in class; often seeming unhappy, sad, or depressed; and worrying about things for a long time. The Withdrawn behavior scale score ranges from 0 to 14, with higher scores representing more frequent or severe withdrawn behavior.

The reliabilities (internal consistency) for these measures for Fall 2000, Spring 2001, Spring 2002 (Head Start), and Spring 2002/2003 (Kindergarten) were as follows: Total Problem Behaviors - .86, .86, .87, and .86; Aggressive behavior - .83, .85, .83, and .85; Hyperactive behavior - .72, .72, .75, and .74; and Withdrawn behavior - .77, .76, .76, and .77.

The teacher is then asked to rate the child's problem solving skills and initiative, social relationships, creative representations, music/movement skills, and language/math skills. The teacher is asked to rate the child's highest level of behavior in each of the above domains observed in the past week. Scale points for each item are described on paper and there is a glossary that provides concrete examples of each anchor point. For the purpose of FACES, 14 items from the Child Observation Record (COR; High/Scope Educational Research Foundation, 1992) were selected with a demonstrated reliability of .94 for both Fall 2000 and Spring 2001, and .93 for Spring 2002 (Head Start). These 14 items were further divided into the following scales: social relationships, creative representations, music and movement, and cognitive.

A.5.3.3 Social Relationships (3 Items)

A composite score was based on teacher's ratings of how well the child makes friends, works with other children, and understands and expresses feelings. Each item is rated on a five-point scale with higher scores representing greater skill in coping with social situations and expressing feelings appropriately. The summary score is the average of the three items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .83 for both Fall 2000 and Spring 2001, and .80 for Spring 2002 (Head Start).

A.5.3.4 Creative Representations (3 Items)

A composite score was based on the teacher's ratings of how well the child uses creative materials for self-expression in making and building things, drawing and painting, and engaging in pretend play. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the three items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .80 for both Fall 2000 and Spring 2002 (Head Start), and .81 for Spring 2001.

A.5.3.5 Music and Movement (4 Items)

A composite score was based on teacher’s ratings of how well the child can imitate movements to a steady beat, follow music and movement directions, exhibit body coordination, and manipulate small objects and perform precise actions. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the four items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .88 for both Fall 2000 and Spring 2001, and .86 for Spring 2002 (Head Start).

A.5.3.6 Cognitive (4 Items)

A composite score was based on teacher’s ratings of how well the child can solve problems, engage in complex play, show interest in reading, and exhibit classification skills by sorting objects. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the four items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .82 for Fall 2000, .83 for Spring 2001, and .80 for Spring 2002 (Head Start).

A.5.4 Parent Interview

Data from the FACES Parent Interview, administered in Fall 2000, Spring 2001, Spring 2002, and Spring 2003, provide Head Start with a comprehensive understanding of the families it serves or served, including the characteristics of households and household members, levels and types of participation in the program and in other community services, involvement with their children, and understanding of their children's development.

Parents were also asked to rate their child on a set of behaviors that assessed the child’s basic social skills and behavior problems. In this section, the parent is asked to indicate the extent to which a given statement (e.g., “makes friends easily”) is characteristic of the child, from 1 “not true” to 3 “very true or often true.” The items making up these ratings were drawn from two well-known measures of children's positive behavior and behavior problems: the Entwisle Scale of Personal Maturity (Entwisle, Alexander, Cadigan, and Pallis, 1987) and the Child Behavior Checklist for Preschool-Aged Children

(Achenbach, Edelbrock, and Howell, 1987). Two scales were formed to assess children's social competence.

A.5.4.1 Social Skills and Positive Approaches to Learning

Parents were asked to rate their child's social skills and positive approaches to learning by describing their children's skills in making friends and accepting their ideas, as well as enjoying learning and trying new things. A summary score based on the scale's seven items ranges from 0 to 14, with higher scores representing more positive behavior. Tables A-23 through A-30 show the reliabilities for the Social Skills measure for Fall 2000, Spring 2001, and Spring 2002 (Head Start), and Spring 2002/2003 (Kindergarten).

A.5.4.2 Total Problem Behaviors

Parents were also asked to rate their children on negative behaviors that are relatively common among preschool children and that are associated with adjustment problems in elementary school. Parents were asked about three domains of problem behavior: hyperactive behavior, aggressive behavior, and depressed or withdrawn behavior. The 12 behavior items were combined in a summary score ranging from 0 to 24, with higher scores representing more frequent or severe negative behavior.

The 12 total problem behavior items were grouped into three subscales for usage in analyses of children's behavior, each pertaining to different types of problem behavior: Aggressive, Hyperactive, and Withdrawn behavior. The Aggressive behavior subscale, comprised of four items, assesses the frequency of having temper tantrums, hitting or fighting with others, not getting along with other kids, and being disobedient at home. The Aggressive behavior scale score ranges from 0 to 8, with higher scores representing more frequent or severe aggressive behavior. The Hyperactive behavior subscale, comprised of three items, assesses the frequency of being unable to concentrate or pay attention for long; being very restless; and being nervous, high strung, or tense. The Hyperactive behavior scale score ranges from 0 to 6, with higher scores representing more frequent or severe hyperactive behavior. The Withdrawn behavior subscale, comprised of five items, assesses the frequency of being unhappy, sad, or

A.5.4.3 Other Parent Interview Scales/Measures Referenced in the Report

Other parent interview scales/measures referenced in the report are listed below.

Names and sources for other parent interview scales/measures referenced in the report		
Name	Source	Description
Pearlin Mastery Scale (Locus of Control)	Pearlin, L.I. and Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior, 22,* 337-356.	Seven items measured the degree to which parents feel they have control over their own lives and their self-confidence in their abilities to solve life's problems.
CES-D Depression Scale	Radloff, L.S. (1977). The CES-D: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.	Twelve items measured levels of depression among primary caregivers.
Family Activities with Children	National Household Education Survey - FACES Research Team	Eleven items asked respondents to indicate which family activities (such as telling a story; teaching letters, words, or numbers; teaching songs; or going on errands) were undertaken by the family members with the children in the past week. Another set of 11 item asked respondents to indicate which types of outings, such as visiting a library, a zoo, or a mall, they participated in with their children in the past month.
Parental Involvement in Head Start	Head Start Quality Research Consortium (QRC)	Asked in the Spring parent interview, 15 items asked parents to indicate how often they participated in various activities (e.g., volunteered in classroom, prepared food or materials for special events) at their child's Head Start center.
Exposure to Violence	FACES Research Team	Five items tapped the extent that the respondent has been exposed to violent crime in his/her neighborhood or home.

Names and sources for other parent interview scales/measures referenced in the report		
Name	Source	Description
Domestic Violence Screener	Feldous, K.M., Koziol-McLain, J., Amsbury, H.L. et. al. (1997). Accuracy of three brief screening questions for detecting partner violence in the emergency room. *JAMA, 227(17),* 1357.	Three items ask the respondent to indicate whether he/she has been a victim of domestic violence, and whether the child has either been a victim or witness of domestic violence.
Substance Abuse Screener	Administration for Children and Families (1997). National Impact Evaluation of the Comprehensive Child Development Program. Washington, DC: U.S. Department of Health and Human Services.	Three items ask the respondent to indicate the number of times a member of the household, who uses alcohol, (including self) has gotten into trouble with family or friends, has gotten in trouble with the police, or has missed work or school because of alcohol use. The same three items were asked for drug use if there is a member of the household who uses drugs.
Involvement with Criminal Justice System	FACES Research Team	Three items ask the respondent to indicate whether he/she or another household member have been arrested or charged with any crime since the child was born; and whether he/she/another household member spent time in jail.
Parenting Style	National Longitudinal Study of Youth (NLSY), Early Head Start Evaluation (EHS), QRC	Parents were also asked to rate a series of statements that address how they were raising their child at home. Four statements formed a scale that assesses whether the parent had an *authoritative* parenting style. Three statements form a scale that assesses whether the parent has an *authoritarian* parenting style.

depressed; worrying about things for a long time; and feeling worthless and inferior. The Withdrawn behavior scale score ranges from 0 to 10, with higher scores representing more frequent or severe withdrawn behavior.

Tables A-23, A-27, and A-28 show the reliabilities for both parent and teacher reported behavior problem measures for Fall 2000, Spring 2001, Spring 2002 (Head Start), and Spring 2002/2003 (Kindergarten).

A.6 Field Staff Training

A weeklong training was conducted prior to each data collection period to prepare field staff for successful completion of data collection. The field staff members consisted of experienced, professionally trained staff working for the FACES data collection efforts. The training included a wide variety of activities covering all the procedures, techniques, and contents required to carry out successful data collection in the Head Start centers, over the telephone, and elsewhere, e.g., homes:

- Lecture, incorporating slides, overheads, and videotapes;
- Exercises that simulate various procedures such as assessing classroom scheduling;
- Video demonstration of assessment techniques and components of classroom scoring procedures;
- Exercises to achieve pre-established levels of inter-rater reliability;
- Participatory involvement of all trainees in small groups so that trainers may evaluate individual performance;
- Multiple occasions of practice in real classroom settings that simulate what they are expected to do in the field, with the presence of a trainer and a small group of trainees to discuss the classroom ratings and provide valuable guidance on scoring reliability and agreement; and
- One-on-one practice and role-play in the administration of child assessment procedures under supervision of training staff.

The field procedures manual contained information about working with a research team, appropriate behaviors within Head Start classrooms and children's homes, and how to orchestrate Head Start center or other visits. Moreover, the manual covered an overview of all data collection instruments and administrative procedures. Complete scoring rules and question-by-question specifications for the child assessment, parent interviews, and classroom observation instruments were also discussed in the manual.

During the training, trainees were introduced to the purpose and goals of the study and provided background information on Head Start. Trainees were also introduced to the data collection materials and general issues regarding children and early childhood learning environments. Each day of training included a question and answer period. For administering child assessments in Spanish, a special training for English-Spanish speaking bilingual trainees was held. The bilingual trainees had an opportunity to practice assessments with Spanish-speaking children.

A.7 Data Collection Procedures

A.7.1 Site Visit Arrangements

The FACES research team obtained feasible dates for the 2-week site visit from each of the sampled Head Start programs. Site visit dates for each program were coordinated within the data collection period and programs were notified about the visit dates. Three weeks before the site visit, a scheduling packet which contained the final visit schedule, a list of sampled children by classroom, a reminder list, and a request for maps and directions to aid the research team was sent to the on-site coordinator (OSC). OSC's are members of the Head Start program staff specially designated to coordinate the data collection efforts by scheduling parent interviews, classroom visits with teachers and obtaining consent forms.

When a sampled child was a kindergartner (either Spring 2002 or Spring 2003), a field staff member would contact the parent(s), in advance, to arrange for a time to conduct the Parent Interview over the telephone. An arrangement also would be made for conducting the child assessment at a location chosen by the parent(s), usually the home. If the Parent Interview could not been conducted over the telephone, for whatever reasons, the field staff member would try to administer the Parent Interview during the visit to conduct the child assessment. Teacher-Child Report and Lead Teacher questionnaires would be sent to the child's school to be filled out and returned by the child's lead teacher.

A.8 Quality Control Visits

In FACES, Quality Control (QC) visits were built into every step of the data collection to ensure the highest quality data possible. The QC visitors consisted of the FACES research project staff who had been involved in designing the instruments, preparing the training materials, and conducting the training. The QC visitors were trained in both observation and assessment data collection and also served as technical consultants in the field. During the Fall 2000 data collection, one 3-day QC visit to several selected program sites was made.

A.9 Data Preparation and Data File Creation

A.9.1 Data Entry

Key entry and verification were performed on the study instruments using a sophisticated production data entry system. This system provides entry form layout, application of edit specifications, data verification control, and provides data entry quality and production reports.

A.9.2 Frequency Review

The frequencies of responses to all data items (both individually and in conjunction with related data items) were reviewed to ensure that appropriate skip patterns were followed. Members of the data preparation team checked each item to make sure the correct number of responses was represented for all items. If a discrepancy was discovered, the problem case was identified and reviewed.

A.9.3 Data Edit

To code and edit questionnaire data, an integrated collection of software was utilized. Through this system of software, coding manuals and codebooks were developed, data editing was performed, and SAS source code was generated.

A.9.4 Data file Creation

Data files were created and analyses performed to provide summaries and assessments of Head Start children and their families during this period and to assess the reliability and validity of information contained within the data collection instruments. Numerous derived variables were created to increase the magnitude and scope of analytical capabilities. The coding for these derived variables may be obtained upon request.

A.10 Reliability and Data Summary

In FACES, various data collection instruments were used to assess the accomplishments and behaviors of children in Head Start programs, as well as the educational and familial support that is provided to them. As noted in Section IV: Data Collection Instruments, these instruments are widely used and report mostly high reliabilities. The reliabilities for each data collection instrument and summaries for these data collection instruments are provided in the Appendix A Table A-2 through Table A-30.

Table A-2. Reliability of Fall 2000 FACES child assessment data—English language and Spanish language assessments

Scales	Fall 2000 (English language assessments)			Fall 2000 (Spanish language assessments)		
	Number of items	Number of cases	Cronbach alphas	Number of items	Number of cases	Cronbach alphas
Social Awareness	5	2,068	.63	5	385	.36
PPVT-III*[1]	144	2,508	.97	-	-	-
TVIP*	-	-	-	125	392	.92
McCarthy: Draw-A-Design	9	2,068	.58	9	375	.57
Color Names	10	2,055	.95	10	378	.92
WJR: Letter-Word Identification*[†]	23	1,273	.84	-	-	-
WM: Letter-Word Identification*	-	-	-	18	219	.75
WJR: Applied Problems*	23	1,054	.90	-	-	-
WM: Applied Problems*	-	-	-	23	219	.85
WJR Dictation*	12	1,054	.77	-	-	-
WM: Dictation*	-	-	-	12	219	.77
Story and Print Concepts: Print Conventions	2	2,116	.73	2	392	.59
Story and Print Concepts: Book Knowledge	5	2,116	.57	5	392	.43
Story and Print Concepts: Comprehension	2	2,116	.43	2	392	.39
Interviewer Rating: Assessment Behavior	8	2,021	.82	8	372	.77

[*] Raw scores were used.

[1] For Fall 2000, Spanish-speaking English Language Learners were administered the PPVT-III in English and their results are incorporated with those of other children as reflected under Fall 2000 (English Language Assessments).

[†] For Fall 2000, Spanish-speaking English Language Learners were administered the WJR: Letter-Word Identification in English and their results are incorporated with those of other children as reflected under Fall 2000 (English Language Assessments).

Table A-3. Reliability of Spring 2001 and Spring 2002 (Head Start) FACES child assessment data—English language assessments including additional Spanish language measures for Spanish-speaking English language learner children

Scales	Fall 2000 (English language assessments)			Fall 2000 (Spanish language assessments)		
	Number of items	Number of cases	Cronbach alphas	Number of items	Number of cases	Cronbach alphas
Social Awareness	5	2,304	.62	5	931	.62
PPVT-III*	144	2,344	.97	144	956	.96
TVIP*[1]	125	364	.92	125	113	.94
McCarthy: Draw-A-Design	9	2,298	.70	9	928	.72
Leiter-R AS – Age 3	4	406	.71	-	-	-
Leiter-R AS - Ages 4 to 5	4	1,763	.81	-	-	-
Leiter-R AS - Ages 4 to 6	-	-	-	4	920	.80
Color Names	10	2,298	.94	10	942	.90
WJR: Letter-Word Identification*	23	1,902	.86	23	955	.86
WM: Letter-Word Identification*[†]	18	307	.78	18	113	.83
WJR: Applied Problems*	23	1,902	.91	23	955	.89
WJR Dictation*	12	1,902	.77	12	955	.71
Story and Print Concepts: Print Conventions	2	2,344	.75	2	956	.84
Story and Print Concepts: Book Knowledge	5	2,344	.59	5	956	.61
Story and Print Concepts: Comprehension	2	2,344	.42	2	956	.40
Interviewer Rating: Assessment Behavior	8	2,254	.80	8	914	.70

[*] Raw scores were used.

[1] For Spring 2001 and Spring 2002 (Head Start), Spanish-speaking English Language Learners were administered the TVIP in Spanish to gauge progress in Spanish receptive vocabulary knowledge from earlier data collection periods.

[†] For Spring 2001 and Spring 2002 (Head Start), Spanish-speaking English Language Learners were administered the WM: Letter-Word Identification in Spanish to gauge progress in Spanish letter/word knowledge from earlier data collection periods.

Table A-4. Reliability of Spring 2002/2003 (kindergarten) FACES child assessment data—English language assessments

Scales	Spring 2002/2003 (kindergarten)		
	Number of items	Number of cases	Cronbach alphas
Social Awareness	7	1,833	.65
PPVT-III*	204	1,913	.97
TOLD – Phonemic Analysis	14	1,913	.96
Reading – ECLS-K Routing*	20	1,913	.87
Red (Low Form)*	18	1,913	.95
Yellow (Middle Form)*	29	1,913	.94
Blue (High Form)*	29	1,913	.94
WJR: Applied Problems*	31	1,912	.88
WJR: Dictation*	20	1,912	.77
Writing Name	1	Not Applicable	Not Applicable
General Knowledge – ECLS-K Routing**	16	1,913	.77
Interviewer Rating: Assessment Behavior	8	1,816	.75

*Raw scores were used.

**Scale scores were used.

Table A-5. Summary statistics for Fall 2000 FACES child assessment data—English language and Spanish language assessments

Scales	Fall 2000 (English language assessments)					Fall 2000 (Spanish language assessments)				
	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Social Awareness	2,101	3.36	1.69	0 - 6	0 - 6	390	2.62	1.21	0 - 6	0 - 6
PPVT-III*1	2,379	31.86	18.42	0 - 98	0 - 144	-	-	-	-	-
TVIP*	-	-	-	-	-	369	11.34	8.38	1 - 47	0 - 125
McCarthy: Draw-A-Design	2,112	2.92	1.33	0 - 13	0 - 19	392	3.37	1.34	0 - 13	0 - 19
Color Names	2,101	11.32	7.37	0 - 20	0 - 20	386	8.90	6.62	0 - 20	0 - 20
WJR: Letter-Word Identification*2	1,144	5.13	2.49	0 - 21	0 - 23	-	-	-	-	-
WM: Letter-Word Identification*	-	-	-	-	-	195	4.37	1.20	0 - 10	0 - 18
WJR: Applied Problems*	963	7.52	4.36	0 - 21	0 - 23	-	-	-	-	-
WM: Applied Problems*	-	-	-	-	-	200	5.29	3.40	0 - 14	0 - 23
WJR: Dictation*	916	5.11	1.83	0 - 12	0 - 12	-	-	-	-	-
WM: Dictation*	-	-	-	-	-	188	4.99	1.28	1 - 11	0 - 12
Story and Print Concepts: Print Conventions	2,089	0.23	0.57	0 - 2	0 - 2	391	0.17	0.53	0 - 3	0 - 2
Story and Print Concepts: Book Knowledge	2,087	1.62	1.27	0 - 5	0 - 5	376	1.25	1.13	0 - 5	0 - 5
Story and Print Concepts: Comprehension	2,102	0.54	0.70	0 - 2	0 - 2	386	0.47	0.67	0 - 2	0 - 2
Interviewer Rating: Assessment Behavior	2,094	17.14	5.02	0 - 24	0 - 24	383	17.54	4.41	0 - 24	0 - 24

[*] Raw scores were used.

[1] See Table Note 1 under Table A-2a.

[2] See Table Note 2 under Table A-2a.

Table A-6. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment data—children who were assessed in English in both Fall and Spring

	Fall 2000					Spring 2001				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
PPVT-III*	1,811	35.23	17.73	0 – 98	0 – 144	1,811	45.80	18.62	3 – 98	0 – 144
WJR: Letter-Word Identification*	839	5.32	2.60	0 – 21	0 – 23	839	7.26	3.35	0 – 22	0 – 23
WJR: Applied Problems*	867	7.58	4.35	0 – 21	0 – 23	867	10.47	4.65	0 – 22	0 – 23
WJR: Dictation*	805	5.13	1.84	0 – 12	0 – 12	805	6.36	2.11	0 – 12	0 – 12
McCarthy: Draw-A-Design	1,925	2.93	1.34	0 – 13	0 – 19	1,925	3.54	1.70	0 – 15	0 – 19
Story and Print Concepts: Book Knowledge	1,886	1.63	1.28	0 – 5	0 – 5	1,886	2.43	1.30	0 – 5	0 – 5
Story and Print Concepts: Comprehension	1,906	0.54	0.70	0 – 2	0 – 2	1,906	0.72	0.75	0 – 2	0 – 2
Story and Print Concepts: Print Knowledge	1,895	0.24	0.58	0 – 2	0 – 2	1,895	0.38	0.70	0 – 2	0 – 2
Social Awareness	1,904	3.38	1.69	0 – 6	0 – 6	1,904	4.01	1.57	0 – 6	0 – 6
Color Names	1,904	11.38	7.38	0 – 20	0 – 20	1,904	15.69	5.89	0 – 20	0 – 20
One-to-One Counting	1,859	2.52	1.24	1 – 5	1 – 5	1,859	3.27	1.38	1 – 5	1 – 5
Interviewer Rating: Assessment Behavior	1,878	17.13	5.03	0 – 24	0 – 24	1,878	19.12	4.31	0 – 24	0 – 24

*Raw scores were used.

Table A-7. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—children who were assessed in English in both Fall and Spring

Scales	Fall 2000			Spring 2001		
	Number of cases	Mean	SD	Number of cases	Mean	SD
PPVT-III Standard Score	1,811	85.36	14.69	1,811	89.41	14.28
WJR: Letter-Word Identification Standard Score	837	92.39	9.39	837	92.64	11.59
WJR: Applied Problems Standard Score	863	88.16	15.29	863	89.33	17.23
WJR: Dictation Standard Score	803	85.10	13.93	803	87.29	14.23
PPVT-III W Ability Score	1,811	64.85	12.60	1,811	71.83	11.36
WJR: Letter-Word Identification W Ability Score	839	356.94	14.36	839	367.45	17.77
WJR: Applied Problems W Ability Score	867	402.10	24.58	867	415.73	21.84
WJR: Dictation W Ability Score	805	348.89	33.93	805	369.62	34.31

Table A-8. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—3-year old children who were assessed in English in both Fall and Spring

Scales	Fall 2000 Number of cases	Fall 2000 Mean	Fall 2000 SD	Spring 2001 Number of cases	Spring 2001 Mean	Spring 2001 SD
PPVT-III Standard Score	922	82.53	14.33	922	87.17	14.30
WJR: Letter-Word Identification Standard Score	0	--	--	0	--	--
WJR: Applied Problems Standard Score	0	--	--	0	--	--
WJR: Dictation Standard Score	0	--	--	0	--	--
PPVT-III W Ability Score	922	59.08	11.54	922	66.74	10.44
WJR: Letter-Word Identification W Ability Score	1	376.00	--	1	386.00	--
WJR: Applied Problems W Ability Score	1	441.00	--	1	435.00	--
WJR: Dictation W Ability Score	1	416.00	--	1	406.00	--

Table A-9. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—4-year old or older children who were assessed in English in both Fall and Spring

Scales	Fall 2000			Spring 2001		
	Number of cases	Mean	SD	Number of cases	Mean	SD
PPVT-III Standard Score	889	88.30	14.50	889	91.73	13.90
WJR: Letter-Word Identification Standard Score	837	92.39	9.39	837	92.64	11.59
WJR: Applied Problems Standard Score	863	88.16	15.29	863	89.33	17.23
WJR: Dictation Standard Score	803	85.10	13.93	803	87.29	14.23
PPVT-III W Ability Score	889	70.83	10.73	889	77.12	9.75
WJR: Letter-Word Identification W Ability Score	838	356.92	14.35	838	367.43	17.77
WJR: Applied Problems W Ability Score	866	402.05	24.55	866	415.71	21.84
WJR: Dictation W Ability Score	804	348.80	33.87	804	369.57	34.31

Table A-10. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001

Scales	Fall 2000 Number of cases	Fall 2000 Mean	Fall 2000 SD	Fall 2000 Reported response range	Fall 2000 Possible response range	Spring 2001 Number of cases	Spring 2001 Mean	Spring 2001 SD	Spring 2001 Reported response range	Spring 2001 Possible response range
TVIP *	300	10.82	7.90	1 – 43	0 – 125	300	16.10	9.80	1 – 44	0 – 125
WM: Letter-Word Identification*	175	4.35	1.18	0 – 10	0 – 23	175	5.15	1.76	3 – 12	0 – 23
WM: Applied Problems*	184	5.37	3.36	0 – 14	0 – 23	--	--	--	--	--
WM: Dictation*	172	4.96	1.25	1 – 11	0 – 12	--	--	--	--	--
McCarthy: Draw-A-Design (SP)	361	3.38	1.32	0 – 13	0 – 19	--	--	--	--	--
Story and Print Concepts: Book Knowledge (SP)	338	1.24	1.15	0 – 5	0 – 5	--	--	--	--	--
Story and Print Concepts: Comprehension	352	0.48	0.68	0 – 2	0 – 2	--	--	--	--	--
Story and Print Concepts: Print Knowledge	356	0.17	0.55	0 – 3	0 – 3	--	--	--	--	--
Social Awareness (SP)	355	2.63	1.21	0 – 6	0 – 6	--	--	--	--	--
Color Names (SP)	354	9.05	6.61	0 – 20	0 – 20	--	--	--	--	--
One-to-One Counting (SP)	338	2.22	1.12	1 – 5	1 – 5	--	--	--	--	--
PPVT-III*	306	13.46	9.72	0 – 49	0 – 144	306	23.28	13.26	1 – 67	0 – 144

Table A-10. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001 (continued)

	Fall 2000					Spring 2001				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
WJR: Letter-Word Identification*	179	4.32	1.49	0 – 13	0 – 23	179	5.50	2.29	3 – 14	0 – 23
WJR: Applied Problems*	--	--	--	--	--	291	5.83	4.13	0 – 19	0 – 23
WJR: Dictation*	--	--	--	--	--	290	5.72	1.74	0 – 11	0 – 12
McCarthy: Draw-A-Design (ENG)	--	--	--	--	--	361	4.06	1.91	0 – 12	0 – 19
Story and Print Concepts: Book Knowledge (ENG)	--	--	--	--	--	338	1.72	1.13	0 – 5	0 – 5
Story and Print Concepts: Comprehension	--	--	--	--	--	352	0.50	0.68	0 – 2	0 – 2
Story and Print Concepts: Print Knowledge	--	--	--	--	--	356	0.14	0.46	0 – 2	0 – 2
Social Awareness (ENG)	--	--	--	--	--	355	2.57	1.30	0 – 6	0 – 6
Color Names (ENG)	--	--	--	--	--	354	13.49	6.47	0 – 20	0 – 20
One-to-One Counting (ENG)	--	--	--	--	--	338	3.03	1.44	1 – 5	1 – 5
Interviewer Rating: Assessment Behavior	348	17.66	4.43	0 – 24	0 – 24	348	18.01	3.49	3 – 24	0 – 24

*Raw scores were used.

(SP) = Spanish translation; (ENG) = English translation.

Table A-11. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001

	Fall 2000			Spring 2001		
Scales	Number of cases	Mean	SD	Number of cases	Mean	SD
TVIP Standard Score	300	85.14	11.81	300	84.19	13.04
WM: Letter-Word Identification Standard Score	174	89.67	5.54	174	86.18	6.58
WM: Applied Problems Standard Score	184	81.86	13.70	--	--	--
WM: Dictation Standard Score	172	85.31	8.81	--	--	--
WM: Letter-Word Identification W Ability Score	175	351.78	7.27	175	356.40	9.26
WM: Applied Problems W Ability Score	184	391.53	22.29	--	--	--
WM: Dictation W Ability Score	172	347.01	20.28	--	--	--
PPVT-III Standard Score	306	59.80	13.55	306	66.75	14.37
WJR: Letter-Word Identification Standard Score	178	89.53	6.37	178	87.38	8.00
WJR: Applied Problems Standard Score	--	--	--	290	76.03	19.93
WJR: Dictation Standard Score	--	--	--	289	86.66	11.80
PPVT-III W Ability Score	306	45.80	12.58	306	56.01	11.71
WJR: Letter-Word Identification W Ability Score	179	351.45	9.05	179	358.18	12.12
WJR: Applied Problems W Ability Score	--	--	--	291	391.65	27.12
WJR: Dictation W Ability Score	--	--	--	290	358.90	29.40

Table A-12. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—3-year old children who were assessed in Spanish in Fall 2000 and English in Spring 2001

Scales	Fall 2000			Spring 2001		
	Number of cases	Mean	SD	Number of cases	Mean	SD
TVIP Standard Score	142	89.34	9.84	142	86.15	11.86
WM: Letter-Word Identification Standard Score	--	--	--	--	--	--
WM: Applied Problems Standard Score	--	--	--	--	--	--
WM: Dictation Standard Score	--	--	--	--	--	--
WM: Letter-Word Identification W Ability Score	--	--	--	--	--	--
WM: Applied Problems W Ability Score	--	--	--	--	--	--
WM: Dictation W Ability Score	--	--	--	--	--	--
PPVT-III Standard Score	144	59.58	13.88	144	67.21	14.79
WJR: Letter-Word Identification Standard Score	--	--	--	--	--	--
WJR: Applied Problems Standard Score	--	--	--	--	--	--
WJR: Dictation Standard Score	--	--	--	--	--	--
PPVT-III W Ability Score	144	41.07	13.34	144	52.23	12.03
WJR: Letter-Word Identification W Ability Score	--	--	--	--	--	--
WJR: Applied Problems W Ability Score	--	--	--	--	--	--
WJR: Dictation W Ability Score	--	--	--	--	--	--

Table A-13. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—4-year old and older children who were assessed in Spanish in Fall 2000 and English in Spring 2001

Scales	Fall 2000			Spring 2001		
	Number of cases	Mean	SD	Number of cases	Mean	SD
TVIP Standard Score	158	81.37	12.18	158	82.42	13.80
WM: Letter-Word Identification Standard Score	174	89.67	5.54	174	86.18	6.58
WM: Applied Problems Standard Score	184	81.86	13.70	--	--	--
WM: Dictation Standard Score	172	85.31	8.81	--	--	--
WM: Letter-Word Identification W Ability Score	175	351.78	7.27	175	356.40	9.26
WM: Applied Problems W Ability Score	184	391.53	22.29	--	--	--
WM: Dictation W Ability Score	172	347.01	20.28	--	--	--
PPVT-III Standard Score	162	60.01	13.28	162	66.33	14.02
WJR: Letter-Word Identification Standard Score	178	89.53	6.37	178	87.38	8.00
WJR: Applied Problems Standard Score	--	--	--	199	74.05	21.68
WJR: Dictation Standard Score	--	--	--	199	86.03	12.63
PPVT-III W Ability Score	162	50.00	10.20	162	59.37	10.35
WJR: Letter-Word Identification W Ability Score	179	351.45	9.05	179	358.18	12.12
WJR: Applied Problems W Ability Score	--	--	--	200	395.63	27.16
WJR: Dictation W Ability Score	--	--	--	200	363.86	30.65

Table A-14. Summary statistics for Spring 2001 and Spring 2002 (Head Start) FACES child assessment data—English language assessments including additional Spanish language measures for Spanish-Speaking English language learner children

	Spring 2001					Spring 2002 (Head Start)				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Social Awareness	2,327	3.76	1.63	0 - 6	0 - 6	950	4.55	1.35	0 - 6	0 - 6
PPVT-III*	2,278	41.88	19.72	1 - 98	0 - 144	925	53.93	17.77	4 - 109	0 - 144
TVIP*[1]	322	16.27	10.04	1 - 48	0 - 125	88	23.01	13.04	4 - 62	0 - 125
McCarthy: Draw-A-Design	2,344	3.61	1.74	0 - 15	0 - 19	955	4.55	1.96	1 - 13	0 - 19
Leiter-R AS - Age 3	420	41.71	12.20	1 - 64	0 - 71	-	-	-	-	-
Leiter-R AS - Ages 4 to 5	1,833	40.49	10.43	4 - 70	0 - 71	-	-	-	-	-
Leiter-R AS - Ages 4 to 6	-	-	-	-	-	933	45.92	10.19	6 - 71	0 - 71
Color Names	2,331	15.26	6.11	0 - 20	0 - 20	954	18.24	3.56	0 - 20	0 - 20
WJR: Letter-Word Identification*	1,808	6.37	3.09	0 - 22	0 - 23	940	7.95	3.48	0 - 23	0 - 23
WM: Letter-Word Identification*[2]	295	5.01	1.68	0 - 12	0 - 18	111	5.85	2.32	1 - 16	0 - 18
WJR: Applied Problems*	1,836	8.47	4.76	0 - 22	0 - 23	949	10.90	4.47	0 - 23	0 - 23
WJR: Dictation*	1,784	5.65	2.05	0 - 12	0 - 12	934	6.42	2.08	0 - 12	0 - 12
Story and Print Concepts: Print Conventions	2,328	0.34	0.67	0 - 2	0 - 2	949	0.58	0.84	0 - 2	0 - 2
Story and Print Concepts: Book Knowledge	2,316	2.30	1.30	0 - 5	0 - 5	950	2.74	1.43	0 - 5	0 - 5
Story and Print Concepts: Comprehension	2,327	0.68	0.74	0 - 2	0 - 2	952	0.93	0.78	0 - 2	0 - 2
Interviewer Rating: Assessment Behavior	2,310	18.89	4.27	0 - 24	0 - 24	938	20.12	3.24	4 - 24	0 - 24

* Raw scores were used.
[1] See Table Note 1 under Table A-2b.
[2] See Table Note 2 under Table A-2b.

Table A-15. Summary statistics for Spring 2002/2003 (kindergarten) FACES child assessment data—English language assessments

Scales	Spring 2002/2003 (kindergarten)				
	Number of cases	Mean	SD	Reported response range	Possible response range
Social Awareness	1,884	5.98	1.68	0 - 8	0 - 8
PPVT-III*	1,856	71.93	17.63	14 - 136	0 - 204
TOLD – Phonemic Analysis	1,911	8.54	5.27	0 - 14	0 - 14
Reading – ECLS-K – Routing and Second Stage Sections**	1,908	30.69	9.96	9.14 - 68.56	0 - 72
WJR: Applied Problems*	1,867	17.34	4.53	1 - 29	0 - 31
WJR: Dictation*	1,842	11.48	2.75	1 - 19	0 - 20
Writing Name	1,894	1.89	0.35	0 - 2	0 - 2
General Knowledge – ECLS-K Routing**	1,887	24.30	8.28	8.47 - 46.77	0 - 51
Interviewer Rating: Assessment Behavior	1,858	19.63	3.68	1 - 24	0 - 24

* Raw scores were used.

** Scale scores were used.

Table A-16. Summary Statistics for Fall 2000 and Spring 2001 FACES child assessment data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001

	Fall 2000					Spring 2001				
Scales	number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
TVIP *	300	10.82	7.90	1 – 43	0 – 125	300	16.10	9.80	1 – 44	0 – 125
WM: Letter-Word Identification*	175	4.35	1.18	0 – 10	0 – 23	175	5.15	1.76	3 – 12	0 – 23
WM: Applied Problems*	184	5.37	3.36	0 – 14	0 – 23	--	--	--	--	--
WM: Dictation*	172	4.96	1.25	1 – 11	0 – 12	--	--	--	--	--
McCarthy: Draw-A-Design (SP)	361	3.38	1.32	0 – 13	0 – 19	--	--	--	--	--
Story and Print Concepts: Book Knowledge (SP)	338	1.24	1.15	0 – 5	0 – 5	--	--	--	--	--
Story and Print Concepts: Comprehension	352	0.48	0.68	0 – 2	0 – 2	--	--	--	--	--
Story and Print Concepts: Print Knowledge	356	0.17	0.55	0 – 3	0 – 3	--	--	--	--	--
Social Awareness (SP)	355	2.63	1.21	0 – 6	0 – 6	--	--	--	--	--
Color Names (SP)	354	9.05	6.61	0 – 20	0 – 20	--	--	--	--	--
One-to-One Counting (SP)	338	2.22	1.12	1 – 5	1 – 5	--	--	--	--	--
PPVT-III*	306	13.46	9.72	0 – 49	0 – 144	306	23.28	13.26	1 – 67	0 – 144

Table A-17. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001 (continued)

	Fall 2000					Spring 2001				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
WJR: Letter-Word Identification*	179	4.32	1.49	0 – 13	0 – 23	179	5.50	2.29	3 – 14	0 – 23
WJR: Applied Problems*	--	--	--	--	--	291	5.83	4.13	0 – 19	0 – 23
WJR: Dictation*	--	--	--	--	--	290	5.72	1.74	0 – 11	0 – 12
McCarthy: Draw-A-Design (ENG)	--	--	--	--	--	361	4.06	1.91	0 – 12	0 – 19
Story and Print Concepts: Book Knowledge (ENG)	--	--	--	--	--	338	1.72	1.13	0 – 5	0 – 5
Story and Print Concepts: Comprehension	--	--	--	--	--	352	0.50	0.68	0 – 2	0 – 2
Story and Print Concepts: Print Knowledge	--	--	--	--	--	356	0.14	0.46	0 – 2	0 – 2
Social Awareness (ENG)	--	--	--	--	--	355	2.57	1.30	0 – 6	0 – 6
Color Names (ENG)	--	--	--	--	--	354	13.49	6.47	0 – 20	0 – 20
One-to-One Counting (ENG)	--	--	--	--	--	338	3.03	1.44	1 – 5	1 – 5
Interviewer Rating: Assessment Behavior	348	17.66	4.43	0 – 24	0 – 24	348	18.01	3.49	3 – 24	0 – 24

*Raw scores were used.

(SP) = Spanish translation; (ENG) = English translation.

Table A-18. Summary statistics for Fall 2000 and Spring 2001 FACES child assessment norm-referenced data—children who were assessed in Spanish in Fall 2000 and English in Spring 2001

	Fall 2000			Spring 2001		
Scales	Number of cases	Mean	SD	Number of cases	Mean	SD
TVIP Standard Score	300	85.14	11.81	300	84.19	13.04
WM: Letter-Word Identification Standard Score	174	89.67	5.54	174	86.18	6.58
WM: Applied Problems Standard Score	184	81.86	13.70	--	--	--
WM: Dictation Standard Score	172	85.31	8.81	--	--	--
WM: Letter-Word Identification W Ability Score	175	351.78	7.27	175	356.40	9.26
WM: Applied Problems W Ability Score	184	391.53	22.29	--	--	--
WM: Dictation W Ability Score	172	347.01	20.28	--	--	--
PPVT-III Standard Score	306	59.80	13.55	306	66.75	14.37
WJR: Letter-Word Identification Standard Score	178	89.53	6.37	178	87.38	8.00
WJR: Applied Problems Standard Score	--	--	--	290	76.03	19.93
WJR: Dictation Standard Score	--	--	--	289	86.66	11.80
PPVT-III W Ability Score	306	45.80	12.58	306	56.01	11.71
WJR: Letter-Word Identification W Ability Score	179	351.45	9.05	179	358.18	12.12
WJR: Applied Problems W Ability Score	--	--	--	291	391.65	27.12
WJR: Dictation W Ability Score	--	--	--	290	358.90	29.40

Table A-19. Reliability of Fall 2000 and Spring 2001 FACES classroom observation data selected measures

Scales	Fall 2000 Number of items	Fall 2000 Number of cases	Fall 2000 Cronbach alphas	Spring 2001 Number of items	Spring 2001 Number of cases	Spring 2001 Cronbach alphas
Assessment Profile: Scheduling	14	227	.89	14	243	.87
Assessment Profile: Learning Environment	18	228	.68	18	228	.77
Assessment Profile: Individualizing	5	250	.50	5	250	.54
ECERS (Total)	37	270	.92	37	235	.92
Personal Care	6	146	.73	6	269	.70
Furnishings	4	263	.52	4	263	.60
Language	4	260	.77	4	272	.76
Motor Skills	4	248	.67	4	274	.64
Creative	6	253	.60	6	262	.71
Social	4	264	.86	4	269	.91
Program Structure	4	256	.60	4	261	.69
Arnett Scale of Caregiver Behavior: Lead Teacher (Total)	30	256	.94	30	258	.94
Sensitivity	10	262	.94	10	266	.94
Harshness	9	265	.83	9	271	.82
Detachment	4	266	.71	4	272	.82
Permissiveness	3	266	.52	3	274	.53
Independence	4	263	.58	4	269	.56

Table A-20. Reliability of Spring 2002 (Head Start) FACES classroom observation data selected measures

Scales	Spring 2002 (Head Start)		
	Number of items	Number of cases	Cronbach alphas
Assessment Profile: Scheduling	14	178	.82
Assessment Profile: Learning Environment	18	180	.65
Assessment Profile: Individualizing	5	195	.44
ECERS (Total)	37	191	.89
Personal Care	6	200	.62
Furnishings	4	202	.50
Language	4	204	.74
Motor Skills	4	202	.59
Creative	6	202	.73
Social	4	202	.88
Program Structure	4	200	.67
Arnett Scale of Caregiver Behavior: Lead Teacher (Total)	30	191	.93
Sensitivity	10	199	.93
Harshness	9	197	.76
Detachment	4	200	.71
Permissiveness	3	200	.53
Independence	4	197	.59

Table A-21. Summary statistics for Fall 2000 and Spring 2001 FACES classroom observation data selected measures

	Fall 2000					Spring 2001				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Assessment Profile: Scheduling	270	11.12	3.20	0 - 14	0 - 14	270	11.00	3.15	3 - 14	0 - 14
Assessment Profile: Learning Environment	270	14.44	2.57	5 - 18	0 - 18	274	14.22	2.92	4 - 18	0 - 18
Assessment Profile: Individualizing	260	3.58	1.15	0 - 5	0 - 5	271	3.42	1.11	0 - 5	0 - 5
ECERS (Total) (Mean)	258	4.84	0.87	1.8 - 6.7	1.0 - 7.0	265	4.91	1.00	1.9 - 6.9	1.0 - 7.0
Personal Care (Mean)	261	5.22	1.33	1.0 - 7.0	1.0 - 7.0	273	4.84	1.63	1.0 - 7.0	1.0 -7.0
Furnishings (Mean)	268	5.48	1.01	1.3 - 7.0	1.0- 7.0	266	5.52	1.02	2.3 - 7.0	1.0 -7.0
Language (Mean)	270	4.86	1.20	1.0 - 7.0	1.0 -7.0	274	5.01	1.27	1.0 - 7.0	1.0 -7.0
Motor Skills (Mean)	264	4.79	1.27	1.3 - 7.0	1.0 -7.0	277	4.89	1.34	1.3 - 7.0	1.0 -7.0
Creative (Mean)	268	4.31	0.90	2.0 - 6.3	1.0 -7.0	277	4.26	1.02	1.8 - 6.7	1.0 -7.0
Social (Mean)	266	5.39	1.38	1.3 - 7.0	1.0 -7.0	276	5.59	1.56	1.0 - 7.0	1.0 -7.0
Program Structure (Mean)	268	4.87	1.24	1.0 - 7.0	1.0 -7.0	276	5.12	1.28	1.0 - 7.0	1.0 -7.0
Arnett Scale of Caregiver Behavior: Lead Teacher (Total)	267	71.48	12.42	26 - 90	0 - 90	274	72.84	12.99	20 - 89	0 - 90
Sensitivity	262	20.99	6.52	1 - 30	0 - 30	266	21.74	6.42	3 - 30	0 - 30
Harshness	265	24.28	3.22	4 - 27	0 - 27	271	24.34	3.16	6 - 27	0 - 27
Detachment	266	11.06	1.54	4 - 12	0 - 12	272	11.08	1.79	2 - 12	0 - 12
Permissiveness	266	7.47	1.37	1 – 9	0 -9	274	7.39	1.45	2 - 9	0 - 9
Independence	263	7.77	2.31	1 - 12	0 -12	269	8.39	2.32	1 - 12	0 - 12

Table A-22. Summary statistics for Spring 2002 (Head Start) FACES classroom observation data selected measures

	Spring 2002 (Head Start)				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range
Assessment Profile: Scheduling	198	11.39	2.56	4 - 14	0 - 14
Assessment Profile: Learning Environment	203	14.63	2.48	5 - 18	0 - 18
Assessment Profile: Individualizing	202	3.52	0.96	0 - 5	0 - 5
ECERS (Total) (Mean)	202	4.84	0.86	1.8 - 6.7	1.0 -7.0
Personal Care (Mean)	203	4.82	1.44	1.0 - 7.0	1.0 -7.0
Furnishings (Mean)	202	5.55	0.92	2.3 - 7.0	1.0 -7.0
Language (Mean)	204	5.06	1.13	1.0 - 7.0	1.0 -7.0
Motor Skills (Mean)	204	4.70	1.27	1.0 - 7.0	1.0 -7.0
Creative (Mean)	204	4.12	0.97	1.7 - 6.7	1.0 -7.0
Social (Mean)	204	5.58	1.39	1.0 - 7.0	1.0 -7.0
Program Structure (Mean)	203	5.02	1.17	1.0 - 7.0	1.0 -7.0
Arnett Scale of Caregiver Behavior: Lead Teacher (Total)	201	73.65	11.81	22 - 90	0 - 90
Sensitivity	199	21.95	6.29	1 - 30	0 - 30
Harshness	197	24.43	2.83	10 - 27	0 - 27
Detachment	200	11.12	1.51	3 - 12	0 - 12
Permissiveness	200	7.60	1.35	0 - 9	0 - 9
Independence	197	8.66	2.13	2 - 12	0 - 12

Table A-23. Reliability of Fall 2000 and Spring 2001 FACES teacher's child report data selected measures

Scales	Fall 2000 Number of items	Fall 2000 Number of cases	Fall 2000 Cronbach alphas	Spring 2001 Number of items	Spring 2001 Number of cases	Spring 2001 Cronbach alphas
Social Skills	12	2,522	.88	12	2,254	.88
Behavioral Problems (Total)	14	2,522	.86	14	2,254	.86
Withdrawn	7	2,522	.77	7	2,254	.76
Aggressive	4	2,522	.83	4	2,254	.85
Hyperactive	3	2,522	.72	3	2,254	.72
Child Observation Record (Total)	14	2,522	.94	14	2,254	.94
Social Relationships	3	2,522	.83	3	2,254	.83
Creative Representations	3	2,522	.80	3	2,254	.81
Music and Movement	4	2,522	.88	4	2,254	.88
Cognitive	4	2,522	.82	4	2,254	.83

Table A-24. Reliability of Spring 2002 (Head Start) and Spring 2002/2003 (kindergarten) FACES teacher's child report data selected measures

	Spring 2002 (Head Start)			Spring 2002/2003 (kindergarten)		
Scales	Number of items	Number of cases	Cronbach alphas	Number of items	Number of cases	Cronbach alphas
Social Skills	12	963	.87	12	1,612	.88
Behavioral Problems (Total)	14	963	.87	14	1,612	.86
Withdrawn	7	963	.76	7	1,612	.77
Aggressive	4	963	.83	4	1,612	.85
Hyperactive	3	963	.75	3	1,612	.74
Child Observation Record (Total)	14	963	.93	-	-	-
Social Relationships	3	963	.80	-	-	-
Creative Representations	3	963	.80	-	-	-
Music and Movement	4	963	.86	-	-	-
Cognitive	4	963	.80	-	-	-
Literacy	-	-	-	8	1,533	.95

Table A-25. Summary statistics for Fall 2000 and Spring 2001 FACES teacher's child report data selected measures

	Fall 2000					Spring 2001				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Social Skills	2,518	14.52	4.85	0 - 24	0 - 24	2,251	16.58	4.63	0 - 24	0 - 24
Behavioral Problems (Total)	2,484	5.76	5.06	0 - 28	0 - 28	2,215	5.30	4.89	0 - 27	0 - 28
Withdrawn	2,453	2.59	2.65	0 - 14	0 - 14	2,186	2.34	2.49	0 -14	0 - 14
Aggressive	2,449	1.80	2.04	0 - 8	0 - 8	2,189	1.74	2.03	0 - 8	0 - 8
Hyperactive	2,466	1.37	1.53	0 - 6	0 - 6	2,194	1.21	1.47	0 - 6	0 - 6
Child Observation Record (Total) (Mean)	2,403	2.82	0.77	1.0 - 5.0	1.0 - 5.0	2,174	3.50	0.81	1.0 - 5.0	1.0 - 5.0
Social Relationships (Mean)	2,464	2.83	1.01	1.0 - 5.0	1.0 - 5.0	2,230	3.51	0.99	1.0 - 5.0	1.0 - 5.0
Creative Representations (Mean)	2,459	2.83	0.83	1.0 - 5.0	1.0 - 5.0	2,230	3.52	0.87	1.0 - 5.0	1.0 - 5.0
Music and Movement (Mean)	2,446	2.95	0.85	1.0 - 5.0	1.0 - 5.0	2,217	3.65	0.89	1.0 - 5.0	1.0 - 5.0
Cognitive (Mean)	2,403	2.75	0.82	1.0 - 5.0	1.0 - 5.0	2,184	3.41	0.89	1.0 - 5.0	1.0 - 5.0

Table A-26. Summary statistics for Spring 2002 (Head Start) and Spring 2002/2003 (kindergarten) FACES teacher's child report data selected measures

	Spring 2002 (Head Start)					Spring 2002/2003 (kindergarten)				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Social Skills	961	18.12	4.28	0 - 24	0 - 24	1,602	17.58	4.68	0 - 24	0 - 24
Behavioral Problems (Total)	951	4.52	4.72	0 - 27	0 - 28	1,605	5.69	5.30	0 - 24	0 - 28
Withdrawn	937	2.05	2.40	0 - 14	0 - 14	1,585	2.60	2.77	0 - 14	0 - 14
Aggressive	937	1.49	1.93	0 - 8	0 - 8	1,594	1.62	2.04	0 – 8	0 - 8
Hyperactive	945	0.97	1.40	0 - 6	0 - 6	1,600	1.46	1.69	0 – 6	0 - 6
Child Observation Record (Total) (Mean)	932	4.06	0.70	1.1 - 5.0	1.0 - 5.0	-	-	-	-	-
Social Relationships (Mean)	954	4.00	0.89	1.0 - 5.0	1.0 - 5.0	-	-	-	-	-
Creative Representations (Mean)	958	4.06	0.79	1.0 - 5.0	1.0 - 5.0	-	-	-	-	-
Music and Movement (Mean)	952	4.24	0.75	1.0 - 5.0	1.0 - 5.0	-	-	-	-	-
Cognitive (Mean)	939	3.96	0.80	1.3 - 5.0	1.0 - 5.0	-	-	-	-	-
Literacy (Mean)	-	-	-	-	-	1,602	2.58	1.12	0.0 – 4.0	0.0 – 4.0

Table A-27. Reliability of Fall 2000 and Spring 2001 FACES parent interview data selected measures

	Fall 2000			Spring 2001		
Scales	Number of items	Number of cases	Cronbach alphas	Number of items	Number of cases	Cronbach alphas
Pearlin Mastery Scale (Locus of Control)	7	2,485	.76	7	2,290	.78
CES-D Depression Scale	12	2,485	.87	12	2,290	.86
Positive Approaches to Learning	7	2,485	.61	7	2,290	.62
Behavior Problems Index (Total)	12	2,485	.71	12	2,290	.75
Behavior Problems (Aggressive)	4	2,485	.59	4	2,290	.65
Behavior Problems (Hyperactive)	3	2,485	.55	3	2,290	.59
Behavior Problems (Withdrawn)	5	2,485	.36	5	2,290	.40
Combined Activities Scale	22	2,419	.69	22	2,210	.69
Weekly Activities Subscale	11	2,441	.60	11	2,231	.62
Monthly Activities Subscale	11	2,459	.58	11	2,267	.58
Emerging Literacy Scale	5	2,469	.60	5	2,273	.62

Table A-28. Reliability of Spring 2002 (Head Start) and Spring 2002/2003 (kindergarten) FACES parent interview data selected measures

Scales	Spring 2002 (Head Start)			Spring 2002/2003 (kindergarten)		
	Number of items	Number of cases	Cronbach alphas	Number of items	Number of cases	Cronbach alphas
Pearlin Mastery Scale (Locus of Control)	7	906	.76	7	2,035	.80
CES-D Depression Scale	12	907	.87	12	2,035	.87
Positive Approaches to Learning	7	907	.66	13	1,999	.83
Behavior Problems Index (Total)	12	907	.76	13	2,015	.80
Behavior Problems (Aggressive)	4	907	.63	5	2,032	.76
Behavior Problems (Hyperactive)	3	907	.61	3	2,030	.56
Behavior Problems (Withdrawn)	3	907	.39	5	2,021	.56
Combined Activities Scale	22	873	.71	17	2,024	.66
Weekly Activities Subscale	11	884	.61	7	2,031	.52
Monthly Activities Subscale	11	895	.64	10	2,028	.57
Emerging Literacy Scale	5	902	.57	5	2,036	.56

Table A-29. Summary statistics for Fall 2000 and Spring 2001 FACES parent interview data selected measures

Scales	Fall 2000 Number of cases	Fall 2000 Mean	Fall 2000 SD	Fall 2000 Reported response range	Fall 2000 Possible response range	Spring 2001 Number of cases	Spring 2001 Mean	Spring 2001 SD	Spring 2001 Reported response range	Spring 2001 Possible response range
Pearlin Mastery Scale (Locus of Control)	2,471	14.80	3.28	0 - 21	0 - 21	2,284	15.30	3.31	3 - 21	0 - 21
CES-D Depression Scale	2,473	6.81	6.72	0 - 36	0 - 36	2,285	6.59	6.36	0 - 36	0 - 36
Positive Approaches to Learning	2,479	12.08	1.76	3 - 14	0 - 14	2,287	12.15	1.75	3 - 14	0 - 14
Behavior Problems Index (Total)	2,480	6.16	3.55	0 - 22	0 - 24	2,287	5.64	3.57	0 - 21	0 - 24
Behavior Problems (Aggressive)	2,468	3.16	1.74	0 - 8	0 - 8	2,275	2.87	1.74	0 - 8	0 - 8
Behavior Problems (Hyperactive)	2,471	1.85	1.50	0 - 6	0 - 6	2,275	1.63	1.45	0 - 6	0 - 6
Behavior Problems (Withdrawn)	2,459	0.61	0.93	0 - 6	0 - 6	2,267	0.63	0.94	0 - 6	0 - 6
Combined Activities Scale	2,469	10.67	3.76	0 - 22	0 - 22	2,285	11.19	3.80	1 - 22	0 - 22
Weekly Activities Subscale	2,473	6.28	2.52	0 - 11	0 - 11	2,290	6.23	2.56	0 - 11	0 - 11
Monthly Activities Subscale	2,477	4.39	2.09	0 - 11	0 - 11	2,285	4.95	2.14	0 - 11	0 - 11
Emerging Literacy Scale	2,469	1.94	1.44	0 - 5	0 - 5	2,290	2.95	1.51	0 - 5	0 - 5

Table A-30. Summary statistics for Spring 2002 (Head Start) and Spring 2002/2003 (kindergarten) FACES parent interview data selected measures

	Spring 2002 (Head Start)					Spring 2002/2003 (kindergarten)				
Scales	Number of cases	Mean	SD	Reported response range	Possible response range	Number of cases	Mean	SD	Reported response range	Possible response range
Pearlin Mastery Scale (Locus of Control)	906	15.32	3.24	5 - 21	0 - 21	2,026	15.24	3.19	4 - 21	0 - 21
CES-D Depression Scale	906	5.14	6.04	0 - 36	0 - 36	896	5.19	5.80	0 - 34	0 - 36
Social Score Positive	905	12.49	1.69	5 - 14	0 - 14	2,030	26.23	5.88	8 - 39	0 - 39
Behavior Problems Index (Total)	905	5.20	3.70	0 - 22	0 - 24	2,030	9.63	5.39	0 - 33	0 - 39
Behavior Problems (Aggressive)	900	2.61	1.80	0 - 8	0 - 8	2,030	4.74	2.80	0 - 15	0 - 15
Behavior Problems (Hyperactive)	900	1.45	1.47	0 - 6	0 - 6	2,024	2.22	1.88	0 - 9	0 - 9
Behavior Problems (Withdrawn)	894	0.62	0.95	0 - 6	0 - 6	2,030	2.66	2.08	0 - 14	0 - 15
Combined Activities Scale	902	11.87	3.95	1 - 22	0 - 22	2,028	8.47	3.09	0 - 17	0 - 17
Weekly Activities Subscale	903	6.46	2.64	0 - 11	0 - 11	2,029	3.61	1.80	0 - 7	0 - 7
Monthly Activities Subscale	905	5.42	2.32	0 - 11	0 - 11	2,030	4.86	2.05	0 - 10	0 - 10
Emerging Literacy Scale	902	3.98	1.21	0 - 5	0 - 5	2,032	4.74	0.66	0 - 5	0 - 5