What Role Do Peer Effects Play in Early Childhood Education?

Evidence from the 2003 Head Start Family and Child Experiences Survey (FACES)

By

Hilary Marie Shager

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Public Policy)

at the

UNIVERSITY OF WISCONSIN-MADISON

2012

Date of final oral examination: 6/05/12

This dissertation is approved by the following members of the Final Oral Committee:

Katherine Magnuson, Associate Professor, Social Work
Maria Cancian, Professor, Social Work
Adam Gamoran, Professor, Sociology
Douglas Harris, Associate Professor, Education Policy Studies
Carolyn Heinrich, Professor, Public Affairs, University of Texas-Austin

ABSTRACT

Previous research suggests that the peer composition of a student's class may influence educational outcomes, yet little attention has been paid to the role peers may play in affecting children's preschool experiences. There are developmental differences between preschoolers and older children, as well as differences in the pedagogical structure of early education compared with later schooling, which suggest that one should not simply extrapolate findings from peer effects research with older children to preschool settings.

This study uses longitudinal data from the 2003 Head Start Family and Child Experiences Survey (FACES), a nationally representative sample of Head Start children, to estimate associations between peers' pre-academic and behavior skills and students' pre-academic and behavior outcomes after one year in the program. Value-added (lagged dependent variable) modeling and a rich set of controls for child, family, and classroom characteristics are used to reduce omitted variable bias, and multilevel modeling is used to account for the data structure (students nested in classrooms). Additionally, Ordinary Least Squares (OLS) regression is used to explore associations between average classroom pre-academic and behavior skills and measures of classroom process quality.

Findings suggest that, indeed, preschool peers may influence children's pre-academic and behavior skills development, as well as preschool classroom process quality. In the pre-academic domain, I find robust, positive associations between peer and individual reading skills, as well as a weaker positive link between peer and individual vocabulary skills. In the behavior domain, robust, positive associations exist between positive peer and individual behaviors (specifically, for learning behaviors and pro-social skills). I also find a negative association

between classroom-level behavior problems and classroom process quality, in terms of teacher behavior.

Taken together, these findings suggest that policy makers should consider peer composition in their calculus of potential tradeoffs between targeted or universal preschool programs, which are likely to differ in terms of student heterogeneity and average baseline skills. Furthermore, even if program composition is largely unalterable, knowing that peers contribute to preschoolers' learning recommends teacher training regarding the use of proven collaborative learning and behavior management techniques that maximize the benefits of peer effects within the classroom.

ACKNOWLEDGEMENTS

I would like to acknowledge the incredible support that I have received from so many individuals during my graduate school journey. Thank you to:

My advisor, Dr. Katherine Magnuson, who has always gone well above and beyond her advisor duties, and provided me with challenging opportunities to think, learn, and act. I could not have asked for a better mentor.

My dissertation committee: Dr. Magnuson, Dr. Carolyn Heinrich, Dr. Maria Cancian, Dr. Adam Gamoran, and Dr. Doug Harris. Thank you for your time and guidance, and for hanging in there as I transitioned from full-time graduate student to full-time employee before finishing my dissertation. A special thank you to Carolyn, for encouraging me to take this journey in the first place, and for making it happen.

My supervisors at the Wisconsin Department of Children and Families, Melissa Wavelet and Becca Schwei, for supporting me as I finished writing and providing me with a great work environment in which to apply what I've learned.

The staff and students in the Social Work Doctoral Program, Interdisciplinary Training Program (ITP), and at the Institute for Research on Poverty (IRP), for their outstanding training and support. In particular, thank you to Sarah Bruch and Meghan Condon for your friendship, for challenging me to be a better researcher, and for making me work on my dissertation when I didn't want to.

My family: First and foremost to my husband, Erik, for his unwavering support even when I didn't deserve it; to my son, Jens, and my daughter, Ava, for their patience and independence; to my parents, Susie Sieg and Boyd Groth, for telling me so often that I could do anything I put my mind to, that I actually believed it; to my siblings, Travis, April, and Clayton, and their partners for being the best friends a person could have; to my parents-in-law, Donna and Connie Shager, for all their support and help with the kids; and to my much-missed grandmother, Neva Kjellander, who always kept me laughing and grounded.

My generous financial supporters: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through award # R305C050055 to the University of Wisconsin-Madison, and by a grant from the American Educational Research Association, which receives funds for its "AERA Grants Program" from the National Science Foundation under Grant #DRL-094104. The opinions expressed are those of the author and do not necessarily reflect those of the granting agencies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Purpose of Dissertation

Previous research suggests that the composition of a student's class, in terms of peers' academic achievement, behavior, gender, race, and socio-economic status, may be an important influence on educational outcomes (Hanushek et al., 2003; Hoxby, 2000). Indeed, the importance of academic peers motivates educational policies such as ability tracking, school desegregation, school choice, and redistributive school financing (Harris, 2010). Yet, despite increasingly high levels of participation in early childhood education and recognition of the importance of early skill building to future development, most extant research on peer effects has focused only on traditional K-12 classrooms and schools. As a result, the following question remains unanswered: what role do peers play in affecting children's preschool experiences and learning?

There are distinct developmental differences between preschoolers and older children, as well as differences in the pedagogical structure of early education programs compared with later schooling. These differences suggest that one should not simply extrapolate findings from peer effects research with older children to preschool settings. My dissertation research attempts to address this gap in the literature. I use longitudinal data from a survey of nationally representative first year Head Start children and families to explore the relationship between peer group pre-academic and behavioral skills and individual children's language, math, and behavioral outcomes. I use value-added (lagged dependent variable) modeling and a rich set of controls for child, family, and classroom characteristics to reduce omitted variable bias, and

multilevel modeling to account for the nested data structure (students nested in classrooms). Furthermore, I explore associations between classroom-level measures of skills and demographics and classroom-level quality outcomes.

Given the current scope and scale of public investment in early education, as well as its potential to improve the educational outcomes of vulnerable children, it is important to generate a knowledge base that creates a fuller and more accurate understanding of classroom processes, which in turn can inform intervention strategies. Findings from this dissertation demonstrate how peers may matter in preschool; and, thereby, illuminates ways in which such influence can be leveraged to improve children's educational outcomes.

Policy Context

Early Childhood Education (ECE) in the United States

The influx of mothers of children under age six into the workforce combined with a growing body of research suggesting the importance of early learning experiences for later development has increased the demand for high quality early childhood education (ECE) (Gormley, 2005; Cunha and Heckman, 2007; Shonkoff & Phillips, 2000). Today, attending ECE is a normative experience for children in the United States; for example, approximately 70 percent of four-year-olds attend some kind of ECE program the year before kindergarten (Cook, 2006). Of particular interest, in light of the federal No Child Left Behind Act (NCLB), which requires academic proficiency from *all* students, including those in at-risk groups, are programs that could potentially serve to improve the "school readiness" of children who have historically begun school behind their same-age peers (e.g., poor, African-American, and Hispanic children) (Neuman, 2003). Additionally, ECE has been presented as a potential policy tool in the fight

against social problems such as poverty, crime, and social immobility (Duncan, Magnuson, & Ludwig, 2007; Yoshikawa, 1995; Barnett and Belfield, 2006).

Traditionally, the United States government has engaged in a targeted policy approach to ECE, offering federally funded programs such as Head Start for children from low-income families, services for preschool-aged children with disabilities, and childcare subsidies included as a component in state-level welfare block grants (Scrivner and Wolfe, 2003). Concerns regarding whether access to such programs is sufficient and equitable remain, however; as do doubts about the programs' effectiveness in closing achievement gaps (Bainbridge et al., 2005; Currie and Thomas, 1995). Despite this, Head Start in particular remains a popular program that served over 904,000 children in Fiscal Year 2009, with an appropriation of over $7.1 billion (U.S. Department of Health and Human Services, 2011). The federal government's commitment to early education is also reflected in its recent Race to the Top Early Learning Challenge grant competition, which provided $500 million in funding for early education infrastructure building to nine states in 2011. Again, the focus of the competition was on providing high quality early learning experiences, particularly for low-income and disadvantaged children. A second round slated for 2012 will offer $133 million for an additional five eligible states.

Recently, support for a more "universal" approach to ECE has also increased, resulting in the proliferation of publicly funded pre-kindergarten (pre-k) programs, subsequently defined as school or center-based programs primarily funded and administered by states, and developed with the explicit goal of improving school readiness. Today, 39 states and the District of Columbia have pre-k initiatives that serve over 1.3 million children, or approximately 28 percent of the nation's four-year-olds and 4 percent of its three-year-olds. Fewer than half of these programs existed 20 years ago (Barnett et al., 2007). Despite tight budgetary times and

decreases in pre-k spending over the past few years, states invested approximately $5.49 billion in pre-k in 20010-2011 (Barnett, Carolan, Fitzgerald, & Squires, 2011). Although many pre-k programs also remain targeted to at-risk children, a growing number of states either offer, or are currently considering, expansions of their programs to make them available to all four-year-olds and, in some cases, three-year-olds (Christina & Nicholson-Goodman, 2005). Theoretically, compared to targeted preschool programs, such universal programs would enroll a more diverse population of students, particularly in terms of socioeconomic status (SES) and skill level.

Another important development in ECE policy is the proliferation of state Quality Rating and Improvement Systems (QRISs), which, in addition to providing publicly available quality ratings of ECE settings, offer feedback and technical assistance to providers, in order to motivate and facilitate quality improvements (Mitchell, 2009). Currently, 19 states and the District of Columbia have developed QRISs, which, importantly, may also be tied to state reimbursement rates for childcare subsidies. Studies of current QRISs show variability in states' strategies for measuring quality, as well as providing incentives and financial support to providers to make improvements (Zellman & Perlman, 2008). QRIS advocates, however, see the potential for such programs to provide a unifying framework for a currently fractured ECE service system, if they are able to include all ECE settings, and align system components such as funding, program standard development, and evaluations (Mitchell, 2009).

The Relevance of Preschool Peers to ECE Policy

Embedded in each of these policy approaches are remaining questions about how best to structure ECE experiences. Although evaluations of a number of ECE programs have shown promising results across a number of domains, particularly for children at-risk of poor

educational outcomes (e.g., for reviews of this literature, see Barnett, 1995; Gilliam & Zigler, 2000; Currie, 2001; Gormley, Phillips, & Gayer, 2008), most frame ECE as a "black box" treatment either experienced or not experienced by children, and do little to explore the mechanisms by which outcomes are attained (Magnuson & Shager, 2010). In particular, most evaluations ignore the potentially important variation of the peer group in the classroom when the treatment is delivered.

Discovering whether and how peer effects matter in ECE allows for exploration of one possible policy lever by which to improve results for such programs. If having higher skilled peers improves learning in ECE, then public policies that increase the average ability level of a child's peers may be warranted. Alternatively, if positive peer effects are not present at the preschool level, or if diversifying the preschool classroom diminishes results, a more targeted approach may be better justified. Also, knowing whether children contribute to preschool quality may raise questions about using quality ratings for high stakes purposes like tiered reimbursement. Additionally, understanding how preschool peers contribute to learning may recommend particular professional development, curricular, or practice approaches.

Theoretical Framework

The Importance of Peers: A Developmental Perspective

A number of developmental theories, broadly applied across the life course, suggest that peers may be important influences on developmental outcomes. For example, Ecological Systems theory suggests that children's interactions with their environment have a profound impact on their development (Bronfenbrenner, 1979). In particular, experiences with peers comprise important and unique developmental contexts in which children acquire skills,

attitudes, and behaviors that affect social, emotional, and cognitive functioning throughout the life span (Rubin, Bukowski & Parker, 2006).

Vygotsky (1978) suggested that cognitive growth and development occur as a result of interpersonal exchanges that allow children to do more than they would have been able to do independently. More specifically, he described a "zone of proximal development," in which cognitive advances occur as a result of collaboration with more skilled partners in a scaffolding process. Vygotsky typically characterized the parent or teacher as the key collaborator in learning experiences, and subsequent research suggests that positive interactions with caregivers, such as those characterized by rich and varied use of language, turn taking, and contingent and focused attention, are associated with improved cognitive outcomes for preschool-aged children (See, e.g., Bradley et al., 1989; Hart & Risley, 1995; Hoff-Ginsburg, 1991; Katz & Snow, 2000; Tomasello & Farrar, 1986). Other researchers, such as Tudge (1992; Hogan & Tudge, 1999) and Rogoff (1997), have suggested that a child's peers might also fill this collaborative teaching role.

Mechanisms of Peer Influence

Although much evidence suggests that peers are important, *how* they influence children's development remains open to debate (Harris, 2010). A number of studies model "peer effects" using classroom composition measures of demographics, such as percent of students of a particular race or ethnicity, or of high or low SES. Such research often finds negative associations between higher percentages of minority students or students from low SES backgrounds and student outcomes. Explaining these findings, scholars point to the way in which the aggregation of disadvantaged students may create lower quality schools; for example, by reducing the likelihood of attracting high quality teachers. Thus, one way in which classroom

composition may be associated with learning outcomes is by shaping the quality of teachers students experience or other quality dimensions of the schools that they attend (Jencks & Mayer, 1990).

Holding school or program quality constant, however, the developmental theories described previously suggest three ways in which peer skills within the classroom might influence children's outcomes through other mechanisms. First, peers provide a context for language, math, and behavior learning, and having more highly skilled peers may create better learning opportunities (or, conversely, having lesser skilled peers may inhibit learning). Second, having more highly or lesser skilled classmates may influence children's early development by shaping teacher practices, both in terms of the quality of interactions with students and the teacher's ability to provide certain types of learning experiences for students. Finally, studies with older children suggest a potential third peer mechanism: peers may change the beliefs and values of other children (for a review, see Harris, 2010). This final mechanism, however, seems less salient for preschoolers, as developmental research suggests that prior to age eight, children's beliefs are unreliable predictors of future behaviors (Davis-Kean et al., 2008).

Although each of these mechanisms generally suggests that having more advantaged peers is beneficial for disadvantaged students, it is also possible that having more advantaged peers may be harmful for disadvantaged students (Harris, 2010). For example, according to relative deprivation theory (Jencks & Meyer, 1990), disadvantaged children, sensing their lower relative skills or social position may become frustrated or and put forth less effort in the company of more disadvantaged peers. Again, the salience of such theories for preschoolers remains questionable, given that they may not be capable of making such assessments of their own and others' skills and characteristics (Davis-Kean et al., 2008). Understanding the specific

peer influence mechanisms at work in the preschool classroom then, remains a topic ripe for exploration.

The Importance of Peers in the Preschool Context

Preschool children's experiences with "peers" are fundamentally different than older children's experiences. For example, preschool "friendships" are more fluid, and interactions within the classroom are more structured by teachers (Rubin et al., 2006). Interactions with peers outside the classroom are also generally more structured by adults; i.e., parents generally choose out of school activities and organize children's play time with others. For these reasons, in this dissertation, a "peer" in the preschool context is more specifically defined as a "classmate," and the mechanisms by which one might expect preschool peers to provide influence are limited to those that might occur within the classroom.

Certainly, given their developmental skills, preschool children have the potential to influence one another's academic and behavioral progress. As children age, their peer experiences increase in frequency, diversity, and complexity, and become more integrated with development. Between ages three and five, a number of important advances occur (Rubin et al., 2006). Although preschool children continue to engage in some solitary and parallel play, they also increase their social activity; in particular, the frequency with which they participate in more cognitively challenging social activities such as sociodramatic play and games with rules (Rubin, Watson & Jambor, 1978; Goncu, Patt & Kouba, 2002). Compared to younger children, preschoolers' social interactions also involve longer sequences or turns (Goncu, 1993). Prosocial caring, sharing, and helping behaviors increase, and aggression generally declines (Benenson, Markovits, Roy & Denko, 2003; Dodge, Coie & Lynam, 2006). Furthermore, compared to

toddlerhood, preschool conflict tends to be more about differences of opinion regarding play and other ideas, rather than the control of toys and other resources (Chen, Fein & Tam, 2001; Laursen & Hartup, 1989). These advances reflect children's increased ability to pay attention to and consider others' ideas, attitudes, and opinions.

Preschoolers also spend a lot of time conversing with classmates, and conversations may reflect specific goals such as negotiating roles and rules, as well as arguing and agreeing (Hay, Payne & Chadwick, 2004). Preschoolers also show increased levels of social-communicative competence; for example, the ability to adjust speech to the needs of particular listeners, as well as the increased use of attention focusing-statements and responsiveness to conversation partners (Shatz & Gelman, 1973; Cooper, 1980). Findings from observational studies suggest that over 60 percent of preschoolers' speech with peers is socially directed, comprehensible, and elicits appropriate responses (Levin & Rubin, 1983; Mueller, 1972; Schuele, Rice & Wilcox, 1995). The quality of such verbal exchanges may vary, however, and social goals are more likely to be met successfully by children with stronger conversational skills; for example, those who verify listener attention, utter comprehensible language, and deliver their messages within arms' length of conversation partners (Mueller, 1972).

Despite the fact that preschool children are generally quite capable communicators who regularly engage in meaningful interactions with their peers, questions remain regarding preschool peers' role in advancing developmental skills, particularly in terms of academic achievement. On the one hand, peer interactions may provide unique learning opportunities (compared to interactions with adults), in the sense that children are more able to explore concepts, emotions, and actions without fear of adult sanction or criticism (Rogoff, 1998). On

the other hand, children may be more skeptical of information provided by their peers if they do not perceive these children as possessing expertise (Gelman, 2009).

Peer influence on behavior in early childhood may be even more likely, given that preschool children's behavior is more variable and potentially responsive to their environment (Hanish et al., 2008). For example, in the preschool classroom, children may shape each other's behavior in a transactional way; i.e., children's predisposing characteristics may affect both their own behavior and the types of social interactions they seek out (e.g., niche picking), as well as the nature and quality of peers' responses. This, in turn, may affect subsequent individual behavior and peer interactions (Martin & Fabes, 2001; Parker, Rubin, Price, & DeRosier, 1995). Classmates may therefore model or directly teach new behaviors, as well as create a "peer culture" that supports or discourages particular behaviors (Hanish et al., 2008). Teachers' responses to particular behaviors may also be important reinforcements of children's behavioral learning and responses (Raver et al., 2008).

Despite the recognized importance of "quality" in early care and education, little theoretical and empirical attention has been paid to the potential role that children and their peers may play in creating the quality of early learning environments. To some extent, this lack of research may be reflective of an earlier era in which children spent more time at home or in part-time early childhood education programs, and, thus, less time with same-age children. Today's increased attendance of early childhood education programs and time spent with peers, however, suggests that these potential influences should be considered.

Ignoring the direct contribution of preschool peers to children's learning may be unwise, considering that early childhood education environments differ substantially from later school environments in ways that may *promote* peer interaction. For example, a recent study

investigating language acquisition in preschool-aged children, in which the authors audio-taped students, revealed that four-year-olds spent an equivalent amount of time talking with their peers (18 percent of the time) as with their teachers (17 percent of the time) (Dickinson & Tabors, 2001). Observational studies of three- and four-year-old children attending preschools suggest that they spend over a third of their free-play time in cooperative play with other children (Berk, 2000), and such play helps build children's cognitive and social skills (Nicolopoulou, 1993; Singer & Singer, 1990). The use of child-initiated instruction, an approach that de-emphasizes whole group and teacher-directed instruction, and encourages children to choose their own activities and playmates, is the cornerstone of developmentally appropriate practice (Marcon, 1999).

Children may also affect general classroom quality, which is typically measured two different ways in early education settings. Structural quality refers to structural aspects of the program, such as the child to teacher ratio or level of teacher education, markers that have the advantage of being easy to measure and regulate. These factors are expected to indirectly affect children by shaping their day-to-day experiences in classrooms (NICHD, 2002), and are more likely to be established prior to children's assignment to classrooms and remain unchanged as a result of classroom composition.

The second set of quality indicators, measuring process or global quality, is designed to directly capture children's day-to-day experiences and interactions. These measures typically involve observing children in the classroom and rating several dimensions of their experiences. Process quality may indeed be influenced by children's pre-academic skills and behaviors, and resulting interactions with peers and teachers, as demonstrated by the central role these factors play in assessing process quality in early education. For example, the Early Childhood

Environmental Rating Scale (ECERS), an instrument often used to measure preschool quality, rates higher classrooms in which children have access to toys and other instructional resources for long periods of time throughout the day, based, at least in part, on the theory that such an arrangement will provide children with opportunities to enhance their skills through play interactions with other children (Harms, Clifford, & Cryer, 1998). The importance of such peer interaction is also reflected in the Head Start Program Performance Standards; for example, Section 1304.21, outlining the development and educational approach for children, calls for "a balanced daily program of child-initiated and adult-directed activities, including individual and small group activities," as well as, "promoting interaction and language use among children" (USDHHS, ACF, Office of Head Start, 2009).

The influence of children in the preschool classroom might also operate via educational resource allocation and other within-school processes (Jencks & Mayer, 1990). In economic terms, one might consider the "public good" aspects of classroom education (Lazear, 2001). For example, whether children engage in cooperative or disruptive behaviors may affect scheduled routines and the delivery of structured learning activities, the warmth of teachers' interactions with children, and the intentionality and type of instructional feedback that teachers provide (Raver et al., 2008). These factors also play an important role in defining process quality in early education.

In sum, one must consider the unique developmental and pedagogical context when thinking about the ways in which peers might influence preschool children's learning experiences. This dissertation focuses specifically on the direct and indirect role of preschool children's pre-academic and behavioral skills in shaping their classmates' learning experiences

within the preschool classroom. This more precise research question helps elucidate more specific policy and practice approaches that might be used to improve ECE programming.

Organization of the Dissertation

This dissertation includes three empirical chapters, each utilizing longitudinal data from the 2003 Head Start Family and Child Experiences Survey (FACES), a nationally representative survey of children who entered Head Start in 2003, funded by the United States Department of Health and Human Services, Administration for Children. The FACES data provide one of the most detailed pictures of preschool participants and classrooms currently available, including information on attendees' background characteristics, experiences, and educational outcomes, garnered from interviews with parents, teachers, center directors, and education coordinators. Importantly, FACES also includes results from direct assessments of children's pre-academic skills, teacher reports of behavioral skills, and observations of classrooms.

Chapter 2 examines associations between peers' pre-academic and learning behavior skills and children's individual vocabulary, reading, and math skills. The analyses employ value-added regression (lagged dependent variable modeling), a method that controls for both observed and unobserved factors affecting skill development; and, if the effects of unobserved characteristics are constant over time, provides a powerful way to reduce omitted variable bias (NICHD-ECCRN & Duncan, 2003). Multi-level modeling is used to account for the nested data structure (children nested in classrooms) (Raudenbush & Bryk, 2002). A series of robustness checks, including models using different methodology, samples, and specifications of peer skills are also conducted. Additional analyses explore sub-group differences; for example, whether the

associations between peer skills and individual pre-academic outcomes vary based on child characteristics (e.g., initial skill level, sociability, gender, or race).

Chapter 3 uses similar methodology to examine associations between peers' behavioral and pre-academic skills and children's individual behavior outcomes. Specifically, analyses utilize measures of preschool learning behaviors, pro-social skills, and problem behaviors. Again, a series of robustness checks and sub-group analyses are also explored.

Chapter 4 employs Ordinary Least Squares (OLS) regression to measure associations between average classroom pre-academic and behavioral skills and measures of classroom process quality, including overall classroom environmental quality and the quality of teacher/child interactions. This chapter also includes an exploration of such associations in relation to a series of alternative measures of process quality (e.g., subscale measures of the classroom language and social skills learning environments, and teacher satisfaction), as well as robustness checks regarding methodology, sample, and skill specification.

Finally, in Chapter 5, I conclude with a summary of the findings from each chapter, as well as a discussion of implications for policy and practice. Discussions of limitations of the present study and ideas for future research are also included.

CHAPTER II

THE INFLUENCE OF PRESCHOOL PEERS ON CHILDREN'S PRE-ACADEMIC OUTCOMES: EVIDENCE FROM THE 2003 HEAD START FAMILY AND CHILD EXPERIENCES SURVEY (FACES)

Background

Developmental theory points to a number of arguments recommending that preschool peers may influence the development of children's pre-academic skills. Instructional approaches that allow for frequent interactions between children may particularly impact language development, a strong predictor of later school achievement (Duncan et al., 2007). Previous research indicates a strong link between the amount, variety, and quality of speech to which a child is exposed and her/his vocabulary growth (Hart & Risley, 1995; Huttenlocher et al., 1991) and grammatical development (Tomasello, 2006). Positive links between features of teachers' language use and preschoolers' language growth have also been found (See, e.g., Connor, Morrison & Slominiski, 2006; Girolametto & Weitzman, 2002; Justice, Mashburn, Pence & Wiggins, 2008). In other words, the "linguistic input" children receive is associated with language skill development. Since other children in the preschool classroom also provide linguistic input via conversations and play situations, the language skills of these peers may also be an important contributor to the language learning environment, if their language is sufficiently varied and complex.

Changes in mathematical understanding at preschool age include the acquisition of number concepts, number words, and counting procedures. In addition, preschoolers develop an

understanding of the use of numbers and counting in terms of their use for cardinality, ordinality, and measuring things (Geary, 2006). Mathematical learning during the preschool years may also be strongly connected to language skills; for example, many early math learning experiences and assessments involve attaching comparative verbal labels to different amounts and sizes, as well as verbally exploring and explaining calculation strategies (Berk, 2000). Furthermore, a recent study by Klibanoff et al. (2006) found a significant relationship between the amount of teachers' math-related talk and growth of preschoolers' mathematical knowledge over the school year. If peers also produce math-related talk, these findings suggest that they may play a role in advancing other children's mathematical skills via both their mathematical and language knowledge (Geary, 2006).

Previous research with elementary school children suggests that math skill development may be particularly sensitive to formal instruction in school (Christian, Morrison, Frazier & Massetti, 2000; Entwisle & Alexander, 1992); however, recent observational studies suggest that very little formal math instruction occurs in most early childhood education classrooms (Pianta, 2008). There is also evidence that children who are exposed to opportunities to learn and practice counting and number identification acquire basic math concepts sooner (Geary, 1994). These findings suggest two salient ideas related to the study of peer effects on math outcomes: 1) math teaching practices, curriculum, and materials available to preschool students should be considered when measuring the impact of peer effects on math outcomes, and 2) the introduction of any peer initiated math activities or learning opportunities may be important.

Literature Review

Despite the fact that preschool programs provide potentially important socialization opportunities for children and that direct peer effects on academic outcomes have been documented among elementary school children (e.g., See Hanushek et al., 2003; Hoxby, 2000), there is a dearth of research examining peer effects in early educational settings. The studies that have tried to measure the influence of peers on cognitive development have modeled the associations between students' pre-academic skills as well as their social characteristics, and findings from studies present mixed results.

In one of the first studies of contextual effects in preschool, Lee et al. (1998) found significant variation in learning between classrooms in a sample of 677 four-year-olds attending Title 1 prekindergarten programs for at-risk students in 5 states. Using hierarchical linear modeling (HLM) to specify contextual effects at the classroom level, and controlling for program entry test scores and a host of individual and classroom level covariates, the authors found that students in classrooms with high percentages of minorities, children with special needs, recent immigrants, and children whose mothers have low levels education scored lower on a general cognitive test (The Preschool Inventory) given at the end of the academic year. Contrary to findings from studies of older students, however, average cognitive skill level of the class was not associated with gains in cognitive skills.

More recently, Henry and Rickman (2007) identified direct, positive peer effects, using longitudinal data from Georgia, which included a more diverse and advantaged probability sample of four-year-olds who attended Head Start, the state's publicly subsidized pre-kindergarten, or private preschool. Theorizing that early child development may draw from a broader (versus subject-specific) base of skills, the analysts created a composite measure of peer

ability by averaging tested classmates' scores across a variety of academic assessments. Controlling for individual children's pre-academic skills measured at the beginning of preschool as well as other important characteristics, the authors found direct and positive effects of classroom ability level on children's math, pre-reading, and expressive language skills, with effect sizes ranging from .02 to .36. An important limitation of Henry and Rickman's (2007) study, however, is that such composites were based on complete fall preschool assessment information for between four and six children per class, and such a small number of observations per classroom may create measurement error, which is likely to attenuate estimates of peer effects. An additional problem is that the student sample was drawn from different programs (Head Start, pre-k, and private preschool). This introduces the problem of selection bias: selection into early childhood education programs is often related to child and family characteristics that also predict child outcomes; peer abilities are likely to differ by program; and programs may have differential effects on skill growth (NICHD-ECCRN & Duncan, 2003).

In another recent study of approximately 1,800 children in pre-kindergarten programs in 11 states, Mashburn and colleagues (2009) found positive associations (effect size = .05) between higher peer expressive language abilities and children's development of receptive and expressive language during the academic school year. The authors used value-added modeling similar to that of Henry and Rickman (2007), but also employed HLM to account for clustering within classrooms. Contrary to expectations, the positive association between peer skills and children's individual outcomes was stronger for children with higher initial receptive language skills and within classrooms characterized as having better classroom management. Again, however, the study suffered from several shortcomings. Only four children per classroom were sampled; therefore, peer skill measures were derived using only three peer scores. Child level

covariates were limited to gender, race, and years of maternal education, although some classroom characteristics (e.g., class size, teacher-child ratio, full- or part-time status, and quality of teacher interactions) were also controlled for in the analyses.

Contrary to Mashburn et al. (2009), a more recent study of 338 children in 49 publicly funded preschool programs found evidence of significant interactions between children's language skills measured in the fall of the academic year and their classmates' language skills suggesting that peer effects were strongest for children with *low* initial language skills (Justice et al., 2011). This study benefitted from a larger sample of children from each classroom by which to construct peer scores (on average, 44 percent of each classroom's enrollment); however, the overall sample remained small, covariates were limited, and outcomes were confined to the language domain.

Evidence suggestive of direct peer effects can also be found in a study of a state-funded, locally managed effort to integrate low-income children into private preschools (Schechter & Bye, 2007). Pre-tests taken in the fall showed significant differences between low-income children and their mid- to upper-income peers. Post-tests in the spring demonstrated significantly greater gains for low-income children in the economically integrated preschool programs, compared to low-income children attending programs with only low-income peers. Although it is not clear that the social composition or skill level of the classroom yielded these associations, the pattern of associations is consistent with some form of peer effect.

The differences in these studies, in terms of population, modeling, and findings present a number of remaining research questions. For example, are the non-significant findings regarding cognitive peer effects in Lee et al. (1998) unique to the low-skilled population in the study or due to the inclusion of aggregated demographic effects? Are the "peer effects" found in Henry and

Rickman (2007), Mashburn et al. (2009), Justice et al. (2011), and Schechter & Bye (2007) really proxies for unmeasured inputs that affect skill development, such as parenting or teacher quality? Higher quality data and replication are necessary to answer these questions. Given theory and previous research, I hypothesize that after controlling for initial individual skill levels and an extensive set of child, family, and classroom-level covariates, the initial pre-academic skill levels of peers will be positively associated with individual children's spring language and math scores.

Previous research and theory also recommend the importance of investigating potential initial skill level and sociability interactions, to test for differential associations between peer and individual behaviors between these groups of children. On one hand, the "skills beget skills" theory suggests that higher skilled children might best be able to gain knowledge from interactions with other higher skilled children (Cunha and Heckman, 2007). On the other hand, children with low initial skill levels might have more opportunity for growth resulting from such interactions. Empirical evidence supporting both theories has been found in preschool settings (Mashburn et al., 2009; Justice et al., 2011). Evidence suggesting generally higher pre-academic scores for girls and non-minority preschool students also recommend the exploration of race and gender interactions. Again, however, the theoretical direction of such effects remains ambiguous (Braza et al., 2009)

Data and Methods

Data

In this study, I use data from the 2003 Head Start Family and Child Experiences (FACES) study, a nationally representative survey of children who entered Head Start in 2003, funded by the U.S. Department of Health and Human Services' Administration for Children and

Families (ACF). These longitudinal data follow children and families served by Head Start from program entrance to the end of kindergarten. FACES data include detailed information on Head Start attendees' background characteristics, experiences, and developmental outcomes. At the beginning and end of the Head Start program year and again in kindergarten, high quality child assessments and parent interviews were completed. Teacher surveys were also conducted at the beginning and end of the Head Start year. Finally, center directors and education coordinators were also interviewed in the fall of the Head Start year, and classrooms were observed for overall quality, types of activities, and interactions between staff and children at the beginning and end of the Head Start year.

The FACES data provide several advantages over data used in the few previous explorations of preschool peer effects. First, all families represented in FACES are participants in the same early childhood education program, Head Start. Therefore, the problem of selection bias based on *type* of program (children of differing abilities selecting into different types of preschool programs) found in prior studies (e.g., Henry & Rickman, 2007) is eliminated. Second, a multitude of high quality child assessments of both pre-academic and behavioral skills, as well as observational assessments of classrooms, provide information about previously unmeasured child and classroom outcomes. Third, rich information regarding important, yet previously omitted control variables, including teacher and classroom quality, as well as parenting, is also available. Fourth, up to 16 children per classroom are assessed upon entry into Head Start (out of an average class size of 15), providing a more accurate estimate of the true peer skill level, compared to previous studies based on measures from as few as 3 students per classroom (e.g., Mashburn et al., 2009).[1] Fifth, although previous peer effect studies have been

[1] Sojourner (2011) notes the potential bias in widely used peer-effects estimators in the extant literature, which do not account for the "missing data" of students not included in such estimators.

limited to data from a few sites or states, the FACES sample is relatively large and nationally representative of first year Head Start students; thus, increasing the study's external validity, particularly for preschool programs targeted to children from low-income families. Finally, given that these data were released in 2008, FACES provides one of the most up-to-date snapshots of children and families served by the country's largest publicly funded compensatory preschool program.

Sample

The FACES sampling frame was drawn from the 2002-03 Head Start Program Information Report (PIR) file, and the sample was stratified on region, urban/rural status, percent minority enrollment, auspice type (school-based, other), and percentage of non-English speaking children in the program.[2] Response rates for children, parents, and teachers in the Head Start year were quite high, ranging from 93 to 97 percent in the initial round of surveys in the fall of 2003, and 86 to 92 percent at the end of the Head Start year, in the spring of 2004 (Westat et al., 2008). The FACES data file contains information on 2,387 children in 373 different Head Start classrooms.

Although levels of missing data were quite low (almost always less than five percent at the item level), to maximize statistical power and minimize any bias due to missing data, I retained cases with missing information at the item level by using multiple imputation techniques (Graham, 2009; Allison, 2002). Specifically, I used STATA's *ice* program to impute 10 data sets and the *mim* command to conduct analyses across datasets (Royston, 2004). Covariates from full regression models were used in the imputation regressions.

[2] Head Start programs in U.S. territories or Puerto Rico, Early Head Start programs, Migrant and Seasonal Head Start programs, American Indian Head Start programs, and Head Start programs already selected as part of the Head Start Quality Research Center samples were not included in the FACES study.

Multiple imputation methods assume that data are missing at random; in other words, conditional on the observed variables in the analysis, the likelihood of missing data on a particular variable is unrelated to its value. Since this assumption does not hold for the English test scores of students tested in Spanish, I chose not to use imputed English test scores for these students; thus, decreasing the sample to 2,076 students in 367 classrooms. As a robustness check, I conducted separate analyses substituting Spanish speaking children's Spanish test scores (vs. excluding these children from the sample).

Furthermore, to increase the validity and reliability of the peer skill measures, I limited analyses to classrooms with at least four fall test scores by which to construct a measure of peer skills. Given this restriction, the final analytic sample size is 1,917 children in 292 classrooms. Descriptive statistics for the analytic sample are provided in Table 2.1.

Not surprisingly, these descriptive statistics reveal a disadvantaged population. Approximately 36 percent of children are black; 22 percent are Hispanic, and 34 percent are white. Upon entry to the program, 9 percent of children in the sample were identified as dual language learners, and 16 percent were identified as having disabilities. At the initial parent interview only 38 percent of mothers reported being married, and over two-thirds of families reported incomes below the federal poverty line. About 29 percent of mothers reported having less than a high school diploma. Children also demonstrated low levels of pre-academic skills upon program entry; on average, scoring an entire standard deviation below national norms for vocabulary; about three-quarters of a standard deviation for math, and half a standard deviation for reading. Despite this, there is evidence of substantial variation in peer skill levels for each subject.

Measures

Dependent Variables. Children's pre-academic skills were assessed in the fall and spring of the Head Start year. Measures from the spring of the year are used as outcomes, and measures from the fall of the year are used as covariates. For all pre-academic skills, nationally normed standard scores are used.[3] The Peabody Picture Vocabulary Test (PPVT-III) assesses children's knowledge of word meanings (Dunn & Dunn, 1997). A shortened version of the test was developed and given in the FACES survey, using Item Response Theory (IRT). This adaptive version retains all of the psychometric qualities of the original test, including high reliability (Cronbach's alpha=.91 for fall and .89 for spring) (Westat et al., 2008). In addition, several subtests of the Woodcock-Johnson Psycho-Educational Battery, Third Edition (WJ-III) were given, including the Letter-Word Identification subtest, which measures children's early reading skills (specifically, their ability to identify isolated letters and words); and the Applied Problems subtest, which measures children's early math skills. High reliabilities are reported for all WJ-III subtests (Cronbach's alpha range=.78-.91) (Westat et al., 2008). For ease of interpretation, in regressions, the dependent variables were standardized to have a mean of 0 and standard deviation of 1.

Peer Skill Measures. Measures of students' average peer pre-academic and behavioral skills were constructed using standardized scores of all children assessed in a particular classroom, excluding a child's individual test score. In the analytic sample, an average of 7.14 fall scores per classroom are available (range=4 to 16), and the average class includes 14.44 students. Average peer skill measures were created for the PPVT–III and the two WJ-III subtests described above, as well as the Preschool Learning Behavior Scale (PLBS) described at the end

[3] Spanish versions of each of the pre-academic skills assessments were given to Spanish-speaking children who did not pass a language screener test (N=273 in fall of 2003 and N=104 in spring of 2004). These students are not included in the primary specifications, but are considered in robustness checks.

of this section. Again, for ease of interpretation, in regressions, these variables were standardized to have a mean of 0 and standard deviation of 1; therefore, coefficients may be interpreted as effect sizes. Pairwise correlations between each of these measures are provided in Appendix 2.A.

Child and Family Covariates. In order to control for important confounds, I selected measures of individual child and family background that are likely to be correlated with a student's own achievement and that of their peers. Of particular importance is the inclusion of measures of each child's initial pre-academic and behavioral skills, assessed upon entry to Head Start. Additionally, demographic and other individual characteristics found in previous studies to be associated with preschool children's achievement include age, gender, race/ethnicity, disability and language learning status, and prior childcare experience. Family background characteristics such as number of children, family structure, income, and particularly maternal education, may also be predictive (NICHD-ECCRN & Duncan, 2003). The quality of preschool children's home environment, in terms of opportunities for cognitive stimulation and caregiver's responsiveness and warmth, may also have an important impact on children's development (NICHD-ECCRN & Duncan, 2003). Measures of such confounding factors have been largely absent from previous studies of preschool peer effects, but are available in the FACES data. Details about these individual child and family covariates and how they were coded are presented in Appendix 2.B.

Classroom Characteristics. In order to isolate the effects of preschool peers, it is also important to control for early education quality, which may also impact student achievement (NICHD-ECCRN & Duncan, 2003). The FACES data contain information about both structural quality (e.g., child to staff ratio and teacher education) and process quality (e.g., the quality of

staff/child interactions and learning opportunities). Structural quality measures include class size, child/staff ratio, and teacher education. Process quality measures are based on observational data collected in the fall of the Head Start year. These include the revised version of the Early Childhood Environment Rating Scale (ECERS-R), which consists of 37 items and 7 subscales measuring the quality of personal care routines, furnishings, program structure, and opportunities to develop language skills, motor skills, social skills, and creativity (Harms, Clifford, & Cryer, 1998); and the Arnett Scale of Caregiver Behavior, which consists of 30 items assessing 5 areas of teacher behavior: sensitivity, punitiveness, detachment, permissiveness, and prosocial interaction (Arnett, 1989). Westat et al. (2008) report high reliabilities for both process quality measures (Cronbach's Alpha=.92 and .93, respectively). A more detailed description of how each classroom characteristic is coded is provided in Appendix 2.B.

Behavior Measures. Although previous investigations of preschool peer effects have largely considered the influence of pre-academic and behavior skills separately, some research suggests that behavioral skills such as attention may be predictive of achievement (Duncan et al., 2007). I therefore utilize the rich information available in the FACES data, and in the most complex models, include individual and peer measures of the Preschool Learning Behavior Scale (PLBS), an instrument that includes 29 items assessing learning-related behaviors (e.g., motivation, attention/persistence, and attitudes toward learning) (McDermott, Green, Francis, & Stott, 2000). Again, for ease of interpretation, in regressions, these variables were standardized to have a mean of 0 and standard deviation of 1; therefore, coefficients may be interpreted as effect sizes.

Analytical Approach

Value-Added Models of Achievement. As pointed out by numerous researchers, empirical analysis of peer effects is difficult, due to conceptual and data problems (See, e.g., Brock and Durlauf, 2001; Manski, 1993; Moffitt, 2001). When studying peer effects, researchers typically model children's achievement as a function of current family, school, and peer interactions, as well as prior experiences (Hanushek et al., 2003). The key concern is that what one identifies as a "peer effect" (i.e., what is measured) may actually be a proxy for other omitted or poorly measured factors that affect children's development (Moffit, 2001; Manski, 1993). For example, peer composition may be influenced by parental choices of neighborhood and school, as well as school policies dictating attendance and classroom placement. Parents may advocate to place their children in certain classes or schools because they perceive such groupings as advantageous, and schools may "track" students based on ability (Moffit, 2001; Manski, 1993). The end result is that students with similar skills are grouped together for a variety of reasons. Models that fail to separate out these other factors may produce upward biased estimates of peer effects on child outcomes.

Hanushek et al. (2003) argued that unmeasured prior experiences are particularly problematic when trying to identify peer effects, because peers are likely to have had similar school and neighborhood experiences, and spurious correlations been peer skills and abilities and an individual child's skills and abilities are created by their similar (unmeasured) past experiences. Although this may be less of a problem for preschoolers, who have encountered less formal schooling, the influences of prior family, neighborhood, and childcare experiences must be accounted for in the models.

The use of panel data, like those available in the 2003 FACES survey, which include multiple observations of an outcome over time, provides one opportunity to address the identification problems described above. Such data enable researchers to estimate a value-added (lagged dependent variable) specification. Adjusting for initial level of skill is a particularly powerful analytic approach, because it reduces the probability of omitted variable bias. In these models, the initial level of achievement essentially serves as a proxy for the cumulative effects of current characteristics and prior experiences that have constant (time invariant) effects on children's achievement. Unfortunately, unmeasured characteristics that vary systematically over time or exert time varying effects on children's development may still result in omitted variable bias (NICHD-ECCRN & Duncan, 2003).

A related methodological approach is to use a gain score (measured as the difference between outcomes at two time points) as the dependent variable. In this approach, the correlation between the outcome at time two and time one is assumed to be 1.0, and any deviation from this attributed to measurement error. This differs from the lagged dependent variable approach, which assumes that there may be other reasons for deviation from this 1.0 correlation. This distinction of what "other" factors may be in play is important. For example, a potential problem with the change score approach is that gain scores are often negatively correlated with initial scores, thus introducing measurement error (lower reliability) in the gain score and increasing imprecision in the estimates. This is a particular concern in analyses of young children's achievement, given the likelihood that initial levels and later gains will be highly correlated (NICHD-ECCRN & Duncan, 2003; Lee et al., 1998). Measurement error in the dependent variables may also bias the regression coefficients. This may be the case if the measurement error in the change model is correlated with the true levels of the dependent and

independent variables at fall or spring (Bound, Brown, Duncan, & Rodgers, 1994). According to NICHD-ECCRN & Duncan (2003, p. 1459), likely causes of such bias may include, "failure to include relevant interactions, differences in the impact of the omitted variable at the two time points, or use of assessment tools in which error is related to ability," any of which may be theoretically possible in these particular pre-academic skill change score models. Nevertheless, such models are considered as a robustness check.

Multilevel Modeling. In order to study classroom peer effects, it is necessary to have data in which multiple children are observed within the same classroom. This creates a problem, however, in that students' individual outcomes within such contexts may not be statistically independent (i.e., the error terms may be correlated). This type of data structure is suited to multilevel modeling, a generalization of linear regression that accounts for clustering by including a unique random effect for each organizational unit (in this case, classrooms) (Hedges, 2007; Raudenbush & Bryk, 2002). There may also be substantive reasons to use multilevel modeling if one is interested in explaining the variation in outcomes between classrooms, as well as cross-level interactions between individual child and contextual characteristics.

One way to test whether multilevel modeling is an appropriate strategy is to partition the variance to determine whether there is significant variation in each outcome measure at the classroom level. By calculating the intraclass correlation for each measure (i.e., the proportion of variance in test scores that exists between classrooms), it is possible to test the hypothesis that significant variance in child outcomes exists between classrooms. Analysis of the 2003 FACES data yields quite high intraclass correlations among the primary outcomes of interest at the end of the Head Start year; for example, 25 percent of the variance in vocabulary scores exists between classrooms. The intraclass correlations for math and reading scores are lower (for both,

ρ=0.11), but still statistically significant. These results suggest that multilevel modeling is an appropriate analytical strategy, and point to the potential importance of classroom factors in predicting developmental outcomes.

Analytic Models

In the analytic models, the pre-academic outcome at time 2 (spring) for child i in classroom j is regressed on the average peer skill measures, as well as each child's initial skill level and an increasingly rich set of child and family characteristics:

$$\text{Spring Outcome}_{ij2} = \beta_{0j} + \beta_{1j}(\text{Peer Skills})_{ij1} + \beta_{2j}(\text{Child's Initial Skills})_{ij1} +$$

$$\beta_{3j}(\text{Child Covariates})_{ij1} + \beta_{4j}(\text{Family Covariates})_{ij1} + r_{ij}$$

In this child-level model, the coefficients of interest (β_{1j}) represent the magnitude of the association between peers' skills at time 1 (Head Start entry) and students' outcomes at time 2 (spring of the Head Start year), holding constant individual skill levels and an extensive set of child and family covariates, all measured at time 1. The individual error term, r_{ij}, represents the remaining unexplained child-level residual variance.

In the classroom-level model, each of the coefficients from the child-level model becomes an outcome variable:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Classroom Covariates})_{j1} + u_{0j}$$

$$\beta_{1j(\text{peer skills})} = \gamma_{10}$$

$$\beta_{2j\ (\text{ind. skills})} = \gamma_{20}$$

$$\beta_{3j(\text{child})} = \gamma_{30}$$

$$\beta_{4j(\text{family})} = \gamma_{40}$$

Here, the average student outcome for the class (β_{0j}) is predicted by a class-level intercept (γ_{00}), a vector of classroom characteristics (e.g., classroom and teacher quality) measured at time 1,

and a random error term (u_{0j}). Coefficients from the first level (β_{2j}, β_{3j}, and β_{4j}) are modeled as "fixed," so that the coefficient is assumed to have the same value in all classrooms.

I begin by estimating a basic bivariate model, regressing the relevant peer skill measure on each child's spring outcome, and then controlling for each child's relevant fall score. Although the child's fall score is a powerful covariate, I remain concerned about omitted variable bias.

Therefore, for each spring outcome (vocabulary, reading, and math), I estimate a series of seven models with an increasingly rich set of controls. To place this study in the context of existing research, Model 1 approximates Mashburn et al.'s (2009) study of preschool peer effects, and includes basic child and family characteristics (child age, gender, race, and maternal education), as well as basic classroom structural characteristics (class size, child/staff ratio, teacher education), which are important potential confounders for which to control (NICHD-ECCRN & Duncan, 2003).

I next exploit the richness of the FACES data and include an extended set of child characteristics, including English language learner and disability status, and whether the child attended center based child care prior to attending Head Start; an extended set of family characteristics, including income, poverty level, family structure, number of children, home literacy activity, parental warmth and depression; and a set of controls for month of pre-test and number of days between pre- and post-test (Model 2). These models test whether the "peer effects" found in previous studies may have been spurious due to failure to control for these other factors that may be associated with behavior outcomes.

Next, I add in an extended set of classroom measures, including measures of process quality (ECERS and Arnett, Model 3). Research suggesting links between children's skills and

process quality, particularly in terms of time spent on instruction and the provision of certain learning activities (Harris, 2010) recommends isolating these measures, considering that their inclusion could amount to over-controlling.

Next, I test whether early child development may draw from a broader (versus subject-specific) base of skills by including children's other two fall test scores as controls (Model 4), as well as measures of peer skills in those two domains (Model 5). Finally, I consider whether behavior plays an explanatory role in children's academic outcomes, by controlling for children's individual fall Preschool Learning Behavior Scale scores (Model 6) and an average peer skill measure using the same scale (Model 7).

Results

Results from the first set of bivariate regressions are displayed in Table 2.2. As expected, these regressions show positive associations between subject-specific peer skills in each of the three pre-academic areas (vocabulary, reading, and math) and children's individual spring outcomes. Adding in the child's individual fall score attenuates the association for each subject. These findings confirm the casual observer's intuition that peer skills and children's individual outcomes are positively correlated, but also point to the importance of controlling for other factors that may affect achievement.

Next, I turn to the set of value-added, multilevel models with increasingly rich controls.[4] Table 2.3 summarizes the results for children's receptive vocabulary skills (PPVT). With Model 1, including controls for basic child and family demographics and basic classroom characteristics, in line with previous studies of preschool peer effects, I find a positive association between peer vocabulary skills and children's individual vocabulary outcomes,

[4] Complete results, for all covariates, in the most complex models for each outcome, can be found in Appendix 2.C.

(effect size = .05).[5] Although the association remains positive after adding in an extended set of controls for family and child characteristics (Model 2) and an extended set of classroom characteristics (Model 3), it is reduced and no longer statistically significant.

I next test whether children's other pre-academic skills are associated with spring vocabulary scores, by adding in children's fall reading and math scores (Model 4). This regression suggests a statistically significant relationship between children's initial math skills and gains in vocabulary skills, but no such link between reading and vocabulary skills. The peer vocabulary measure remains statistically insignificant. Coefficients for peer reading and math skills are also insignificant (Model 5). The final regressions, including individual (Model 6) and peer (Model 7) behavior scores suggest a significant association between individual behavior and vocabulary gains, but no such link for peer behavior. Furthermore, inclusion of these behavior measures has no impact on the coefficients for pre-academic peer skill measures.

Overall, these findings provide some evidence of a positive relationship between peers' initial vocabulary skills and children's individual vocabulary gains; however, the effect size is small and the specification matters, in that models with the most rich set of controls suggest little or no meaningful association.

Results presented in Table 2.4 suggest a stronger and more robust relationship between peer reading skills and children's individual gains in reading skills over the Head Start year. While all models suggest a positive relationship between peer reading skills and individual gains in reading, the model with the richest set of control variables (Model 7) yields the largest effect size (.06). Although individual math and vocabulary skills are significantly related to children's

[5] Effect sizes can be interpreted as the percent increase in standard deviation of the outcome related to a one standard deviation increase in the peer skill score.

gains in reading, peer skills in these subject areas are not. Furthermore, neither individual nor peer behavior is consistently predictive of children's reading outcomes.

Table 2.5 summarizes the results from regressions of peer skills on children's math skills in the spring of the Head Start year. These findings suggest that there is no statistically significant association between peers' math skills and children's individual math outcomes, regardless of specification. Individual math, reading, and learning behavior skills are predictive of math outcomes; however, peer skill measures in these subjects are not.

Robustness Checks and Sub-group Analyses

To test the robustness of these findings, I conduct a series of alternative specifications and theoretically based sub-group analyses. As previously acknowledged, gain score analysis is an alternative methodological approach made viable with the FACES longitudinal data, although concerns remain regarding measurement error (NICHD-ECCRN & Duncan, 2003; Lee et al., 1998). In models using an otherwise complete set of covariates (i.e., mimicking Model 7 in Tables 2.3 through 2.5), but using the change between spring and fall test scores as the outcome instead of the spring score with a lagged dependent variable as covariate, I find qualitatively similar results with a few changes in significance (See Table 2.6). For example, in the reading change score outcome model, the peer reading coefficient is less precisely estimated, leading to only marginal significance. The negative association between peer math scores and reading and math outcomes is slightly stronger and marginally significant in these models, as is the association between peer learning behavior and math outcomes.

As a second robustness check, I substitute Spanish version scores for missing test scores, to test whether the peer effect findings hold when Spanish-speaking Head Start students are

included in the sample. These models include all of the covariates included in Model 7 of Tables 2.3 through 2.5, plus dummy variables indicating whether students took pre-academic tests in Spanish in the spring and/or fall. Peer skill measures are also recalculated to include the Spanish test scores. The results, reported in Table 2.7, show a robust, larger association between peer reading skills and children's reading growth (effect size =.09). The negative association between peer vocabulary and individual reading skills is also statistically significant (effect size = .07).

As an additional robustness check, I also consider alternative specifications of the peer skills measures. The primary analytic models, using average peer skills (also known as "linear in means" models), assume that introducing a single student who increases peer average achievement by *x* points will have the same effect as introducing several students that collectively increase the average by the same amount. This may not be the correct theoretical mechanism, if, for example, having even one child with particularly high or low skills changes the dynamic of instruction, or having a "critical mass" of children with particularly high or low skills is necessary to truly impact classmates' learning.

To test whether having a student with a particularly high or low test score is associated with individual children's pre-academic outcomes, I construct a "peer maximum" and "peer minimum" score within each classroom, for each pre-academic and learning behavior measure, using fall scores.[6] I then substitute these measures for the average peer skill measures within the primary models. Results, shown in Table 2.8, using the class maximum (Panel 1) and minimum (Panel 2) score as the peer skill measure for each subject, suggest that the main finding regarding a relationship between peer reading scores and individual reading outcomes is robust to the

[6] I also considered class median as a specification of each peer skill measure. This may be important, for example, if the average peer effect is skewed by a particularly high or low score, but the true mechanism of the peer effect is a result of the median (i.e., children will be less likely to have or be impacted by interactions with one particular child, or classroom teachers are more likely to teach to the median than the extremes). Each peer median score, however, is so similar to the peer average, that I do not expect any difference in the results.

maximum score specification (effect size = .08), but not the minimum score specification. In other words, having a higher maximum reading score is associated with higher individual reading outcomes. There is also a marginally significant relationship between maximum and minimum peer math score and individual math outcomes (effect size = .05, for both).

Next, I estimate models that measure whether having a higher percentage of peers scoring in the highest or lowest quartile of each skill area is associated with improved individual outcomes. These models recognize that the "smoothing" effect of an average peer score may not truly represent how children's skills or behaviors impact classroom processes. Three different reference groups are considered: for each outcome, Model 1 uses middle- and low-scoring students as the reference group; Model 2 uses middle- and high-scoring students, and Model 3 uses middle-scoring students only. The results, reported in Table 2.9, show little evidence of a relationship between having a higher percentage of peers with low scores and individual scores, in any subject area. However, there is evidence that having a higher percentage of peers scoring in the highest quartile is associated with higher subject specific individual scores for both reading (effect size = .09) and vocabulary (effect size = .06). There is also some evidence of a positive relationship between having a high percentage of peers with high reading scores and individual math scores (effect size = .06).

I also test whether the associations between peer skills and individual pre-academic outcomes vary based on child characteristics. I use interactions between such characteristics and the peer skill measures in models with a full set of covariates, mimicking model 7 in Tables 2.3 through 2.5. First, I include only interactions between individual characteristics and the peer skill measure matching the outcome (where main effects were found in the primary

specifications). Then, to test for cross-domain influence, I run a model with a full set of interactions (the particular child characteristic interacted with each peer skill measure).

I first test whether the associations between peer skills and children's outcomes vary by children's initial pre-academic and learning behavior skill levels. I develop a series of dummy variables indicating whether children have "high" initial skill levels (meaning their fall test score lies within the 75^{th} percentile of each domain) or "low" initial skill levels (meaning their fall test score lies within the 25^{th} percentile of each domain). I test four models for each outcome, starting with interactions between initial test scores and the relevant peer skill measure, using different reference groups: in Model 1, the reference group includes children with test scores in the low and middle quartiles; in Model 2 the reference group includes children with test scores in the high and middle quartiles; and in Model 3, the reference group includes children with test scores only in the middle quartiles. To check for cross-domain peer effects, I also consider a model in which both high and low score dummies are interacted with each relevant peer skill measure. I find virtually no evidence of differential associations in any of the specifications; however, the main peer reading skill association remains robust (See Tables 2.10 through 2.12).

Using the same methodological strategy described in the previous paragraph, I test whether the associations between peer skills and children's outcomes vary by children's initial pro-social skill levels. I use children's fall scores on the Social Skills Scale, an instrument that includes 12 items that measure how often children engage in cooperative behavior. Again, I construct a "high" and "low" pro-social skill score based on quartiles. These dummy variables are then interacted with each of the peer skill measures. Although coefficients are generally in the expected direction (positive for high pro-social skill/peer skill interactions and negative for

low pro-social skill/peer skill interactions), virtually none of them is statistically significant (See Tables 2.13-2.15).

I next look at interactions between peer skills and gender. Given the theoretical ambiguity regarding hypothesized direction of such interactions, it is perhaps not surprising that I find no consistent pattern regarding associations between gender by peer skills interactions and any of the pre-academic outcomes examined (See Table 2.16). Only the negative coefficient on the male by peer reading interaction is marginally significant (effect size = .07). Adding the full set of gender by peer skill interactions does not qualitatively change the pattern of results for the main peer skill measures.

Given its salience in discussions of educational outcomes, I also examine race by initial skill interactions; although, again, the expected direction of such effects remains ambiguous. In models testing black- and Hispanic-peer skill interactions (using white and other race as the reference category), I find some suggestive evidence of a negative interaction between black and language peer skills; however, many of the coefficients are imprecisely estimated, and statistical significance remains elusive (See Table 2.17). These findings are consistent with the descriptive evidence suggesting that the white children in the sample generally have higher test scores than the black or Hispanic children, although the theoretical mechanisms driving such findings are unclear.

Discussion

Despite the fact that experiences with preschool peers provide potentially important learning opportunities for children, only a handful of empirical studies have examined this phenomenon. Consistent with prior research, the present study finds statistically significant

associations between peers' language skills and children's individual language outcomes after one academic year in Head Start. Moreover, this finding, specifically for reading skills, is robust to the inclusion of a multitude of previously unmeasured child, family, and classroom covariates, and across a number of alternative peer skill specifications. The study also shows some evidence of a weaker link between peer vocabulary skills and children's vocabulary outcomes. There is no evidence, however, of a relationship between peer math skills and individual math outcomes, nor any statistically significant associations between peer learning behavior measures and any of the pre-academic outcomes. The findings also point to cross-domain skill influence at the individual level (i.e., individual scores in other pre-academic subject areas and behavior are often predictive of individual outcomes); however, there appear to be few cross-domain relationships at the peer skill level.

Why do I find results for reading, but not other skills? Theoretically, given the pedagogical structure of most early education programs, which encourages peer interaction, more skilled peers may provide classmates with richer learning opportunities, independent of the teacher's efforts. Therefore, the stronger finding for reading versus vocabulary is perhaps not surprising, given previous research suggesting that reading skills such as letter and word identification are more "teachable" than receptive vocabulary (Christian et al., 2000). This same research suggests that math skills are also sensitive to instruction; however, observational studies indicating that few math learning opportunities occur in preschool classrooms may explain the non-significant findings regarding the influence of peer math skills (Pianta, 2008).

Also of note is that in this sample, children's reading skills are, on average, stronger than their math or vocabulary skills. Perhaps Head Start students lack exposure to children with vocabulary and math skills strong enough to impact learning. The results for the alternative

specifications suggesting that children in classrooms with a higher percentage of children with higher vocabulary scores *do* realize some benefit to their individual vocabulary scores also provides some support for this line of thinking. These findings are also consistent with findings from the recent experimental evaluation of Head Start, which show growth in participants' letter recognition skills, but not math or vocabulary skills (Puma et al., 2005). Unfortunately, in these models, I cannot differentiate whether such an effect is driven by children's initial skill levels or by teaching practices that may emphasize one skill over the other.

Although peer reading skills are predictive of individual reading skills, the effect is substantively small (effect size =.06). In general, estimates in this study are similar to those found by Mashburn et al. (2009), but smaller than those reported by Henry & Rickman (2007). Again, one explanation for the lower estimates in this study may be the initially lower skill levels of Head Start children compared to those sampled in previous studies. This argument is consistent with findings from Lee et al.'s (1998) previous study of at-risk preschool children, which found no link between the cognitive skill level of the class and individual cognitive outcomes. Furthermore, although one might hypothesize that children with lower abilities might have the most to gain from exposure to higher skilled peers, Mashburn et al. (2009) found the opposite. In their study, higher skilled children seemed better able to take advantage of opportunities to interact with other high skilled children. Since the FACES sample contains fewer high skilled children, this may also help to explain the smaller effect size findings.

I find little evidence of variation in the relationship between peer skills and children's pre-academic outcomes based on a child's initial pre-academic and learning behavior skill levels, pro-social skills, gender, or race. This is perhaps not surprising given the theoretical ambiguity, as well as mixed empirical evidence, regarding the potential direction of such underlying

mechanisms. The main positive association between peer reading skills and growth in children's reading skills, however, remains robust in these alternative specifications.

A unique contribution of this study is the concurrent consideration of the influence of children's behavior skills on pre-academic outcomes. Unlike previous studies, I am able to include direct assessments of children's behavior, and find that although individual behavior skills are predictive of children's vocabulary and math scores, peer behavior skills are not predictive of any of the outcomes of interest. I also find that the inclusion of these behavior measures does not affect the coefficient of any of the pre-academic peer skill measures.

These findings are also presented with the acknowledgement of several limitations. First, I recognize that these are non-experimental analyses, and any resulting associations may not be causal in nature. The analytic approaches employed attempt to reduce omitted variable bias by controlling not only for important child, family, and classroom characteristics, but also for children's prior achievement. Nevertheless, it is possible that bias due to unmeasured factors affecting achievement remains.

A second important limitation to this study is that the models assume a homogeneous treatment effect; in other words, that each student has an equal effect on each other student. This assumption has been refuted by some researchers (See Harris, 2010; Hoxby & Weingarth, 2005), who argue that evidence from studies with older children suggests that certain classmates (e.g., those of the same race or with higher initial levels of achievement) may be more influential than others. Theoretically, this concern is at least partially mitigated by the structure of Head Start pedagogy, in which teachers may be more controlling of peer groupings, and there is an emphasis on inclusion and interaction with all fellow students (USDHHS, ACF, Office of Head Start, 2009). Furthermore, there is generally less stability in preschool friendships compared to

the friendships of older children (Rubin et al., 2006). Developmental research does suggest, however, that preschool-aged children appear to be more attracted to peers who are similar; for example, in terms of age, sex, or behavioral tendencies (Rubin et al., 2006). Unfortunately, given the sample size in each classroom, such hypotheses are not testable using the FACES data.

Finally, the homogeneity of the FACES population is both a blessing and a curse. On the one hand, homogeneity in population and programming reduces concerns presented by selection bias, in that program quality, a potential omitted variable, is likely to be relatively more similar. On the other hand, it means that variation in outcomes and peer skill levels are decreased, particularly at the high end of the skill distribution; thus, making it more difficult to detect peer effects. The fact that the FACES data include only first-year Head Start students, and therefore exclude measures of students participating in their second year of the program may also attenuate peer effect estimates, assuming Head Start is an effective intervention that increases skill levels.

In sum, the findings from this study contribute to the mounting evidence that preschool peers may indeed contribute to gains in children's pre-academic skills; however, in Head Start, such contributions are limited to language skills and are substantively small. If, as these findings suggest, lower-skilled children may learn more from opportunities to interact with higher-skilled children, then policy makers should consider peer composition in their calculus of potential tradeoffs between targeted or universal preschool programs, which are likely to differ in terms of student heterogeneity and average baseline skills. Furthermore, even if program composition is largely unalterable, knowing that peers contribute to preschoolers' learning recommends teacher training regarding the use of proven collaborative learning techniques that maximize the benefits of peer effects within a given classroom (Wilkinson, Parr, Fung, Hattie, & Townsend, 2002).

Table 2.1: Descriptive Statistics

	Min.	Max.	Mean	S.D.
Basic Child Characteristics				
Child Gender (1=Male, 0=Female)	0	1	0.48	0.50
Race/Ethnicity				
White, Non-Hispanic (Referent)	0	1	0.34	0.47
African American	0	1	0.36	0.48
Hispanic	0	1	0.22	0.41
Other race/ethnicity	0	1	0.08	0.28
Child Age at post-test (months)	41	69	54.1	6.66
Maternal Education				
Less than high school diploma	0	1	0.29	0.45
High school diploma (Referent)	0	1	0.39	0.49
Any post-secondary education	0	1	0.33	0.47
Extended Child and Family Characteristics				
English Language Learner	0	1	0.09	0.29
Disability	0	1	0.16	0.37
Center-based care before Head Start	0	1	0.2	0.4
Log of family income	4.32	11.61	9.43	0.86
Income below poverty line	0	1	0.68	0.47
Number of children	1	13	2.64	1.38
Married mother	0	1	0.38	0.49
Frequency read to child	0	7	4.71	2.32
Parental warmth	2.8	5	4.38	0.42
Parental depression	0	36	6.98	6.43
Basic Classroom Characteristics				
Child/Staff Ratio	2	15	6.63	2.12
Class size	4	32	14.42	3.32
Arnett Lead Teacher Behavior Score	12	90	73.25	11.98
Extended Classroom Characteristics				
Teacher Education				
No post-secondary education (Referent)	0	1	0.05	0.23
Associate's degree or some college	0	1	0.56	0.50
Bachelor's or advanced degree	0	1	0.38	0.49
ECERS-R average score	2.06	6.60	4.66	0.95

Table 2.1: Descriptive Statistics, Continued

	Min.	Max.	Mean	S.D.
Timing of Testing				
October pre-test	0	1	0.31	0.46
November pre-test	0	1	0.42	0.49
December pre-test	0	1	0.26	0.44
Days between pre- and post-test	111	249	170.02	21.03
Children's Fall Test Scores				
PPVT (Vocabulary)	37.28	120.15	84.79	11.6
WJ-Letters & Dictation (Reading)	30.19	169	92.45	16.99
WJ-Applied Problems (Math)	26.69	143	88.64	17.70
Preschool Learning Behavior Scale t-score	12	64	50.54	10.43
Peer (Fall) Skills Scores				
Peer-PPVT (Vocabulary)	54.56	102.53	84.79	6.82
Peer-WJ-Letters & Dictation (Reading)	60.87	123	92.45	8.36
Peer-WJ-Applied Problems (Math)	56.76	117.83	88.64	9.4
Peer-Preschool Learning Behavior Scale (t-score)	32.33	63.88	50.54	6.35
Children's Spring Test Scores (Outcomes)				
PPVT (Vocabulary)	45.57	126.35	87.33	11.31
WJ-Letters & Dictation (Reading)	39.31	159.79	98.07	16.02
WJ-Applied Problems (Math)	38.02	156.04	91.74	16.11

Note: N=19,170 (across 10 imputed data sets); descriptives provided for children in classrooms with at least 4 fall academic test socres.

Table 2.2: Bivariate Regressions of Peer Skills on Children's Pre-Academic Outcomes

	Vocabulary		Reading		Math	
Peer Skills	0.31**	0.09**	0.14**	0.04[t]	0.16**	0.02
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Fall Score		0.65**		0.58**		0.62**
		(0.02)		(0.02)		(0.02)
N	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** $p<0.01$, * $p<0.05$; [t] $p<0.10$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 2.3: Summary of Results from Regressions of Average Peer Skills on Children's Receptive Vocabulary Skills in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Vocabulary	0.05*	0.03	0.03	0.03	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Fall Vocabulary	0.62**	0.61**	0.61**	0.52**	0.52**	0.51**	0.51**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Child Fall Test Scores							
Fall Reading				0.03	0.03	0.02	0.02
				(0.02)	(0.02)	(0.02)	(0.02)
Fall Math				0.16**	0.16**	0.14**	0.14**
				(0.02)	(0.02)	(0.02)	(0.02)
Other Peer Skills Scores							
Peer Reading					-0.02	-0.02	-0.02
					(0.02)	(0.02)	(0.02)
Peer Math					0.01	0.01	0.01
					(0.03)	(0.03)	(0.03)
Behavior Measures							
Ind. Preschool Learning Behavior						0.06*	0.06*
						(0.02)	(0.02)
Peer Preschool Learning Behavior							0.00
							(0.02)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 2.4: Summary of Results from Regressions of Average Peer Skills on Children's Reading Skills in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Reading	0.04	0.04^{t}	0.04^{t}	0.04	0.06*	0.06*	0.06*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Fall Reading	0.55**	0.55**	0.55**	0.47**	0.47**	0.47**	0.47**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Child Fall Test Scores							
Fall Vocabulary				0.07*	0.07*	0.07*	0.07*
				(0.03)	(0.03)	(0.03)	(0.03)
Fall Math				0.15**	0.16**	0.15**	0.15**
				(0.03)	(0.03)	(0.03)	(0.03)
Other Peer Skills Scores							
Peer Vocabulary					-0.01	-0.01	-0.01
					(0.03)	(0.03)	(0.03)
Peer Math					-0.05	-0.05	-0.05
					(0.03)	(0.03)	(0.03)
Behavior Measures							
Ind. Preschool Learning Behavior						0.04^{t}	0.04
						(0.02)	(0.02)
Peer Preschool Learning Behavior							0.01
							(0.03)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 2.5: Summary of Results from Regressions of Average Peer Skills on Children's Math Skills in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Math	-0.01	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Fall Math	0.59**	0.59**	0.59**	0.44**	0.44**	0.42**	0.42**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Child Fall Test Scores							
Fall Vocabulary				0.20**	0.20**	0.19**	0.19**
				(0.03)	(0.03)	(0.03)	(0.03)
Fall Reading				0.16**	0.16**	0.15**	0.15**
				(0.02)	(0.02)	(0.02)	(0.02)
Other Peer Skills Scores							
Peer Vocabulary					-0.00	-0.00	-0.00
					(0.03)	(0.03)	(0.03)
Peer Reading					0.02	0.03	0.03
					(0.03)	(0.03)	(0.03)
Behavior Measures							
Ind. Preschool Learning Behavior						0.10**	0.10**
						(0.02)	(0.02)
Peer Preschool Learning Behavior							0.00
							(0.02)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 2.6: Summary of Results from Regressions of Average Peer Skills on Children's Changes in Pre-Academic Skills between Fall and Spring of the Head Start Year

	Chage Score Outcomes		
	Vocabulary	Reading	Math
Peer Vocabulary	0.02	0.02	0.00
	(0.03)	(0.04)	(0.04)
Peer Reading	-0.01	0.06^{t}	0.02
	(0.02)	(0.04)	(0.03)
Peer Math	0.00	-0.07^{t}	-0.08^{t}
	(0.03)	(0.04)	(0.04)
Peer Learning Behavior	0.02	0.03	0.05^{t}
	(0.02)	(0.03)	(0.03)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, excluding the related fall score, which is instead captured in the outcome (spring score - fall score).

Table 2.7: Summary of Results from Regressions of Average Peer Skills on Children's Pre-Academic Skills in the Spring of the Head Start Year, Including Spanish Test Scores

	Spring Outcomes		
	Vocabulary	Reading	Math
Peer Vocabulary	0.03	-0.07*	0.02
	(0.03)	(0.03)	(0.03)
Peer Reading	-0.02	0.09**	0.02
	(0.02)	(0.04)	(0.02)
Peer Math	0.03	0.03	-0.05
	(0.03)	(0.03)	(0.03)
Peer Learning Behavior	-0.01	-0.04	-0.01
	(0.02)	(0.03)	(0.03)
N	2,190	2,190	2,190

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. These models use reconstructed peer scores, substituting Spanish test scores for Spanish speaking students who failed a language screener in the fall or spring. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus dummies indicating whether students took tests in Spanish in the fall and/or spring.

Table 2.8, Panel 1: Summary of Results from Regressions of Maximum Peer Skills on Children's Behavior Skills in Spring of the Head Start Year

	Spring Outcomes		
	Vocabulary	Reading	Math
Max Peer Vocabulary	0.02	-0.02	-0.03
	(0.03)	(0.03)	(0.03)
Max Peer Reading	-0.01	0.08*	0.01
	(0.02)	(0.03)	(0.03)
Max Peer Math	0.02	0.00	0.05[t]
	(0.03)	(0.03)	(0.03)
Max Peer Learning Behavior	-0.00	-0.03	-0.03
	(0.03)	(0.03)	(0.03)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5.

Table 2.8, Panel 2: Summary of Results from Regressions of Minimum Peer Skills on Children's Behavior Skills in Spring of the Head Start Year

	Spring Outcomes		
	Vocabulary	Reading	Math
Min Peer Vocabulary	-0.00	-0.02	-0.01
	(0.03)	(0.03)	(0.03)
Min Peer Reading	-0.00	-0.01	0.02
	(0.02)	(0.03)	(0.02)
Min Peer Math	0.02	-0.04	-0.05[t]
	(0.03)	(0.03)	(0.03)
Min Peer Learning Behavior	-0.00	0.01	0.01
	(0.02)	(0.03)	(0.02)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5.

Table 2.9: Summary of Results from Regressions of Percent of Class with Scores in Highest and Lowest Quartiles on Children's Pre-Academic Skills in Spring of the Head Start Year

	Vocabulary			Reading			Math		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
% Hi Quart Vocabulary	0.05^{t}		0.06*	0.03		0.04	0.01		0.02
	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)
% Hi Quart Reading	-0.03		0.01	0.08*		0.09*	0.05^{t}		0.06*
	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)
% Hi Quart Math	-0.00		0.00	-0.03		-0.03	-0.00		0.01
	(0.02)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)
% Hi Quart Learning Behaviors	-0.01		-0.00	0.01		0.01	0.00		-0.00
	(0.03)		(0.02)	(0.03)		(0.03)	(0.02)		(0.02)
% Lo Quart Vocabulary		0.01	0.03		0.01	0.04		0.02	0.03
		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)
% Lo Quart Reading		0.02	0.02		0.01	0.03		0.01	0.02
		(0.02)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)
% Lo Quart Math		0.04	0.02		0.01	-0.02		0.03	0.00
		(0.02)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)
% Lo Quart Learning Behaviors		0.00	0.00		-0.01	-0.00		-0.02	-0.02
		(0.02)	(0.02)		(0.03)	(0.03)		(0.03)	(0.03)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. All models include a full set of covariates (mimicking Model 7 in Tables 2.3-2.5).

Table 2.10: Summary of Results from Regressions of Average Peer Skills on Children's Vocabulary Skills in the Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Vocabulary Scores			
	Model 1	Model 2	Model 3	Model 4
Hi Ind. Voc.*Peer Voc.	0.03		0.03	0.03
	(0.04)		(0.05)	(0.05)
Lo Ind. Voc.*Peer Voc.		-0.02	-0.01	-0.01
		(0.05)	(0.05)	(0.05)
Peer Vocabulary	0.02	0.04	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Ind. Read.*Peer Read.				0.02
				(0.04)
Lo Ind. Read.*Peer Read.				-0.02
				(0.05)
Peer Reading	-0.02	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Hi Ind. Math*Peer Math				0.00
				(0.05)
Lo Ind. Math*Peer Math				-0.02
				(0.05)
Peer Math	0.01	0.01	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Ind. Beh.*Peer Beh.				-0.06
				(0.05)
Lo Ind. Beh.*Peer Beh.				-0.01
				(0.05)
Peer Learning Behavior	0.00	0.00	0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.03)
N	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.11: Summary of Results from Regressions of Average Peer Skills on Children's Reading Skills in the Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Reading Scores			
	Model 1	Model 2	Model 3	Model 4
Hi Ind. Read.*Peer Read.	0.01		0.01	0.01
	(0.05)		(0.05)	(0.05)
Lo Ind. Read.*Peer Read.		0.00	0.01	0.01
		(0.05)	(0.05)	(0.05)
Peer Reading	0.06*	0.06*	0.05^{t}	0.06^{t}
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Ind. Voc.*Peer Voc.				0.05
				(0.06)
Lo Ind. Voc.*Peer Voc.				-0.06
				(0.06)
Peer Vocabulary	-0.02	-0.01	-0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Ind. Math*Peer Math				-0.01
				(0.05)
Lo Ind. Math*Peer Math				0.05
				(0.05)
Peer Math	-0.05	-0.04	-0.04	-0.05
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Ind. Beh.*Peer Beh.				-0.03
				(0.05)
Lo Ind. Beh.*Peer Beh.				0.03
				(0.05)
Peer Learning Behavior	0.01	0.00	0.00	0.01
	(0.02)	(0.02)	(0.02)	(0.03)
N	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.12: Summary of Results from Regressions of Average Peer Skills on Children's Math Skills in the Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Math Scores			
	Model 1	Model 2	Model 3	Model 4
Hi Ind. Math*Peer Math	0.03		0.05	0.05
	(0.05)		(0.05)	(0.05)
Lo Ind. Math*Peer Math		0.04	0.05	0.05
		(0.05)	(0.05)	(0.05)
Peer Math	-0.04	-0.04	-0.06	-0.06
	(0.03)	(0.03)	(0.04)	(0.04)
Hi Ind. Read.*Peer Read.				0.01
				(0.05)
Lo Ind. Read.*Peer Read.				0.01
				(0.05)
Peer Reading	0.02	0.03	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Ind. Voc.*Peer Voc.				-0.01
				(0.05)
Lo Ind. Voc.*Peer Voc.				-0.00
				(0.05)
Peer Vocabulary	-0.00	-0.00	-0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Ind. Beh.*Peer Beh.				0.02
				(0.05)
Lo Ind. Beh.*Peer Beh.				-0.00
				(0.05)
Peer Learning Behavior	0.01	0.01	0.01	-0.00
	(0.02)	(0.02)	(0.02)	(0.03)
N	1,917	1,917	1,917	1,917

NOTE: ** $p<0.01$, * $p<0.05$; ^{t} $p<0.10$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.13: Summary of Results from Regressions of Average Peer Skills on Children's Vocabulary Skills in the Spring of the Head Start Year, Including Initial Pro-Social Skill Level Interactions

	Spring Vocabulary Scores			
	Model 1	Model 2	Model 3	Model 4
Hi Pro-Social*Peer Voc.	0.03		0.03	0.01
	(0.04)		(0.05)	(0.06)
Lo Pro-Social*Peer Voc.		-0.00	0.00	0.00
		(0.05)	(0.05)	(0.06)
Peer Vocabulary	0.03	0.03	0.02	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Pro-Social*Peer Read.				0.01
				(0.05)
Lo Pro-Social*Peer Read.				0.03
				(0.05)
Peer Reading	-0.02	-0.02	-0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.03)
Hi Pro-Social*Peer Math				0.02
				(0.06)
Lo Pro-Social*Peer Math				-0.02
				(0.06)
Peer Math	0.01	0.01	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Pro-Social.*Peer Beh.				0.02
				(0.04)
Lo Pro-Social*Peer Beh.				0.04
				(0.05)
Peer Learning Behavior	0.00	0.00	0.00	0.06*
	(0.02)	(0.02)	(0.02)	(0.02)
N	1,917	1,917	1,917	1,917

NOTE: ** $p<0.01$, * $p<0.05$; ^{t} $p<0.10$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.14: Summary of Results from Regressions of Average Peer Skills on Children's Reading Skills in the Spring of the Head Start Year, Including Initial Pro-Social Skill Level Interactions

	Spring Reading Scores			
	Model 1	Model 2	Model 3	Model 4
Hi Pro-Social*Peer Read.	0.04		0.02	-0.02
	(0.03)		(0.05)	(0.06)
Lo Pro-Social*Peer Read.		-0.07	-0.07	-0.08
		(0.06)	(0.06)	(0.06)
Peer Reading	0.05[t]	0.08*	0.07*	0.08*
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Pro-Social*Peer Voc.				-0.01
				(0.07)
Lo Pro-Social*Peer Voc.				0.09
				(0.07)
Peer Vocabulary	-0.01	-0.01	-0.01	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Pro-Social*Peer Math				0.09
				(0.07)
Lo Pro-Social*Peer Math				-0.09
				(0.07)
Peer Math	-0.05	-0.05	-0.05	-0.05
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Pro-Social*Peer Beh.				0.04
				(0.05)
Lo Pro-Social*Peer Beh.				0.06
				(0.06)
Peer Learning Behavior	0.02	0.01	0.01	0.04
	(0.02)	(0.03)	(0.03)	(0.02)
N	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.15: Summary of Results from Regressions of Average Peer Skills on Children's Math Skills in the Spring of the Head Start Year, Including Initial Pro-Social Skill Level Interactions

	Spring Math Scores			
	Model 1	Model 2	Model 3	Model 3
Hi Pro-Social*Peer Math	0.02		0.04	0.03
	(0.05)		(0.05)	(0.07)
Lo Pro-Social*Peer Math		0.04	0.05	0.02
		(0.05)	(0.05)	(0.07)
Peer Math	-0.04	-0.04	-0.05	-0.04
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Pro-Social*Peer Read.				0.01
				(0.05)
Lo Pro-Social*Peer Read.				0.02
				(0.06)
Peer Reading	0.03	0.03	0.03	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
Hi Pro-Social*Peer Voc.				-0.05
				(0.06)
Lo Pro-Social.*Peer Voc.				0.04
				(0.06)
Peer Vocabulary	-0.00	-0.00	-0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.04)
Hi Pro-Social*Peer Beh.				0.09[t]
				(0.05)
Lo Pro-Social*Peer Beh.				0.02
				(0.05)
Peer Learning Behavior	0.00	0.00	0.00	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
N	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.16: Summary of Results from Regressions of Average Peer Skills on Children's Pre-Academic Skills in the Spring of the Head Start Year, Including Gender Interactions

	Spring Pre-Academic Scores					
	Vocabulary		Reading		Math	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Male*Peer Vocabulary	-0.04	-0.03		0.01		0.02
	(0.04)	(0.05)		(0.05)		(0.05)
Peer Vocabulary	0.05	0.05	-0.01	-0.02	-0.00	-0.01
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Male*Peer Reading		-0.07^{t}	-0.02	-0.00		-0.01
		(0.04)	(0.04)	(0.05)		(0.04)
Peer Reading	-0.02	0.01	0.07*	0.06^{t}	0.03	0.03
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Male*Peer Math		0.03		-0.05	0.01	0.00
		(0.05)		(0.06)	(0.04)	(0.05)
Peer Math	0.01	-0.00	-0.05	-0.02	-0.04	-0.03
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Male*Peer Learning Beh.		-0.06		0.00		0.00
		(0.04)		(0.04)		(0.04)
Peer Learning Behavior	0.00	0.03	0.01	0.01	0.00	0.00
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
N	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Table 2.17: Summary of Results from Regressions of Average Peer Skills on Children's Pre-Academic Skills in the Spring of the Head Start Year, Including Race Interactions

	Spring Pre-Academic Scores					
	Vocabulary		Reading		Math	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Black*Peer Vocabulary	-0.10[t]	-0.08		-0.09		0.08
	(0.05)	(0.07)		(0.08)		(0.07)
Hispanic*Peer Vocabulary	-0.05	-0.04		-0.02		-0.05
	(0.05)	(0.06)		(0.07)		(0.05)
Peer Vocabulary	0.07*	0.06	-0.01	0.02	-0.01	-0.01
	(0.04)	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
Black*Peer Reading		-0.05	-0.06	-0.05		-0.05
		(0.05)	(0.05)	(0.06)		(0.05)
Hispanic*Peer Reading		-0.06	-0.08	-0.05		-0.02
		(0.06)	(0.05)	(0.06)		(0.06)
Peer Reading	-0.02	0.01	0.10*	0.09*	0.03	0.06
	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Black*Peer Math		-0.02		0.03	-0.07	-0.07
		(0.06)		(0.07)	(0.05)	(0.07)
Hispanic*Peer Math		0.04		-0.04	-0.01	0.05
		(0.07)		(0.08)	(0.05)	(0.07)
Peer Math	0.01	0.01	-0.05	-0.05	-0.00	-0.02
	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)
Black*Peer Learning Beh.		0.03		-0.03		-0.05
		(0.05)		(0.05)		(0.05)
Hispanic*Peer Learning Beh.		-0.01		-0.06		-0.07
		(0.05)		(0.06)		(0.05)
Peer Learning Behavior	0.00	-0.01	0.01	0.04	0.00	0.03
	(0.02)	(0.04)	(0.03)	(0.04)	(0.02)	(0.04)
N	1917	1917	1917	1917	1917	1917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 2.3-2.5, plus the interactions indicated in the first column.

Appendix 2.A : Pairwise Correlations of Average Peer Skills in Fall of Head Start Year

Variables	1	2	3	4
1. Peer-PPVT (Vocabulary)	--			
2. Peer-WJ-Letters & Dictation (Reading)	0.28*	--		
3. Peer-WJ-Applied Problems (Math)	0.58*	0.40*	--	
4. Peer-Preschool Learning Behavior Scale t-score	0.12*	0.18*	0.24*	--

NOTE: N=19,170 (across 10 imputed data sets); * $p<0.05$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Appendix 2.B: Child, Family, and Classroom Covariate Description and Coding

	Concept	Coding
Individual Child Characteristics	Initial skill levels	Continuous variables; child's individual standardized test scores (Academic skills=PPVT, reading, math; Behavior skills=learning related behaviors)
	Age at assessment	Continuous variable; months
	Gender	Dummy variable; 1=male
	Race/Ethnicity	Series of dummy variables for black, Hispanic, other race (omitted=white)
	Disability	Dummy variable; 1=has disability
	Language learning status	Dummy variable; 1=English language learner
	Prior childcare experience	Dummy variable; 1=previously attended center-based care
Individual Family Characteristics	Number of children	Continuous variable
	Family structure	Dummy variable; 1=mother married
	Maternal education	Series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent)
	Income	Continuous variable; natural log annual family income
	Literacy behavior	Continuous variable; how often child has been read to in the last week (scale of 0, "not at all" to 7, "every day")
	Maternal depression	Continuous variable derived from12 items from CES-D Depression Scale
	Parent warmth and sensitivity	Continuous variable derived from 5 items asking about parenting behavior
Classroom Characteristics	Child/staff ratio	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Class size	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Teacher Education	Series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent)
	Classroom process (global) quality	Continuous variable; ECERS-R mean score
	Teacher behavior	Continuous variable; Arnett Scale of Caregiver Behavior (lead teacher)

Appendix 2.C: Complete Results from Regressions of Average Peer Skills on Children's Pre-Academic Skills in the Spring of the Head Start Year (Model 7, Tables 2.3-2.5)

	Vocabulary	Reading	Math
Peer Vocabulary	0.03	-0.01	-0.00
	(0.03)	(0.03)	(0.03)
Fall Vocabulary	0.51**	0.07*	0.19**
	(0.02)	(0.03)	(0.03)
Peer Reading	-0.02	0.06*	0.03
	(0.02)	(0.03)	(0.03)
Fall Reading	0.02	0.47**	0.15**
	(0.02)	(0.02)	(0.02)
Peer Math	0.01	-0.05	-0.03
	(0.03)	(0.03)	(0.03)
Fall Math	0.14**	0.15**	0.42**
	(0.02)	(0.03)	(0.02)
Peer Preschool Learning Behavior	0.06*	0.04	0.10**
	(0.02)	(0.02)	(0.02)
Fall Preschool Learning Behavior	0.00	0.01	0.02
	(0.02)	(0.03)	(0.01)
Age	0.01^{t}	-0.02**	-0.01**
	(0.00)	(0.00)	(0.00)
Male	-0.00	-0.15**	0.02
	(0.04)	(0.04)	(0.04)
Black	-0.28**	0.15*	-0.21**
	(0.05)	(0.06)	(0.06)
Hispanic	-0.08	0.03	-0.06
	(0.06)	(0.07)	(0.06)
Other Race	-0.07	0.15^{t}	-0.10
	(0.07)	(0.09)	(0.08)
Mom Ed. <High School	-0.05	-0.02	-0.01
	(0.04)	(0.05)	(0.05)
Mom Ed. Any Postsecondary	0.10*	0.10*	0.05
	(0.04)	(0.05)	(0.04)
Child/Staff Ratio	-0.01	0.02	0.02^{t}
	(0.01)	(0.01)	(0.01)
Class Size	-0.00	0.01	-0.02*
	(0.01)	(0.01)	(0.01)
Teacher Sensitivity (Arnett)	0.00^{t}	-0.00	0.00
	(0.00)	(0.00)	(0.00)
English Language Learner	-0.18*	0.09	-0.04
	(0.07)	(0.09)	(0.08)
Disability	-0.10^{t}	0.06	-0.06
	(0.05)	(0.05)	(0.05)

Appendix 2.C: Complete Results from Regressions of Average Peer Skills on Children's Pre-Academic Skills in the Spring of the Head Start Year (Model 7, Tables 2.3-2.5), Continued

	Vocabulary	Reading	Math
Center Care before Head Start	-0.01	-0.01	0.01
	(0.04)	(0.05)	(0.05)
Family Income	0.03	-0.01	-0.03
	(0.03)	(0.03)	(0.03)
Poverty	0.03	-0.02	-0.01
	(0.05)	(0.06)	(0.06)
Number of Children	-0.02	-0.03^{t}	0.00
	(0.01)	(0.01)	(0.02)
Married	-0.02	0.02	0.03
	(0.04)	(0.05)	(0.04)
Parental Reading Scale	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)
Parental Warmth	0.01	-0.06	-0.02
	(0.04)	(0.05)	(0.05)
Parental Depression	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)
November Pre-test	0.10*	0.05	0.07
	(0.05)	(0.06)	(0.05)
December Pre-test	0.12^{t}	-0.12	0.02
	(0.07)	(0.08)	(0.07)
Days between Pre- and Post-test	0.00*	0.00	0.00*
	(0.00)	(0.00)	(0.00)
Classroom Quality (ECERS-R)	-0.02	0.07^{t}	0.03
	(0.03)	(0.04)	(0.03)
Teacher has Bachelor's Degree	0.01	0.12	-0.10
	(0.09)	(0.11)	(0.10)
Teacher has Associate's Degree	-0.00	0.09	-0.13
	(0.09)	(0.10)	(0.10)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

CHAPTER III

THE INFLUENCE OF PRESCHOOL PEERS ON CHILDREN'S BEHAVIORAL OUTCOMES: EVIDENCE FROM THE 2003 HEAD START FAMILY AND CHILD EXPERIENCES SURVEY (FACES)

Background

Although preference in the early childhood education literature is often given to program impacts on cognitive skills, Head Start was founded on the idea that social and emotional skills, such as attention and problem behavior, may also be important predictors of later scholastic success. The importance of children's socio-emotional development is also clearly reflected in the Head Start Program Performance Standards (USDHHS, ACF, Office of Head Start, 2009). Recent reports by the National Research Council also emphasize the important links between emotional and behavioral adjustment, and learning and socialization experiences in early childhood classrooms (National Research Council, 2001; National Research Council and Institute of Medicine, 2001). Some studies suggest that the development of social and emotional skills may be important for early school engagement and classroom adjustment (Denham, 2006; Knitzer, 2003; Ladd, Herald, & Kochel, 2006; Raver, 2002; Thompson & Raikes, 2007). Others document links between early behavior problems and social and academic difficulties in preschool and during the transition to kindergarten, as well as problems forming relationships with peers and teachers (Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003; Fantuzzo, Bulotsky-Shearer, Fusco, & McWayne, 2005; Huffman, Mehlinger, & Kerivan, 2000).

Important social-cognitive developmental milestones that occur during the preschool years include the development of a more explicit sense of self, the emergence of achievement-

related attributions, improved delay of gratification, internalization of many pro-social standards and prohibitions, and the increased development of social problem-solving strategies.

Developmentalists argue that such skills develop at least in part through children's interactions with their environment (Berk, 2000). In addition to children's interactions with adult caregivers and teachers, their experiences in early childhood education classrooms with peers, in terms of exposure to pro- and anti-social behaviors, as well as opportunities to practice particular skills, are also important (Rubin et al., 2006). Given that in early childhood classrooms, "play" is often considered equivalent to "learning" (Singer, Golinkoff, & Hirsch-Pasek, 2006), successful engagement within the preschool peer context may promote the development of both social and academic skills including emotional regulation, problem solving, empathy, cognitive, and language skills (Coolahan, Fantuzzo, Mendez, & McDermott, 2000; McClelland & Morrison, 2003). Social learning theory also suggests that peers may encourage the development of social skills by rewarding or positively reinforcing culturally acceptable behaviors, and by punishing or ignoring non-normative behaviors (Rubin et al., 2006).

A rarely investigated piece of the peer effects puzzle is the potential impact of early childhood education peer contexts on long-term peer relational skills and other social behavior in children. Howes and Lee (2006) argue that peer contexts vary significantly across early educational environments, and that this is often overlooked by researchers. The authors' own longitudinal work suggests that social competency in peer relations at the preschool level (e.g., engagement in complex play), may predict increased pro-social behavior, lower levels of aggression, and less social withdrawal through middle childhood (Howes & Lee, 2006). If such relationships reflect a causal effect of peer interactions on children's development of pro-social skills, the results suggest that children who develop pro-social skills early on may continue to

build them throughout their childhood, pointing to the importance of early skills as building blocks of later social skills. More specifically, Howes (1988) suggests that it may benefit children to enter peer groups at a younger age, and to maintain familiar, stable peer networks over time.

Literature Review

Despite the potentially important theoretical role of peers in preschool children's socio-emotional development, however, little research beyond a body of work focusing on early gender socialization (e.g., see Maccoby & Jacklin, 1987) exists. Although that research suggests that preschool children's behaviors and play choices may indeed be significantly influenced by peer preferences, it does not necessarily consider the influential role that peers' *skills* may play. The few studies of potential peer effects on preschool children's behavior outcomes are even more limited regarding outcomes and covariates than those investigating pre-academic outcomes.

Given studies of older children that suggest that peers may reinforce antisocial and delinquent behaviors (e.g., see Dishion, Eddy, Haas, Li, & Spracklen, 1997), it is perhaps not surprising that most peer behavior studies in the preschool context have focused on children's exposure to problematic behaviors. For example, Anthony, Anthony, Morrel, and Acosta (2005) examined associations between the prevalence of problem behaviors in the classroom and individual student scores on teacher behavior ratings. Using a sample of 336 three- and four-year-old African-American children attending a Baltimore Head Start program, the authors found significant negative correlations between children's Social Competence and Behavior Evaluation scores and the percentage of students in the classroom rated in the problem categories on the internalizing and externalizing scales.

Some research suggests potential gender differences in associations between peer behavior and individual behavior. For example, in a sample of low-risk preschoolers and kindergarteners, Hanish, Martin, Fabes, Leonard, and Herzog (2005) found associations between exposure to externalizing peers and increased rates of problematic behaviors for girls, but not boys. This study and others documenting greater levels of peer play interaction for preschool girls versus boys (e.g., Coolahan et al., 2000) point to the importance of testing for gender interactions within the FACES sample.

Previous research also suggests that children's behavior may color their interactions with peers, suggesting the importance of testing for initial skill level and peer skill interactions. For example, in the case of externalizing behavior, research suggests that exposure to such behavior may not be equally distributed across all children. Aggressive children are more likely to seek out other externalizing peers, and aggressive children may also be actively excluded from play with non-aggressive children (Hanish et al., 2005; Hanish, Martin, Fabes, & Barcelo, 2008).

There is also reason to believe that the same theoretically ambiguous situation regarding potentially different associations peer effects and outcomes based on initial skill levels found in the pre-academic domain might also apply to behavior. Again, if one adopts a "skills beget skills" philosophy, children with more positive behavior skills or fewer problematic behaviors may be best situated to gain knowledge from interactions with other higher skilled children (Cunha & Heckman, 2007). However, children with lower pro-social skills or higher levels of problematic behavior might have more opportunity for growth resulting from interactions with better behaved classmates. Although empirical evidence supporting both theories has been found in preschool settings for language outcomes (Mashburn et al., 2009; Justice et al., 2011), there has been little research testing such potential interactions in the preschool behavior domain.

The limited nature of the existing studies regarding peer effects on behavior in the preschool classroom leaves much room for additional research. Most studies have focused on children's exposure to problematic behavior; yet, developmental theory highlights the importance of pro-social skill development, and positive learning behaviors are predictive of future academic success. The FACES data provide the opportunity to examine whether peer behavior matters in the development of both negative and positive behaviors. Previous research and theory also recommend the importance of investigating potential gender and initial skill level interactions, to test for differential associations between peer and individual behaviors in these groups of children.

Data and Methods

Data

In this study, I use data from the 2003 Head Start Family and Child Experiences (FACES) study, a nationally representative survey of children who entered Head Start in 2003, funded by the U.S. Department of Health and Human Services' Administration for Children and Families (ACF). These longitudinal data follow children and families served by Head Start from program entrance to the end of kindergarten. FACES data include detailed information on Head Start attendees' background characteristics, experiences, and developmental outcomes. At the beginning and end of the Head Start program year and again in kindergarten, high quality child assessments and parent interviews were completed. Teacher surveys were also conducted at the beginning and end of the Head Start year. Finally, center directors and education coordinators were also interviewed in the fall of the Head Start year, and classrooms were observed for overall quality, types of activities, and interactions between staff and children at the beginning and end of the Head Start year.

The FACES data provide several advantages over data used in the few previous explorations of preschool peer effects. First, all families represented in FACES are participants in the same early childhood education program, Head Start. Therefore, the problem of selection bias based on *type* of program (children of differing abilities selecting into different types of preschool programs) found in prior studies (e.g., Henry & Rickman, 2007) is eliminated. Second, a multitude of high quality child assessments of both pre-academic and behavioral skills, as well as observational assessments of classrooms, provide information about previously unmeasured child and classroom outcomes. Third, rich information regarding important, yet previously omitted control variables, including teacher and classroom quality, as well as parenting, is also available. Fourth, up to 16 children per classroom are assessed upon entry into Head Start (out of an average class size of 15), providing a more accurate estimate of the true peer skill level, compared to previous studies based on measures from as few as 3 students per classroom (e.g., Mashburn et al., 2009).[7] Fifth, although previous peer effect studies have been limited to data from a few sites or states, the FACES sample is relatively large and nationally representative of first year Head Start students; thus, increasing the study's external validity, particularly for preschool programs targeted to children from low-income families. Finally, given that these data were released in 2008, FACES provides one of the most up-to-date snapshots of children and families served by the country's largest publicly funded compensatory preschool program.

[7] Sojourner (2011) notes the potential bias in widely used peer-effects estimators in the extant literature, which do not account for the "missing data" of students not included in such estimators.

Sample

The FACES sampling frame was drawn from the 2002-03 Head Start Program Information Report (PIR) file, and the sample was stratified on region, urban/rural status, percent minority enrollment, auspice type (school-based, other), and percentage of non-English speaking children in the program.[8] Response rates for children, parents, and teachers in the Head Start year were quite high, ranging from 93 to 97 percent in the initial round of surveys in the fall of 2003, and 86 to 92 percent at the end of the Head Start year, in the spring of 2004 (Westat et al., 2008). The FACES data file contains information on 2,387 children in 373 different Head Start classrooms.

Although levels of missing data were quite low (almost always less than five percent at the item level), to maximize statistical power and minimize any bias due to missing data, I retain cases with missing information at the item level by using multiple imputation techniques (Graham, 2009; Allison, 2002). Specifically, I used STATA's *ice* program to impute 10 data sets and the *mim* command to conduct analyses across datasets (Royston, 2004). Covariates from full regression models were used in the imputation regressions.

Multiple imputation methods assume that data are missing at random; in other words, conditional on the observed variables in the analysis, the likelihood of missing data on a particular variable is unrelated to its value. Since this assumption does not hold for the English test scores of students tested in Spanish, I chose not to use imputed English test scores for these students; thus, decreasing the sample to 2,076 students in 367 classrooms. As a robustness check, I conduct separate analyses substituting Spanish speaking children's Spanish test scores (vs. excluding these children from the sample).

[8] Head Start programs in U.S. territories or Puerto Rico, Early Head Start programs, Migrant and Seasonal Head Start programs, American Indian Head Start programs, and Head Start programs already selected as part of the Head Start Quality Research Center samples were not included in the FACES study.

Furthermore, to increase the validity and reliability of the peer skill measures, I limit analyses to classrooms with at least four fall test scores by which to construct a measure of peer skills. Given this restriction, the final analytic sample size is 1,917 children in 292 classrooms. Descriptive statistics for the analytic sample are provided in Table 3.1.

Not surprisingly, these descriptive statistics reveal a disadvantaged population. Approximately 36 percent of children are black; 22 percent are Hispanic, and 34 percent are white. Upon entry to the program, 9 percent of children in the sample were identified as dual language learners, and 16 percent were identified as having disabilities. At the initial parent interview only 38 percent of mothers reported being married, and over two-thirds of families reported incomes below the federal poverty line. About 29 percent of mothers reported having less than a high school diploma. Children also demonstrate low levels of pre-academic skills upon program entry; on average, scoring an entire standard deviation below national norms for vocabulary; about three-quarters of a standard deviation for math, and half a standard deviation for reading. Despite this, there is evidence of substantial variation in peer skill levels for each subject.

Measures

Dependent Variables. A strength of the FACES survey is its assessment of multiple behavioral dimensions, including learning-related behaviors and pro-social skills, as well as problematic behaviors. Behavior outcomes are based on teacher reports of children's social skills and classroom behavior. Specifically, teachers were asked to indicate the extent to which a given statement (e.g., "follows teacher's directions") described the child, on a scale of 1 ("never") to 3 ("very often"). The Preschool Learning Behavior Scale (PLBS) includes 29 items

assessing learning-related behaviors (e.g., motivation, attention/persistence, and attitudes toward learning) (McDermott, Green, Francis, & Stott, 2000). Westat et al. (2008) report a Cronbach's alpha of .90 for the PLBS in both fall and spring of the Head Start year. The Social Skills scale includes 12 items drawn from the Personal Maturity Scale (Alexander & Entwisle, 1988) and the Social Skills Rating System (Elliot, Gresham, Freeman, & McCloskey, 1988) that measure how often children engage in cooperative classroom behavior (e.g., following directions and rules, helping put things away, and complimenting classmates). High reliabilities are reported for the measure at both fall (Cronbach's alpha=.88) and spring (Cronbach's alpha =.89) of the Head Start year. The Behavior Problems scale includes 14 items drawn from the Personal Maturity Scale (Alexander & Entwisle, 1988), the Child Behavior Checklist for Preschool-Aged Children, the Teacher Report (Achenbach, Edelbrock, & Howell, 1987), and the Behavior Problems Index (Zill, 1990) that measure negative child behaviors associated with later learning problems and grade retention (e.g., aggression, hyperactivity, anxiety, and withdrawal). High reliabilities are reported for this measure, also (Cronbach's alpha =.85 at both fall and spring) (Westat et al., 2008). Descriptive statistics for these measures are provided in Table 3.1. Again, for ease of interpretation, in regressions, the dependent variables are standardized to have a mean of 0 and standard deviation of 1.

Peer Skill Measures. Measures of students' average peer behavioral and pre-academic skills are constructed using standardized scores of all children assessed in a particular classroom, excluding a child's individual test score. In the analytic sample, an average of 7.14 fall scores per classroom are available (range=4 to 16), and the average class includes 14.44 students. Average peer skill measures were created for the PLBS, as well as the Social Skills and Behavior Problems Scales. Given research suggesting links between cognitive skills and behavior (e.g.,

See Braza et al., 2009), average peer pre-academic skill measures for vocabulary, reading, and math (described previously) are also included in the most complex models. Again, for ease of interpretation, in regressions, these variables are standardized to have a mean of 0 and standard deviation of 1; therefore, coefficients may be interpreted as effect sizes. Pairwise correlations between each of the measures are provided in Appendix 3.A.

Child and Family Covariates. In order to control for important confounds, I selected measures of individual child and family background that are likely to be correlated with a student's own behavior and that of their peers. Of particular importance is the inclusion of measures of each child's initial behavioral and pre-academic skills, assessed upon entry to Head Start. Additionally, demographic and other individual characteristics found in previous studies to be associated with preschool children's behavior include age, gender, race/ethnicity, disability and language learning status, and prior childcare experience. Family background characteristics such as number of children, family structure, income, and particularly maternal education, may also be predictive. The quality of preschool children's home environment, in terms of opportunities for cognitive stimulation and caregiver's responsiveness and warmth, may also have an important impact on children's development (Campbell, 1994; Belsky et al., 2007). Measures of such confounding factors have been largely absent from previous studies of preschool peer effects, but are available in the FACES data. Details about these individual child and family covariates and how they were coded are presented in Appendix 3.B.

Classroom Characteristics. In order to isolate the effects of preschool peers, it is also important to control for early education quality, which may also impact student behavioral development (Belsky et al., 2007; Peisner-Feinberg et al., 2001; Votruba-Drzal et al., 2010). Inclusion of such covariates, measured in the fall, importantly control for initial "levels" of

various proxies for quality; however, their inclusion does not rule out a potential indirect peer effect mechanism in which changes in socioemotional learning opportunities or teacher sensitivity due to children's classroom behaviors concurrently impact child behavior outcomes.[9]

The FACES data contain information about both structural quality (e.g., child to staff ratio and teacher education) and process quality (e.g., the quality of staff/child interactions and learning opportunities). Structural quality measures include class size, child/staff ratio, and teacher education. Process quality measures are based on observational data collected in the fall of the Head Start year. These include the revised version of the Early Childhood Environment Rating Scale (ECERS-R), which consists of 37 items and 7 subscales measuring the quality of personal care routines, furnishings, program structure, and opportunities to develop language skills, motor skills, social skills, and creativity (Harms, Clifford, & Cryer, 1998); and the Arnett Scale of Caregiver Behavior, which consists of 30 items assessing 5 areas of teacher behavior: sensitivity, punitiveness, detachment, permissiveness, and prosocial interaction (Arnett, 1989). Westat et al. (2008) report high reliabilities for both process quality measures (Cronbach's Alpha=.92 and .93, respectively). A more detailed description of how each classroom characteristic is coded is provided in Appendix 3.B.

Analytical Approach

Value-Added Models of Achievement. As pointed out by numerous researchers, empirical analysis of peer effects is difficult, due to conceptual and data problems (See, e.g., Brock and Durlauf, 2001; Manski, 1993; Moffitt, 2001). When studying peer effects, researchers typically model children's achievement as a function of current family, school, and peer interactions, as

[9] Whether such changes are present in the FACES data, and whether they are associated with classroom demographics and skill levels are topics of investigation in the final chapter of this dissertation.

well as prior experiences (Hanushek et al., 2003). The key concern is that what one identifies as a "peer effect" (i.e., what is measured) may actually be a proxy for other omitted or poorly measured factors that affect children's development (Moffit, 2001; Manski, 1993). For example, peer composition may be influenced by parental choices of neighborhood and school, as well as school policies dictating attendance and classroom placement. Parents may advocate to place their children in certain classes or schools because they perceive such groupings as advantageous, and schools may "track" students based on ability (Moffit, 2001; Manski, 1993). The end result is that students with similar skills are grouped together for a variety of reasons. Models that fail to separate out these other factors may produce upward biased estimates of peer effects on child outcomes.

Hanushek et al. (2003) argued that unmeasured prior experiences are particularly problematic when trying to identify peer effects, because peers are likely to have had similar school and neighborhood experiences, and spurious correlations between peer skills and abilities and an individual child's skills and abilities are created by their similar (unmeasured) past experiences. Although this may be less of a problem for preschoolers, who have encountered less formal schooling, the influences of prior family, neighborhood, and childcare experiences must be accounted for in the models.

The use of panel data, like those available in the 2003 FACES survey, which include multiple observations of an outcome over time, provides one opportunity to address the identification problems described above. Such data enable researchers to estimate a value-added (lagged dependent variable) specification. Adjusting for initial level of skill is a particularly powerful analytic approach, because it reduces the probability of omitted variable bias. In these models, the initial level of behavior essentially serves as a proxy for the cumulative effects of

current characteristics and prior experiences that have constant (time invariant) effects on children's behavior. Unfortunately, unmeasured characteristics that vary systematically over time or exert time varying effects on children's development may still result in omitted variable bias (NICHD-ECCRN & Duncan, 2003).

A related methodological approach is to use a gain score (measured as the difference between outcomes at two time points) as the dependent variable. In this approach, the correlation between the outcome at time two and time one is assumed to be 1.0, and any deviation from this attributed to measurement error. This differs from the lagged dependent variable approach, which assumes that there may be other reasons for deviation from this 1.0 correlation. This distinction of what "other" factors may be in play is important. For example, a potential problem with the change score approach is that gain scores are often negatively correlated with initial scores, thus introducing measurement error (lower reliability) in the gain score and increasing imprecision in the estimates. This is a particular concern in analyses of young children's behavior, given the likelihood that initial levels and later gains will be highly correlated (NICHD-ECCRN & Duncan, 2003; Lee et al., 1998). Measurement error in the dependent variables may also bias the regression coefficients. This may be the case if the measurement error in the change model is correlated with the true levels of the dependent and independent variables at fall or spring (Bound, Brown, Duncan, & Rodgers, 1994). According to NICHD-ECCRN & Duncan (2003, p. 1459), likely causes of such bias may include, "failure to include relevant interactions, differences in the impact of the omitted variable at the two time points, or use of assessment tools in which error is related to ability," any of which may be theoretically possible in these models, particularly given that the outcomes are teacher reports. Nevertheless, such models are considered as a robustness check.

Multilevel Modeling. In order to study classroom peer effects, it is necessary to have data in which multiple children are observed within the same classroom. This creates a problem, however, in that students' individual outcomes within such contexts may not be statistically independent (i.e., the error terms may be correlated). This type of data structure is suited to multilevel modeling, a generalization of linear regression that accounts for clustering by including a unique random effect for each organizational unit (in this case, classrooms) (Hedges, 2007; Raudenbush & Bryk, 2002). There may also be substantive reasons to use multilevel modeling if one is interested in explaining the variation in outcomes between classrooms, as well as cross-level interactions between individual child and contextual characteristics.

One way to test whether multilevel modeling is an appropriate strategy is to partition the variance to determine whether there is significant variation in each outcome measure at the classroom level. By calculating the intraclass correlation for each measure (i.e., the proportion of variance in test scores that exists between classrooms), it is possible to test the hypothesis that significant variance in child outcomes exists between classrooms. Analysis of the 2003 FACES data yields quite high, statistically significant intraclass correlations for each of the primary behavior outcomes of interest at the end of the Head Start year. Approximately 30 percent of the variance in behavior problem scores exists between classrooms; 35 percent for social skills scores, and 30 percent for learning behavior scores. These results suggest that multilevel modeling is an appropriate analytical strategy, and point to the potential importance of classroom factors in predicting developmental outcomes.

Analytic Models

In the analytic models, the behavior outcome at time 2 (spring) for child i in classroom j is regressed on the average peer skill measures, as well as each child's initial skill level and an increasingly rich set of child and family characteristics:

$$\text{Spring Outcome}_{ij2} = \beta_{0j} + \beta_{1j}(\text{Peer Skills})_{ij1} + \beta_{2j}(\text{Child's Initial Skills})_{ij1} + \beta_{3j}(\text{Child Covariates})_{ij1} + \beta_{4j}(\text{Family Covariates})_{ij1} + r_{ij}$$

In this child-level model, the coefficients of interest (β_{1j}) represent the magnitude of the association between peers' skills at time 1 (Head Start entry) and students' outcomes at time 2 (spring of the Head Start year), holding constant individual skill levels and an extensive set of child and family covariates, all measured at time 1. The individual error term, r_{ij}, represents the remaining unexplained child-level residual variance.

In the classroom-level model, each of the coefficients from the child-level model becomes an outcome variable:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Classroom Covariates})_{j1} + u_{0j}$$

$$\beta_{1j(\text{peer skills})} = \gamma_{10}$$

$$\beta_{2j\ (\text{ind. skills})} = \gamma_{20}$$

$$\beta_{3j(\text{child})} = \gamma_{30}$$

$$\beta_{4j(\text{family})} = \gamma_{40}$$

Here, the average student outcome for the class (β_{0j}) is predicted by a class-level intercept (γ_{00}), a vector of classroom characteristics (e.g., classroom and teacher quality) measured at time 1, and a random error term (u_{0j}). Coefficients from the first level (β_{2j}, β_{3j}, and β_{4j}) are modeled as "fixed," so that the coefficient is assumed to have the same value in all classrooms.

I begin by estimating a basic bivariate model, regressing the relevant peer skill measure on each child's spring behavior outcome, and then controlling for each child's relevant fall score.

Although the child's fall score is a powerful covariate, I remain concerned about omitted variable bias.

Therefore, for each spring outcome (learning behavior, social skills, and problem behaviors), I estimate a series of seven models with an increasingly rich set of controls. Adding any covariates beyond gender to the models signifies an improvement over much previous research in the behavior domain; however, the richness of the FACES data also allows me to include almost all of the covariates used in previous pre-academic outcome studies, which may also have implications for children's behavior. To place this study in the context of existing research, Model 1 approximates Mashburn et al.'s (2009) study of preschool peer effects, and includes basic child and family characteristics (child age, gender, race, and maternal education), as well as basic classroom structural characteristics (class size, child/staff ratio, teacher education), which are important potential confounders for which to control (NICHD-ECCRN & Duncan, 2003).

I next exploit the richness of the FACES data and include an extended set of child characteristics, including English language learner and disability status, and whether the child attended center based child care prior to attending Head Start; an extended set of family characteristics, including income, poverty level, family structure, number of children, home literacy activity, parental warmth and depression; and a set of controls for month of pre-test and number of days between pre- and post-test (Model 2). These models test whether the "peer effects" found in previous studies may have been spurious due to failure to control for these other factors that may be associated with behavior outcomes.

Next, I add in an extended set of classroom measures, including measures of process quality (ECERS and Arnett, Model 3). Research suggesting links between children's

(particularly negative) behavior and teacher's responses, in terms of warmth of interactions or time spent on instruction, (Morris, Raver, Millenky, Jones, & Lloyd, 2010; Raver et al., 2008), recommends isolating these measures, considering that their inclusion could amount to over-controlling.

Next, I test whether early child development may draw from a broader base of behavioral skills by including children's other two fall test scores as controls (Model 4), as well as measures of peer skills in those two behavior domains (Model 5). Finally, I consider whether pre-academic skills play an explanatory role in children's behavioral outcomes, by controlling for children's individual fall vocabulary, reading, and math scores (Model 6) and an average peer skill measure for each subject (Model 7).

Results

Results from the first set of bivariate regressions are displayed in Table 3.2. As expected, these regressions show positive associations between domain-specific peer skills in each of the three behavior areas (learning behaviors, social skills, and problem behaviors) and children's individual spring behavior outcomes. Adding in the child's individual fall score attenuates the association for each behavior domain, to the point of non-significance for problem behaviors. These findings confirm the casual observer's intuition that peer skills and children's individual outcomes are positively correlated, but also point to the importance of controlling for other factors that may affect behavior. These findings also suggest that the relationship between peer skills and individual behavior outcomes may be different for positive and negative behaviors.

Next, I turn to the set of value-added, multilevel models with increasingly rich controls.[10] Table 3.3 summarizes the results for children's Preschool Learning Behaviors Scale (PLBS). With Model 1, including controls for basic child and family demographics and basic classroom characteristics, as expected, I find a positive, statistically significant association between average peer preschool learning behavior skills and children's individual preschool learning behavior outcomes (effect size = .06).[11] The association remains robust, and the effect size increases (to .07) after adding in an extended set of controls for family and child characteristics (Model 2) and an extended set of classroom characteristics (Model 3).

I next test whether children's other behavior skills are associated with spring preschool learning behavior scores, by adding in children's fall social skills and problem behavior scores (Model 4). This regression suggests that there is indeed a statistically significant relationship between both types of behavior skills and children's spring PLBS scores: individual social skills measured in the fall are positively associated (effect size = .14), and individual behavior problems are negatively associated (effect size = .20). The peer PLBS measure also remains statistically significant and increases in size (effect size = .09). Adding in peer measures of social skills and problem behaviors increases the size of the PLBS coefficient (.15); however, there is not consistent evidence of a relationship between either of these peer behavior measures and spring PLBS outcomes.

The final models test for links between spring PLBS outcomes and individual (Model 6) and peer (Model 7) pre-academic skills. These regressions suggest that individual fall reading and math skills are predictive of PLBS scores; however, vocabulary skills are not. Peer reading

[10] Complete results, for all covariates, in the most complex models for each behavior outcome, can be found in Appendix 3.C.

[11] Effect sizes can be interpreted as the percent increase in standard deviation of the outcome related to a one standard deviation increase in the peer skill score.

skills are negatively associated with individual spring PLBS scores (effect size = .06), but there is no statistically significant relationship between PLBS scores and measures of peer math and vocabulary skills. The statistically significant relationship between peer and individual PLBS scores remains robust throughout these models (effect size = .16).

Results regarding the relationship between measures of average peer social skills and children's individual social skills outcomes, shown in Table 3.4, follow a pattern similar to those of the PLBS outcomes. Models 1 through 3, utilizing an increasingly rich set of child, family, and classroom covariates, demonstrate a significant positive relationship between the fall measure of average peer social skills and individual spring Social Skills Rating System scores (effect size range = .06 to .08). Model 4 also shows a significant negative relationship between individual fall problem behavior scores and spring social skills scores; however, there is no statistically significant relationship between individual learning behavior skills and individual social skills. There is also no relationship between measures of peer learning behaviors and behavior problems and individual social skills outcomes (Model 5), although the relationship between peer and individual social skills remains robust (effect size = .13).

The final models once again show a positive relationship between individual reading and math skills and growth in individual social skills; however, there is no such relationship between social skills and vocabulary skills (Model 6). The only pre-academic peer measure that is predictive of individual social skills is peer math, and only marginally so (effect size = .06). As with the learning behavior outcomes, the relationship between peer social skills and individual social skills remains robust even in the most complex model (effect size = .12).

Results for problem behavior outcomes are shown in Table 3.5. Although coefficients for the measure of average peer behavior problems are in the expected direction, none is

significant, regardless of the specification. Individual social skills, reading, and math skills are predictive of spring behavior problem scores (higher fall scores in these areas are associated with lower spring behavior problem scores); however, none of the other peer skill measure coefficients is statistically significant.

Robustness Checks and Sub-group Analyses

To test the robustness of these findings, I conduct a series of alternative specifications and theoretically based sub-group analyses. As previously acknowledged, gain score analysis is an alternative methodological approach made viable with the FACES longitudinal data, although concerns remain regarding measurement error (NICHD-ECCRN & Duncan, 2003; Lee et al., 1998). Unlike the lagged dependent variable models, models using an otherwise complete set of covariates (i.e., mimicking Model 7 in Tables 3.3 through 3.5), but using the change between spring and fall test scores as the outcome, show few significant relationships between average peer behavior skills measured in the fall and actual changes in behavior scores between fall and spring (See Table 3.6).

Although theoretically one might expect change score analyses to tell a story similar to lagged dependent variable analyses, the differences in results, particularly for these teacher reported behavior measures, are perhaps not surprising, given the limitations of change score analysis. Educators and psychologists often warn against using simple change scores versus level scores due to their greater measurement error, and therefore, decreased reliability (NICHD-ECCRN & Duncan, 2003). Change scores are particularly susceptible to measurement error when the original two scores are highly correlated, which is the case with each of the fall and spring FACES behavior measures. As demonstrated here, using the less reliable change score

outcomes results in less precise peer skill estimates (i.e., with larger standard errors). This measurement error may also bias the regression coefficients. In this case, there is concern, for example, that a teacher may rate everyone particularly high or particularly low, or that his/her rating may be impacted by a particular student's extremely "good" or "bad" behavior (i.e., everyone else may look relatively good or bad compared to this student, despite the "true" level of their behavior).

As a second robustness check, I substitute Spanish version scores for missing pre-academic test scores, to test whether the peer effect findings for behavior hold when Spanish-speaking Head Start students are included in the sample. Note that behavior scores for these children do not change, given that they are based on teacher reports, but sample size and composition do. These models include all of the covariates included in Model 7 of Tables 3.3 through 3.5, plus dummy variables indicating whether students took tests in Spanish in the spring and/or fall. Peer skill measures are also recalculated to include the Spanish test scores. The results, reported in Table 3.7, show robust, larger associations between peer learning behavior and social skills and their respective individual scores in these domains (effect sizes =.21 and .15, respectively). The negative association between peer math and individual learning behaviors is no longer significant, although the positive relationship between peer math and individual social skills is strengthened (effect size = .08). The fact that these results are qualitatively similar to those yielded by models without the Spanish-speaking students allays fears that excluding these students from the sample changes results.

As an additional robustness check, I also consider alternative specifications of the peer skills measures. The primary analytic models, using average peer skills (also known as "linear in means" models), assume that introducing a single student who increases a classroom's average

achievement by *x* points will have the same effect as introducing several students that collectively increase the average by the same amount. This may not be the correct theoretical mechanism, if, for example, having even one child with particularly high or low skills changes the dynamic of instruction, or having a "critical mass" of children with particularly high or low skills is necessary to truly impact classmates' learning.

To test these assumptions, I construct a "peer maximum" and "peer minimum" score within each classroom, for each behavior and pre-academic skill, as measured in the fall.[12] I then substitute these measures for the average peer skill measures within the primary models. Table 3.8, Panels 1 and 2, show that the main findings regarding a relationship between peer social skills and preschool learning behaviors, and matching individual outcomes, are robust to these alternative specifications (i.e., the higher the minimum or maximum score on these measures within a classroom, the higher the individual outcome). As with the primary specifications, no such relationship is found for problem behaviors. There is also no evidence of "cross-domain" peer effects for any of the additional behavior or pre-academic skill measures.

Next, I estimate models that measure whether having a higher percentage of peers scoring in the highest or lowest quartile of each skill area is associated with improved individual behavior outcomes. These models recognize that the "smoothing" effect of an average peer score may not truly represent how children's behaviors or pre-academic skills impact classroom processes. Three different reference groups are considered: for each outcome, Model 1 uses middle- and low-scoring students as the reference group; Model 2 uses middle- and high-scoring students, and Model 3 uses middle-scoring students only. The results, reported in Table 3.9, are

[12] I also considered class median as a specification of each peer skill measure. This may be important, for example, if the average peer effect is skewed by a particularly high or low score, but the true mechanism of the peer effect is a result of the median (i.e., children will be less likely to have or be impacted by interactions with one particular child, or classroom teachers are more likely to teach to the median than the extremes). Each peer median score, however, is so similar to the peer average, that I do not expect any difference in the results.

qualitatively similar to the main specification findings. There are positive associations between the percentage of high-scoring peers for both learning behaviors and social skills (effect size = .11 and .07, respectively), but the negative associations between having a higher percentage of lower-scoring peers in these domains are smaller and only marginally or non-significant (effect size = .07 and .06, respectively). There is also some evidence of a negative relationship between having a high percentage of peers with low reading scores and individual social skills scores (effect size = .06). As in the main specifications, there are no significant associations between any of the peer skills and individual problem behavior.

Given research suggesting potential differences in the operation of behavior peer effects within particular sub-groups, I also test whether the associations between peer skills and individual pre-academic outcomes vary based on child characteristics. For each set of sub-groups, I use interactions between such characteristics and the peer skill measures in models with a full set of covariates, mimicking model 7 in Tables 3.3 through 3.5.

First, I test whether the associations between peer skills and children's outcomes vary by children's initial skill levels. I develop a series of dummy variables indicating whether children have "high" initial skill levels (meaning their fall test score lies within the 75^{th} percentile of each domain) or "low" initial skill levels (meaning their fall test score lies within the 25^{th} percentile of each domain). I test three models for each outcome: In Model 1, I include interactions between initial scores and the relevant peer skill measure (where main effects were found in the primary specifications). The next models test for cross-domain influences: in Model 2, I include interactions between initial scores and the other behavior peer skill measures; and, in Model 3, I add interactions between initial scores and the pre-academic skill measures. The reference group for each model includes children with test scores in the middle quartiles.

Results for these models, shown in Tables 3.10-3.12, show little evidence of such interactions; however, they do suggest that having low social skills makes children less likely to benefit from peer social skills (Panel 2). There is also evidence of a positive interaction between high individual vocabulary skills and peer vocabulary skills, which is associated with higher individual PLBS outcomes (Panel 1). The main findings from the primary specifications (positive associations between peer and individual PLBS and social skills, no association between peer and individual problem behaviors) remain robust throughout the interacted models.

Another potentially important interaction is between gender and peer behavior skills; although, as noted previously, the theoretical direction of such effects is ambiguous. Consistent with findings from previous research, in this sample, girls' positive behavior scores (both PLBS and social skills) are higher than boys' scores, and boys' problem behavior scores are higher than girls'. Contrary to previous studies, however, I find little evidence of gender by peer skill interactions (See Table 3.13). There are no main gender interactions, although the male by peer behavior problems interaction is negatively associated with individual PLBS outcomes. Also for this outcome, the male by peer vocabulary interaction is positive, but the male by peer reading interaction is negative. Again, the main effects for each behavior outcomes remain intact throughout the gender interacted models.

Discussion

Despite the acknowledgement that early behavioral skill development may be important for later school success and that experiences with preschool peers may help shape such skill development, only a handful of empirical studies have looked for associations between such experiences and individual behavior outcomes. A novel finding from this study is that positive

peer behavior skills (preschool learning behaviors and pro-social skills) are indeed predictive of children's spring skills in these areas, holding constant a rich series of covariates previously unavailable in most preschool data sets. Effect sizes for these measures are moderate (.16 and .12, respectively), although larger than those found for pre-academic outcomes. Somewhat contrary to prior research, the present study does not find a statistically significant association between peers' average behavior problem scores and children's individual behavior problem outcomes after one year in Head Start. These findings are also robust across a number of alternative specifications of the peer skill score.

Another unique contribution of this study is its investigation of potential "cross-domain" influences on behavioral skill development, in terms of different kinds of behaviors and pre-academic skill influence. Although the findings point to cross-behavioral skill influence at the individual level (e.g., higher individual social skills are predictive of higher individual learning behaviors and higher levels of problem behaviors are predictive of lower social skills and PLBS scores); there appear to be few cross-behavior relationships at the peer skill level. Findings are similar for pre-academic skills. Although individual pre-academic skills are often predictive of spring behavior outcomes, there is little consistent evidence of peer pre-academic skill influence on these outcomes.

Why do I find results for the positive behavior outcomes, but not problem behaviors? It may be that the level of problem behavior in the FACES sample is generally low, and that the average peer skill measure construction does not adequately capture the disruptive nature of a child with clinical level behavior problems. The alternative peer measure specifications attempt to get at different types of relationships (e.g., having at least one highly disruptive student, or a

higher percentage of students with high levels of behavior problems), but perhaps there are better measures or specifications to be tested.

Another factor noted earlier in this paper is that children's contact with disruptive peers is not necessarily evenly distributed throughout the preschool classroom (Hanish et al., 2005; Hanish et al., 2008). Children who exhibit more behavior problems may have fewer interactions with peers in general, or teachers may shield other children from such interactions. These mechanisms would tend to attenuate the impact of negative peer behavior on individual outcomes. Unfortunately, in the FACES data, there is no information regarding which peers children interact with; therefore, these hypotheses cannot be tested.

This study also presents some evidence of variation in the relationship between peer skills and children's behavior outcomes based on a child's initial level of social skills. Specifically, children with lower social skills are less likely to benefit from higher skilled peers in this behavior area. I find little evidence of interactions between gender and peer skills, which, although contrary to some preschool behavior studies (e.g., Hanish et al., 2005) is perhaps not surprising given the theoretical (and empirical) ambiguity of whether higher skills (displayed by girls) put children in a better position to gain from peer interactions, or with less room to improve. Regardless of the interactions included in the models, the main positive associations between peer PLBS and social skills and related individual outcomes remain robust, as does the lack of such a link for problem behaviors.

These findings are presented with the acknowledgement of several limitations. First, these are non-experimental analyses, and any resulting associations may not be causal in nature. The analytic approaches employed attempt to reduce omitted variable bias by controlling not only for important child, family, and classroom characteristics, but also for children's prior

behavior. Nevertheless, it is possible that bias due to unmeasured factors affecting behavior remains.

A second important limitation to this study is that the models assume a homogeneous treatment effect; in other words, that each student has an equal effect on each other student. This assumption has been refuted by some researchers (See Harris, 2010; Hoxby & Weingarth, 2005), who argue that evidence from studies with older children suggests that certain classmates (e.g., those of the same race or with higher initial levels of achievement) may be more influential than others. In the preschool realm, this assumption may be particularly problematic in relation to children with behavior problems.

Theoretically, though, this concern is at least partially mitigated by the structure of Head Start pedagogy, in which teachers may be more controlling of peer groupings, and there is an emphasis on inclusion and interaction with all fellow students (USDHHS, ACF, Office of Head Start, 2009). Furthermore, there is generally less stability in preschool friendships compared to the friendships of older children (Rubin et al., 2006). Developmental research does suggest, however, that preschool-aged children appear to be more attracted to peers who are similar; for example, in terms of age, sex, or behavioral tendencies (Rubin et al., 2006). Unfortunately, given the sample size in each classroom and lack of a true measure of peer interaction, such hypotheses are not testable using the FACES data.

Finally, the homogeneity of the FACES population is both a blessing and a curse. On the one hand, homogeneity in population and programming reduces concerns presented by selection bias, in that program quality, a potential omitted variable, is likely to be relatively more similar. On the other hand, it means that variation in outcomes and peer skill levels are decreased, particularly at the high end of the skill distribution; thus, making it more difficult to detect peer

effects. The fact that the FACES data include only first-year Head Start students, and therefore exclude measures of students participating in their second year of the program may also attenuate peer effect estimates, assuming Head Start is an effective intervention that increases skill levels.

In sum, the findings from this study contribute to the mounting evidence that preschool peers may indeed contribute to the development of individual children's behavior skills. On a particularly positive note, in Head Start, children appear to show growth in individual positive behavior skills associated with peers' positive behavior skills; however, they do not appear influenced by peers' problematic behaviors. If, as these findings suggest, lower-skilled children may learn more from opportunities to interact with higher-skilled children, then policy makers must consider peer composition in their calculus of potential tradeoffs between targeted or universal preschool programs, which are likely to differ in terms of student heterogeneity and average baseline skills. Furthermore, even if program composition is largely unalterable, knowing that peers contribute to preschoolers' learning recommends teacher training regarding the use of proven collaborative learning and behavior management techniques that maximize the benefits of peer effects within a given classroom (Snyder et al., 2011; Wilkinson, Parr, Fung, Hattie, & Townsend, 2002).

Table 3.1: Descriptive Statistics

	Min.	Max.	Mean	S.D.
Basic Child Characteristics				
Child Gender (1=Male, 0=Female)	0	1	0.48	0.50
Race/Ethnicity				
White, Non-Hispanic (Referent)	0	1	0.34	0.47
African American	0	1	0.36	0.48
Hispanic	0	1	0.22	0.41
Other race/ethnicity	0	1	0.08	0.28
Child Age at post-test (months)	41	69	54.1	6.66
Maternal Education				
Less than high school diploma	0	1	0.29	0.45
High school diploma (Referent)	0	1	0.39	0.49
Any post-secondary education	0	1	0.33	0.47
Extended Child and Family Characteristics				
English Language Learner	0	1	0.09	0.29
Disability	0	1	0.16	0.37
Center-based care before Head Start	0	1	0.2	0.4
Log of family income	4.32	11.61	9.43	0.86
Income below poverty line	0	1	0.68	0.47
Number of children	1	13	2.64	1.38
Married mother	0	1	0.38	0.49
Frequency read to child	0	7	4.71	2.32
Parental warmth	2.8	5	4.38	0.42
Parental depression	0	36	6.98	6.43
Basic Classroom Characteristics				
Child/Staff Ratio	2	15	6.63	2.12
Class size	4	32	14.42	3.32
Arnett Teacher Behavior Score	12	90	73.25	11.98
Extended Classroom Characteristics				
Teacher Education				
No post-secondary education (Referent)	0	1	0.05	0.23
Associate's degree or some college	0	1	0.56	0.50
Bachelor's or advanced degree	0	1	0.38	0.49
ECERS-R Score	2.06	6.60	4.66	0.95

Table 3.1: Descriptive Statistics, Continued

	Min.	Max.	Mean	S.D.
Timing of Testing				
October pre-test (Referent)	0	1	0.31	0.46
November pre-test	0	1	0.42	0.49
December pre-test	0	1	0.26	0.44
Days between pre- and post-test	111	249	170.02	21.03
Children's Fall Pre-Academic Test Scores				
PPVT (Vocabulary)	37.28	120.15	84.79	11.6
WJ-Letters & Dictation (Reading)	30.19	169	92.45	16.99
WJ-Applied Problems (Math)	26.69	143	88.64	17.70
Peer (Fall) Pre-Academic Skills Scores				
Peer-PPVT (Vocabulary)	54.56	102.53	84.79	6.82
Peer-WJ-Letters & Dictation (Reading)	60.87	123	92.45	8.36
Peer-WJ-Applied Problems (Math)	56.76	117.83	88.64	9.4
Children's Spring Pre-Academic Test Scores (Outcomes)				
PPVT (Vocabulary)	45.57	126.35	87.33	11.31
WJ-Letters & Dictation (Reading)	39.31	159.79	98.07	16.02
WJ-Applied Problems (Math)	38.02	156.04	91.74	16.11
Children's Fall Behavior Test Scores				
Preschool Learning Behavior Scale (t-score)	12.00	64.00	50.54	10.43
Social Skills	0.00	24.00	15.50	4.67
Problem Behaviors	0.00	24.00	4.73	4.64
Peer (Fall) Behavior Skills Scores				
Peer-Preschool Learning Behavior Scale (t-score)	32.33	63.88	50.54	6.35
Peer-Social Skills	6.80	23.00	15.50	3.05
Peer-Problem Behaviors	0.00	15.00	4.73	2.63
Children's Spring Behavior Scores (Outcomes)				
Preschool Learning Behavior Scale (t-score)	14.00	64.00	52.39	9.94
Social Skills	1.00	24.00	17.27	4.61
Problem Behaviors	0.00	26.00	4.39	4.54

Note: N=19,170 (across 10 imputed data sets); descriptives provided for children in classrooms with at least 4 fall academic test socres.

Table 3.2: Bivariate Regressions of Peer Skills on Children's Behavior Outcomes

	Preschool Learning		Social Skills		Problem Behaviors	
Peer Skills	0.29**	0.05*	0.34**	0.06*	0.22**	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Fall Score		0.61**		0.61**		0.67**
		(0.03)		(0.02)		(0.02)
N	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t}p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 3.3: Summary of Results from Regressions of Average Peer Skills on Children's Preschool Learning Behavior Skills in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Preschool Learning Behavior	0.06*	0.07*	0.07*	0.09*	0.15*	0.17**	0.16**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Fall Preschool Learning Behavior	0.57**	0.55**	0.55**	0.30**	0.29**	0.28**	0.27**
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Individual Behavior Measures							
Individual Social Skills				0.14**	0.14**	0.12**	0.12**
				(0.03)	(0.03)	(0.03)	(0.03)
Individual Behavior Problems				-0.20**	-0.21**	-0.18**	-0.18**
				(0.04)	(0.04)	(0.03)	(0.03)
Other Peer Behavior Measures							
Peer Social Skills					-0.01	-0.02	-0.01
					(0.03)	(0.03)	(0.03)
Peer Problem Behaviors					0.07[t]	0.07[t]	0.06
					(0.04)	(0.04)	(0.04)
Child Fall Pre-Academic Test Scores							
Fall Reading						0.08**	0.07**
						(0.02)	(0.02)
Fall Math						0.07*	0.08**
						(0.02)	(0.02)
Fall Vocabulary						0.00	0.00
						(0.03)	(0.03)
Peer Pre-Academic Skills Scores							
Peer Reading							-0.06*
							(0.03)
Peer Math							0.02
							(0.03)
Peer Vocabulary							0.01
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 3.4: Summary of Results from Regressions of Average Peer Skills on Children's Social Skills in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Social Skills	0.06*	0.08*	0.08*	0.10*	0.13**	0.13*	0.12*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Fall Social Skills	0.56**	0.54**	0.54**	0.41**	0.40**	0.38**	0.38**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Individual Behavior Measures							
Individual Preschool Learning Behaviors				0.02	0.02	0.01	0.01
				(0.03)	(0.03)	(0.03)	(0.03)
Individual Behavior Problems				-0.18**	-0.18**	-0.16**	-0.15**
				(0.03)	(0.03)	(0.03)	(0.03)
Other Peer Behavior Measures							
Peer Preschool Learning Behaviors					-0.01	0.01	0.01
					(0.05)	(0.05)	(0.05)
Peer Problem Behaviors					0.05	0.05	0.06
					(0.04)	(0.04)	(0.04)
Child Fall Pre-Academic Test Scores							
Fall Reading						0.07**	0.08**
						(0.02)	(0.02)
Fall Math						0.08**	0.09**
						(0.02)	(0.02)
Fall Vocabulary						0.01	0.00
						(0.02)	(0.02)
Peer Pre-Academic Skills Scores							
Peer Reading							0.01
							(0.03)
Peer Math							0.06[t]
							(0.04)
Peer Vocabulary							-0.04
							(0.04)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 3.5: Summary of Results from Regressions of Average Peer Skills on Children's Behavior Problems in the Spring of the Head Start Year

	Models						
	1	2	3	4	5	6	7
Peer Behavior Problems	0.02	0.03	0.02	0.03	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)
Fall Behavior Problems	0.64**	0.61**	0.61**	0.50**	0.50**	0.48**	0.48**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Basic Child & Family Characteristics	yes	yes	yes	yes	yes	yes	yes
Basic Classroom Characteristics	yes	yes	yes	yes	yes	yes	yes
Extended Child & Family Characteristics		yes	yes	yes	yes	yes	yes
Extended Classroom Characteristics			yes	yes	yes	yes	yes
Other Individual Behavior Measures							
Individual Preschool Learning Behaviors				-0.04	-0.04	-0.03	-0.03
				(0.03)	(0.04)	(0.03)	(0.03)
Individual Social Skills				-0.13**	-0.13**	-0.11**	-0.11**
				(0.03)	(0.03)	(0.03)	(0.03)
Other Peer Behavior Measures							
Peer Preschool Learning Behaviors						-0.01	-0.01
						(0.04)	(0.04)
Peer Social Skills						0.01	0.00
						(0.03)	(0.03)
Child Fall Pre-Academic Test Scores							
Fall Reading						-0.08**	-0.07*
						(0.02)	(0.02)
Fall Math						-0.05*	-0.05*
						(0.02)	(0.02)
Fall Vocabulary						-0.01	-0.01
						(0.02)	(0.02)
Peer Pre-Academic Skills Scores							
Peer Reading							0.04
							(0.03)
Peer Math							0.00
							(0.03)
Peer Vocabulary							0.00
							(0.03)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 3.6: Summary of Results from Regressions of Average Peer Skills on Children's Changes in Behavior Skills between Fall and Spring of the Head Start Year

	Chage Score Outcomes		
	Learning Behavior	Social Skills	Behavior Problems
Peer Learning Behavior	-0.02	0.06	-0.09*
	(0.05)	(0.05)	(0.04)
Peer Social Skills	0.04	-0.02	-0.03
	(0.04)	(0.04)	(0.04)
Peer Behavior Problems	-0.04	0.02	-0.06
	(0.04)	(0.05)	(0.04)
Peer Vocabulary	-0.00	-0.04	-0.02
	(0.04)	(0.04)	(0.03)
Peer Reading	-0.04	-0.02	0.04
	(0.03)	(0.03)	(0.03)
Peer Math	0.03	0.06	0.01
	(0.04)	(0.04)	(0.03)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, excluding the related fall score, which is instead captured in the outcome (spring score - fall score).

Table 3.7: Summary of Results from Regressions of Average Peer Skills on Children's Behavior Skills in the Spring of the Head Start Year, Including Spanish Test Scores

	Spring Outcomes		
	Learning Behavior	Social Skills	Problem Behavior
Peer Learning Behavior	0.21**	0.01	-0.02
	(0.05)	(0.05)	(0.04)
Peer Social Skills	-0.01	0.15*	0.02
	(0.04)	(0.04)	(0.04)
Peer Behavior Problems	0.05	0.07	0.04
	(0.04)	(0.05)	(0.04)
Peer Vocabulary	-0.03	-0.06	-0.01
	(0.03)	(0.04)	(0.03)
Peer Reading	-0.04	0.02	0.04
	(0.03)	(0.03)	(0.03)
Peer Math	0.04	0.08[t]	-0.01
	(0.04)	(0.04)	(0.03)
N	2,190	2,190	2,190

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, plus dummies indicating whether students took tests in Spanish in the fall and/or spring. These models include reconstructed peer scores, including Spanish pre-academic test scores for Spanish speaking students.

Table 3.8, Panel 1: Summary of Results from Regressions of Maximum Peer Skills on Children's Behavior Skills in Spring of the Head Start Year

	Spring Outcomes		
	Learning Behavior	Social Skills	Behavior Problems
Max Peer Learning Behavior	0.09*	-0.00	-0.03
	(0.04)	(0.04)	(0.03)
Max Peer Social Skills	-0.01	0.10*	-0.00
	(0.03)	(0.04)	(0.03)
Max Peer Behavior Problems	-0.02	-0.01	0.02
	(0.03)	(0.03)	(0.03)
Max Peer Vocabulary	-0.03	-0.04	0.02
	(0.03)	(0.04)	(0.03)
Max Peer Reading	-0.01	-0.01	0.01
	(0.03)	(0.03)	(0.03)
Max Peer Math	0.01	0.04	0.01
	(0.04)	(0.04)	(0.03)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5.

Table 3.8, Panel 2: Summary of Results from Regressions of Minimum Peer Skills on Children's Behavior Skills in Spring of the Head Start Year

	Spring Outcomes		
	Learning Behavior	Social Skills	Behavior Problems
Min Peer Learning Behavior	0.10*	-0.02	-0.03
	(0.04)	(0.04)	(0.03)
Min Peer Social Skills	-0.01	0.09*	0
	(0.04)	(0.04)	(0.04)
Min Peer Behavior Problems	-0.04	-0.00	0.05
	(0.04)	(0.04)	(0.03)
Min Peer Vocabulary	0.03	-0.00	-0.01
	(0.03)	(0.03)	(0.03)
Min Peer Reading	-0.01	-0.00	0.01
	(0.03)	(0.03)	(0.03)
Min Peer Math	-0.05	0.05	0.03
	(0.03)	(0.04)	(0.03)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5.

Table 3.9: Summary of Results from Regressions of Percent of Class with Scores in Highest and Lowest Quartiles on Children's Behavior Skills in Spring of the Head Start Year

	Learning Behaviors			Social Skills			Behavior Problems		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
% Hi Quart Learning Behaviors	0.11*		0.11*	-0.01		-0.00	-0.04		-0.03
	(0.04)		(0.04)	(0.04)		(0.04)	(0.03)		(0.03)
% Hi Quart Social Skills	-0.04		-0.03	0.09*		0.07^{t}	0.02		0.03
	(0.03)		(0.04)	(0.04)		(0.04)	(0.03)		(0.03)
% Hi Quart Behavior Problems	-0.01		0.03	-0.02		0.02	0.02		0.01
	(0.03)		(0.04)	(0.03)		(0.04)	(0.03)		(0.04)
% Hi Quart Vocabulary	0.02		0.01	0.00		-'0.01	-0.02		-0.01
	(0.03)		(0.04)	(0.04)		(0.04)	(0.03)		(0.03)
% Hi Quart Reading	-0.03		-0.02	0.03		0.00	0.02		0.02
	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)
% Hi Quart Math	0.00		0.01	-0.03		-0.03	0.04		0.04
	(0.03)		(0.03)	(0.04)		(0.04)	(0.03)		(0.03)
% Lo Quart Learning Behaviors		-0.07*	-0.07^{t}		-0.02	-0.02		0.01	-0.00
		(0.03)	(0.04)		(0.03)	(0.04)		(0.03)	(0.04)
% Lo Quart Social Skills		-0.02	-0.03		-0.07^{t}	-0.06		0.02	0.02
		(0.03)	(0.03)		(0.04)	(0.04)		(0.03)	(0.03)
% LoQuart Behavior Problems		0.01	-0.02		-0.00	-0.02		-0.01	-0.00
		(0.03)	(0.03)		(0.03)	(0.04)		(0.03)	(0.03)
% Lo Quart Vocabulary		-0.01	-0.00		-0.03	-0.02		0.01	0.01
		(0.03)	(0.04)		(0.03)	(0.04)		(0.03)	(0.03)
% Lo Quart Reading		0.05	0.03		-0.07*	-0.06^{t}		-0.00	0.01
		(0.03)	(0.03)		(0.03)	(0.04)		(0.03)	(0.03)
% LoQuart Math		0.00	-0.00		-0.05	-0.05		0.05	0.04
		(0.03)	(0.03)		(0.03)	(0.04)		(0.03)	(0.03)
N	1,917	1,917	1,917	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. All models include a full set of covariates (mimicking Model 7 in Tables 3.3-3.5).

Table 3.10: Summary of Results from Regressions of Average Peer Skills on Children's Learning Behavior Skills in Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Preschool Learning Behavior Scores		
	Model 1	Model 2	Model 3
Hi Ind. Learning Beh.*Peer Learning Beh.	0.01	0.02	0.03
	(0.05)	(0.05)	(0.05)
Lo Ind. Learning Beh*Peer Learning Beh.	-0.00	-0.01	-0.00
	(0.05)	(0.06)	(0.06)
Peer Preschool Learning Behavior	0.16*	0.16*	0.16*
	(0.05)	(0.05)	(0.05)
Hi Ind. Social Skills*Peer Social Skills		0.03	0.03
		(0.06)	(0.06)
Lo Ind. Social Skills*Peer Social Skills		0.01	0.00
		(0.06)	(0.07)
Peer Social Skills	-0.02	-0.03	-0.03
	(0.03)	(0.04)	(0.04)
Hi Ind. Beh. Problems*Peer Beh. Problems		-0.03	-0.03
		(0.05)	(0.05)
Lo Ind. Beh. Problems*Peer Beh. Problems		0.04	0.04
		(0.05)	(0.05)
Peer Behavior Problems	0.06	0.06	0.06
	(0.04)	(0.05)	(0.05)
Hi Ind. Vocabulary*Peer Vocabulary			0.11*
			(0.05)
Lo Ind. Vocabulary*Peer Vocabulary			0.07
			(0.05)
Peer Vocabulary	0.01	0.01	-0.01
	(0.03)	(0.03)	(0.04)
Hi Ind. Reading*Peer Reading			0.06
			(0.05)
Lo Ind. Reading*Peer Reading			0.07
			(0.05)
Peer Reading	-0.06*	-0.06*	-0.09*
	(0.03)	(0.03)	(0.03)
Hi Ind. Math*Peer Math			-0.02
			(0.05)
Lo Ind. Math*Peer Math			0.02
			(0.05)
Peer Math	0.02	0.01	0.01
	(0.03)	(0.03)	(0.04)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, plus the interactions indicated in the first column.

Table 3.11: Summary of Results from Regressions of Ave. Peer Skills on Children's Social Skills in Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Preschool Social Skills Scores		
	Model 1	Model 2	Model 3
Hi Ind. Learning Beh.*Peer Learning Beh.		-0.04	-0.05
		(0.05)	(0.05)
Lo Ind. Learning Beh*Peer Learning Beh.		-0.04	-0.04
		(0.05)	(0.06)
Peer Preschool Learning Behavior	0.00	0.03	0.03
	(0.02)	(0.05)	(0.05)
Hi Ind. Social Skills*Peer Social Skills	0.03	0.05	0.05
	(0.05)	(0.05)	(0.05)
Lo Ind. Social Skills*Peer Social Skills	-0.11[t]	-0.12[t]	-0.12*
	(0.06)	(0.06)	(0.06)
Peer Social Skills	0.13*	0.13*	0.13*
	(0.04)	(0.04)	(0.04)
Hi Ind. Beh. Problems*Peer Beh. Problems		-0.05	-0.04
		(0.05)	(0.05)
Lo Ind. Beh. Problems*Peer Beh. Problems		0.04	0.03
		(0.05)	(0.06)
Peer Behavior Problems		0.06	0.06
		(0.05)	(0.05)
Hi Ind. Vocabulary*Peer Vocabulary			0.06
			(0.05)
Lo Ind. Vocabulary*Peer Vocabulary			0.03
			(0.05)
Peer Vocabulary	-0.04	-0.04	-0.07
	(0.04)	(0.04)	(0.04)
Hi Ind. Reading*Peer Reading			-0.01
			(0.05)
Lo Ind. Reading*Peer Reading			0.06
			(0.05)
Peer Reading	0.01	0.01	-0.00
	(0.03)	(0.03)	(0.04)
Hi Ind. Math*Peer Math			-0.02
			(0.05)
Lo Ind. Math*Peer Math			0.01
			(0.05)
Peer Math	0.06	0.06[t]	0.07
	(0.04)	(0.04)	(0.04)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, plus the interactions indicated in the first column.

Table 3.12: Summary of Results from Regressions of Ave. Peer Skills on Children's Problem Behaviors in Spring of the Head Start Year, Including Initial Skill Level Interactions

	Spring Problem Behavior Scores		
	Model 1	Model 2	Model 3
Hi Ind. Learning Beh.*Peer Learning Beh.		0.02	0.01
		(0.05)	(0.05)
Lo Ind. Learning Beh*Peer Learning Beh.		0.09	0.09
		(0.06)	(0.06)
Peer Preschool Learning Behavior	-0.01	-0.04	0.00
	(0.04)	(0.05)	(0.02)
Hi Ind. Social Skills*Peer Social Skills		-0.00	-0.01
		(0.05)	(0.05)
Lo Ind. Social Skills*Peer Social Skills		-0.03	-0.03
		(0.06)	(0.06)
Peer Social Skills	0.00	0.01	0.02
	(0.03)	(0.04)	(0.04)
Hi Ind. Beh. Problems*Peer Beh. Problems	-0.07	-0.04	-0.04
	(0.05)	(0.05)	(0.05)
Lo Ind. Beh. Problems*Peer Beh. Problems	-0.04	-0.04	-0.04
	(0.05)	(0.05)	(0.05)
Peer Behavior Problems	0.07	0.05	0.05
	(0.04)	(0.04)	(0.04)
Hi Ind. Vocabulary*Peer Vocabulary			-0.06
			(0.05)
Lo Ind. Vocabulary*Peer Vocabulary			0.00
			(0.05)
Peer Vocabulary	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.04)
Hi Ind. Reading*Peer Reading			-0.05
			(0.05)
Lo Ind. Reading*Peer Reading			-0.06
			(0.05)
Peer Reading	0.04	0.04	0.06*
	(0.03)	(0.03)	(0.03)
Hi Ind. Math*Peer Math			0.07
			(0.05)
Lo Ind. Math*Peer Math			-'0.07
			(0.05)
Peer Math	0.00	0.00	0.01
	(0.03)	(0.03)	(0.04)
N	1,917	1,917	1,917

NOTE: ** $p<0.01$, * $p<0.05$; ^{t} $p<0.10$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, plus the interactions indicated in the first column.

Table 3.13: Summary of Results from Regressions of Average Peer Skills on Children's Behavior Skills in the Spring of the Head Start Year, Including Gender Interactions

	Spring Behavior Scores					
	Learning Behaviors		Social Skills		Problem Behaviors	
	Model 1	Model 3	Model 1	Model 3	Model 1	Model 3
Male*Peer Learning Beh.	0.05	-0.07		0.03		0.03
	(0.04)	(0.06)		(0.06)		(0.06)
Peer Learning Behavior	0.14*	0.19**	0.01	-0.01	-0.01	-0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
Male* Peer Social Skills		0.03	0.03	-0.00		-'0.00
		(0.05)	(0.03)	(0.05)		(0.04)
Peer Social Skills	-0.01	-0.03	0.10*	0.12*	0.00	0.00
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Male*Peer Beh. Problems		-0.15*		-0.01	0.05	0.08
		(0.06)		(0.05)	(0.03)	(0.05)
Peer Behavior Problems	0.06	0.12*	0.06	0.06	0.01	0.01
	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Male*Peer Vocabulary		0.09*		0.03		-0.04
		(0.05)		(0.04)		(0.04)
Peer Vocabulary	0.01	-0.03	-0.04	-0.06	0.00	0.02
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
Male*Peer Reading		-0.08*		-0.04		0.03
		(0.04)		(0.04)		(0.04)
Peer Reading	-0.06*	-0.02	0.01	0.03	0.04	0.02
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Male*Peer Math		-0.01		0.04		-0.04
		(0.05)		(0.05)		(0.05)
Peer Math	0.02	0.02	0.06[t]	0.04	-0.00	0.02
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
N	1,917	1,917	1,917	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Covariates include all of those included in Model 7, Tables 3.3-3.5, plus the interactions indicated in the first column.

Appendix 3.A : Pairwise Correlations of Average Peer Skills in Fall of Head Start Year

Variables	1	2	3	4	5	6
1. Peer-PPVT (Vocabulary)	--					
2. Peer-WJ-Letters & Dictation (Reading)	0.28*	--				
3. Peer-WJ-Applied Problems (Math)	0.58*	0.40*	--			
4. Peer-Preschool Learning Behavior Scale t-score	0.12*	0.18*	0.24*	--		
5. Peer-Social Skills	0.16*	0.25*	0.30*	0.61*	--	
6. Peer-Behavior Problems	-0.03*	-0.17*	-0.22*	-0.76*	-0.56*	--

NOTE: N=19,170 (across 10 imputed data sets); * $p<0.05$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Appendix 3.B: Child, Family, and Classroom Covariate Description and Coding

	Concept	Coding
Individual Child Characteristics	Initial skill levels	Continuous variables; child's individual standardized test scores (Academic skills=PPVT, reading, math; Behavior skills=learning related behaviors, pro-social skills, and behavior problems)
	Age at assessment	Continuous variable; months
	Gender	Dummy variable; 1=male
	Race/Ethnicity	Series of dummy variables for black, Hispanic, other race (omitted=white)
	Disability	Dummy variable; 1=has disability
	Language learning status	Dummy variable; 1=English language learner
	Prior childcare experience	Dummy variable; 1=previously attended center-based care
Individual Family Characteristics	Number of children	Continuous variable
	Family structure	Dummy variable; 1=mother married
	Maternal education	Series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent)
	Income	Continuous variable; natural log annual family income
	Literacy behavior	Continuous variable; how often child has been read to in the last week (scale of 0, "not at all" to 7, "every day")
	Maternal depression	Continuous variable derived from12 items from CES-D Depression Scale
	Parent warmth and sensitivity	Continuous variable derived from 5 items asking about parenting behavior
Classroom Characteristics	Child/staff ratio	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Class size	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Teacher Education	Series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent)
	Classroom process (global) quality	Continuous variable; ECERS-R mean score
	Teacher behavior	Continuous variable; Arnett Scale of Caregiver Behavior (lead teacher)

Appendix 3.C: Complete Results from Regressions of Average Peer Skills on Children's Behavior Skills in the Spring of the Head Start Year (Model 7, Tables 3.3-3.5)

	Preschool Learning Behaviors	Social Skills	Problem Behaviors
Peer Preschool Learning Behavior	0.16**	0.01	-0.01
	(0.04)	(0.05)	(0.04)
Fall Preschool Learning Behavior	0.27**	0.01	-0.03
	(0.04)	(0.03)	(0.03)
Peer Social Skills	-0.01	0.12*	0.00
	(0.03)	(0.04)	(0.03)
Fall Social Skills	0.12**	0.38**	-0.11**
	(0.03)	(0.03)	(0.03)
Peer Problem Behavior	0.06	0.06	0.03
	(0.04)	(0.04)	(0.04)
Fall Problem Behavior	-0.18**	-0.15**	0.48**
	(0.03)	(0.03)	(0.03)
Peer Vocabulary	0.01	-0.04	0.00
	(0.03)	(0.04)	(0.03)
Fall Vocabulary	0.00	0.00	-0.01
	(0.03)	(0.02)	(0.02)
Peer Reading	-0.06*	0.01	-0.04
	(0.03)	(0.03)	(0.03)
Fall Reading	0.07*	0.06*	-0.07*
	(0.02)	(0.02)	(0.02)
Peer Math	0.02	0.06^{t}	0.00
	(0.03)	(0.04)	(0.03)
Fall Math	0.08*	0.09**	-0.05*
	(0.02)	(0.02)	(0.02)
Age	0.01^{t}	0.01**	-0.00
	(0.00)	(0.00)	(0.00)
Male	-0.12*	-0.14**	0.14**
	(0.04)	(0.03)	(0.04)
Black	0.13*	-0.01	-0.08
	(0.06)	(0.06)	(0.06)
Hispanic	0.04	0.06	-0.10^{t}
	(0.06)	(0.06)	(0.06)
Other Race	0.16*	0.04	-0.05
	(0.08)	(0.08)	(0.07)
Mom Ed. <High School	0.03	0.04	0.01
	(0.05)	(0.04)	(0.04)
Mom Ed. Any Postsecondary	0.04	-0.03	-0.02
	(0.04)	(0.04)	(0.04)

Appendix 3.C: Complete Results from Regressions of Average Peer Skills on Children's Behavior Skills in the Spring of the Head Start Year (Model 7, Tables 3.3-3.5), Continued

	Learning Behaviors	Social Skills	Problem Behaviors
Child/Staff Ratio	0.03[t]	0.01	-0.03[t]
	(0.01)	(0.02)	(0.01)
Class Size	-0.01	-0.00	0.02*
	(0.01)	(0.01)	(0.01)
Teacher Sensitivity (Arnett)	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)
English Language Learner	0.07	-0.05	-0.03
	(0.08)	(0.08)	(0.08)
Disability	-0.05	-0.03	0.20**
	(0.03)	(0.05)	(0.06)
Center Care before Head Start	-0.04	-0.03	0.17**
	(0.05)	(0.05)	(0.05)
Family Income	0.01	0.01	-0.03
	(0.03)	(0.03)	(0.03)
Poverty	0.05	0.02	-0.01
	(0.05)	(0.05)	(0.05)
Number of Children	0.03*	0.01	-0.02[t]
	(0.01)	(0.01)	(0.01)
Married	0.08*	0.04	-0.02
	(0.04)	(0.04)	(0.04)
Parental Reading Scale	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)
Parental Warmth	-0.00	0.01	0.05
	(0.04)	(0.04)	(0.04)
Parental Depression	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)
November Pre-test	0.00	-0.06	-0.04
	(0.06)	(0.06)	(0.06)
December Pre-test	-0.06	-0.12	0
	(0.09)	(0.09)	(0.08)
Days between Pre- and Post-test	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)
Classroom Quality (ECERS-R)	0.02	0.02	0.01
	(0.04)	(0.05)	(0.04)
Teacher has Bachelor's Degree	0.09	0.09	0.01
	(0.14)	(0.14)	(0.13)
Teacher has Associate's Degree	0.07	0.07	-0.04
	(0.13)	(0.14)	(0.12)
N	1,917	1,917	1,917

NOTE: ** p<0.01, * p<0.05; [t] p<0.10; standard errors in parentheses. Analyses limited to classrooms w/ at least 4 fall test scores.

CHAPTER IV

THE INFLUENCE OF PRESCHOOL CHILDREN ON CLASSROOM PROCESS QUALITY OUTCOMES: EVIDENCE FROM THE 2003 HEAD START FAMILY AND CHILD EXPERIENCES SURVEY (FACES)

Background

Discussion of classroom quality in early education generally focuses on two different concepts. Structural quality, which refers to structural aspects of programs, such as the child to teacher ratio or level of teacher education, is expected to indirectly affect children by shaping their day-to-day experiences. Process quality, which commonly refers the nature of children's day-to-day experiences and interactions, is expected to more directly affect children. Although teachers are often seen as drivers of process quality, the potential contributions of *children* are rarely considered.

Yet, certainly, one important classroom process that could be affected by class composition is teaching, particularly if teachers determine or change their practices based on which children are present in the classroom. For example, a group of children with more advanced academic skills may encourage teachers to increase the level of curricular content, or the pace at which it is introduced. Teachers' pedagogical beliefs, in terms of *how* instruction should be delivered and *which* skills are most important to develop, may also drive practice, and may be related to their assessment of children's needs based on characteristics such as socioeconomic status (SES) and initial skill levels (Charlesworth et al., 1993; Stipek & Byler, 1997).

Children's behavior may also influence classroom process quality. At the preschool level, given alarming rates of expulsion uncovered in studies of state prekindergarten programs (e.g., see Gilliam, 2005; Gilliam & Shahar, 2006), growing attention has been paid to the ways in

which children's behavior may influence learning opportunities and the emotional climate of the classroom. For example, whether children engage in cooperative or disruptive behaviors may affect scheduled routines and the delivery of structured learning activities, the warmth of teachers' interactions with children, and the intentionality and type of instructional feedback that teachers provide (Raver et al., 2008).

Understanding such influences on early childhood teaching and classroom environment is important, as studies have consistently found links between measures of process quality in preschool programs and children's achievement and behavior. For example, Duncan and colleagues found that a one standard deviation increase in program process quality was associated with .06 standard deviation increase in cognitive development (NICHD-ECCRN & Duncan, 2003). A recent meta-analysis by Burchinal and colleagues (in press) found that overall quality was associated with a .11 effect size, with stronger associations for achievement outcomes (.12) than for behavioral outcomes (.08). The authors' analysis of five preschool datasets confirmed such modest effects sizes, although they noted in several instances slightly larger and more consistent effects when the program measure and outcome measure were more closely aligned (e.g., quality of language environment predicting language outcomes).

Identifying which classroom processes are dependent on children's pre-academic and behavior skill levels is complicated, however. Research with school-age children has identified classrooms, and, in particular, unobserved differences in teachers, as important confounds with peer effects (Vigdor & Necheyba, 2005). Consideration of these processes in the preschool context, however, has remained virtually unexplored. This chapter attempts to begin to fill this empirical gap, using the rich Family and Child Experiences (FACES) data to investigate

associations between classroom-level pre-academic and behavioral skills and classroom-level process quality outcomes.

Literature Review

Few studies have used actual measures of children's skills to examine their link to teacher practices and process quality. In one of the only such studies, Barr and Dreeben (1983), investigating first grade reading instruction, found that having higher percentages of low-ability readers in the classroom made organizing effective reading instruction more difficult.

More common are studies using children's socioeconomic status (SES) as a proxy for skills. For example, Pianta et al. (2002) found that kindergarten classrooms with high proportions of low-income children were less child-centered, and a recent study from the National Center for Education Statistics (NCES, 1999) found that teachers in schools with high levels of poverty used more "routine" pedagogy such as lectures and worksheets. Although there is evidence that children from lower SES backgrounds enter school with lower skill levels (Lee & Burkam, 2002), the above studies are not able to disentangle the influence of children's skills on classroom processes from other influences strongly associated with SES (e.g., classroom resources and parental involvement).

The few extant studies at the preschool level also suggest links between children's socioeconomic background and teaching beliefs and practices. For example, in one qualitative study of preschool teachers' beliefs about appropriate early literacy and mathematics instruction, researchers found significant differences in pedagogical beliefs related to whether teachers taught primarily children with low-SES or middle- to high-SES (Lee & Ginsburg, 2007). Teachers of low-SES children were more focused on preparing children for kindergarten via

literacy and mathematics instruction, and favored more didactic teaching methods than teachers of middle- to high-SES children, who showed greater preferences for classroom activities and teaching methods that took into consideration individual child preferences and emphasized nonacademic areas. Similarly, in their study of child engagement in prekindergarten classrooms, Chien et al. (2010) found that poor children were more likely to be taught using an "individual instruction" model, emphasizing strong teacher instructional support, large amounts of literacy instruction, high quality teacher feedback, and teacher-led discussions that elicit cognitive skills, compared to alternative approaches that emphasized child-directed exploration (free play), group instruction, and scaffolded learning. Yet, whether such differences in instruction truly lead to "better" or "worse" process quality, and potentially, better or worse child outcomes remain unanswered questions.

Classroom composition, particularly in terms of behavior, may also be related to preschool teachers' feelings of frustration and maladaptive behavior management practices. For example, research suggests a link between repeated conflict with children who are disruptive, overly needy, or hard to manage and teachers' feelings of emotional distress, which can lead to "burnout," marked by "emotional exhaustion" and "depersonalization" (Brouwers & Tomic, 2000; Morris-Rothschild & Brassard, 2006). This same line of research suggests that teachers and children can easily become caught up in cycles of negative interactions, in which teachers inadvertently exacerbate children's disruptive behavior through harsh and ineffective behavior management techniques (Morris et al., 2010).

Although theory and the little extant research on the topic suggests that children's skills may influence preschool teaching and classroom quality, a number of unanswered questions remain. I hypothesize that after controlling for an extensive set of classroom contextual

measures, initial average classroom pre-academic (vocabulary, reading, and math) and positive behavioral skill (social skills, learning behaviors) levels will be positively associated with measures of teacher and environmental quality. Alternatively, I hypothesize that the level of classroom problem behaviors will be negatively associated with such process quality measures.

Data and Methods

Data

In this study, I use data from the 2003 Head Start Family and Child Experiences (FACES) study, a nationally representative survey of children who entered Head Start in 2003, funded by the U.S. Department of Health and Human Services' Administration for Children and Families (ACF). These longitudinal data follow children and families served by Head Start from program entrance to the end of kindergarten. FACES data include detailed information on Head Start attendees' background characteristics, experiences, and developmental outcomes. At the beginning and end of the Head Start program year and again in kindergarten, high quality child assessments and parent interviews were completed. Teacher surveys were also conducted at the beginning and end of the Head Start year. Finally, center directors and education coordinators were also interviewed in the fall of the Head Start year, and classrooms were observed for overall quality, types of activities, and interactions between staff and children at the beginning and end of the Head Start year.

The FACES data provide several advantages over data used in the few previous explorations of preschool classroom composition effects. First, all families represented in FACES are participants in the same early childhood education program, Head Start. Therefore, the problem of selection bias based on *type* of program (children of differing abilities selecting

into different types of preschool programs) is eliminated. Second, a multitude of high quality child assessments of both pre-academic and behavioral skills, as well as observational assessments of classrooms, provide information about previously unmeasured child skills and classroom outcomes. Third, rich information regarding important, yet previously omitted control variables, including parenting, is also available. Fourth, up to 12 children per classroom are assessed upon entry into Head Start (out of an average class size of 15), providing a more accurate estimate of true classroom skill levels.[13] Fifth, although previous preschool classroom composition studies have been limited to data from a few sites or states, the FACES sample is relatively large and nationally representative of first year Head Start students; thus, increasing the study's external validity, particularly for preschool programs targeted to children from low-income families. Finally, given that these data were released in 2008, FACES provides one of the most up-to-date snapshots of the country's largest publicly funded compensatory preschool program.

Sample

The FACES sampling frame was drawn from the 2002-03 Head Start Program Information Report (PIR) file, and the sample was stratified on region, urban/rural status, percent minority enrollment, auspice type (school-based, other), and percentage of non-English speaking children in the program.[14] Response rates for children, parents, and teachers in the Head Start year were quite high, ranging from 93 to 97 percent in the initial round of surveys in the fall of 2003, and 86 to 92 percent at the end of the Head Start year, in the spring of 2004 (Westat et al.,

[13] Sojourner (2011) notes the potential bias in widely used peer-effects estimators in the extant literature, which do not account for the "missing data" of students not included in such estimators.

[14] Head Start programs in U.S. territories or Puerto Rico, Early Head Start programs, Migrant and Seasonal Head Start programs, American Indian Head Start programs, and Head Start programs already selected as part of the Head Start Quality Research Center samples were not included in the FACES study.

2008). The FACES data file contains information on 2,387 children in 373 different Head Start classrooms.

Although levels of missing data were quite low (almost always less than five percent at the item level), to maximize statistical power and minimize any bias due to missing data, I retain cases with missing information at the item level by using multiple imputation techniques (Graham, 2009; Allison, 2002). Specifically, I employed STATA's *ice* program to impute 10 data sets, using covariates from full regression models in the imputation regressions (Royston, 2004). Then, within each imputed dataset, I collapsed all variables to the classroom level and used the *mim* command to conduct analyses across datasets (See Appendix 4.A: Child, Family, and Classroom Covariate Description and Coding for further detail about how each measure was coded and collapsed).

Multiple imputation methods assume that data are missing at random; in other words, conditional on the observed variables in the analysis, the likelihood of missing data on a particular variable is unrelated to its value. Since this assumption does not hold for the English test scores of students tested in Spanish (N=311), I chose not to use imputed English test scores for these students; thus, decreasing the individual child sample to 2,076 students in 367 classrooms. As a robustness check, I conduct separate analyses substituting Spanish speaking children's Spanish test scores (vs. excluding these children from the sample).

Furthermore, to increase the validity and reliability of the peer skill measures, I limit analyses to classrooms with at least four fall test scores by which to construct a measure of peer skills. Given this restriction, the final analytic sample includes 292 classrooms. Descriptive statistics for the analytic sample are provided in Table 4.1.

Not surprisingly, these descriptive statistics reveal classrooms with a disadvantaged population. On average, approximately 33 percent of children sampled in each classroom are black; 24 percent are Hispanic, and 35 percent are white. Upon entry to the program, about 11 percent of children sampled in each classroom were identified as dual language learners, and 16 percent were identified as having disabilities. At the initial parent interview, on average, only 38 percent of mothers in each classroom reported being married, and over two-thirds of families in each classroom reported incomes below the federal poverty line. About 29 percent of mothers of children sampled in each classroom reported having less than a high school diploma.

Sample children also demonstrate low levels of pre-academic skills upon program entry. The classroom average score for vocabulary is an entire standard deviation below the national norm; about three-quarters of a standard deviation below the national norm for math, and half a standard deviation below the national norm for reading. Despite this, there is evidence of substantial variation in average classroom skill levels for each pre-academic subject area. Variation in average classroom behavior skills is also present.

Measures

Dependent Variables. The dependent variables of interest in this study are measures of classroom process quality, based on observational data collected in the fall and spring of the Head Start year. One measure is the revised version of the Early Childhood Environment Rating Scale (ECERS-R), which consists of 37 items and 7 subscales measuring the quality of personal care routines, furnishings, program structure, and opportunities to develop language skills, motor skills, social skills, and creativity (Harms, Clifford, & Cryer, 1998). For each question, observers rate the classroom environment based on a scale of 1 to 7, indicating “inadequate” to

"excellent" conditions. Westat et al. (2008) report high reliability for this measure (Cronbach's Alpha=.92).

Although the ECERS-R provides a common measure of overall classroom process quality, one might expect particular subscales to be more or less sensitive to the kinds of changes that teachers might make based on their assessments of children's skills or responses to children's behaviors. For example, the language skills subscale measures the availability and use of books and pictures, opportunities that encourage children to communicate, use of language to develop reasoning skills, and informal use of language in the classroom. One might predict that teachers would adjust the amount of time spent or level of language learning opportunities based on children's initial skills or behavior. The predicted direction of such effects is ambiguous, however. For example, teachers who are teaching particularly low-skilled students may increase the number of language learning opportunities, but lower the level or change the type of instruction (e.g., didactic vs. scaffolded) in ways that could yield a positive or negative relationship. Problem behaviors that disrupt language learning opportunities or strain communication between children and their teachers and peers may also have a negative impact on this type of measure, just as positive behaviors that encourage learning and communication may have positive impacts. Likewise, the social skills subscale, which measures the quality of the general supervision of children, discipline, staff-child interactions, and interactions among children, may be particularly sensitive to children's initial behavior skills. Therefore, alternative models using each of these subscales as outcomes are also considered.

The Arnett Scale of Caregiver Behavior consists of 30 items assessing 5 areas of lead teacher behavior: sensitivity, punitiveness, detachment, permissiveness, and prosocial interaction (Arnett, 1989). Westat et al. (2008) reports a high reliability for this measure as well

(Cronbach's Alpha=.93, respectively). Given that teacher stress may also be affected by the challenge of dealing with needy or disruptive children, in an alternative specification, I also consider a measure of teacher satisfaction, using a scale of 0 to 5, which is derived from teachers' answers to questions regarding: 1) how much they enjoy their present teaching job; 2) how certain they are that they are making a difference in the lives of the children they teach; 3) if they could start over, whether they would choose teaching again as their career; and 4) how likely they are to continue working for Head Start through the next year.

For ease of interpretation, in all regressions, the dependent variables are standardized to have a mean of 0 and standard deviation of 1. Descriptive information for each of the main process quality measures is provided in Table 4.1.

Classroom Level Skill Measures. The key independent variables in this study are classroom level measures of students' average pre-academic and behavioral skills upon entry to Head Start. These averages were constructed using standardized scores of all children assessed in a particular classroom. In the analytic sample, an average of 7.14 fall scores per classroom were available (range=4 to 16), and the average class included 14.37 students.

Average classroom pre-academic skill measures were created using nationally normed standard scores for three common early childhood assessments completed in the fall of the Head Stat year.[15] The Peabody Picture Vocabulary Test (PPVT-III) assesses children's knowledge of word meanings (Dunn & Dunn, 1997). A shortened version of the test was developed and given in the FACES survey, using Item Response Theory (IRT). This adaptive version retains all of the psychometric qualities of the original test, including high reliability (Cronbach's alpha=.91) (Westat et al., 2008). In addition, several subtests of the Woodcock-Johnson Psycho-Educational

[15] Spanish versions of each of the pre-academic skills assessments were given to Spanish-speaking children who did not pass a language screener test (N=273 in fall of 2003 and N=104 in spring of 2004). These students are not included in the primary specifications, but are considered in robustness checks.

Battery, Third Edition (WJ-III) were given, including the Letter-Word Identification subtest, which measures children's early reading skills (specifically, their ability to identify isolated letters and words); and the Applied Problems subtest, which measures children's early math skills. High reliabilities are reported for all WJ-III subtests (Cronbach's alpha range=.78-.91) (Westat et al., 2008).

Classroom average behavior skills measures are based on teacher reports of children's social skills and classroom behaviors, also assessed in the fall of the Head Start year. Specifically, across three separate scales, each measuring different types of behavior, teachers were asked to indicate the extent to which a given statement (e.g., "follows teacher's directions") described the child, on a scale of 1 ("never") to 3 ("very often"). The Behavior Problems scale includes 14 items drawn from the Personal Maturity Scale (Alexander & Entwisle, 1988), the Child Behavior Checklist for Preschool-Aged Children, the Teacher Report (Achenbach, Edelbrock, & Howell, 1987), and the Behavior Problems Index (Zill, 1990), which measure negative child behaviors associated with later learning problems and grade retention (e.g., aggression, hyperactivity, anxiety, and withdrawal). High reliabilities are reported for this measure (Cronbach's alpha =.85) (Westat et al., 2008). The Social Skills scale includes 12 items drawn from the Personal Maturity Scale (Alexander & Entwisle, 1988) and the Social Skills Rating System (Elliot, Gresham, Freeman, & McCloskey, 1988) that measure how often children engage in cooperative classroom behavior (e.g., following directions and rules, helping put things away, and complimenting classmates). High reliabilities are reported for this measure as well (Cronbach's alpha=.88). The Preschool Learning Behavior Scale (PLBS) includes 29 items assessing learning-related behaviors (e.g., motivation, attention/persistence, and attitudes toward learning) (McDermott, Green, Francis, & Stott, 2000). Westat et al. (2008) report a

Cronbach's alpha of .90 for this measure. Descriptive statistics for these measures are provided in Table 4.1.

Although each of the scales measure a distinct behavioral construct, the classroom level problem behavior and learning behavior measures are relatively highly correlated (r = -0.74). Given the particular theoretical importance of behavior problems and their potential direct impact on the classroom process quality measures at hand, this measure is included in all regressions. The other two behavior measures are entered one by one in the most complex models. Pairwise correlations for each of the classroom level skill measures are provided in Appendix 4.B. Again, for ease of interpretation, in regressions, these variables are standardized to have a mean of 0 and standard deviation of 1; therefore, coefficients may be interpreted as effect sizes.

Classroom Demographics (Child and Family Characteristics). Previous research linking classroom composition to process quality has focused primarily on the socioeconomic background of children, without the benefit of actual measures of children's skills. In order to isolate the unique contribution of these skills, I also use the rich FACES data to control for a number of potentially confounding classroom level demographic characteristics, which may be correlated with both process quality (assuming children from lower SES backgrounds attend lower quality preschool programs) and children's skills. I collapse individual-level information for individual students sampled in each classroom, by classroom for each measure.[16] This yields a number of classroom-level demographic and socioeconomic measures, including racial/ethnic make-up, percentage of boys in the classroom, average child age at post-test, average level of maternal education, average number of children in the family, percentage of children with

[16] The exception is percent of boys in the classroom, which is based on an average of two classroom observations, and, therefore, is a more reliable representation of the actual percentage of boys than the collapsed individual student-level information.

married mothers, average log of family income, percentage of children in poverty, and average number of days between pre- and post-tests.

I also include an extended set of classroom level measures of other characteristics found in previous studies to be associated with classroom outcomes, including: percentage of children who are English Language Learners (Cho, 2010), percentage of children with a disability (Fletcher, 2009), and percentage of children with prior center-based childcare experience (Neidell & Waldfogel, 2011). The FACES data also allow for construction of classroom averages of measures of the quality of children's home environments, including parental warmth, literacy behaviors, and maternal depression, which have rarely been included in previous studies of classroom quality, but are known to affect child developmental outcomes (e.g., see NICHD-ECCRN & Duncan, 2003).

Classroom Structural Quality. I also include a set of variables that measure the structural quality of Head Start classrooms, including class size, teacher/student ratio, and teacher education. Generally, one would not expect children's skills to impact such measures of quality, which are largely determined prior to a child's classroom assignment and are likely to remain relatively static throughout the year. Such measures may, however, be predictive of classroom process quality, and, thus, children's outcomes (Clarke-Stewart, Lowe Vandell, Burchinal, O'Brien, & McCartney, 2002; Burchinal, Cryer, Clifford, & Howes, 2002; Burchinal et al., 2000; Blau 1997, 2000; Ghazvini & Mullis, 2002; NICHD-ECCRN 1999; NICHD-ECCRN 2002). Therefore, they are included as controls in the more complex analytic models.

Analytical Approach

In this chapter, I use ordinary least squares (OLS) regression to measure associations between average classroom child pre-academic and behavioral skills and classroom level process quality outcomes. Primary models take the following form:

$$Y_{j2} = \beta_0 + \beta_1(CS_{j1}) + \beta_2(CF_{j1}) + \beta_3(SQ_{j1}) + e_j$$

In this model, which depicts a spring outcome measure, for each classroom, *j*, Y_{j2} is the measure of the outcome of interest at time 2, spring of the Head Start year[17]; β_0 is a constant; CS_{j1} is a vector of classroom-level pre-academic and behavioral skills measured at time 1, fall of the Head Start year; and β_1 is a vector of regression coefficients, the estimates of particular interest assessing the magnitude of the association between class skills and classroom process quality. CF_{j1} is a vector of child and family background characteristics aggregated at the classroom level and measured in the fall; SQ_{j1} is a vector of classroom structural quality measures measured in the fall, and e_j is the error term.

I begin by estimating a series of basic bivariate models, regressing each classroom skill measure on each of the process quality measures. In the primary analyses, I estimate a series of six models for each outcome, with an increasingly rich set of controls. I begin by including all three classroom level pre-academic and the problem behavior skills measures (Model 1). Next, I include a set of basic classroom-level demographic measures: racial/ethnic make-up, percentage of boys in the classroom, average child age at post-test, average level of maternal education, average number of children in the family, percentage of children with married mothers, average log of family income, percentage of children in poverty, and average number of days between pre- and post-tests (Model 2). Next, I also include the set of classroom structural quality

[17] If classroom process quality and teaching are highly sensitive to classroom skill level effects, the fall scores provide estimates that are less likely to be biased by omitted variables, as spurious correlations between children's skills and the quality indicators are likely to be minimal. If, however, as theory and empirical evidence suggests, adjustments in process quality occur gradually over the school year, the spring measures allow for more complete estimation of the associations between classroom composition and classroom-level outcomes.

measures: class size, teacher/student ratio, and teacher education (Model 3). Next, I include an extended set of classroom-level demographic measures: percentage of children who are English Language Learners, percentage of children with a disability, percentage of children with prior center-based childcare experience, and averages of measures of the quality of children's home environments, including parental warmth, literacy behaviors, and maternal depression (Model 4). Finally, in the last two models, I include the other classroom behavior measures available: social skills (Model 5) and preschool learning behaviors (Model 6).

Robustness Checks

The analytic model described above presumes that classroom skills upon entry are uncorrelated with initial classroom process quality (or the ability of teachers to change quality); therefore, any association between initial skills and later (spring) measures of process quality are causal in nature, not spurious. For example, an initial correlation could arise from either a selection process in which poorly skilled children are aggregated in classrooms with less skilled teachers or if children's fall skills are already affected by their short time in the Head Start program. Using OLS regressions in which change scores (spring process quality minus fall process quality) are used as dependent variables provides one alternative modeling strategy that does not presume that aggregated pre-academic skills and behaviors are independent. These models implicitly hold constant variation in initial levels of quality, and estimate associations between classroom level skills and changes in quality from fall to spring.

To test whether findings hold for classrooms with Spanish speaking Head Start students, I run models in which I substitute Spanish version scores for missing test scores. I also explore a number of alternative process quality measures, including language and social subscales of the

ECERS-R and a measure of Head Start teacher satisfaction. I also consider alternative specifications of classroom skills, including minimum classroom scores, maximum classroom scores, and percentage of students with scores in the highest and lowest quartiles.

Results

Results from the bivariate regressions are displayed in Table 4.2. These regressions show little evidence of significant associations between classroom level skills and classroom process quality as measured by the ECERS-R. Only the relationship between classroom level math skills and the average spring ECERS-R score is marginally significant (effect size = .11).[18] Insignificant associations between classroom vocabulary, learning behavior, and behavior problems and ECERS-R measures are in the expected direction, but relationships between classroom reading and social skills and this process quality measure are not.

Results of bivariate regressions for the Arnett Caregiver Behavior Scale show more sensitivity to classroom skill levels. In these simple models, there are, as expected, significant positive associations between classroom math and learning behavior skills and process quality (effect sizes = .19 and .12, respectively), as well as an expected negative relationship between classroom behavior problems and this process quality measure (effect size= .12). Insignificant relationships between vocabulary and social skills and teacher behavior scores are in the expected direction, but the relationship between classroom level reading skills and this measure is not.

Overall, these findings suggest that there may indeed be some links between classroom level pre-academic and behavior skills and classroom process quality; however, the non-intuitive

[18] Effect sizes can be interpreted as the percent increase in standard deviation of the outcome related to a one standard deviation increase in the classroom skill score.

nature of some of the bivariate relationships suggest that more controls are needed. These initial regressions also suggest that teaching interactions may be more influenced by children's skill levels than overall classroom environment quality.

Next, I turn to the set of models with increasingly rich controls.[19] Table 4.3 summarizes the results for the ECERS-R. Consistent with the bivariate results, in Model 1, including all classroom level average pre-academic skills and problem behavior measures, I find a marginally significant positive relationship between classroom level math skills and the spring ECERS-R measure (effect size = .15), but also an unexpected negative relationship between classroom level reading skills and this process quality measure (effect size = .15). Adding in controls for basic classroom demographics (Model 2) reduces the size of these coefficients and renders them statistically insignificant. Including measures of classroom structural quality (Model 3), the extended set of classroom demographics (Model 4), and the additional behavioral skills measures (Models 5 and 6) also yields no significant associations between any of the classroom level skill measures and the ECERS-R.

Table 4.4 summarizes the results for the Arnett Caregiver Behavior Scale. Again, in Model 1, which includes all classroom level average pre-academic skills and problem behavior measures, I find a significant positive relationship between classroom level math skills and the spring ECERS-R measure (effect size = .22), and an unexpected negative relationship between classroom level reading skills and this process quality measure (effect size = .13). The statistically significant relationship between behavior problems and the Arnett measure present in the bivariate models is no longer present here. With the inclusion of basic classroom demographics (Model 2), the negative coefficient on the classroom level reading skills is reduced and rendered insignificant. The positive coefficient on classroom level math skills is also

[19] Complete results, for all covariates, in the most complex models for each outcome, can be found in Appendix 4.C.

reduced (effect size =.16), but remains marginally significant. Conversely, the coefficient on classroom behavior problems increases (effect size =.10), suggesting a marginally significant negative relationship between classroom level problematic behavior and teaching quality.

The relationship between classroom math skills and the Arnett measure, as well as the relationship between remains sensitive to model specification. Both associations remain in tact in Model 3, which adds classroom structural quality measures, and the effect size for behavior problems increases (.12). In Model 4, however, which includes the extended set of classroom demographics, the effect size for math skills decreases and the standard error increases, rendering the coefficient statistically insignificant. These numbers remain stable in the last two models (5 and 6).

Conversely, the effect size for behavior problem increases (.13) in Model 4, while the standard error remains stable; thus, the statistical significance of the coefficient is no longer marginal in nature. Adding the classroom level social skills measure in Model 5 also increases the behavior problems coefficient (.15). Ultimately, in Model 6, however, including the classroom level learning behaviors measure both decreases the behavior problems coefficient (.11) and increases the standard error; thus, rendering the association statistically insignificant.

Robustness Checks

To test the robustness of these findings, I conduct a series of alternative specifications, considering differences in sample, methodology, classroom skill measurement, and process quality measures. Given the results from the main analyses and the knowledge that the behavior problems and learning behavior measures are highly correlated, for each alternative specification, I employ models with a full set of demographic and structural quality controls, but

vary inclusion of just the behavior problem measure (Model 1 in Tables 4.5 to 4.10) or all three behavior measures (Model 2 in Tables 4.5 to 4.10).

First, I test whether the results hold after including Spanish speaking children who did not pass an English language screener in the fall. As previously noted, these children are excluded from the main analyses, as one cannot confidently impute their English test scores. However, these children were given Spanish versions of each of the pre-academic tests, which can be substituted for English scores. Results from these analyses, which mimic the most complex models in Tables 4.3 and 4.4, are shown in Table 4.5. They are very similar to the main analyses findings, with the exception that the positive relationship between classroom math skills and the Arnett measure is larger, and maintains statistical significance in the more complex models (effect size = .18).

As previously acknowledged, gain score analysis is an alternative methodological approach made viable with the FACES longitudinal data, although concerns remain regarding measurement error (NICHD-ECCRN & Duncan, 2003). As shown in Table 4.6, like the OLS models, models using a complete set of covariates (Model 2), but using the change between spring and fall process quality measures as outcomes, show no significant relationships between classroom average pre-academic and behavior skills measured in the fall and actual changes in process quality measures between fall and spring. The model excluding the social skills and learning behavior measures fails to yield the statistically significant negative relationship between classroom level behavior problems and the Arnett measure of caregiver interactions present in the spring outcome regression; however, the coefficient is in the same direction. The fact that most coefficients in these analyses are in the same general direction and of the same magnitude as in previous analyses (albeit with larger standard errors), allays some concern that

ignoring potential correlations between classroom skills and fall process quality (by excluding the fall process quality measures in those models) could bias the resulting estimates.

As noted in the dependent measures section of this chapter, I also consider alternative measures of classroom process quality. Results shown in Table 4.7 (columns 1 & 2) provide no evidence of a link between classroom-level pre-academic or behavioral skills and the ECERS-R language subscale. Column 3, as hypothesized, shows a marginally significant negative relationship (effect size = .11) between behavior problems and the social skills subscale, although the same coefficient is imprecisely estimated and does not rise to the level of statistical significance in the most complex model. The same is true of the behavior problems estimate in both models using the teacher satisfaction measure as the outcome (Table 4.7, Columns 5 and 6). Results for this model do, however, suggest a positive link between classroom-level math skills and teacher satisfaction (effect size = .22).

Models summarized in Table 4.8, Panels 1 and 2 explore whether having one particularly high or low-skilled student might be enough to impact classroom quality, by using the classroom maximum (Panel 1) or minimum (Panel 2) as alternative specifications of classroom skill levels. Almost all estimates are close to zero, suggesting that this is not the case, with the exception of the maximum behavior problem coefficient in relation to the Arnett outcome (Panel 1, Column 3). The coefficient is statistically significant in Model 1 (effect size = .03), providing some suggestion that having even a single student with a particularly high level of problematic behaviors could lower the overall quality of interactions between teachers and children in the classroom.

Pushing this line of thinking further, I estimate models that measure whether having a higher percentage of students scoring in the highest or lowest quartile of each skill area is

associated with improved or reduced classroom process quality. These models recognize that the "smoothing" effect of an average classroom score may not truly represent how children's skills or behaviors impact classroom processes. The predicted direction of such effects remains ambiguous, however. For example, a teacher with a large percentage of children with very high language skills may adopt a "skills beget skills" philosophy and increase the language learning opportunities within the classroom (thus, affecting the ECERS-R score) (Cunha & Heckman, 2007). Alternatively, a different teacher in a similar situation may *decrease* language learning opportunities, perhaps focusing efforts on a different skill area where s/he perceives weaknesses. The potential impact of having high percentages of positive or problem behavior may be more straightforward; in particular, one might expect the strongest relationship between such skill levels and teacher interactions.

Results, shown in Table 4.9 for the ECERS-R outcome and Table 4.10 for the Arnett measure, reflect some of this anticipated ambiguity. Three different reference groups are considered: Columns 1 and 2 use middle- and low-scoring students as the reference group; Columns 3 and 4 use middle- and high-scoring students, and Columns 5 and 6 use middle-scoring students only. The results suggest little evidence of a relationship between having a higher percentage of high- or low-scoring students in any of the pre-academic or behavior areas and the ECERS-R measure (Table 4.9). There is also little evidence of a relationship between having a higher percentage of low-scoring students and the Arnett measure of caregiver interactions (Table 4.10). Having a higher percentage of children with greater behavior problems is, as expected, negatively associated with the quality of child/teacher interactions (Table 4.10, Columns 1 and 2, effect size = .14 to .15). A rather unexpected result, however—a marginally significant negative relationship between having a higher percentage of children with

high vocabulary scores and the Arnett—is shown in Columns 1, 2, 5, and 6 (effect size = .15 to .16).

Discussion

Although research suggests that preschool process quality matters for child outcomes, few studies have explored how classroom composition is linked to the quality of child/teacher interactions and learning environment that drive these measures. In particular, few studies at the preschool level have used actual measures of child skills to identify such associations. Understanding these relationships is important in order to realize the full potential of early education programs to impact child outcomes. If class composition matters, attention should be paid to the way in which children are grouped in early education classrooms, and professional development that provides teachers with the skills needed to best adapt their instruction and interactions to meet their classroom's needs should be encouraged.

The results from this study, one of the first of its kind, suggest that, indeed, such links may exist. The most consistent evidence suggests a negative association between children's behavior problems and the quality of child/teacher interactions at the classroom level. Although somewhat sensitive to specification, this finding was present in several different models, including those using classroom average level behavior problems, percent of children in the class with the highest level of behavior problems, and the maximum behavior problem score in a classroom as the independent variables of interest. There is also suggestion of a relationship between class behavior problems and the ECERS-R social skills subscale, a measure of classroom climate.

What are the implications for preschool programs? Recent studies of professional development programs designed to improve preschool teachers' behavior management skills, and, thus, the emotional climate in the classroom, suggest that such negative patterns can be altered (Morris et al., 2010). For example, Raver and colleagues' (2008) randomized study of the Chicago School Readiness Project, a behavior management training program designed to improve teachers' emotional supportiveness within Head Start programs, found significantly higher levels of positive classroom climate, teacher sensitivity, and behavior management in treatment classrooms (effect sizes ranging from .52 to .89). Specifically, teachers in treatment classrooms demonstrated greater enthusiasm w/students, more responsiveness to students' needs, and lower use of harsh or emotionally negative practices (Raver et al, 2008). In general, training that provides teachers with strategies to refrain from engaging in emotionally negative cycles of interactions with disruptive students may be helpful.

The results regarding the relationship between pre-academic skills and process quality are more ambiguous. I find a positive association between classroom math skills and teacher satisfaction, and the same type of link between math skills and child/teacher interactions in the less complex main analyses. It is difficult to explain why such links appear only for math skills, when one might expect more consistent results across all academic areas. Despite research suggesting that math skills are more "teachable" than, for example, vocabulary skills (Christian et al., 2000), there is widely documented lack of math teaching at the preschool level (Pianta, 2008), documented misconceptions among preschool teachers regarding young children's ability to learn math, and few required courses in math education in early education teacher preparation programs (Lee & Ginsburg, 2008). Perhaps, somewhat ironically, preschool teachers with students more skilled in math feel more confident and effective without having to make the kinds

of adjustments they feel compelled to make in language instruction. Additionally, as noted earlier in this paper, perhaps the different adjustments teachers make in language instruction work in different directions or "cancel" each other out when aggregated across many classrooms, as in these analyses. Certainly, more research in this area is needed, although the findings are worth notice, given other recent research suggesting that early math skills are the best predictors of later achievement (Duncan et al., 2007).

I find little evidence of a relationship between children's skills and overall classroom environmental quality, as measured by the ECERS-R. Again, this is perhaps not surprising given that the ECERS-R assesses quality across a wide swath of subscales, some of which we expect to be less affected by children's skills (e.g., furnishings) than others (e.g., social skills). The lack of relationship between children's language skills and the language subscale is somewhat surprising; however, as noted earlier, teachers' responses to their assessment of children's language skills could result in an enrichment of the language learning environment, or a diversion of efforts to perceived needs in other areas. More detailed research that is able to illuminate what the "best" response for children at different skill levels may be needed to make clearer use of these findings.

The findings are also offered with a number of limitations. Although this study provides one of the first opportunities to explore associations between preschool classroom composition and process quality using a rich set of child skill measures and background information on children, families, and classrooms, the analyses are non-experimental in nature, and any resulting associations may not be causal. Although I attempt to reduce omitted variable bias by controlling for important child, family, and classroom characteristics, there may be other factors affecting process quality (e.g., funding, management), that remain unmeasured. Although the

availability of fall measures of the independent variables of interest and control variables is helpful, reverse causality also remains a concern that cannot be ruled out in these OLS analyses.

The study of this phenomenon solely within the context of Head Start is helpful, in that given the homogeneity in population and programming, one can be less concerned about selection bias. Indeed, perhaps the general non-findings in this study tell a positive story, in that children in Head Start do *not* appear to be sorted in the way one might predict or be concerned about—i.e., lower skilled children receiving lower quality programming. However, the generalizability of such findings to other preschool programs (e.g., state pre-k programs or private preschool) is perhaps limited. Head Start is a program with well-defined standards and a mission to serve low-skilled children. It may be difficult to detect the kinds of changes in process quality sought here given this program structure and more limited range of child skills. Perhaps these processes are more dynamic in private preschool and state pre-k programs, as suggested in research that shows much higher levels of expulsion in those types of programs (Gilliam, 2005; Gilliam & Shahar, 2006).

Even with such limitations, this research presents some of the first quantitative evidence of associations between preschool classroom composition and process classroom quality. The results suggesting the existence of such links, particularly between problem behaviors and the quality of child/teacher interactions, recommend careful consideration of where children with high levels of behavior problems are placed, and the advancement of professional development opportunities for preschool teachers that help them more effectively navigate their responses to and relationships with disruptive students. In addition, the findings raise potential questions about states' use of process quality measures for high stakes purposes, such as tiered reimbursement tied to Quality and Rating Improvement Systems (QRISs).

Table 4.1: Descriptive Statistics

	Min.	Max.	Mean	S.D.
Basic Classroom Demographics (Child and Family Characteristics)				
Percent boys	0.13	0.89	0.50	0.13
Race/Ethnicity				
Percent White, Non-Hispanic (Referent)	0	1	0.35	0.38
Percent African American	0	1	0.33	0.38
Percent Hispanic	0	1	0.24	0.31
Percent Other race/ethnicity	0	1	0.08	0.16
Average child age at post-test (months)	42.40	64.17	54.22	4.31
Maternal Education				
Percent less than high school diploma	0	1	0.29	0.20
Percent high school diploma (Referent)	0	1	0.38	0.21
Percent any post-secondary education	0	1	0.33	0.22
Extended Classroom Demographics (Child and Family Characteristics)				
Percent English Language Learners	0	1	0.11	0.21
Percent of children with a disability	0	0.80	0.16	0.16
Percent of children who attended center-based care before Head Start	0	1	0.20	0.20
Average log of family income	8.34	10.80	9.44	0.39
Percent of families with income below poverty line	0	1	0.68	0.20
Average number of children in family	1.25	5.57	2.63	0.64
Percent of children with married mother	0	1	0.38	0.23
Average measure of frequency parents read to child	1.50	7	4.77	1.08
Average measure of parental warmth	3.80	5	4.39	0.18
Average measure of parental depression	0	21	6.96	2.90
Average number of days between pre- and post-test	117.11	234.75	170.82	17.65
Classroom Structural Quality Measures				
Child/Staff Ratio	2	15	6.64	2.05
Class size	4	27	14.39	3.21
Teacher Education				
No post-secondary education (Referent)	0	1	0.06	0.22
Associate's degree or some college	0	1	0.57	0.48
Bachelor's or advanced degree	0	1	0.37	0.47

Table 4.1: Descriptive Statistics, Continued

	Min.	Max.	Mean	S.D.
Classroom Average Fall Pre-Academic Test Scores				
PPVT (Vocabulary)	58.65	100.57	84.69	6.82
WJ-Letters & Dictation (Reading)	67.19	114	92.25	7.93
WJ-Applied Problems (Math)	61.96	112.84	88.64	9.24
Classroom Average Fall Behavior Skills Scores				
Preschool Learning Behavior Scale (t-score)	34.33	63.50	50.47	6.20
Social Skills	7.17	22.14	15.45	2.99
Problem Behaviors	0	12.75	4.73	2.55
Classroom Spring Process Quality Measures (Outcomes)				
Arnett Teacher Behavior Score	21.38	90	69.71	12.16
ECERS-R Average Score	1.83	6.39	4.18	0.74

Note: N=2,920 classrooms across 10 imputed data sets; descriptives provided for classrooms with at least 4 fall pre-academic test scores.

Table 4.2: Bivariate Regressions of Average Classroom Skills on Classroom Process Quality Outcomes

	ECERS-R	Arnett
Classroom Vocabulary	0.08	0.08
	(0.06)	(0.06)
Classroom Reading	-0.08	-0.03
	(0.06)	(0.06)
Classroom Math	0.11^{t}	0.19*
	(0.06)	(0.06)
Classroom Learning Behavior	0.02	0.12*
	(0.06)	(0.06)
Classroom Social Skills	-0.02	0.05
	(0.06)	(0.06)
Classroom Behavior Problems	-0.01	-0.12^{t}
	(0.06)	(0.06)
N	292	292

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 4.3: Summary of Results from Regressions of Average Classroom Skills on Classroom Process Quality (ECERS-R) in the Spring of the Head Start Year

	Models					
	1	2	3	4	5	6
Classroom Vocabulary	0.04	-0.04	-0.07	-0.02	-0.03	-0.03
	(0.07)	(0.09)	(0.09)	(0.11)	(0.11)	(0.11)
Classroom Reading	-0.15*	-0.05	-0.06	-0.06	-0.06	-0.06
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Classroom Math	0.15^{t}	0.06	0.06	0.06	0.06	0.06
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Classroom Behavior Problems	0.00	-0.02	-0.02	-0.03	-0.02	-0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.10)
Classroom Social Skills					0.02	0.02
					(0.08)	(0.08)
Classroom Learning Behavior						0.01
						(0.10)
Basic Classroom Demographics		yes	yes	yes	yes	yes
Classroom Structural Quality			yes	yes	yes	yes
Extended Classroom Demographics				yes	yes	yes
N	292	292	292	292	292	292

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 4.4: Summary of Results from Regressions of Average Classrom Skills on Classroom Process Quality (Arnett) in the Spring of the Head Start Year

	Models					
	1	2	3	4	5	6
Classroom Vocabulary	-0.00	-0.10	-0.09	-0.10	-0.10	-0.09
	(0.07)	(0.09)	(0.09)	(0.11)	(0.11)	(0.11)
Classroom Reading	-0.13*	-0.03	-0.04	-0.02	-0.01	-0.01
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Classroom Math	0.22*	0.16^{t}	0.16^{t}	0.14	0.14	0.14
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Classroom Behavior Problems	-0.09	-0.10^{t}	-0.12^{t}	-0.13*	-0.15*	-0.11
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.10)
Classroom Social Skills					-0.05	-0.07
					(0.08)	(0.08)
Classroom Learning Behavior						0.06
						(0.10)
Basic Classroom Demographics		yes	yes	yes	yes	yes
Classroom Structural Quality			yes	yes	yes	yes
Extended Classroom Demographics				yes	yes	yes
N	292	292	292	292	292	292

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Table 4.5: Summary of Results from Regressions of Maximum Skills on Measures of Classroom Process Quality in Spring of the Head Start Year, Including Spanish Scores

	Spring Outcomes			
	ECERS-R Model 1	ECERS-R Model 2	Arnett Model 1	Arnett Model 2
Classroom Vocabulary	-0.03	-0.02	-0.05	-0.04
	(0.08)	(0.08)	(0.08)	(0.08)
Classroom Reading	-0.06	-0.06	-0.04	-0.04
	(0.06)	(0.06)	(0.06)	(0.06)
Classroom Math	0.10	0.11	0.17*	0.18*
	(0.07)	(0.07)	(0.07)	(0.08)
Classroom Behavior Problems	-0.01	-0.03	-0.12*	-0.12
	(0.06)	(0.09)	(0.06)	(0.09)
Classroom Social Skills		-0.02		-0.07
		(0.08)		(0.08)
Classroom Learning Behaviors		-0.02		0.04
		(0.09)		(0.09)
N	333	333	333	333

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses.
Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 4.3 & 4.4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors. Spanish scores are substituted as appropriate.

Table 4.6: Summary of Results from Regressions of Classroom Average Skills on Changes in Classroom Process Quality Measures between Fall and Spring of the Head Start Year

	Chage Score Outcomes			
	ECERS-R Model 1	ECERS-R Model 2	Arnett Model 1	Arnett Model 2
Classroom Vocabulary	-0.12	-0.11	-0.15	-0.14
	(0.11)	(0.11)	(0.11)	(0.11)
Classroom Reading	-0.03	-0.03	-0.05	-0.04
	(0.07)	(0.07)	(0.07)	(0.07)
Classroom Math	-0.05	-0.04	0.04	0.04
	(0.09)	(0.09)	(0.09)	(0.09)
Classroom Behavior Problems	-0.06	-0.01	-0.08	-0.05
	(0.07)	(0.10)	(0.07)	(0.10)
Classroom Social Skills		-0.07		-0.08
		(0.09)		(0.09)
Classroom Learning Behaviors		0.14		0.07
		(0.10)		(0.10)
N	292	292	292	292

NOTE: ** p<0.01, * p<0.05; ^{t} p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 4.3 & 4.4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors.

Table 4.7: Summary of Results from Regressions of Classroom Average Skills on Alternative Measures of Classroom Process Quality in Spring of the Head Start Year

	Spring Outcomes					
	ECERS-R Language Model 1	ECERS-R Language Model 2	ECERS-R Social Model 1	ECERS-R Social Model 2	Teacher Satisfaction Model 1	Teacher Satisfaction Model 2
Classroom Vocabulary	-0.09	-0.09	-0.12	-0.12	-0.12	-0.12
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Classroom Reading	0.01	-0.09	0.02	0.02	-0.03	-0.03
	(0.07)	(0.11)	(0.07)	(0.07)	(0.07)	(0.07)
Classroom Math	0.12	0.12	0.11	0.11	0.22*	0.22*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Classroom Behavior Problems	-0.06	-0.05	-0.11^{t}	-0.09	-0.08	-0.07
	(0.06)	(0.10)	(0.06)	(0.10)	(0.07)	(0.10)
Classroom Social Skills		0.02		0.01		-0.00
		(0.09)		(0.09)		(0.09)
Classroom Learning Behaviors		0.01		0.02		0.02
		(0.10)		(0.10)		(0.10)
N	292	292	292	292	292	292

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 4.3 & 4..4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors.

Table 4.8, Panel 1: Summary of Results from Regressions of Maximum Skills on Measures of Classroom Process Quality in Spring of the Head Start Year

	Spring Outcomes			
	ECERS-R Model 1	ECERS-R Model 2	Arnett Model 1	Arnett Model 2
Max Classroom Vocabulary	0.00	0.00	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.00)
Max Classroom Reading	-0.01	-0.01	-0.00	-0.00
	(0.00)	(0.00)	(0.01)	(0.00)
Max Classroom Math	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Max Classroom Behavior Problems	-0.01	-0.01	-0.03*	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Max Classroom Social Skills		-0.01		-0.01
		(0.02)		(0.02)
Max Classroom Learning Behaviors		-0.01		-0.01
		(0.01)		(0.02)
N	292	292	292	292

Table 4.8, Panel 2: Summary of Results from Regressions of Minimum Skills on Measures of Classroom Process Quality in Spring of the Head Start Year

	Spring Outcomes			
	ECERS-R Model 1	ECERS-R Model 2	Arnett Model 1	Arnett Model 2
Min Classroom Vocabulary	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Min Classroom Reading	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Min Classroom Math	0.00	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Min Classroom Behavior Problems	0.05	0.06	0.02	0.04
	(0.05)	(0.05)	(0.05)	(0.05)
Min Classroom Social Skills		0.01		0.01
		(0.02)		(0.02)
Min Classroom Learning Behaviors		0.00		0.01
		(0.01)		(0.01)
N	292	292	292	292

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 4.3 & 4.4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors.

Table 4.9: Summary of Results from Regressions of Percent of Class with Scores in Highest and Lowest Quartiles on Classroom Process Quality Measure (ECERS-R) in Spring of the Head Start Year

				ECERS-R		
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
% Hi Quartile Vocabulary	0.00	0.01			0.01	0.01
	(0.08)	(0.08)			(0.08)	(0.09)
% Hi Quartile Reading	-0.03	-0.03			-0.03	-0.03
	(0.07)	(0.07)			(0.07)	(0.07)
% Hi Quartile Math	0.08	0.07			0.07	0.07
	(0.08)	(0.08)			(0.08)	(0.08)
% Hi Quartile Problem Behaviors	-0.01	-0.02			-0.04	-0.06
	(0.06)	(0.07)			(0.08)	(0.09)
% Hi Quartile Social Skills		-0.04				-0.03
		(0.08)				(0.09)
% Hi Quartile Learning Behaviors		0.02				0.05
		(0.08)				(0.09)
% Lo Quartile Vocabulary			0.06	0.06	0.06	0.05
			(0.09)	(0.09)	(0.09)	(0.09)
% Lo Quartile Reading			-0.01	-0.01	-0.00	-0.01
			(0.06)	(0.06)	(0.07)	(0.07)
% Lo Quartile Math			-0.04	-0.04	-0.01	-0.01
			(0.08)	(0.08)	(0.08)	(0.08)
% Lo Quartile Problem Behaviors			-0.02	-0.02	-0.04	-0.05
			(0.06)	(0.07)	(0.07)	(0.09)
% Lo Quartile Social Skills				-0.03		-0.03
				(0.07)		(0.08)
% Lo Quartile Learning Behaviors				0.03		0.06
				(0.07)		(0.08)
N	292	292	292	292	292	292

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 3 & 4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors.

Table 4.10: Summary of Results from Regressions of Percent of Class with Scores in Highest and Lowest Quartiles on Classroom Process Quality Measure (Arnett) in Spring of the Head Start Year

				Arnett		
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
% Hi Quartile Vocabulary	-0.16[t]	-0.15[t]			-0.16[t]	-0.15[t]
	(0.08)	(0.08)			(0.08)	(0.08)
% Hi Quartile Reading	-0.00	-0.00			-0.02	-0.02
	(0.07)	(0.07)			(0.07)	(0.07)
% Hi Quartile Math	0.10	0.10			0.05	0.05
	(0.08)	(0.08)			(0.08)	(0.08)
% Hi Quartile Problem Behaviors	-0.14*	-0.15*			-0.12	-0.13
	(0.06)	(0.07)			(0.07)	(0.09)
% Hi Quartile Social Skills		-0.01				-0.03
		(0.08)				(0.09)
% Hi Quartile Learning Behaviors		-0.01				0.02
		(0.08)				(0.08)
% Lo Quartile Vocabulary			0.02	0.02	-0.01	-0.01
			(0.09)	(0.09)	(0.09)	(0.09)
% Lo Quartile Reading			-0.04	-0.04	-0.03	-0.03
			(0.07)	(0.07)	(0.07)	(0.07)
% Lo Quartile Math			-0.13	-0.12	-0.11	-0.11
			(0.08)	(0.08)	(0.08)	(0.08)
% Lo Quartile Problem Behaviors			0.07	0.06	0.01	0.03
			(0.06)	(0.07)	(0.07)	(0.09)
% Lo Quartile Social Skills				0.02		0.01
				(0.08)		(0.08)
% Lo Quartile Learning Behaviors				-0.04		0.06
				(0.07)		(0.08)
N	292	292	292	292	292	292

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores. Model 1 mimics Model 4 in Tables 3 & 4, including all classroom-level pre-academic skills, behavior problems, basic and extended demographics, and structural quality. Model 2 mimics Model 6 in Tables 4.3 & 4.4, adding classroom-level social skills and learning behaviors.

Appendix 4.A: Child, Family, and Classroom Covariate Description and Coding

	Concept	Coding
Child Characteristics	Initial skill levels	Continuous variables; child's individual standardized test scores (Academic skills=PPVT, reading, math; Behavior skills=learning related behaviors, pro-social skills, and behavior problems); averaged by classroom
	Age at assessment	Continuous variable; individual child age in months, averaged by classroom
	Gender	Average of two classroom observations (1 in fall, 1 in spring), percent boys
	Race/Ethnicity	Individual child; series of dummy variables for black, Hispanic, other race (omitted=white), collapsed to represent percent of each in classroom
	Disability	Individual child; dummy variable; 1=has disability; collapsed to represent percent in each classroom
	Language learning status	Individual child; dummy variable; 1=English language learner; collapsed to represent percent in each classroom
	Prior childcare experience	Individual child; dummy variable; 1= previously attended center-based care; collapsed to represent percent in each classroom
Family Characteristics	Number of children	Individual child; continuous variable; averaged by classroom
	Family structure	Dummy variable; 1=mother married
	Maternal education	Individual child; series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent); collapsed to represent percent in each classroom
	Income	Individual child; continuous variable; natural log annual family income; averaged by classroom
	Literacy behavior	Individual child; continuous variable; how often child has been read to in the last week (scale of 0, "not at all" to 7, "every day"); averaged by classroom
	Maternal depression	Individual child; continuous variable derived from12 items from CES-D Depression Scale; averaged by classroom
	Parent warmth and sensitivity	Individual child; continuous variable derived from 5 items asking about parenting behavior; averaged by classroom

Classroom Characteristics	Child/staff ratio	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Class size	Continuous variable based on average of two observations (1 in fall, 1 in spring)
	Teacher Education	Series of dummy variables for <high school, some post-secondary education (omitted= high school diploma or equivalent)
	Classroom process (global) quality	Continuous variable; ECERS-R mean score
	Teacher sensitivity	Continuous variable; Arnett Scale of Caregiver Behavior (lead teacher score)

Appendix 4.B : Pairwise Correlations of Average Classroom Skills in Fall of Head Start Year

Variables	1	2	3	4	5	6
1. Classroom-PPVT (Vocabulary)	--					
2. Classroom-WJ-Letters & Dictation (Reading)	0.29*	--				
3. Classroom-WJ-Applied Problems (Math)	0.56*	0.39*	--			
4. Classroom-Preschool Learning Behavior Scale	0.07*	0.12*	0.21*	--		
5. Classroom-Social Skills	0.13*	0.24*	0.30*	0.60*	--	
6. Classroom-Behavior Problems	0.02	-0.12*	-0.20*	-0.74*	-0.54*	--

NOTE: N=2,920 classrooms across 10 imputed data sets; * $p<0.05$; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

Appendix 4.C: Complete Results from Regressions of Average Classroom Skills on Classroom Process Quality in the Spring of the Head Start Year (Models 4 & 6, Tables 4.3-4.4)

	ECERS-R Model 4	ECERS-R Model 6	Arnett Model 4	Arnett Model 6
Classroom Vocabulary	-0.02	-0.03	-0.10	-0.09
	(0.11)	(0.11)	(0.11)	(0.11)
Classroom Reading	-0.06	-0.06	-0.02	-0.01
	(0.07)	(0.07)	(0.07)	(0.07)
Classroom Math	0.06	0.06	0.14	0.14
	(0.09)	(0.09)	(0.09)	(0.09)
Classroom Behavior Problems	-0.03	-0.01	-0.13*	-0.11
	(0.06)	(0.10)	(0.06)	(0.10)
Classroom Social Skills		0.02		-0.07
		(0.08)		(0.08)
Classroom Learning Behaviors		0.01		0.06
		(0.10)		(0.10)
Percent African-American	-0.63*	-0.63*	-0.80*	-0.77*
	(0.27)	(0.27)	(0.28)	(0.28)
Percent Hispanic	-0.03	-0.04	-0.34	-0.32
	(0.30)	(0.31)	(0.31)	(0.31)
Percent Other Race	-0.38	-0.38	-0.68	-0.72
	(0.42)	(0.43)	(0.43)	(0.44)
Percent Mom Ed. < High School	-0.05	-0.06	0.16	0.19
	(0.35)	(0.35)	(0.36)	(0.36)
Percent Mom Ed. Any Post-Secondary	-0.20	-0.20	-0.14	-0.10
	(0.34)	(0.34)	(0.35)	(0.35)
Average Child Age at Post-Test	-0.01	-0.01	-0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Percent Boys	0.21	0.22	-0.18	-0.20
	(0.47)	(0.48)	(0.48)	(0.48)
Average # of Days between Pre-/Post Test	0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Average (Log) Family Income	0.07	0.06	0.14	0.15
	(0.27)	(0.27)	(0.27)	(0.28)
Percent Poverty	0.34	0.32	0.82[t]	0.87[t]
	(0.46)	(0.47)	(0.47)	(0.47)
Average Number of Children in Family	-0.07	-0.07	-0.02	-0.02
	(0.11)	(0.11)	(0.11)	(0.11)
Percent Children with Married Mothers	0.03	0.03	0.21	0.20
	(0.32)	(0.32)	(0.32)	(0.32)

Appendix 4.C: Complete Results from Regressions of Average Classroom Skills on Classroom Process Quality in the Spring of the Head Start Year (Models 4 & 6, Tables 4.3-4.4), Continued

	ECERS-R Model 4	ECERS-R Model 6	Arnett Model 4	Arnett Model 6
Teacher/Student Ratio	-0.09*	-0.09*	-0.03	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)
Class Size	0.01	0.01	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Teacher has Associate's Degree	-0.15	-0.16	-0.21	-0.20
	(0.27)	(0.27)	(0.27)	(0.27)
Teacher has Bachelor's or Advanced Degree	-0.22	-0.23	-0.03	-0.03
	(0.28)	(0.27)	(0.28)	(0.29)
Percent Children with Disabilities	0.19	0.18	0.09	0.10
	(0.41)	(0.41)	(0.40)	(0.40)
Percent Children who are English Language Learners	0.53	0.54	0.12	0.14
	(0.47)	(0.48)	(0.47)	(0.48)
Percent Children Previously Attended Center-Based Care	-0.02	-0.02	0.10	0.09
	(0.33)	(0.33)	(0.33)	(0.33)
Average Measure of Home Literacy Behavior	0.08	0.08	0.07	0.07
	(0.06)	(0.06)	(0.07)	(0.07)
Average Measure of Parental Warmth	-0.13	-0.12	-0.73*	-0.74*
	(0.35)	(0.35)	(0.35)	(0.36)
Average Measure of Maternal Depression	0.05*	0.05*	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
N	292	292	292	292

NOTE: ** p<0.01, * p<0.05; t p<0.10; standard errors in parentheses. Analyses limited to classrooms with at least 4 fall test scores.

CONCLUSION

Policy Implications

Understanding the role that peer effects play in early academic and behavioral skill development, and in explaining variation across early education programs, is important for several reasons. Substantive gaps in pre-academic and behavior skills are evident upon kindergarten entry, and bode poorly for disadvantaged children's later academic success (Magnuson & Duncan, 2005). Experts agree that early childhood education has great potential to improve children's school readiness and subsequent educational outcomes; yet, many evaluations of early childhood education frame each program as a "black box" treatment either experienced or not experienced by children, and do little to explore the mechanisms by which outcomes are attained (Magnuson & Shager, 2010).

This study attempts to shed light inside that black box, helping to identify classroom processes that may be amenable to intervention; for example, via more optimal grouping of children, specialized curricula, or professional development for teachers. In addition, the failure to identify the contribution of peer effects in previous studies may have led to biased estimates of the impact of general early education program quality or other structural quality indicators, such as teacher education; thus, pointing to incorrect policy solutions.

The findings from this dissertation add to the growing evidence suggesting that, indeed, preschool peers may influence children's pre-academic and behavioral skills development, as well as preschool classroom process quality. In the pre-academic domain, I find robust positive associations between peer and individual reading skills, as well as a suggestion of a weaker positive link between peer and individual vocabulary skills. In the behavior domain, robust positive associations exist between productive peer and individual behaviors (specifically, for

learning behaviors and pro-social skills). I also find a negative association between the level of classroom behavior problems and classroom process quality, specifically in terms of teacher/child interactions.

Taken together, these findings present a number of implications for early education policy and practice. For example, knowing that lower-skilled children may learn more from opportunities to interact with higher-skilled children suggests the need to consider such tradeoffs involved in choosing between targeted or universal preschool programs, which are likely to differ in terms of student heterogeneity and average baseline skills.

As far back as the 1960s, Ed Zigler, one of the developers of Head Start, questioned the wisdom of creating a preschool program exclusively for poor children:

> "The civil rights movement of the time was trying to end racial segregation. I thought it was equally inappropriate to segregate children by socioeconomic status. I argued that Head Start centers should include middle-class children. Not only would heterogeneous enrollment be ethical, but social modeling theory espoused by scholars such as Bandura and others suggested that poor and middle-class children could learn important skills from one another" (Zigler & Styfco, 2010, p.43).

Zigler proposed that one-third of Head Start slots be made available to middle-class children; however, other developers opposed the idea. A compromise was reached, allowing 10 percent of slots to be filled by children from families with income above the poverty level, which is still in place today, although more as a statement than a regular enrollment practice (Zigler & Styfco, 2010).[20]

The present findings, however, do not necessarily automatically recommend such heterogeneous grouping. After all, introducing lesser skilled children into higher achieving classrooms will lower the classroom average skill level. Although the alternative specifications

[20] Recent legislation also gives programs some discretion to serve children near the federal poverty level (USDHHS, ACF, OHS, 2008).

in this dissertation suggesting that minimum scores or having a higher percentage of lower skilled peers are not necessarily driving associations, careful consideration of what "optimal grouping" really looks like in early education is needed. Questions remain regarding for which groups of children we should "weight" the findings most. In other words, is it enough to say that if policies that would improve the skills of peer groups for disadvantaged children at least "do no harm" to advantaged children, are such policies worth enacting? Or must results be "positive" for both groups of children?

Other factors make thoughtful preschool grouping challenging. Lack of assessment prior to and during preschool, and, thus, a lack of information about children's incoming skills, presents one formidable challenge. Furthermore, particularly for low-income parents, early education and childcare program choices are often based on factors like convenience and affordability, thus limiting flexibility to alter the demographic make-up of local classrooms (Henly & Lyons, 2000). Additionally, it is less clear that applying a universal approach to preschool enrollment would impact behavior skill distribution, given that the "gap" between the behavior skills of children from high- and low-SES backgrounds is smaller than the gap for academic skills (Lee & Burkam, 2000).

Even if program composition is largely unalterable, however, knowing that peers contribute to preschoolers' learning might recommend teacher training regarding the use of proven collaborative learning techniques that maximize the benefits of peer effects within a given classroom (Wilkinson, Parr, Fung, Hattie, & Townsend, 2002). In addition, knowing that higher levels of problem behavior negatively affect students' learning would suggest the need for behavior management training for early childhood education teachers. Current research suggesting high levels of emotional and behavioral problems among preschoolers (Campbell,

1995), but low levels of professional development and educational attainment among early educators, suggests that preschool teachers may be particularly ill-prepared in this regard (Morris et al., 2010; Raver et al., 2008). The findings may also recommend interventions with parents to help raise skill levels of children upon preschool entry. Evidence-based home visiting presents one such promising approach (Paulsell et al., 2010). Finally, the fact that children may contribute to process quality should give states pause when potentially applying such measures to high stakes policies such as child care subsidy reimbursement in their Quality Rating and Improvement Systems (QRISs).

Limitations and Directions for Future Research

Although the rich FACES data and rigorous methodology employed in this research present improvements over previous studies of preschool peer effects, a number of limitations remain. First and foremost, these are non-experimental analyses, and any resulting associations may not be causal in nature. Ideally, one would conduct an experiment in which children were randomly assigned to preschool classrooms in which the only variation was initial peer skill levels. Although this type of experiment is highly unlikely in the current preschool policy context, an ideal research situation would include purposeful randomization of children into classrooms within a universal preschool program in a mixed income neighborhood. Full classroom information would allay concerns regarding missing data.

A second important limitation to this study is that the models assume a homogeneous treatment effect; in other words, that each student has an equal effect on each other student. This assumption has been refuted by some researchers (See Harris, 2010; Hoxby & Weingarth, 2005), who argue that evidence from studies with older children suggests that certain classmates (e.g.,

those of the same race or with higher initial levels of achievement) may be more influential than others. Theoretically, this concern is at least partially mitigated by the structure of Head Start pedagogy, in which teachers may be more controlling of peer groupings, and there is an emphasis on inclusion and interaction with all fellow students (USDHHS, ACF, Office of Head Start, 2009). Furthermore, there is generally less stability in preschool friendships compared to the friendships of older children (Rubin et al., 2006). Developmental research does suggest, however, that preschool-aged children appear to be more attracted to peers who are similar; for example, in terms of age, sex, or behavioral tendencies (Rubin et al., 2006). Unfortunately, given the sample size in each classroom, such hypotheses are not testable using the FACES data.

While the above discussion presents a benefit of using Head Start data, the external validity of study findings employing such a sample is questionable. On the one hand, homogeneity in population and programming reduces concerns presented by selection bias, in that program quality, a potential omitted variable, is likely to be relatively more similar. On the other hand, it means that variation in outcomes and peer skill levels are decreased, particularly at the high end of the skill distribution; thus, making it more difficult to detect peer effects. The fact that the FACES data include only first-year Head Start students, and therefore exclude measures of students participating in their second year of the program may also attenuate peer effect estimates, assuming Head Start is an effective intervention that increases skill levels. In other words, the current findings may present a "lower bound" for preschool peer effects.

Given the nascent nature of research in this area, it will be important to track efforts to replicate such studies in different preschool settings, using different outcome measures and peer skill specifications. Additionally, qualitative research and observational data detailing exchanges between peers, as well as the ways in which teachers help structure peer interactions and allocate

resources based on their assessments of children's skills would help explain the mechanisms underlying peer effects. Multilevel models using cross-level interactions to examine whether peer effects operate differently in classes with different characteristics might also be illuminating. Also important to investigate are questions regarding whether preschool peer effects persist, and whether effects are cumulative over time. If such research can help inform early education policy and practice, these lines of inquiry are worth pursuing.

REFERENCES

Achenbach, T. M., Edelbrock, C., & 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.

Alexander, K. L., & 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.

Allison, P. (2002). *Missing data.* Thousand Oaks, CA: Sage Publications.

Anthony, B.J., Anthony, L.G., Morrel, T.M. & Acosta, M. (2005). Evidence for social and behavior problems in low-income, urban preschoolers: Effects of site, classroom, and teacher. *Journal of Youth and Adolescence, 34*, 31-39.

Arnett, J. (1989). Caregivers in daycare centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.

Bainbridge, J., Meyers, M.K., Tanaka, S. & Waldfogel, J. (2005). Who gets an early education?: Family income and the enrollment of three- to five-year-olds from 1968-2000. *Social Science Quarterly, 86*, 724-745.

Barnett, W.S., & Belfield, C.R. (2006). Early childhood development and social mobility. *Future of Children, 16*(2), 73-98.

Barnett, W.S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). *The state of preschool: 2011 preschool yearbook.* The National Institute for Early Education Research.

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Belsky, J., Burchinal, M., McCartney, K., Vandell, D. L., Clarke-Stewart, K. A., & Owen, M. T. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681-701.

Benenson, J. F., Markovits, H., Roy, R. & Denko, P. (2003). Behavioural rules underlying learning to share: Effects of development and context. *International Journal of Behavioural Development, 27*, 116-121.

Berk, L.E. (2000). *Child development, 5th edition.* Needham Heights, MA: Allyn & Bacon.

Blau, D. M. (1997). The production of quality in child-care centers. *Journal of Human Resources, 32*(2), 354-387.

Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science, 4*(3), 136-148.

Bound, J., Brown, C., Duncan, G., & Rodgers, W. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics, 12,* 345 – 368.

Bradley, R.H., Caldwell, B.M., Rock, S.L., Ramey, C.T., Barnard, K.E., Gray, C., et al. (1989). Home environment and cognitive development in the first three years: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25,* 217-235.

Braza, F., Azurmendi, A., Munoz, J. M., Carreras, M. R., Braza, P., Garcia, A., Sorozabal, A., & Sanchez-Martin, J. R. (2009). Social cognitive predictors of peer acceptance at age 5 and the moderating effects of gender. *British Journal of Developmental Psychology, 27,* 703-716.

Brock, W.A. & Durlauf, S.N. (2001). Interactions-based models. In *Handbook of Econometrics,* Heckman, J.J. and Leamer, E.E. (Eds). North-Holland: Amsterdam; 3297-3380.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge: Harvard University Press.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16,* 239-253.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes. C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*(1), 2-11.

Burchinal, M., Kainz, K., & Cai, Y. (in press). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow (Ed.) *Reasons to take stock and strengthen our measures of quality.* Baltimore, MD: Brooks Publishing.

Burchinal, M. R., Roberts, J. E., Riggins, Jr., R. Zeisel, S. A., Neebe, E., & Bryant, D. (2000).Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*(2), 338-357.

Campbell, S. (1994). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry, 36,* 113-139.

Charlesworth, R., Hart, C. H., Burts, D. C., Thomason, R. H., Mosley, J., & Fleege, P. O. (1993). Measuring the developmental appropriateness of kindergarten teachers' beliefs and practices. *Early Childhood Research Quarterly, 8, 255-276.*

Chen, D. W., Fein, G. & Tam, H. P. (2001). Peer conflicts of preschool children: Issues, resolution, incidence, and age-related patterns. *Early Education and Development, 12,* 523-544.

Chien, N. C., Howes, C., Pianta, R. C., Burchinal, M., Ritchie, S., Bryant, D. M., Clifford, R. M., Early, D. M., & Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development, 81,* 1534-1549.

Cho, R. M. (2010). Are there peer effects associated with having English Language Learner (ELL) classmates? Evidence from the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K). Unpublished manuscript.

Christian, K., Morrison, F. J., Frazier, J. A. & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development, 1*(4), 429-448.

Christina, R. & Nicholson-Goodman, J. (2005). *Going to scale with high-quality early education: Choices and consequences in universal pre-kindergarten efforts.* RAND Corporation. Available on line at http://www.rand.org/pubs/technical_reports/2005/RAND_TR237.pdf

Clarke-Stewart, K. A., Lowe Vandell, D., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly, 17*(1), 52-86).

Connor, C. M., Morrison, F. J. & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology, 98,* 665-689.

Cook, T.D. 2006. "What Works in Publicly Funded Pre-kindergarten Education?" Presentation at Institute for Policy Research at Northwestern University Policy Briefing: *Children's Achievement: What Does the Evidence Say About Teachers, Pre-K Programs, and Economic Policies?* 5 December. Chicago, IL. Available on line at http://www.northwestern.edu/ipr/events/briefingdec06-cook/slide1.html.

Coolahan, K., Fantuzzo, J., Mendez, J., & McDermott, P. (2000). Preschool peer interactions and readiness to learn: Relationships between classroom peer play and learning behaviors and conduct. *Journal of Educational Psychology, 92,* 458-465.

Cooper, C. R. (1980). Development of collaborative problem sovling among preschool children. *Developmental Psychology, 16,* 433-440.

Cunha, F. & Heckman, J. (2007). The technology of skill formation. *The Economics of Human Development,* 97(2): 31-47.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15*(2), 213-238.

Currie, J. & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review, 85,* 341-364.

Davis-Kean, P. E., Huesmann, R., Jager, J., Collins, W. A., Bates, J. E. & Lansford, J. E. (2008). Changes in the relation of self-efficacy beliefs and behaviors across development. *Child Development, 79*(5), 1257-1269.

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What it is and how do we assess it? *Early Education and Development, 17,* 57-89.

Dickinson, D.K. & Tabors, P.O. (2001). *Beginning literacy with language.* Baltimore, MD: Brooks Publishing.

Dishion, T. J., Eddy, J. M., Haas, E., Li, F., & Spracklen, K. (1997). Friendships and violent behavior during adolescence, *Social Development, 6,* 207-223.

Dodge, K. A., Coie, J. D. & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon, R. Lerner & N. Eisenberg (Eds.), *Handbook of child psychology, volume 3, social, emotional, and personality development, 6th edition,* (Ch. 12). New York: Wiley.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japeli, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6): 1428-1446.

Duncan, G.J., Ludwig, J., and Magnuson, K.A. (2007). Reducing poverty through preschool interventions. *Future of Children, 17*(2), 143-160.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test, Third Edition. Examiner's manual and norms booklet.* Circle Pines, MN: American Guidance Service.

Elliot, S. N., Gresham, F. M., Freeman, R., & 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.

Entwisle, D., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition and mathematics achievement in the first two years of school. *American Sociological Review, 57,* 72-84.

Fantuzzo, J. W., Bulotsky, R., McDermott, P., Mosca, S., & Lutz, M. N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32,* 185-203.

Fantuzzo, J. W., Bulotsky-Shearer, R., Fusco, R. A., & McWayne, C. (2005). An investigation of preschool emotional and behavioral adjustment problems and social-emotional school readiness competencies. *Early Childhood Research Quarterly, 20,* 259-275.

Fletcher, J. (2009). Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management, 29*, 69-83.

Geary, D.C. (2006). Development of mathematical understanding, In D. Kuhn and R. Siegler (Eds.), *Handbook of child psychology, volume 2.* Hoboken, NJ: John Wiley & Sons.

Geary, D.C. (1994). *Children's mathematical development.* Washington, DC: American Psychological Association.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60*, 115-140.

Gilliam, W. S. (2005). Prekindergarteners left behind: Expulsion rates in state prekindergarten systems. New Haven, CT: Yale University Child Study Center.

Gilliam, W. S., & Shahar, G. (2006). Prekindergarten expulsion and suspension: Rates and predictors in one state. Infants and Young Children, 19, 228-245.

Gilliam, W.S. & Zigler, E.F.. (2000). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly, 15*, 441-473.

Girolametto, L. & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools, 33*, 268-281.

Goncu, A. (1993). Development of intersubjectivity in the dyadic play of preschoolers. *Early Childhood Research Quarterly, 8,* 99-116.

Goncu, A., Patt, M. B. & Kouba, E. (2002). Understanding young children's pretend play in context. In P. Smith & C. Hart (Eds.), *Blackwell handbook of childhood social development* (p. 418-437). Malden, MA: Blackwell.

Gormley, W.T. Jr. 2005. "The Universal Pre-K Bandwagon." *Phi Delta Kappan.* 87(3): 246-249.

Gormley, W. T., Jr., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320,* 1723-24.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

Greenburg, M. (2007, April). Discussant comments. In C. P. Li-Grining (Chair), *Getting schools ready for children: Observing and stimulating change in early childhood classroom quality.* Symposium conducted at the meeting of the Society for Research in Child Development, Boston, MA.

Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2007). *Growth models of classroom quality over the course of the year in preschool programs.* Paper presented at the meeting of the Society for Research in Child Development, Boston, MA.

Hanish, L. D., Martin, C. L., Fabes, R. A., & Barcelo, H. (2008). The breadth of peer relationships among preschoolers: An application of the Q-connectivity method to externalizing behavior. *Child Development, 79,* 1119-1136.

Hanish, L. D., Martin, C. L., Fabes, R. A., Leonard, S., & Herzog, M. (2005). Exposure to externalizing peers in early childhood: Homophily and peer contagion processes. *Journal of Abnormal Child Psychology, 33,* 267-281.

Hanushek, E.A., Kain, J.F., Markman, J.M., & Rivkin, S.G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527-544.

Harms, T., Clifford, R., & Cryer, D. (1998). *Early childhood environment rating scale.* New York: Teachers College Press.

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record,* 112(4), 1163-1197.

Hart, B. & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes Publishing Co.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37,* 449-481.

Hay, D. F., Payne, A. & Chadwick, A. (2004). Peer relations in childhood. *Journal of Child Psychology and Psychiatry, 45,* 84-108.

Hedges, L. (2007) *How clustering affects statistical analyses: Experimental designs and robust standard errors.* Presentation to the Interdisciplinary Training Program in Education Sciences Seminar, University of Wisconsin-Madison. 27 April.

Henly, J. R., & Lyons, S. (2000). The negotiation of child care and employment demands among low-income parents. *Journal of Social Issues, 56,* 683–706.

Henry, G.T. & Rickman, D.K. (2007). Do peers influence children's skill development in preschool? *Economics of Education Review, 26,* 100-112.

Hoff-Ginsburg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development, 62,* 782-796.

Hogan, D. & Tudge, J. (1999). Implications of Vygotsky's theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39-65). Mahwah, NJ: Erlbaum.

Howes, C. (1988). Peer interaction in young children. *Monograph of the Society for Research in Child Development, 217*(53, Serial No. 1).

Howes, C. & Lee, L. (2006). Peer relations in young children. In *Child psychology: A handbook of contemporary issues, second edition,* (Eds.) L. Balter and C.S. Tamis-LeMonda (pp. 135-151). NY: Psychology Press, Taylor & Francis Group.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *Working Paper #7867*. Cambridge, MA: National Bureau of Economic Research.

Hoxby, C. M. & Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. *Unpublished manuscript*.

Huffman, L. C., Mehlinger, S. L., & Kerivan, A. S. (2000). *Off to a good start: Research on the risk factors for academic and behavioral problems at the beginning of school.* Washington, D. C.: National Institute of Mental Health.

Huttenlocher, J. Haight, W., Bryk., A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-248.

Jencks, C. & Mayer, S.E. (1990). The social consequences of growing up in a poor neighborhood. In *Inner city poverty in the United States,* (eds.) L.E. Lynn, Jr. and M.G.H. McGeary (pp. 111-186). Washington, DC: National Academy Press.

Justice, L. M., Mashburn, A., Pence, K., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech, Language, and Hearing Research, 51*(4), 1-19.

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development, 82*(6), 1768-1777.

Katz, J.R., & Snow, C.E. (2000). Language development in early childhood: The role of social interaction. In D. Cryer & T. Harms (Eds.), *Infants and toddlers in out-of-home care* (pp. 49-86). Baltimore: Brookes.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M. & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk." *Developmental Psychology, 42*(1), 59-69.

Ladd, G. W., Herald, S. L., & Kochel, K. P. (2006). School readiness: Are there social prerequisites? *Early Education and Development, 17,* 115-150.

Laursen, B. & Hartup, W. W. (1989). The dynamics of preschool children's conflicts. *Merrill-Palmer Quarterly, 35,* 281-297.

Lazear, E.P. (2001). Educational production. *Quarterly Journal of Economics, 116*, 777-803.

Lee, J. S., & Ginsburg, H. P. (2007). Preschool teachers' beliefs about appropriate early literacy and mathematics education for low- and middle-socioeconomic status children. *Early Education and Development, 18*, 111-143.

Lee, J. S., & Ginsburg, H. P. (2009). Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australasian Journal of Early Childhood, 34*, 37-45.

Lee, V. E., & Burkam, D. T. (2002). Inequality at the starting gate: Social background differences in achievement as children begin school. Washington, D.C.: Economic Policy Institute.

Lee, V.E., Loeb, S., & Lubeck, S. 1998. Contextual effects of prekindergarten classrooms for disadvantaged children on cognitive development: The case of chapter 1. *Child Development, 69*, 479-494.

Levin, E. & Rubin, K. H. (1983). Getting others to do what you wanted them to do: The development of children's requestive strategies. In K. Nelson (Ed.), *Child language* (Vol. 4, pp. 157-186). Hillsdale, NJ: Erlbaum.

Maccoby, E. E. & Jacklin, C. N. (1987). Gender segregation in childhood. In H. W. Reese (Ed.) *Advances in child development and behavior* (Vol. 20, pp. 239-287). San Diego, CA: Academic Press.

Magnuson, K., & Duncan, G. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future of Children, 15*(1), 35-54.

Magnuson, K. & Shager, H. (2010). Early education: Progress and promise for children from low-income families. *Children and Youth Services Review, 32*(9), 1186-1198.

Manski, C.F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies, 60*, 531-542.

Marcon, R.A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three cohort study. *Developmental Psychology, 35,* 358-375.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*, 431–446.

Mashburn, A. J., Justice, L. M., Downer, J. T. & Pianta, R. C. Peer effects on children's language achievement during pre-kindergarten. *Child Development, 80*(3), 686-702.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly, 18,* 206-224.

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (2000). *Preschool Learning Behaviors Scale.* Philadelphia: Edumetric and Clinical Science.

Mitchell, A. (2009). Quality rating & improvement systems as the framework for early care and education system reform. Accessed September 18, 2009, from http://www.buildinitiative.org/files/QRIS-Framework.pdf.

Moffit, R.A. (2001). Policy interventions, low-level equilibria, and social interactions. In *Social Dynamics,* S. Durlauf and P. Young, eds. Cambridge, MA: MIT Press.

Morris, P., Raver, C. C., Millenky, M., Jones, S., & Lloyd, C. M. (2010). *Making preschool more productive: How classroom management training can help teachers.* New York, New York: MDRC.

Morris-Rothschild, B., & Brassard, M. R. (2006). Teachers' conflict management styles: The role of attachment styles and classroom management efficacy. *Journal of School Psychology, 44,* 105-121.

Mueller, E. (1972). The maintenance of verbal exchanges between young children. *Child Development, 48,* 854-861.

National Center for Education Statistics (NCES). 1999. *What happens in classrooms? Instructional Practices in Elementary and Secondary Schools, 1994-1995.* Washington, DC: U.S. Department of Education.

National Institute of Child Health and Human Development (NICHD)-Early Childcare Research Network (ECCRN). (2002). Child-care, structure, process, outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*(3), 199-206.

National Institute of Child Health and Human Development Early Child Care Research Network (NICHD-ECCRN) & Duncan, G.J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74*(5), 1454-1475.

National Research Council. (2001). *Eager to learn: Educating our preschoolers.* Committee on Early Childhood Pedagogy, B. T. Bowman, M. S. Donovan, & M. S. Burns (Eds.), Commission on Behavioral and Social Sciences and Education. Washington, D. C.: National Academy Press.

National Research Council and Institute of Medicine. (2000). *From neurons to niehgborhoods: The science of early childhood development.* Committee on Integrating the Science of Early Childhood Development. J. P. Shonkoff and D. A. Phillips (Eds.). Board on Children, youth, and Families, Commission on Behavioral and Social Sciences and Education. Washington, D. C.: National Academy Press.

Neidell, M., & Waldfogel, J. (2011). Cognitive and non-cognitive peer effects in early education. *Review of Economics and Statistics, 92*.

Neuman, S.B. 2003. "From Rhetoric to Reality: The Case for High-Quality Compensatory Prekindergarten Programs." *Phi Delta Kappan.* 85(4): 286-291.

Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. *Human Development, 36,* 1-23.

Parker, J. G., Rubin, K. H., Price, J. M., & DeRosier, M. E. (1995). Peer relationships, child development, and adjustment: A developmental psychopathology perspective. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol.2. Risk, disorder, and adap*tation (pp. 96–161). New York: Wiley.

Paulsell, D., Avellar, S., Sama Martin, E., & Del Grosso, P. (2010). *Home Visiting Evidence of Effectiveness Review: Executive Summary.* Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., 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.

Pianta, R.C. (2008). *Improving impacts of classrooms: Professional development and classroom observation*. Paper presented at the Teacher Quality Conference, Northwestern University, Evanston, IL. 1 May.

Pianta, R.C., Paro, K.M.L., Payne, C., Cox, M.J. & Bradley, R. (2002). The relation of classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal, 102,* 225-238.

Puma, M., Bell, S. Cook, R., Heid, C., & Lopez, M. (2005). *Head Start impact study: First year findings.* Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, second edition.* Thousand Oaks, CA: Sage Publications.

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report, 16,* 1-20.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start Settings. *Early Childhood Research Quarterly, 23*, 10-26.

Rogoff, B. (1997). Evaluating development in the process of participation: Theory, methods and practice building on each other. In E. Amsel & K. A. Renninger (Eds.), *Change and development: Issues in change, method, and application*, (pp. 265-285). Mahwah, NJ: Erlbaum.

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal, 4*, 227-241.

Rubin, K. H., Bukowski, W. M. & Parker, J. G. (2006). Peer interactions, relationships, and groups. In W. Damon, R. Lerner & N. Eisenberg (Eds.), *Handbook of child psychology, volume 3, social, emotional, and personality development, 6th edition*, (571-645). New York: Wiley.

Rubin, K. H., Watson, K. & Jambor, T. (1978). Free play behaviors in pre-school and kindergarten children. *Child Development, 49,* 534-536.

Schechter, C. & Bye, B. (2007). Preliminary evidence for the impact of mixed-income preschools on low-income children's language growth. *Early Childhood Research Quarterly, 22*, 137-146.

Schuele, C. M., Rice, M. L. & Wilcox, K. A. (1995). Redirects: A strategy to increase peer initiations. *Journal of Speech and Hearing Research, 38,* 1319-1333.

Scrivner, S. & Wolfe, B. (2003). Universal preschool: Much to gain but who will pay? Institute for Research on Poverty (IRP) Discussion Paper no. 1271-03. University of Wisconsin-Madison. Available at: http://www.irp.wisc.edu.

Shatz, M. & Gelman, R. (1973). The development of communication skills: Modification sin the speech of young children as a function of the listener. *Monographs of the Society for Reseach in Child Development, 38* (38).

Shonkoff, J.P. & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

Singer, D., Golinkoff, R. M., & Hirsh-Pasek, K. (Eds.). (2006). *Play=learning: How play motivates and enhances children's cognitive and social-emotional growth.* New York: Oxford University Press.

Singer, D.G. & Singer, J.L. (1990). *The house of make-believe.* Cambridge, MA: Harvard University Press.

Snyder, J., Low, S., Schultz, T., Barner, S., Moreno, D., Garst, M., Leiker, R., Swink, N., & Schrepferman, L. (2011). The impact of a brief teacher training on classroom management and child behavior in at-risk preschool settings: Mediators and treatment utility. *Journal of Applied Developmental Psychology*, *32*, 336-345.

Sojourner, A. (2011). Identification of peer effects with missing peer data: Evidence from Project STAR. Institute for the Study of Labor (IZA) Discussion Paper No. 5432.

Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly, 19*, 548-568.

Stipek, D., & Byler, P. (1997). Early childhood teachers: Do they practice what they preach? *Early Childhood Research Quarterly, 12*, 305-325.

Thompson, R. A., & Raikes, H. A. (2007). The social and emotional foundations of school readiness. In D. F. Perry, R. K. Kaufman, & J. Knitzer (Eds.), *Social and emotional health in early childhood: Building bridges between services and systems* (pp. 13-36). Baltimore: Brookes.

Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn and R. Siegler (Eds.), *Handbook of child psychology, volume 2.* Hoboken, NJ: John Wiley & Sons.

Tomasello, M. & Farrar, J. (1986). Joint attention and early language. *Child Development, 57,* 1454-1463.

Tudge, J. (1992). Processes and consequences of peer collaborations: A Vygotskian analysis. *Child Development,63,* 1364-1379.

U.S. Department of Health and Human Services, Administration for Children and Families, Office of Head Start. (2008). About the Office of Head Start. Retrieved September 22, 2008, from http://www.acf.hhs.gov/programs/ohs/about/index.html#reports.

U.S. Department of Health and Human Services, Administration for Children and Families. (2011). Head Start program fact sheet. http://www.acf.hhs.gov/programs/hsb/about/fy2010.html. Accessed June 6, 2012.

United States Department of Health and Human Services, Administration for Children and Families, Office of Head Start. (2009). *Head Start program performance standards.* Retrieved May 22, 2009, from http://eclkc.ohs.acf.hhs.gov/hslc/Program%20Design%20and%20Management/Head%20Start%20Requirements/Head%20Start%20Requirements.

Vigdor, J. & T. Nechyba. (2005). Peer effects in North Carolina public schools. Manuscript. Available at http://www.ksg.harvard.edu/pepg/PDF/events/Munich/PEPG-04-20Nechyba.pdf.

Votruba-Drzal, E., Coley, R. L., Maldonado-Carreno, C., Li-Grining, C. P., & Chase-Lansdale, P. L. (2010). Child care and the development of behavior problems among economically disadvantaged children in middle childhood. *Child Development, 81,* 1460-1474.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Westat, Zill, N., Kim, K., Sorongon, A., Shapiro, G., & Herbison, R. (2008). *Head Start FACES 2003 data user's guide.* Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Wilkinson, I. A. G., Parr, J. M., Fung, I. Y. Y., Hattie, J. A. C., & Townsend, M. A. R. (2002). Discussion: Modeling and maximizing peer effects in school. International Journal of Educational Research, 37, 521-535.

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children, 5,* 51-75.

Zellman, G. L. & Perlman, M. (2008). *Child-care quality rating and improvement systems in five pioneer states: Implementation issues and lessons learned.* Santa Monica, CA: RAND Corporation.

Zigler, E., & Styfco, S. J. (2010). *The hidden history of Head Start.* New York: Oxford University Press.

Zill, N. (1990). Behavior Problems Index based on parent report. Unpublished memorandum. Washington, DC: Child Trends.