Published in final edited form as:
Child Youth Serv Rev. 2012 May 1; 34(5): 946–954. doi:10.1016/j.childyouth.2012.01.026.

Academic Performance of Subsequent Schools and Impacts of Early Interventions: Evidence from a Randomized Controlled Trial in Head Start Settings

Fuhua Zhai, Ph.D. [Assistant Professor]
SUNY Stony Brook University School of Social Welfare L2-093 Health Sciences Center Stony Brook, NY 11794

C. Cybele Raver, Ph.D. [Professor]
Applied Psychology New York University Steinhardt School of Culture, Education, and Human Development 246 Greene Street, #403W New York, NY 10003 Phone: 212-998-5519 cybele.raver@nyu.edu

Stephanie M. Jones, Ph.D. [Assistant Professor]
Harvard University Graduate School of Education Larsen 603, 14 Appian Way Cambridge, MA 02138 Phone: 617-496-2223 stephanie_m_jones@gse.harvard.edu

Abstract

The role of subsequent school contexts in the long-term effects of early childhood interventions has received increasing attention, but has been understudied in the literature. Using data from the Chicago School Readiness Project (CSRP), a cluster-randomized controlled trial conducted in Head Start programs, we investigate whether the intervention had differential effects on academic and behavioral outcomes in kindergarten if children attended high- or low-performing schools subsequent to the preschool intervention year. To address the issue of selection bias, we adopt an innovative method, principal score matching, and control for a set of child, mother, and classroom covariates. We find that exposure to the CSRP intervention in the Head Start year had significant effects on academic and behavioral outcomes in kindergarten for children who subsequently attended high-performing schools, but no significant effects on children attending low-performing schools. Policy implications of the findings are discussed.

Keywords

Head Start; randomized controlled trial; school performance; principal score matching

1. Introduction

The early school years, especially from kindergarten to third grade, are a critical transitional period not only for promoting children's scholastic and psychosocial development but also for preventing the dissipating effects of earlier interventions (Reynolds, Magnuson, & Ou, 2006, 2010). Research has suggested that the benefits of high-quality early interventions can

Fuhua Zhai, Ph.D. (Corresponding Author) Phone: 631-444-3697 Fax: 631-444-7565 fuhua.zhai@stonybrook.edu.

be sustained in later school years and even into adulthood for participants who attend continuing enrichment programs in the early school years, particularly so for those from low-income families, but that benefits tend to fade out by the second or third year of formal schooling for participants who subsequently attend inferior schools (Currie, 2001; Currie & Thomas, 1995, 2000; Lee & Loeb, 1995; Magnuson, Ruhm, & Waldfogel, 2007; Reynolds et al., 2007, 2011; Takanishi & Bogard, 2007). Therefore, in the investigation of long-term effects of early childhood interventions, it is important to take into account the role of participants' subsequent school experiences, especially the quality and performance of the schools they attend. Research as to the mechanisms of how later schools promote or hinder the initial gains fostered by early interventions is likely to have important policy implications for the design and improvement of effective interventions targeting disadvantaged children.

However, largely due to the lack of data on the quality or performance of later schools, few empirical studies have examined directly their role in the long-term effects of early childhood interventions. When the data do exist, selection bias remains an important issue. On the one hand, economically disadvantaged children such as Head Start participants tend to attend low-quality or low-performing schools compared to their counterparts, which could undermine their earlier gains (Currie & Thomas, 1995, 2000; Hastings & Weinstein, 2008; Lee & Loeb, 1995; Pigott & Israel, 2005). On the other hand, children's enrollment in higher versus lower performing schools may be endogenous to, or affected by, exposure to an initial treatment such as preschool interventions (Hong & Raudenbush, 2008). Evidence shows that enrollment in schools of different quality is associated with children's later outcomes (Hastings & Weinstein, 2008). As such, analyses that do not consider the role of subsequent schools may result in biased estimates of the long-term effects of early childhood interventions.

We use data from a cluster-randomized controlled trial conducted in Head Start programs, the Chicago School Readiness Project (CSRP), to investigate whether the academic performance of subsequent schools mattered in sustaining the effects of the CSRP intervention on children's academic and behavioral outcomes in kindergarten. As detailed below, in the analyses we employ a principal score matching method to address the issue of selection bias.

1.1. Background and Prior Research

Studies conducted over the last two decades and beyond have consistently shown that the benefits of many high-quality early interventions, especially cognitive gains from programs targeting economically disadvantaged children such as Head Start, tend to dissipate after only a few years of formal schooling (see research and reviews by Barnett, 1995; Currie & Thomas, 1995, 2000; Lee & Loeb, 1995; Magnuson et al., 2007; U.S. Department of Health and Human Services [USDHHS], 2010). A common explanation for the fade-out of initial gains from early interventions is the low-quality of schools that participants subsequently attend (Currie, 2001; Currie & Thomas, 2000; Lee & Loeb, 1995; Magnuson et al., 2007). Research has suggested that children in economically disadvantaged families who attended high-quality preschool programs such as Head Start were systematically more likely to attend low-quality and low-performing schools compared to their counterparts, which may be due to residential proximity, parental expectations of low return to education, and budget constraints in educational spending (Currie & Thomas, 1995, 2000; Hastings & Weinstein, 2008; Lee & Loeb, 1995; Pigott & Israel, 2005). For example, evidence from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), a nationally representative sample, shows that compared to other children, Head Start participants tended to attend schools with lower average levels of socio-economic status, lower achievement in math and reading, more minority children, and more children eligible for free lunch (Pigott & Israel, 2005).

From the perspectives of life cycle skill formation and human capital accumulation, as well as developmental cascades theory, continuing and enriching school environments can facilitate the ongoing skill acquisition and school achievement of children with skill advantages gained from high-quality early interventions, and can also compensate for the skill deficits experienced by at-risk children at school entry (Cunha, Heckman, Lochner, & Masterov, 2006; Hamre & Pianta, 2005; Magnuson et al., 2007; Masten & Cicchetti, 2010). As a result, high-performing schools that children attend subsequent to preschool interventions may be able to reinforce their initial gains and make early interventions more effective (Cunha et al., 2006). For example, recent results from the Child-Parent Center (CPC) Early Education Program, which included services for low-income children from age 3 to third grade, demonstrated that continuing intervention strengthened learning gains from preschool and was independently associated with school performance leading to adult well-being (Reynolds et al., 2011). In contrast, low-quality learning environments and unchallenging programs may undermine children's earlier gains and the advantages fostered by initial interventions (Currie & Thomas, 2000; Lee & Loeb, 1995).

In these and some other studies on the role of schools, school academic performance has typically been indexed by the percentage of students meeting or exceeding performance benchmarks in standardized tests of math or reading (Currie & Thomas, 2000; Hanushek, 1986; Lee & Loeb, 1995). Families as well as educators recognize the importance of test-score-based ratings of school quality with academic performance of schools serving as one of the most important factors in parents' school choices (Schneider, Teske, Marshall, & Roch, 1998; Weidner & Herrington, 2000). Evidence from experimental studies shows that parents in low-income families were more likely to choose higher-performing schools if they received direct information on school test scores (Hastings & Weinstein, 2008). Other studies have shown that when low school performance information was publicly disseminated, students, especially those from low-income neighborhoods, were more likely to switch their schools (Friesen, Javdani, Smith, & Woodcock, 2011; Howell, 2006). Moreover, high-performing schools appear to make significant differences in the experiences and outcomes of low-income children. For example, high-performing schools, especially those in low-income neighborhoods, have been associated with a more caring and nurturing environment characterized by high expectations for staff and students, effective leadership, committed teachers, and a strong focus on academics, instruction, and student learning (Carter, 2000; Kannapel & Clements, 2005; McGee, 2004; Wolf & Hoople, 2006). Experimental studies found that attending high-performing schools significantly increased low-income students' test scores (Hastings & Weinstein, 2008).

Although the role of school academic performance in the long-term effects of early interventions has received increasing attention recently, few empirical studies have been conducted to investigate their role directly. One common challenge is that the chance of attending higher- versus lower-performing schools may differ for participants and non-participants, which may, in turn, contribute to the long-term outcomes of interest. As described above, the probability of children's subsequent enrollment in higher- versus lower-performing schools may be systematically different for economically disadvantaged children compared to their counterparts, and importantly, may also be a function of the initial intervention. For example, the cognitive and social-emotional benefits that participants gain from high-quality preschool interventions may motivate their parents to pursue better schools. In the CSRP, children in the intervention group had significantly better developmental outcomes than those in the control group at the end of the 9-month intervention during the Head Start year, including improvement in self-regulation and academic skills as well as reductions in emotional and behavioral difficulties (Raver et al., 2009, 2011). In addition, preliminary evidence suggests that children in the CSRP intervention group were significantly more likely to enroll in high-performing schools in kindergarten than children in the control group, even after adjusting for child and mother covariates as well as teacher and classroom characteristics in Head Start year (Zhai & Raver, 2010). Taken together, these results suggest that we run the risk of incorporating considerable bias into our estimates of the long-term effects of the initial intervention unless we also take the performance of those subsequent schools into account in our models. The present study therefore takes careful steps to minimize the risk of this bias, as we outline below.

1.2. The Present Study

In this study, we use data from the CSRP to investigate whether exposure to the intervention during the Head Start year had differential effects on children's academic (i.e., language, literacy, and math) and behavioral outcomes (i.e., internalizing and externalizing behavior problems) in kindergarten if children subsequently attended either high- or low-performing schools. To do this we conduct separate analyses for children who subsequently attended high-performing schools and for those who were enrolled in low-performing schools, and then examine whether the CSRP intervention effects were different between these two groups of children.

Building on prior research (Currie & Thomas, 2000; Hanushek, 1986; Lee & Loeb, 1995), we define low- and high-performing schools based on school-level aggregates of students' standardized test scores. As detailed below, we adopt a principal score matching method to address the issue of selection bias and control for a set of child, mother, and classroom covariates.

2. Method

2.1. Procedure and Participants

The CSRP used a clustered randomized controlled trial (RCT) design and a pairwise matching procedure (Bloom, 2005). Two cohorts of children and teachers from 18 Head Start sites in seven of the most economically disadvantaged neighborhoods of Chicago participated in the CSRP intervention. Cohort One (from 10 Head Start sites) participated from fall to spring in 2004–05 and Cohort Two (from 8 Head Start sites) participated from fall to spring in 2005–06. Nine pairs of matched sites were first identified based on a range of site-level demographic characteristics that were collected by each site and reported annually to the federal government. One site in each matched pair was then randomly assigned to the intervention group and the other to the control group. Two classrooms were randomly selected from each site. After the randomized assignment, one classroom in the control group from Cohort Two left the study due to Head Start funding cuts. The original design and methods have subjected to rigorous review and are described in detail in previous studies (e.g., Raver et al., 2009, 2011).

Overall, a total of 602 children and 90 teachers in 35 classrooms from 18 Head Start sites participated in the CSRP. Children in the CSRP, on average, were 4 years old and about half were boys. Approximately 66% of participating children were non-Hispanic Black, 26% were Hispanic, and 8% were from other racial or ethnic groups. Teachers, on average, were 40 years old and almost all (i.e., 97%) were female. About 70% of teachers were non-Hispanic Black, 20% were Hispanic, and 10% were non-Hispanic White.

As detailed below, we use the school records of the Illinois Standards Achievement Test (ISAT) from the Chicago Public Schools (CPS) to define school performance. Only 60% of children in the original CSRP sample (n = 361) attended schools with available ISAT data and are therefore included in the analysis. For a variety of reasons (e.g., being too young to attend elementary schools, attending private schools, or other reasons), the remaining children (n = 241) were not in schools with available ISAT data and thus are not included in the analysis. As shown in Appendix Table 1, while there did not appear to be differential exclusion from the analysis group by CSRP intervention status (i.e., roughly equal numbers of children were excluded from both groups), there was a fair amount of variation in the type of differences between the included and excluded groups by intervention status. For example, in the CSRP intervention group, compared to children included in the analysis, excluded children were less likely to be non-Hispanic Black, to have low educated mothers, to have mothers working 10 hours or less per week, or to have family income below 50% poverty line; they attended Head Start programs with higher quality but had lower cognitive skills at baseline. In the CSRP control group, compared to children included in the analysis, excluded children were younger and attended Head Start programs with higher scores of emotional climate, but had lower cognitive skills and more behavior problems at baseline.

Therefore, while statistically significant differences exist between children included in and those excluded from the analysis sample, these differences vary and their impacts on the estimate of CSRP intervention effects given the academic performance of subsequent schools are unknown.

2.2. Intervention Services

The CSRP intervention aimed to support low-income, ethnic minority preschoolers' development of self-regulation, reduce their risk of behavioral difficulty, and increase their opportunities for learning through the improvement of teachers' positive emotional support and effective classroom management strategies (Raver et al., 2009). In particular, four components of services were provided in the intervention group. The first was a 30-hour teacher training focusing on behavior management strategies, which were adapted from the Incredible Years teacher training module (Webster-Stratton, Reid, & Hammond, 2004). All intervention-assigned teachers were invited to participate in the five 6-hour training sessions held on Saturdays from September to March during the Head Start year. Paired with the training, the second component was the placement of mental health consultants (MHCs) in intervention classrooms. MHCs attended classes one morning per week to coach teachers in implementing the behavior management strategies as well as to assist teachers with stress reduction. The third component was the stress reduction workshops provided by MHCs. In the second 10 weeks of the intervention, MHCs held a one-day stress reduction workshop for each Head Start site to help teachers relieve work-related stress, reduce burnout, and strengthen their ability and confidence both in managing children's disruptive behaviors and in meeting children's needs. The fourth component included individual mental health consultation services for a small number of children (3–4 children per class) with high emotional and behavioral problems from March to May in the Head Start year.

To ensure that the child-staff ratio was similar across intervention and control classrooms, teachers in the control group were given staffing support by a teacher's aide who only provided additional staffing support during everyday classroom activities for the same amount of time per week as the MHCs in the intervention group.

2.3. Measures

2.3.1. School performance measures—Following from prior research (Currie & Thomas, 2000; Hanushek, 1986; Lee & Loeb, 1995), we use school-level students' overall performance on Illinois standardized tests, the ISAT, as an indicator of school academic performance. The ISAT is designed to measure the achievement of individual students from

3^{rd} to 8^{th} grade relative to the Illinois Learning Standards. Overall, the schools that children in the CSRP attended tended to have slightly lower academic performance than the CPS (e.g., 56% of students meeting or exceeding ISAT reading vs. 61% in the CPS; 63% of students meeting or exceeding ISAT math vs. 69% in the CPS). The schools that CSRP participants attended also had more low-income students than the CPS, overall (e.g., 92% of students eligible for free or reduced-price lunch vs. 83% in the CPS). That said, the range in the academic performance and poverty level of schools in which CSRP students were enrolled was remarkably wide (with about 5% of CSRP students attending schools with more than 75% of students meeting or exceeding ISAT reading and more than 85% of students meeting or exceeding ISAT math, for example). This heterogeneity in school enrollment patterns allowed us to draw the comparisons outlined below.

Based on school-level reports of the percentage of students meeting or exceeding state standards for math or reading from the CPS, we categorize schools into a high- or a low-performing group. High-performing schools are defined as those schools whose percentage of students meeting or exceeding state standards for math or reading was beyond 0.5 standard deviations above the mean in the distribution of schools in the sample (i.e., 62% or more students meeting or exceeding state standards, or about the top quartile of the distribution). Low-performing schools were in the bottom of the distribution, in which the percentage of students meeting or exceeding state standards for math or reading was 0.5 standard deviations below the mean (i.e., 48% or fewer students meeting or exceeding state standards, or approximately the bottom one third of the distribution).

The comparison between the top high-performing schools and the bottom low-performing schools may help tease out the roles of school performance in the investigation of CSRP effects; while schools in the middle of the distribution are more similar in academic performance and thus may not show different impacts on CSRP participants. Descriptive statistics show that CSRP children in low-performing schools tend to be more disadvantaged than those in high-performing schools (e.g., 53% with their mother working 10 hours or less per week compared to 37% of children in high-performing schools, and 56% with family income-to-needs ratios at less than half the federal poverty threshold compared to 39% of children in high-performing schools). This highlights the value of including extensive baseline covariates in the analyses outlined below.

2.3.2. Outcome variables—We focus on CSRP intervention effects on children's academic (i.e., teacher-reported scores in language, literacy, and math) and behavioral outcomes (i.e., teacher-reported internalizing and externalizing behavior problems) in kindergarten. Teachers completed the surveys between November and February in the kindergarten year.

Academic outcomes were measured using a modified version of the Academic Rating Scale (ARS; Rock, Pollack, & Hausken, 2002). The ARS was designed to indirectly assess the process and products of children's learning in school and is meant to be a supplement to direct measures of cognitive outcomes. The ARS is targeted to a specific grade level and items contain explicit objective elements and subjective elements that would correspond to that grade level. Teachers compare the target child to their same age peers on a 1–5 scale (i.e., not yet, beginning, in progress, intermediate, and proficient). We use two aggregated measures as academic outcome variables in kindergarten, including language and literacy (12 items; $\alpha = 0.95$) and mathematical thinking (8 items; $\alpha = 0.95$).

Children's behavioral outcomes in kindergarten were measured using teachers' reports on the Caregiver-Teacher Report Form (C-TRF; Achenbach & Rescorla, 2001). The measure consists of 100 items asking the respondent to rate the child on a scale from 0 to 2 (where 0

= not true, 1 = somewhat or sometimes true, and 2 = very true or often true). Responses are summed into Internalizing (32 items; α = 0.89) and Externalizing (34 items; α = 0.96) subscales.

2.3.3. Baseline covariates—The covariates in the analysis include child and mother characteristics as well as classroom quality-related indicators collected at baseline in the fall of the Head Start year (i.e., pre-intervention). Specifically, child demographics include child gender, age, race/ethnicity (i.e., non-Hispanic Black or not), and pre-intervention scores in academic and behavioral development. The CSRP did not include pre-intervention academic and behavioral data using the same measures employed in the kindergarten year (i.e., language and literacy, mathematical thinking, and C-TRF Internalizing and Externalizing Behavior Problems). We use data on academic and behavioral development collected at baseline with similar measures to control for pre-intervention status for the respective outcome variables. In particular, for the outcome of language and literacy, we control for children's pre-intervention scores on the Peabody Picture Vocabulary Test (PPVT-III). As a 24-item scale (α = 0.78), the PPVT-III was assessed by asking children to identify one out of four pictures that corresponded to the word or action spoken by the assessor (Dunn & Dunn, 1997; Zill, 2003a). For mathematical thinking, we adopt scores in the Early Math Skills at baseline, which consists of 19 items (α = 0.82) that cover basic addition and subtraction (Zill, 2003b). For C-TRF outcome measures, we control for pre-intervention scores on the Behavior Problems Index (BPI) reported by teachers. The BPI was adapted from a 28-item rating scale originally designed for parent report of child behavior problems (Zill, 1990). Following recommendations from the National Longitudinal Survey of Youth (NLSY; Zill, 1990), the items are summed to form two domains: Internalizing (α = 0.80) and Externalizing (α = 0.92).

Mother characteristics include mother's marital status (i.e., whether mother was married at baseline) and family poverty-related risks, including mother holding less than high school diploma, mother working 10 hours or less per week, and family income-to-needs ratio at less than half the federal poverty threshold. Many preventive interventions targeting low-income children have demonstrated the importance of accounting for the potential confounders of child and family demographic characteristics (see for example, Aber, Brown, & Jones, 2003; Schaeffer et al., 2006; Tolan, Gorman-Smith, & Henry, 2004). As shown in previous analyses with large, nationally representative datasets, these factors represent the most reduced and informative set of indicators of families' exposure to deep poverty (Raver, Garner, & Smith-Donald, 2007).

Finally, research suggests that interventions in settings with different institutional resources may work differently (Gottfredson, Jones, & Gore, 2002). Thus, we include a set of classroom covariates at baseline as proxies to represent classroom quality and environment. Classroom quality was assessed using the Early Childhood Environment Rating Scale-R (ECERS-R; Harms, Clifford, & Cryer, 2003) and the Classroom Assessment Scoring System (CLASS; La Paro, Pianta, & Stuhlman, 2004). Based on 43 items, the ECERS-R is a widely used research tool that measures early childhood classroom quality across a wide range of constructs. We use the total ECRS-R scores as an indicator of classroom overall quality in the fall of Head Start year. The CLASS indicators include 7-point Likert scores on classroom emotional climate and teacher behavior management skills. In addition, based on the number of children and adults observed in classrooms during the CSRP data collection, we also include student-staff ratio to control for the potential confounding of differences in class size or staffing ratios.

2.3.4. Missing data—In the analysis sample a small number of children had missing values on the outcome measures and child and mother covariates. Although not completely overlapping, about 2–8% of the children had missing data on mother covariates and 7–18% on pre-intervention and kindergarten outcome measures. Complete case analyses on all valid outcome measures and covariates would reduce the sample to 258 children, a reduction of more than one quarter of the analysis sample (n = 361). Since our analysis sample was already relatively small, such a substantial reduction in size would further limit the statistical power of the analyses and make it harder to conduct principal score matching (Hill, Reiter, & Zanutto, 2004; Little & Rubin, 1987).

We adopt a multiple imputation (MI) method to address the issue of missing data. MI uses multiple predictions for each missing value of certain variables, based on other observed variables, to account for the uncertainty in imputed values (Guo & Fraser, 2009; Hill et al., 2004; Hill, Waldfogel, & Brooks-Gunn, 2002; Rubin, 1987; Schafer, 1997). We include child, mother, and classroom covariates as well as the fixed effects of the original paired Head Start sites in MI to account for the heterogeneity existed between paired Head Start sites. We also use a bootstrap method, which estimates regression coefficients in a bootstrap sample of the non-missing observations and thus has the advantage of robustness (Royston, 2005; van Buuren, Boshuizen, & Knook, 1999). We generate five sets of imputations for missing data and perform principal score matching with each dataset separately (Hill et al., 2004; Little & Rubin, 1987; Rubin, 1987; Schafer, 1997). With five imputed datasets, the expected relative efficiency (RE) for recovering missing values ranges from 98.2% to 99.5% (Rubin, 1987). After obtaining the estimates separately from the analyses of five imputed datasets, we use their means as the final estimates and their standard errors are obtained using Rubin's (1987) rules for combining MI.

2.4. Analytic Strategies

To address the issue of selection bias, as discussed above, we adopt a principal score matching method that considers selection into post-intervention treatments, such as enrolling in high-performing schools, as well as their impacts on the effects of the CSRP intervention on kindergarten outcomes. Principal score matching is a derivative of propensity score matching that builds on recent methodological innovations in principal stratification and subgroup analysis in the context of randomized experiments (Barnard et al., 2003; Frangakis & Rubin, 2002; Gibson, 2003; Hill et al., 2002; Peck, 2003; Zhai et al., 2010). In the analysis, we conduct three stages of principal score matching to identify children in the control group who would have similar probabilities of attending high- or low-performing schools in kindergarten to their peers in the CSRP intervention group had they been assigned to that group initially.

Specifically, in the first stage, we estimate the propensities of children in the CSRP intervention group to attend high- or low-performing schools based on their pre-intervention characteristics as well as a set of mother and classroom covariates, using the logit model specified in Equation (1):

$$\text{logit}\,(Q_i) = \beta_0 + \beta_1 X_i \tag{1}$$

where X_i represents the pre-intervention child, mother, and classroom covariates of child i that possibly influenced his/her propensity of attending high- or low-performing schools in kindergarten. To account for the pairwise matching in the CSRP original design, we also control for the fixed effects of the matched pairs of Head Start sites. After the logit model is applied in the CSRP intervention group, we apply the parameters obtained from Equation (1) to children in the control group to estimate their probabilities of attending high- or low-performing schools if they had been assigned to the intervention group. These probabilities are referred to as principal scores since they are used to stratify the population into mutually exclusive subgroups (i.e., principal strata) based on pre-intervention variables (Frangakis & Rubin, 2002; Hill et al., 2002).

In the second stage, we use the principal scores resulting from the first stage to match children who attend high- or low-performing schools to those in the control group who have similar principal scores, using a one-to-one nearest neighbor matching method with replacement. Matching with replacement can minimize biases in estimates since it allows each treatment unit to be matched with the nearest control unit. Thus it produces higher match quality and is less sensitive to the order of units than matching without replacement (Abadie & Imbens, 2002, 2006; Dehejia & Wahba, 2002; Gibson, 2003; Guo & Fraser, 2009). The RCT design of the CSRP ensured that children in the intervention and control groups overall had similar characteristics. As a result, children in the control group that had similar characteristics to those in the intervention group who attended high- or low-performing schools did exist, which makes it possible to find “matches” in the control group for intervention-assigned children with similar principal scores (Hill et al., 2002; Peck, 2003; Zhai et al., 2010).

Finally in the third stage, we use regression-adjusted differences to estimate the CSRP intervention effects on children's academic and behavioral outcomes after adjusting the performance of kindergarten schools. Specifically, we conduct ordinary least squares (OLS) regressions in the sample of matched children, controlling child, mother, and classroom pre-intervention covariates as presented in Equation (2):

$$O_i=\beta_0+\beta_1 Q_i+\beta_2 X_i+\xi_i \qquad (2)$$

where O_i represents the outcome (i.e., language and literacy, mathematical thinking, and C-TRF Internalizing and Externalizing Behavior Problems) of child i; Q_i stands for a binary variable of school performance (1 = high or low performance; 0 = matched control); denotes child, mother, and classroom characteristics; and ξ_{ijk} is a random error term. Huber-White robust standard errors are adopted to account for the cluster feature of children nested in paired Head Start sites based on the original CSRP design. Since children were matched with replacement in the second stage, we also apply weights to Equation (2), which are calculated as the number of times that matched control units are used (Dehejia & Wahba, 1999, 2002; Hill et al., 2004).

The process of estimating principal scores and matching are performed separately for low- and high-performing schools and for individual outcome variables to control for their corresponding pre-intervention scores.

3. Results

3.1. Descriptive Statistics

Table 1 shows the descriptive statistics (i.e., means) of covariates in the full analysis sample (n = 361) and subsamples by CSRP intervention condition and matching status. It also presents the statistical significance levels from *t*-tests of the mean differences between high- and low-performing schools, respectively, in Column (a) “Intervention” and Column (b) “Full Control” (i.e., control group before matching) with statistical significance levels indicated in Column (b), as well as the mean differences between Column (a) “Intervention” and Column (c) “Matched Control” (i.e., control groups after matching) with statistical significance levels, if any, indicated in Column (c).

As shown in Table 1, among children who attended high-performing schools subsequent to the Head Start year, those in the control group tended to be older, were less likely to be non-

Hispanic Black, and were more likely to be from high-performing classrooms and to have fewer behavior problems at baseline. Similar differences existed between children in the CSRP intervention and control groups who subsequently attended low-performing schools. In addition, compared to those in the intervention group, the mothers of children in the control group who attended low-performing schools were more likely to be married and were less likely to work 10 hours or less per week or to be in poverty.

Table 1 also shows that after principal score matching, there were almost no remaining statistically significant differences between the (c) "Matched Control" and (a) "Intervention" groups. Only two variables at the classroom level still had marginally significant mean differences (at $p < 0.10$) after matching: teacher behavior management skills for high-performing schools and emotional climate for low-performing schools. However, the balance in these two variables has been improved after matching. Therefore, the principal score matching approach employed in our study overall identifies comparable control groups for the high- and low- performing schools based on the observed covariates included in the models. As a result, the analyses employing the matched samples are likely to reduce selection bias on the observed covariates in estimating the CSRP intervention effects.

3.2. CSRP Effects on Children Attending High- and Low-performing Schools

Table 2 presents a summary of the results of the CSRP intervention effects that combined the estimates from the analyses of five datasets generated by MI on children who attended high- and low-performing schools subsequent to the Head Start year. Overall, we find that the CSRP intervention showed significant effects on children who subsequently attended high-performing schools but not on children who attended low-performing schools.

Specifically, children in the CSRP intervention group who attended high-performing schools in kindergarten tended to be 0.58 points higher (i.e., effect size of 0.53 standard deviations [SDs]) in language and literacy scores than children in the matched control group. The CSRP intervention did not show statistically significant effects on mathematical thinking scores of children attending high-performing schools in kindergarten. With regard to the behavioral outcomes, compared to children in the matched CSRP control group, children in the intervention group who attended high-performing schools tended to have C-TRF Internalizing Behavior Problems ratings that were 2.99 points lower (i.e., effect size of −0.45 SDs) and C-TRF Externalizing Behavioral Problems ratings that were 5.36 points lower (i.e., effect size of −0.44 SDs).

In contrast, the CSRP intervention did not show statistically significant effects on the academic or behavioral development of children who subsequently attended low-performing schools. The signs of the coefficients would suggest that children in the CSRP intervention group tended to have *lower* scores in language and literacy as well as mathematical thinking and *higher* scores in C-TRF Internalizing and Externalizing Behavior Problems than the CSRP control-assigned children who also attended low-performing schools in kindergarten. However, none of the coefficients are statistically different from zero and thus the evidence is only suggestive.

4. Discussion

Using a subsample from the original CSRP cluster-randomized evaluation and a principal score matching method to address the issue of selection bias, we find that the CSRP intervention in the Head Start year had significant effects on academic and behavioral outcomes in kindergarten for children who subsequently attended high-performing schools, but no effects on children attending low-performing schools.

How can these different sets of findings be reconciled? Children who were enrolled in CSRP intervention-assigned Head Start classrooms were exposed to a 9-month preschool intervention where the emotional climate and quality of instruction were substantially improved (Raver et al., 2008). Moreover, consistent with the intervention theory, children in these classrooms showed positive gains in self-regulation, behavioral outcomes, and pre-academic skills at the end of the Head Start year (both literacy and math; Raver et al., 2009, 2011). Among children in the CSRP intervention-assigned group, there were no statistically significant differences at the end of the Head Start year between those who later attended low-performing schools and those who later entered high-performing schools. Nevertheless, the findings in this paper indicate that children from those preschool intervention-enrolled classrooms, who then transitioned to higher performing elementary schools where the instructional quality may have been sustained, appeared to have maintained their initial learning gains, when compared to demographically similar children in the control group who also transitioned to higher performing schools.

In contrast, children who had spent 9 months in emotionally and behaviorally supportive intervention-enrolled Head Start classrooms but who then subsequently transitioned to lower performing elementary schools (schools in which the percentage of students meeting or exceeding state standards for math or reading was 0.5 standard deviations below the mean) do not significantly differ from their demographically similar, control-group assigned peers on key learning outcomes. Taken together, these findings suggest that if the “treatment,” in this case, exposure to higher quality instructional practice and more emotionally positive classroom climate, is not sustained across the transition to elementary school, preschool interventions such as the CSRP may be benefiting some children, while potentially placing those children who “sort” (or are enrolled) into lower performing schools at greater risk, relative to their control-assigned peers who transitioned into similar schools. These findings have substantial implications for developmental science as well as policy research. They provide compelling evidence of processes of human capital formation (also conceptualized developmental “cascading” or “canalization”) whereby children's academic outcomes at a given point in time reflect cumulative exposure to a wide array of time-varying risks and opportunities across multiple ecological contexts (Blair & Raver, in press; Masten & Cicchetti, 2010).

These findings are particularly important as we face a period of deep economic uncertainty, of shrinking federal and state budgets for education and social programming, and careful scrutiny of practice. While the field of preschool/pre-kindergarten-based early intervention has made great strides in the last decade, indicating that high-quality early childhood experiences can make a substantial positive difference in children's behavioral and academic school readiness, there exists a relative absence of longitudinal studies describing the cumulative, multiyear impact of interventions of this type, particularly those that span developmental and ecological transitions (i.e., through the transition from preschool to elementary schools; see the body of work on the Chicago CPC for an exception, e.g., Reynolds et al., 2007, 2011). In addition, prevention scientists face mounting evidence that the positive effects of exposure to such high quality preschool experiences fade away when children enter kindergarten and are faced with a new set of social and academic challenges (e.g., Currie & Thomas, 2000; Lee & Loeb, 1995; Reynolds, Magnuson, & Ou, 2006, 2010; USDHHS, 2010).

The CSRP intervention did not provide services to classrooms and children in kindergarten. However, sophisticated methodological tools enabled us to examine the influence on children of multi-year exposure to high quality educational experiences (in Head Start and then in early elementary school) and to examine controlled variation in those experiences (e.g., two years of high-quality experiences compared to one year followed by low quality experience). The work presented here provides support for the value of investing in high quality early childhood educational programming that is *sustained* over time, increasing the likelihood that children maintain successes through challenging transition from preschool to kindergarten. The findings also suggest that, consistent with developmental-ecological theory (Bronfenbrenner & Morris, 2006; Sameroff, 2010), longer-term effects of early childhood intervention reflects a dynamic interplay of child skills and the quality of educational contexts over time. Indeed, the work underscores a view held by many (e.g., Jones & Zigler, 2002; Zigler, Gilliam, & Jones, 2006) that preschool intervention may not be the magic bullet it is frequently believed to be. Moreover, in our case, there is suggestive evidence that the gains achieved after exposure to high quality preschool experiences may quickly be lost if they are subsequently followed by exposure to educational experiences that are of low quality.

The implications of these findings are both important and sensitive. As state and federal policymakers continue to consider investments in early education, our work indicates that it is short-sighted to not simultaneously address equally important efforts to raise and sustain the quality of kindergarten and early schooling. Our work is not definitive with regard to potential tradeoffs, as there is emerging evidence that, even with short-term fade-out of early intervention effects, important long-term positive effects do emerge (e.g., Deming, 2009; Gibbs, Ludwig, & Miller, 2011). In short, our findings suggest that policy makers should weigh the tradeoffs involved in making investments in early education and ways that investments in preschool may not be worth the time and effort if they are not matched by effort to raise and sustain quality in kindergarten.

The findings in our study should be interpreted with care. The CSRP intervention was conducted among a small sample of children who attended Head Start programs located in seven very disadvantaged neighborhoods in Chicago. Moreover, the present study only uses a subsample (n = 361) of the original CSRP sample (n = 602), including only those children in kindergarten schools with available ISAT records. Children who were too young to attend elementary schools or who attended private schools are not included in the analysis. Therefore, the findings presented in this paper should be considered suggestive. They should be replicated with other samples, and should be generalized neither to the population nationwide nor to all the children who originally participated in the CSRP.

In addition, potential biases resulting from the principal score matching method adopted in our study should be noted. As a derivative of the propensity score matching approach, a principal score matching method is subject to the assumption of ignorable treatment or selection on observables, which requires that all confounding covariates related to intervention status are observed (Dehejia & Wahba, 1999, 2002; Gibson, 2003; Hill et al., 2002; Joffe & Rosenbaum, 1999; Rosenbaum & Rubin, 1983). If any important covariates were omitted in the predictive models, then group members could be mismatched and thus the estimates of CSRP intervention effects could be biased. As such, our findings should be interpreted with caution.

Acknowledgments

The project described was supported by Award Number R01HD046160 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health. The Chicago School Readiness Project is not associated with The Chicago School®, which is a trademark of The Chicago School of Professional Psychology.

Appendix

Table 1

A comparison of included and excluded samples for analysis from the CSRP

	Intervention Group		Control Group	
	Included (n = 181)	**Excluded (n = 127)**	**Included (n = 180)**	**Excluded (n = 114)**
Child Covariates				
Boy	0.52 (0.50)	0.48 (0.50)	0.46 (0.50)	0.38 (0.49)
Age	4.23 (0.84)	4.14 (0.69)	4.45 (0.80)	3.88 (0.87) **
Non-Hispanic Black	0.75 (0.43)	0.54 (0.50) **	0.63 (0.48)	0.68 (0.47)
Pre-intervention score				
PPVT-III	0.47 (0.16)	0.42 (0.14) *	0.46 (0.18)	0.38 (0.16) **
Early Math Skills	0.43 (0.18)	0.35 (0.19) **	0.42 (0.21)	0.29 (0.20) **
BPI Internalizing Problems	2.33 (2.54)	2.74 (2.40)	1.56 (2.09)	2.79 (2.74) **
BPI Externalizing Problems	6.33 (6.31)	5.94 (5.63)	4.01 (4.46)	7.65 (6.21) **
Baseline Mother Covariates				
Mother was married	0.20 (0.40)	0.14 (0.34)	0.28 (0.45)	0.20 (0.40)
Less than high school education	0.31 (0.46)	0.19 (0.39) *	0.26 (0.44)	0.24 (0.43)
Working ≤ 10 hours per week	0.48 (0.50)	0.34 (0.47) *	0.40 (0.49)	0.38 (0.49)
Income below 50% poverty line	0.52 (0.50)	0.39 (0.49) *	0.43 (0.50)	0.36 (0.48)
Baseline Classroom Covariates				
Overall quality	4.34 (0.75)	4.62 (0.70) **	4.93 (0.70)	5.04 (0.82)
Emotional climate	15.16 (2.94)	15.75 (2.62) +	16.47 (2.55)	17.15 (2.55) *
Teacher behavior management skills	4.59 (1.06)	4.57 (1.09)	5.09 (0.87)	5.27 (0.93) +
Student-staff ratio	6.95 (1.38)	6.84 (1.43)	7.38 (1.82)	7.48 (1.90)

Notes: means and standard deviations in parentheses;

** $p < 0.01$,

* $p < 0.05$,

+ $p < 0.10$ for two-tailed t-statistics testing the mean differences between the sample included in the analysis and the sample excluded from the analysis by intervention status, with significance levels, if any, indicated in the descriptive statistics of the excluded sample.

References

Abadie A, Imbens G. Simple and bias-corrected matching estimators for average treatment effects. NBER Technical Working Papers #0283. 2002:1–55.

Abadie A, Imbens G. Large sample properties of matching estimators for average treatment effects. Econometrica. 2006; 74:235–267.

Aber JL, Brown JL, Jones SM. Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. Developmental Psychology. 2003; 39:324–348. [PubMed: 12661889]

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms & profiles. University of Vermont Research Center for Children, Youth, & Families; Burlington: 2001.

Barnard J, Frangakis CE, Hill LH, Rubin DB. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. Journal of the American Statistical Association. 2003; 98:299–323.

Barnett WS. Long-term effects of early childhood programs on cognitive and school outcomes. The Future of Children. 1995; 5:25–50.

Blair C, Raver CC. Child development in the context of poverty: Experiential canalization of brain and behavior. American Psychologist. in press.

Bloom, HS. Learning more from social experiments: Evolving analytic approaches. Russell Sage; New York: 2005.

Bronfenbrenner, U.; Morris, PA. The bioecological model of human development. In: Damon, W.; Lerner, RM., editors. Handbook of child psychology, Vol. 1: Theoretical models of human development. John Wiley; New York: 2006. p. 793-828.

Carter, SC. No excuses: Lessons from 21 high-performing, high-poverty schools. Heritage Foundation; Washington, DC: 2000.

Cunha F, Heckman JJ, Lochner L, Masterov DV. Interpreting the evidence on life cycle skill formation. Handbook of the Economics of Education. 2006; 1:697–812.

Currie J. Early childhood intervention programs: What do we know? Journal of Economic Perspectives. 2001; 15:213–238.

Currie J, Thomas D. Does Head Start make a difference? American Economic Review. 1995; 85:341–364.

Currie J, Thomas D. School quality and the longer-term effects of Head Start. Journal of Human Resources. 2000; 35:755–774.

Dehejia RH, Wahba S. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American Statistical Association. 1999; 94:1053–1062.

Dehejia R, Wahba S. Propensity score matching methods for non-experimental causal studies. The Review of Economics and Statistics. 2002; 84:151–161.

Deming D. Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics. 2009; 1:111–134.

Dunn, LM.; Dunn, LM. Examiner's manual and norms booklet. 3rd Ed.. American Guidance Service, Inc; Circle Pines, MN: 1997. Peabody Picture Vocabulary Test.

Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]

Friesen, J.; Javdani, M.; Smith, J.; Woodcock, S. Does public information about school quality lead to flight from low-achieving schools?. Centre for Education Research and Policy, Simon Fraser University; Burnaby, B.C., Canada: 2011.

Gibbs, C.; Ludwig, J.; Miller, D. Does Head Start do any lasting good?. In: Bailey, M.; Danziger, S., editors. *the War on Poverty: A 50-year retrospective*. 2011. NBER Working Paper #17452, and forthcoming

Gibson CM. Privileging the participants: The importance of sub-group analysis in social welfare evaluations. American Journal of Evaluation. 2003; 24:443–469.

Gottfredson GD, Jones EM, Gore TW. Implementation and evaluation of a cognitive-behavioral intervention to prevent problem behavior in a disorganized school. Prevention Science. 2002; 3:43–56. [PubMed: 12002558]

Guo, S.; Fraser, MW. Propensity score analysis: Statistical methods and applications. Sage Publications, Inc; Thousand Oaks, CA: 2009.

Hamre BK, Pianta RC. Can instruction and emotional support in the first-grade classroom make a difference for children at risk of school failure? Child Development. 2005; 76:949–967. [PubMed: 16149994]

Hanushek E. The economics of schooling: Production and efficiency in public schools. Journal of Economic Literature. 1986; 24:1141–1177.

Harms, T.; Clifford, RM.; Cryer, D. Early childhood environment rating scale. revised edition. Teachers College Press; New York: 2003.

Hastings JS, Weinstein JM. Information, school choice, and academic achievement: Evidence from two experiments. The Quarterly Journal of Economics. 2008; 123:1373–1414.

Hill, LJ.; Reiter, JP.; Zanutto, EL. A comparison of experimental and observational data analyses. In: Gelman, A.; Meng, X., editors. Applied Bayesian modeling and causal inference from incomplete-data perspectives: An essential journey with Donald Rubin's statistical family. John Wiley & Sons. Ltd; West Sussex, UK: 2004. p. 49-60.

Hill LJ, Waldfogel J, Brooks-Gunn J. Differential effects of high-quality child care. Journal of Policy Analysis and Management. 2002; 21:601–627.

Hong G, Raudenbush SW. Causal inference for time-varying instructional treatments. Journal of Educational and Behavioral Statistics. 2008; 33:333–362.

Howell W. Switching schools? A closer look at parents' initial interest in and knowledge about the choice provisions of No Child Left Behind. Peabody Journal of Education. 2006; 81:140–179.

Joffe MM, Rosenbaum PR. Invited commentary: Propensity scores. American Journal of Epidemiology. 1999; 150:327–333. [PubMed: 10453808]

Jones SM, Zigler E. The Mozart Effect: Not learning from history. Journal of Applied Developmental Psychology. 2002; 23:355–372.

Kannapel, PJ.; Clements, SK. Inside the black box of high performing high-poverty schools. The Prichard Committee for Academic Excellence; Lexington, KY: 2005.

La Paro KM, Pianta RC, Stuhlman M. The classroom assessment scoring system: Findings from the prekindergarten year. The Elementary School Journal. 2004; 104:409–426.

Lee VE, Loeb S. Where do Head Start attendees end up? One reason why preschool effects fade out. Educational Evaluation and Policy Analysis. 1995; 17:62–82.

Little, RJA.; Rubin, DB. Statistical analysis with missing data. John Wiley & Sons; New York: 1987.

Magnuson K, Ruhm C, Waldfogel J. The persistence of preschool effects: Do subsequent classroom experiences matter? Early Childhood Research Quarterly. 2007; 22:18–38.

Masten AS, Cicchetti D. Developmental cascades [Special Issue, Parts 1 and 2]. Development and Psychopathology. 2010; 22:491–715. 717–983. [PubMed: 20576173]

McGee GW. Closing the achievement gap: Lessons from Illinois' Golden Spike high-poverty high-performing schools. Journal of Education for Students Placed at Risk. 2004; 9:97–125.

Peck LR. Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. American Journal of Evaluation. 2003; 24:157–187.

Pigott TD, Israel MS. Head Start children's transition to kindergarten: Evidence from the Early Childhood Longitudinal Study. Journal of Early Childhood Research. 2005; 3:77–104.

Raver, CC.; Garner, P.; Smith-Donald, R. The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal?. In: Pianta, RC.; Cox, MJ.; Snow, KL., editors. School readiness and the transition to kindergarten in the era of accountability. Brookes Publishing; Baltimore, MD: 2007. p. 121-147.

Raver CC, Jones SM, Li-Grining CP, Metzger M, Smallwood K, Sardin L. Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. Early Childhood Research Quarterly. 2008; 23:10–26.

Raver CC, Jones SM, Li-Grining CP, Zhai F, Bub K, Pressler E. CSRP's impact on low-income preschoolers' pre-academic skills: Self-regulation and teacher-student relationships as two mediating mechanisms. Child Development. 2011; 82:362–387. [PubMed: 21291447]

Raver CC, Jones SM, Li-Grining CP, Zhai F, Metzger MW, Solomon B. Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. Journal of Consulting and Clinical Psychology. 2009; 77:302–316. [PubMed: 19309189]

Reynolds AJ, Magnuson K, Ou S. Preschool-to-third grade programs and practices: A review of research. Children and Youth Services Review. 2010; 32:1121–1131.

Reynolds AJ, Temple JA, Ou S, Robertson DL, Mersky JP, Topitzes JW, et al. Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. Archives of Pediatrics and Adolescent Medicine. 2007; 161:730–739. [PubMed: 17679653]

Reynolds AJ, Temple JA, White BAB, Ou S, Robertson. Age 26 cost-benefit analysis of the Child-Parent Center Early Education Program. Child Development. 2011; 82:379–404. [PubMed: 21291448]

Reynolds, A.; Magnuson, K.; Ou, S. PK-3 education: Programs and practices that work in children's first decade. Foundation for Child Development; New York: 2006.

Rock, DA.; Pollack, JM.; Hausken, EG. Early childhood longitudinal study-Kindergarten class of 1998–99 (ECLS–K): Psychometric report for kindergarten through first grade. U.S. Department of Education, National Center for Education Statistics; Washington, DC: 2002.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

Royston P. Multiple imputation of missing values: Update of ICE. Stata Journal. 2005; 5:527–536.

Rubin, DB. Multiple imputation for nonresponse in surveys. John Wiley & Sons; New York: 1987.

Sameroff A. A unified theory of development: A dialectic integration of nature and nurture. Child Development. 2010; 81:6–22. [PubMed: 20331651]

Schaeffer CM, Petras H, Ialongo N, Masyn KE, Hubbard S, Poduska J, et al. A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction of young adult antisocial outcomes. Journal of Consulting and Clinical Psychology. 2006; 74:500–510. [PubMed: 16822107]

Schafer, JL. Analysis of incomplete multivariate data. Chapman & Hall; London: 1997.

Schneider M, Teske P, Marshall M, Roch C. Shopping for schools: In the land of the blind, the one-eyed parent may be enough. American Journal of Political Science. 1998; 42:769–793.

Takanishi R, Bogard KL. Effective educational programs for young children: What we need to know. Child Development Perspectives. 2007; 1:40–45.

Tolan P, Gorman-Smith D, Henry D. Supporting families in a high-risk setting: Proximal effects of the SAFE Children preventive intervention. Journal of Consulting and Clinical Psychology. 2004; 72:855–869. [PubMed: 15482043]

U.S. Department of Health and Human Services (USDHHS). Administration for Children and Families. Head Start Impact Study: Final report. Washington, DC: 2010.

van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine. 1999; 18:681–694. [PubMed: 10204197]

Webster-Stratton C, Reid MJ, Hammond M. Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. Journal of Clinical Child and Adolescent Psychology. 2004; 33:105–124. [PubMed: 15028546]

Weidner VR, Herrington CD. Are parents informed consumers: Evidence from the Florida McKay scholarship program. Peabody Journal of Education. 2000; 81:27–56.

Wolf PJ, Hoople DS. Looking inside the black box: What school factors explain voucher gains in Washington, DC? Peabody Journal of Education. 2006; 81:7–26.

Zhai, F.; Raver, CC. Can classroom-based interventions improve Head Start children's subsequent school experiences?. Paper presented at the Society for Prevention Research (SPR) 18th Annual Meeting; CO: Denver. June 2010; 2010.

Zhai F, Raver CC, Jones S, Li-Grining C, Pressler E, Gao Q. Dosage effects on school readiness: Evidence from a randomized classroom-based intervention. Social Service Review. 2010; 84:615–654. [PubMed: 21488322]

Zigler, E.; Gilliam, WS.; Jones, SM. A vision for universal preschool education. Cambridge University Press; New York: 2006.

Zill, N. Behavior problem index based on parent report: National health interview survey—Child health supplement. National Center for Health Statistics; Washington, DC: 1990.

Zill, N. Letter naming task. Westat, Inc; Rockville, MD: 2003a.

Zill, N. Early math skills Test. Westat, Inc; Rockville, MD: 2003b.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Highlights

- Using data from a cluster-randomized controlled trial in Head Start programs
- Adopting principal score matching to address the issue of selection bias
- Significant intervention effects on children attending high-performing schools
- No intervention effects on children attending low-performing schools

Table 1

Descriptive statistics by intervention condition and matching status

		High-performing Schools			Low-performing Schools		
	Full Sample (n = 361)	Intervention (a) (n = 85)	Full Control (b) (n = 50)	Matched Control (c) (n = 85)	Intervention (a) (n = 74)	Full Control (b) (n = 104)	Matched Control (c) (n = 74)
Child Covariates							
Boy	0.49 (0.50)	0.53 (0.50)	0.44 (0.50)	0.53 (0.50)	0.51 (0.50)	0.47 (0.50)	0.50 (0.50)
Age	4.34 (0.83)	4.38 (0.95)	4.85^{*} (0.71)	4.32 (0.72)	4.03 (0.76)	4.32^{*} (0.75)	4.07 (0.66)
Non-Hispanic Black	0.69 (0.46)	0.58 (0.50)	0.36^{*} (0.48)	0.56 (0.50)	0.93 (0.25)	0.73^{**} (0.45)	0.86 (0.36)
Pre-intervention score							
PPVT-III	0.46 (0.17)	0.44 (0.17)	0.48 (0.17)	0.46 (0.20)	0.44 (0.18)	0.47 (0.18)	0.42 (0.17)
Early Math Skills	0.43 (0.19)	0.41 (0.18)	0.45 (0.20)	0.43 (0.23)	0.40 (0.18)	0.45^{+} (0.22)	0.42 (0.23)
BPI Internalizing Problems	1.95 (2.35)	2.89 (2.82)	1.84^{*} (2.02)	2.18 (2.47)	2.43 (2.74)	1.44^{**} (2.21)	2.53 (2.32)
BPI Externalizing Problems	5.18 (5.59)	6.58 (5.86)	4.86^{+} (4.46)	5.21 (4.75)	6.91 (6.78)	3.26^{**} (4.30)	7.19 (4.69)
Baseline Mother Covariates							
Mother was married	0.24 (0.43)	0.26 (0.44)	0.38 (0.49)	0.30 (0.47)	0.14 (0.34)	0.30^{**} (0.46)	0.19 (0.43)
Less than high school education	0.28 (0.45)	0.34 (0.48)	0.26 (0.44)	0.28 (0.44)	0.32 (0.47)	0.25 (0.44)	0.29 (0.46)
Working ≤ 10 hours per week	0.44 (0.50)	0.38 (0.49)	0.34 (0.48)	0.39 (0.49)	0.64 (0.48)	0.45^{*} (0.50)	0.63 (0.49)
Income below 50% poverty line	0.48 (0.50)	0.42 (0.50)	0.34 (0.48)	0.42 (0.51)	0.65 (0.48)	0.49^{*} (0.50)	0.63 (0.49)
Baseline Classroom Covariates							
Overall quality	4.63 (0.78)	4.47 (0.66)	4.88^{**} (0.73)	4.62 (0.76)	4.19 (0.80)	4.97^{**} (0.64)	4.31 (0.69)
Emotional climate	15.81 (2.83)	15.91 (1.97)	17.13^{**} (2.33)	16.13 (2.71)	14.85 (3.43)	16.17^{**} (2.58)	15.58^{+} (2.67)
Teacher behavior management	4.84 (1.00)	4.70 (0.70)	5.30^{**} (0.73)	4.91^{**} (1.00)	4.64 (1.31)	4.98^{*} (0.90)	4.50 (0.91)
Student-staff ratio	7.17 (1.63)	7.13 (1.31)	7.01 (1.65)	7.13 (1.58)	6.73 (1.52)	7.60^{*} (1.86)	7.09 (1.54)

Notes: means with standard deviations in parentheses;

$^{**}p < 0.01$,

$^{*}p < 0.05$,

$^{+}p < 0.10$ for two-tailed t-statistics testing the mean differences in high- and low-performing schools, respectively, between Column (a) “Intervention” and Column (b) “Full Control” (i.e., control group before matching) with significance levels indicated in Column (b), as well as the mean differences between Column (a) “Intervention” and Column (c) “Matched Control” (i.e., control groups after matching) with significance levels, if any, indicated in Column (c).

Table 2

CSRP intervention effects on children attending high- and low-performing schools

	High-performing Schools				Low-performing Schools			
	Language and Literacy	Math	C-TRF Internalizing	C-TRF Externalizing	Language and Literacy	Math	C-TRF Internalizing	C-TRF Externalizing
Intervention	0.58** (0.24)	0.07 (0.35)	−2.99* (1.27)	−5.36** (1.66)	−0.16 (0.29)	−0.17 (0.42)	0.91 (2.26)	2.66 (3.39)
Baseline Child Covariates								
Boy	0.53 (0.38)	−0.51+ (0.30)	1.71 (2.28)	1.34 (2.38)	−0.55+ (0.30)	−0.41* (0.20)	−1.62 (2.39)	3.45 (3.20)
Age	−0.01 (0.08)	0.03 (0.03)	−0.03 (0.21)	0.05 (0.27)	−0.3 (0.03)	−0.02 (0.02)	0.10 (0.12)	0.15 (0.24)
Non-Hispanic Black	0.86* (0.38)	0.02 (0.35)	1.45 (1.26)	−3.50 (3.36)	0.50 (0.44)	0.44 (0.67)	0.09 (2.08)	−1.04 (4.89)
Pre-intervention score	2.34* (1.11)	0.39 (1.03)	0.52+ (0.29)	0.80** (0.17)	2.96* (1.32)	1.36* (0.70)	0.48+ (0.26)	0.66** (0.27)
Baseline Mother Covariates								
Mother was married	0.84+ (0.48)	0.33 (0.36)	−2.88 (2.64)	−3.77 (3.37)	0.01 (0.66)	0.14 (0.55)	1.79 (2.68)	−0.84 (3.33)
Less than high school education	0.32 (0.68)	−0.02 (0.54)	−4.09* (2.01)	−0.80 (6.41)	0.58 (0.59)	−0.12 (0.30)	1.75 (2.17)	1.20 (2.60)
Working ≤ 10 hours per week	−0.72+ (0.43)	0.44 (0.67)	3.09+ (1.65)	4.55+ (2.71)	0.25 (0.18)	0.27 (0.22)	−1.50 (1.61)	−4.54 (3.49)
Income below 50% poverty line	0.45 (0.40)	−0.05 (0.28)	−1.35 (1.59)	−0.30 (3.01)	−0.42 (0.28)	−0.36* (0.16)	1.03 (1.39)	−1.08 (3.27)
Baseline Classroom Covariates								
Overall quality	0.16 (0.33)	−0.01 (0.49)	−0.03 (1.63)	0.38 (2.64)	−0.17 (0.46)	−0.07 (0.49)	3.04+ (1.71)	1.91 (4.82)
Emotional climate	−0.08 (0.20)	0.04 (0.22)	0.95 (1.01)	1.36 (1.08)	0.09 (0.13)	0.04 (0.10)	−0.73 (0.60)	−0.98 (1.28)
Teacher behavior management skills	0.22 (0.39)	−0.26 (0.43)	−1.37 (2.26)	−2.57 (3.06)	−0.19 (0.12)	0.05 (0.09)	0.55 (0.67)	1.70 (1.94)
Student-staff ratio	0.00 (0.13)	−0.05 (0.15)	−0.33 (0.66)	−0.76 (0.99)	−0.14 (0.17)	−0.11 (0.09)	−0.43 (0.37)	−1.48* (0.73)
Constant	0.77 (4.95)	1.87 (1.81)	2.84 (2.99)	0.48 (2.98)	4.24 (2.65)	3.22* (1.30)	−4.27 (7.36)	−4.78 (4.25)

Notes: coefficients with standard errors in parentheses combined from the estimates of five datasets generated by multiple imputation;

** $p < 0.01$,

* $p < 0.05$,

+ $p < 0.10$