Early Childhood Research Quarterly 34 (2016) 1–12

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Does Head Start differentially benefit children with risks targeted by the program's service model?[☆]

Elizabeth B. Miller*, George Farkas, Greg J. Duncan

University of California, Irvine, United States

ARTICLE INFO

Article history:
Received 21 July 2014
Received in revised form 17 July 2015
Accepted 5 August 2015
Available online 20 August 2015

Keywords:
Head Start
Risk
Differential effects
Policy

ABSTRACT

Data from the Head Start Impact Study (N = 3540) were used to test for differential benefits of Head Start after one program year and after kindergarten on pre-academic and behavior outcomes for children at risk in the domains targeted by the program's comprehensive services. Although random assignment to Head Start produced positive treatment main effects on children's pre-academic skills and behavior problems, residualized growth models showed that random assignment to Head Start did not differentially benefit the pre-academic skills of children with risk factors targeted by the Head Start service model. The models showed detrimental impacts of Head Start for maternal-reported behavior problems of high-risk children, but slightly more positive impacts for teacher-reported behavior. Policy implications for Head Start are discussed.

Since Head Start's creation in 1965 as part of the War on Poverty, its mission has been to improve the school readiness of low-income children (Zigler & Styfco, 2010). According to the U.S. Department of Health and Human Services (DHHS), school readiness means that "children are ready for school, families are ready to support their children's learning, and schools are ready for children" (U.S. DHHS, Head Start Approach to School Readiness, 2011). To encourage this goal, the Head Start program uses a "whole child" model, which aims to promote children's transition to school by enhancing their development through the provision of educational, health, and nutritional services to children and families. Head Start also engages parents in their children's learning and helps parents with their own educational, literacy, and employment goals with the belief that these too are important in promoting children's preparedness for school (U.S. DHHS, Final Report, 2010b).

Over the course of its nearly fifty year history, Head Start has evolved into a comprehensive service delivery program designed to serve poor children at risk in the targeted domains of cognitive development, socio-emotional development, health, and family functioning (Zigler & Styfco, 2010). This "whole child" approach to school readiness offers multiple services to children and families with the expectation that the accurate targeting of services to needs, and the positive synergy among the services and benefits received, will act together to adequately prepare children for kindergarten.

In the 1998 reauthorization of Head Start, Congress mandated that the U.S. DHHS determine whether the program contributed to key outcomes in children's learning and development. The resulting Head Start Impact Study (HSIS) gathered data from a large, nationally representative sample of children assigned at random to Head Start centers or a comparison group between 2002 and 2006. The Final Report of the HSIS (2010a) found that at the end of the program year, Head Start significantly increased children's pre-academic skills (ES = .19–.22), reduced behavior problems (ES = −.08 to −.14), and provided children access to dental care and improved children's overall health (ES = .11–.33), compared with control group children. Although these effect sizes are small, they indicate that the program is improving children's development in a wide array of areas (U.S. DHHS, Final Report, 2010b) and they are consistent with a recent meta-analysis of 57 Head Start studies from 1965 to 2002, which found the average program-level effect size to be .27 SD (Shager et al., 2013). The HSIS report (2010a, 2010b) also provided some evidence of differential program effects among key subgroups. For example, Head Start impacts were larger for Dual Language Learners than monolingual-English speakers on a measure of receptive vocabulary.

The current study extends the examination of the differential effects of Head Start in novel ways (Barton, Spiker, & Williamson, 2012; Bloom & Weiland, 2015; McCoy, Morris, Connors, Yoshikawa,

☆ Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P01HD065704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

* Corresponding author.
E-mail address: ebmiller@uci.edu (E.B. Miller).

http://dx.doi.org/10.1016/j.ecresq.2015.08.001
0885-2006/© 2015 Elsevier Inc. All rights reserved.

& Gomez, 2015; Miller, Farkas, Vandell, & Duncan, 2014). Specifically, we test whether the program is meeting its conceptual goals of promoting school readiness outcomes for children whose risk profiles matched those targeted by the Head Start service model. Using Head Start's "whole child" approach, we first created ten conceptually based risk factors in four broad areas based on the Head Start program model: (1) children's pre-academic skills; (2) children's behavior problems; (3) children's health; and (4) family functioning. We next estimated a set of regressions in which each baseline risk factor was individually interacted with assignment to Head Start. Then we created a total risk index by summing the ten items and estimated regressions in which this total index was interacted with assignment to Head Start. Positive coefficients on these interaction variables would indicate larger program benefits for children in a given risk group compared with other children.

Since school readiness may be conceptualized as a continuum, and given that children at risk in these domains are typically the farthest away from being ready for kindergarten, it follows that these higher-risk children would stand to gain more from a program (Head Start) that explicitly targets these risk domains compared with an alternative one. That is, because Head Start strives to create an environment that improves the multiple needs of children, individuals that fit the package of services best, i.e., children at greatest risk in the targeted domains, should experience larger positive program impacts than similar children in the control condition.

Theoretical framework. The overall framework of this study draws on two complementary developmental theories. The first, bioecological theory (Bronfenbrenner & Morris, 2006), posits that human development results from the interplay of *Process* × *Person* × *Context* × *Time*. The core of the model is *Process*, which constitutes interactions between an organism and their environment known as proximal processes. The effects of these proximal processes on developmental outcomes systematically vary based on the characteristics of the person and their surrounding environmental context. Consequently, children and their families respond in varying ways to the program treatment environments they encounter (i.e., treatment effect heterogeneity). With an environment like Head Start, and the economically disadvantaged families and children it serves, bioecological theory would predict that the program will not affect all children in the same way since the fit between the child's needs and what the program provides is likely to differ across families, children, and outcomes. Therefore the match between children's characteristics, including prior experiences and needs, and the services offered by Head Start is crucial in determining whether the program is appropriate and successful for a given child, creating a source of heterogeneous treatment effects (Imbens & Angrist, 1994).

The second developmental theory is the cumulative risk model. In recent decades, a great deal of work has been conducted on risk and protective factors for infants and preschool children who grow up in adverse conditions (see Werner, 2001 for a review). Often these factors were based on a single indicator of risk measured at the family level. Examples of such risk factors include economic hardship (Egeland, Carison, & Sroufe, 1993; McLoyd, 1998); parental mental illness (Cicchetti & Toth, 1998; Seifer, Sameroff, Baldwin, & Baldwin, 1992); substance abuse (Werner, 2004); teenage motherhood (Furstenberg, Brooks-Gunn, & Morgan, 1987); and child abuse and neglect (Farber & Egeland, 1987).

Cumulative risk models account for the fact that some children are exposed to multiple dimensions of family-level risk and that the developmental outcomes for these children might be most compromised. These models traditionally include factors such as poverty, single motherhood, low levels of parental education, and unemployment, which tend to cluster within the same families (Masten et al., 1995) and may be conceptually and empirically difficult to examine individually (Burchinal, Roberts, Hooper, & Zeisel, 2000; Evans, Li, & Whipple, 2013). Studies using cumulative risk models have demonstrated that the more risk factors a child is exposed to, the more likely they are to experience a range of developmental problems, with the effects multiplicative rather than additive (Garmezy, Masten, & Tellegen, 1984; Rutter, 1987; Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987; Sameroff, Seifer, Baldwin, & Baldwin, 1993; Sameroff, 2006). The cumulative risk model complements bioecological theory as multiple risk factors may sufficiently disrupt the proximal processes between a person and their environment necessary for healthy development, as well as inhibit alternative sources of these proximal processes (Evans et al., 2013).

Despite the potential for cumulative risk factors to harm children's development, many of the above-mentioned studies emphasized how many high-risk children overcame trying circumstances to have good developmental outcomes through protective mechanisms that buffered their stressful situations. In fact, both Werner (1997) and Sameroff (2006) explicitly mentioned Head Start as a program that could serve as a protective mechanism for children at risk in a wide array of developmental areas, which is why theoretically, we might expect program benefits to be greatest for those most in need. Thus, cumulative risk models have been especially useful in studies examining potential protective factors because they provide a more comprehensive and precise representation of a child's disadvantage than examining each factor individually (Garmezy et al., 1984; Masten et al., 1995; Rutter, 1987; Seifer et al., 1992; Wright, Masten, & Narayan, 2013).

This is the theoretical approach we have taken in our study, which focuses on the premise that effective policies must fit with as many of individual children's developmental needs as possible in order to succeed. Accordingly, using individual risk factors first and then a cumulative risk index, we examine how well Head Start's "whole child" model differentially benefitted children at risk. In particular, we ask whether the program differentially benefits children at greatest risk in the domains targeted by Head Start's comprehensive services. In our empirical work, this translates into expectations of larger treatment effects for these higher-risk children relative to higher-risk children in the control condition.

Cumulative risk models and differential program impacts. There is a rich literature on differential impacts using cumulative risk models. We discuss some recent studies here, which used cumulative risk to examine differential program impacts through experimental variation. These studies, including those based on the HSIS, highlight how a particular program may have served as a protective factor for children at risk, buffering against their adversity.

The Final Report of the HSIS (U.S. DHHS, 2010a, 2010b) and the Early Head Start Research and Evaluation (EHSRE) study (Raikes, Vogel, & Love, 2013) both used a cumulative risk index comprised of five family-level items – receipt of TANF or Food Stamps; neither parent in the household had a high school diploma or a GED; neither parent in household was employed or in school; the child's biological mother or caregiver was a single parent; and the biological mother was teenaged at the child's birth – to test for differential program effects. Families were characterized according to whether they had 0–2 (no/low), 3 (moderate), or 4–5 (high) risk factors. In the HSIS, Head Start children from high-risk households experienced sustained cognitive outcomes through the end of 1st grade relative to high-risk children not offered a Head Start enrollment slot, and Head Start children from moderate-risk households experienced more positive socio-emotional impacts. There were no differential impacts for Head Start children from low-risk households compared with controls.

In the EHSRE study, program impacts were relatively weak at age five for children from homes characterized by no/low risk as well as for children from homes with high risk. Children from homes characterized by moderate risk had the strongest program impacts with positive gains on three socio-emotional outcomes compared with controls.

Additionally, Klebanov and Brooks-Gunn (2006) used data from the Infant Health and Development Program (IHDP) to form a cumulative risk index comprised of human capital risks (e.g., maternal employment and education) and psychological risks (e.g., maternal mental health and social support) to examine program impacts on cognitive outcomes at ages three, five, and eight. They found that treatment effects at age three did not depend on level of risk, whereas at ages five and eight, children with moderate levels of human capital risk, but not psychological risk, had sustained treatment effects. Treatment group children with low and high levels of human capital risk did not sustain IQ gains compared with controls.

Similarly, the evaluation of the Chicago School Readiness Project (CSRP) included an examination of moderated program impacts on behavior by a composite of poverty-related risk factors. Included items were maternal education level, maternal employment, and family income-to-needs ratio. Results showed that children with no or one risk benefitted from the CSRP intervention compared with controls, whereas the intervention effect for children with two or more risks was not statistically significant (Raver et al., 2009).

Thus, the HSIS, EHSRE, IHDP, and CSRP studies all used variations of a cumulative risk model to test for differential program impacts in the context of a random-assignment experiment. Differing effect patterns emerged, sometimes favoring children at highest risk and other times favoring children at more moderate or even low risk, possibly owing to the different definitions among these studies of what constituted risk.

Other methodological considerations. The above-mentioned studies composed their cumulative risk indices from family-level risk factors traditionally included in the literature such as poverty, single motherhood, low levels of parental education, and unemployment. These risk factors may not necessarily be malleable by the intervention services. Another increasingly popular methodological approach to creating risk indices is to use latent class analysis (LCA) to discover combinations of risk factors most commonly found among program-eligible children. A recent paper by Cooper and Lanza (2014) applied LCA to HSIS data in order to identify combinations of baseline home environment characteristics that occurred most frequently among Head Start children. The authors then tested whether the effects of Head Start differed across these subgroups. They found that for some subgroups, Head Start had positive impacts, for other subgroups no impacts, and yet for others mixed positive and negative impacts, particularly with regard to maternal versus teacher reports of behavior. Similarly, Halle, Hair, Wandner, and Chien (2012) used LCA to uncover school readiness developmental profiles along cognitive, socio-emotional and health domains for four-year olds enrolled in Head Start. They found that a substantial proportion of Head Start children moved from a developmental profile including some risk to a strengths profile, suggesting that many Head Start students are improving in their developmental status over the Head Start year. Among other factors, family structure and maternal educational attainment were associated with stability and change in profile membership. Lastly, Bulotsky-Shearer, Wen, Faria, Hahs-Vaughn, and Korfmacher (2012) also used LCA to examine associations between family involvement and Head Start classroom quality on children's cognitive and socio-emotional skills at the end of the program year. Children from homes with high levels of parental involvement and high classroom quality had the best outcomes while children with low degrees of both fared worst on outcomes.

Present study. The recent studies using LCA estimated separate Head Start effects for groups of children possessing characteristics that are most common in the program-eligible population. By contrast, the present study estimates Head Start effects for children at greatest risk of negative outcomes that are specifically targeted by the program's model. We do this by examining interactions with individual risk factors as well as with a cumulative risk index.

One of the most theoretically compelling procedures for choosing variables to include in a cumulative risk model is to consider salient proximal processes for a given developmental outcome and then to consider which risk factors, in combination, likely contribute to children's adversity (Evans et al., 2013). Thus, instead of using family-level risk factors that might not be malleable by program services or using LCA to uncover empirically correlated risk factors, we instead take a more elemental and direct approach consistent with the Head Start program's theory of change. Specifically, our study uses the proximal processes fostered by Head Start's "whole child" model and defines risk by a set of child- and family-level items (see Measures, below) matched to the school readiness domains that the program targets with its comprehensive services. Of particular importance is our examination of individual child-level risk factors such as pre-existing academic, behavioral, and health issues. These domains are targeted explicitly by Head Start and may be important in determining whether the program is successful for a given child. As a consequence, this study is to our knowledge the first to define children's risk based on the goals of the program – the academic, socio-emotional, and health needs of children as well as family functioning – as opposed to simply demographic groupings of risk.

Our principal research question is: Does Head Start differentially benefit children at greatest risk on achievement and behavioral outcomes after one academic year in the program and at the end of kindergarten compared with control group children? On the basis of the program design and the match between children's needs and the services provided by Head Start, we hypothesize a positive interaction between the Head Start treatment and the cumulative risk index, with the largest positive treatment effects for children with the greatest number of risks.

Method

Participants

We analyzed data from the HSIS, a nationally representative sample of 84 Head Start grantee and delegate agencies and nearly 5000 newly entering, eligible three and four-year-old children. Children were randomly assigned to either: (1) a Head Start group that had access to Head Start program services; or (2) a control group that was not eligible to enroll in the Head Start center to which they applied for the lottery, but could enroll in other early childhood programs or services selected by their parents, including other Head Start centers not in the study (U.S. DHHS, Final Report, 2010b).

The study employed a multi-stage sampling process to select a representative group of Head Start programs and children. It began with a list of 1715 grantee and delegate Head Start agencies that were operating in Fiscal Year (FY) 1998–99. This pool was then organized into 161 geographic clusters across 25 strata in order to ensure variation across region of the country, urban and rural location, race and ethnicity, and state pre-kindergarten and child care policies. One cluster was then randomly selected from each of the 25 strata yielding 261 grantee and delegate agencies. Agencies were eliminated that had recently closed, merged, or were serving all eligible children in their communities, and smaller agencies were grouped together. Approximately three grantee and delegate agencies were then randomly selected from each of the 25 strata, yielding a final 84 grantee and delegate agencies.

These 84 Head Start agencies generated lists of 1427 individual centers that were expected to be in operation for the 2002–03 school year. After individual programs were eliminated because they had recently closed, merged, or were serving all eligible children in their communities, and groups of centers were stratified along the same dimensions as the geographical agency clusters, 383 individual centers remained (U.S. DHHS, Final Report, 2010b). An average of four centers were selected from each Head Start agency with a range of 1–7 centers (C. Heid, personal communication, April 10, 2013).

Once the centers were selected, a lottery process was used to determine which children were and were not assigned a place in Head Start. The goal was to randomly select 27 children from each center – 16 to be assigned to Head Start and 11 to the control condition. In total 4442 children were randomly selected – 2646 for Head Start and 1796 for the control condition. Data collection took place from fall 2002, at the time the treatment group entered Head Start, until spring 2006, at the end of first grade (U.S. DHHS, Final Report, 2010b). The resulting sample (N = 3540) was roughly split into thirds by child's race – black, Hispanic, and white or other. Further, about half the sample was male and about a tenth had a disability at baseline. Over a third of the mothers in the sample had less than a high school education, about 15% were teenage mothers, and about a fifth had immigrated to the United States in the past 10 years.

Measures

Outcomes

Pre-academic skills. Prior to program entry in the fall of 2002 and every subsequent spring for the study duration, treatment and control group children were administered a battery of assessments, including the Woodcock Johnson (WJ) III Pre-academic Skills Cluster (Woodcock, McGrew, & Mather, 2001) as a measure of academic achievement. The WJ Pre-academic Skills Cluster is norm-referenced assessment (M = 100, SD = 15), comprised of the WJ Letter–Word, WJ Spelling, and WJ Applied Problems tests. It provides an overall score for a child's pre-reading skills, letter and word identification skills, developing mathematics skills, and skill in written production (α = .97; U.S. DHHS, Technical Report, 2010a). Given the critical importance of these academic domains for later school success (Whitehurst & Lonigan, 2003; Yesil-Dagli, 2011), we used this assessment as a measure of a child's overall pre-academic skills at baseline, after one academic year in Head Start or the control condition, and at the end of kindergarten. Descriptive statistics for each time period are found in Table 1.

Total child behavior problems (maternal-report). Both at baseline and every subsequent spring of the study, the primary caregiver was administered an interview by project staff. The primary caregiver was considered the person living with and most responsible for the care of the child, and in the HSIS, three-quarters of the respondents were the biological or adoptive mothers. In this interview, the primary caregiver was asked to report on children's behavior in three broad areas: aggressive behavior (e.g., "hits and fights with others"), hyperactive behavior (e.g., "is very restless and fidgets a lot"), and withdrawn behavior (e.g., "is unhappy, sad, or depressed"). These items were based on a modified Child Behavior Checklist (CBCL; Achenbach, Edelbrock, & Howell, 1987) and were used as well in the U.S. DHHS's Head Start Family and Child Experiences Survey (FACES). For each item, the parent was asked to judge whether the behavioral description was "not true" (0), "sometimes true" (1), or "very true" (2) of the child. HSIS study administrators then constructed a total child behavior problems scale derived from parent ratings of all 14 items across the three areas, with a total score ranging from 0, all items marked "not true", to 28, all items marked "very true" (α = .77; U.S. DHHS, Technical Report, 2010a). We used this scale as a measure of the maternal-report of a child's total behavior problems at baseline, after one academic year in Head Start or the control condition, and at the end of kindergarten. Descriptive statistics for each time period are found in Table 1.

Total child behavior problems (kindergarten teacher-report). In the spring of the kindergarten year, the head classroom teacher was asked to fill out the Adjustment Scales for Preschool Intervention (ASPI; Lutz, Fantuzzo, & McDermott, 2002) for each study child. The ASPI contains 24 classroom situations that provide 144 descriptions of both typical and problem classroom behavior in five areas: aggressive behavior, withdrawn or low energy behavior, socially reticent behavior, oppositional behavior, and inattentive or hyperactive behavior. The teacher selected all descriptions that matched a child's behavior in a specified classroom situation over the past two months. Raw scores were based on the sum of the checked behavior descriptions. To create a composite kindergarten teacher-report of total behavior problems paralleling the maternal-report, we summed the raw scores from all five behavior areas for a range of scores from 0 items checked to 51 items checked, with about 75% of sample children having ten or fewer behaviors checked (α = .68). Descriptive statistics are found in Table 1.

Risk factors. The comprehensive services model of Head Start provides preschool education; medical, dental, and mental health care; nutrition assistance; and efforts to help parents foster their children's development (U.S. DHHS, Final Report, 2010b). Using these goals, we created ten conceptually-based risk factors in four broad areas based on the Head Start program model: (1) children's pre-academic skills; (2) children's behavior problems; (3) children's health; and (4) family functioning. The specific items chosen had appropriate variables in the HSIS data to represent them, and previous work has demonstrated their potential to negatively impact children's development (see Zigler & Styfco, 2010 for a review). The ten risk factors were all measured at baseline in the fall of 2002. We modeled our conceptualization of risk for each factor after Sameroff (2006) who argued for dichotomous high risk/low risk categorizations, and for continuous variables, the bottom 25% to be considered at risk. Because of the low-income sample of Head Start children, we used 1 SD below national norms as the criterion for high risk on continuous variables rather than the bottom 25%. Thus, a child was coded "1" if they had a given risk factor and "0" if they did not.

To form a cumulative index, we then summed for a total measure of baseline risk from 0, no additional risk factors, to 10, all additional risk factors. Because the basis of the "whole child" model is that optimal development occurs when children's needs in each domain are met (Zigler & Styfco, 2010), each corresponding risk factor was given equal weight. Similar to Bradley (2004) and Bollen (2002), we believe that our cumulative risk measure is the formative result of these ten items, rather than an underlying reflective construct causing these ten items, and therefore internal consistency reporting for our index is inappropriate. Nonetheless, descriptive statistics for each risk factor as well as our cumulative risk index are located in Table 1. The individual factors were determined as follows.

Low receptive vocabulary skills. We classified children as having low receptive vocabulary skills at baseline if they scored more than 1 SD below national norms (below an 85), on the Peabody Picture Vocabulary Test, Third Edition (PPVT; Dunn & Dunn, 1997). Children who scored below an 85 on the PPVT at baseline were given a "1" for this risk factor, and everyone else who scored at or above an 85 was coded "0." Almost a quarter of sample children had this risk factor at baseline according to our classification.

Low reading identification skills. Similarly, we classified children as having low reading identification skills at baseline if they scored more than 1 SD below national norms (below an 85), on the WJ Letter–Word test. Children who scored below an 85 on the WJ Letter–Word test at baseline were given a "1" for this risk factor, and everyone else who scored at or above an 85 was coded "0." About 35% of sample children had this risk factor at baseline.

Low early math skills. Likewise, we classified children as having low early math skills at baseline if they scored more than 1 SD below

Table 1
Descriptive statistics, full sample.

	Treatment			Control		
	N	Mean/% of sample	Standard deviation	*N*	Mean/% of sample	Standard deviation
Outcome variables						
Baseline (Fall 2002)						
Pre-academic skills	2216	90.95	12.14	1324	90.27	11.71
Total behavior problems – maternal report	2216	6.05	3.88	1324	6.12	3.29
End of Head Start (Spring 2003)						
Pre-academic skills	2216	91.58	13.29	1324	88.58***	13.74
Total behavior problems – maternal report	2216	5.67	3.94	1324	6.01*	3.24
End of kindergarten (Spring 2004/2005)						
Pre-academic skills	1572	98.03	11.04	968	98.22	9.26
Total behavior problems – maternal report	1572	4.96	4.11	968	5.07	3.40
Total behavior problems – teacher report	1572	7.13	8.17	968	6.82	6.77
Risk factors – baseline (Fall 2002)						
Individual components						
Child low receptive vocabulary skills	2216	0.23		1324	0.21	
Child low reading ID skills	2216	0.34		1324	0.36	
Child low early math skills	2216	0.33		1324	0.36	
Child high hyperactive behavior	2216	0.15		1324	0.17	
Child high aggressive behavior	2216	0.09		1324	0.08	
Child has never seen dentist	2216	0.21		1324	0.23	
Child suboptimal overall health	2216	0.21		1324	0.21	
Maternal depression	2216	0.21		1324	0.19	
Maternal low literacy skills	1973	0.47		1090	0.48	
Family economic difficulty	2216	0.37		1324	0.35	
Cumulative risk	2216	2.56	1.81	1324	2.58	1.61
Control variables – baseline (Fall 2002)						
Child characteristics						
Age at Spring 2003 assessment in weeks	2216	236.02	31.38	1324	236.13	27.06
Age-3 cohort	2216	0.54		1324	0.55	
Gender – male	2216	0.49		1324	0.49	
Race	2216			1324		
Black		0.32			0.29	
Hispanic		0.36			0.36	
White & other		0.33			0.35	
Spanish as baseline testing language	2216	0.24		1324	0.23	
Disability	2216	0.14		1324	0.11**	
Family characteristics						
Caregiver age	2216	29.37	8.06	1324	29.08	6.31
Maternal education	2216			1324		
Less than high school		0.36			0.39	
High school diploma/GED		0.34			0.33	
Beyond high school		0.31			0.28	
Married mother	2216	0.45		1324	0.46	
Teenage mother	2216	0.15		1324	0.15	
Parents lived together	2216	0.50		1324	0.51	
Immigrant mother	2216	0.19		1324	0.18	
Home language Spanish	2216	0.28		1324	0.29	

Note. p level of treatment/control difference: * $p < .05$. ** $p < .01$. *** $p < .001$.
PPVT = Peabody Picture Vocabulary Test. WJ = Woodcock–Johnson. Pre-academic skills are the average scores of 3 direct standardized assessments: PPVT, WJ Letter–Word test, & WJ Applied Problems test. *t*-Tests for differences in means were conducted for continuous variables, chi-square tests for categorical variables. Weight used = CHSPR2003WTCA for control variables, risk indices and components, and pre-academic skills baseline and end of HS; CHSPR2003WTPI for maternal-report behavior problems baseline and end of HS; CHSPR2005WTCA for pre-academic skills end of K; CHSPR2005WTPI for maternal-report behavior problems end of K; S05TRWTCA_PI_TS for teacher-report behavior problems end of K.

national norms (below an 85), on the WJ Applied Problems test. Children who scored below an 85 on the WJ Applied Problems test at baseline were given a “1” for this risk factor, and everyone else who scored at or above an 85 was coded “0.” About a third of sample children had this risk factor at baseline.

High hyperactive behavior. Using the baseline maternal-report of behavior problems, children were coded “1” for the risk factor of high hyperactive behavior if they were rated a 4 or higher (out of a possible 6) for total hyperactive behavior. Everyone else was coded “0.” About 15% of sample children had this risk factor at baseline.

High aggressive behavior. Similarly, using the baseline maternal-report of behavior problems, children were coded “1” for the risk factor of high aggressive behavior if they were rated a 6 or higher (out of a possible 8) for total aggressive behavior. Everyone else was coded “0.” About 10% of sample children had this risk factor at baseline.

Never seen the dentist. In the baseline parent interview, the primary caregiver was asked to report on the health status of their child. One of the key items of interest was whether sample children had ever seen the dentist prior to the start of the study. Children were coded “1” if their primary caregiver indicated that they had never seen the dentist, and “0” if they had. Almost 25% of sample children had never seen the dentist at baseline.

Suboptimal overall health. In this same interview, the primary caregiver was also asked about the overall health status of their child at baseline. Children were coded “1” if their primary caregiver indicated that their health status was only good, fair, or poor, and “0” if it was rated excellent or very good. Over 20% of sample children were rated as having suboptimal overall health at baseline.

Maternal depression. Study mothers were administered at baseline a shortened form of the Center for Epidemiologic Studies-Depression Scale (CES-D; Radloff, 1977), which measures depressive symptoms (α = .85). HSIS study administrators created four subgroups from the scale: (1) no depressive symptoms (score of 0–4); (2) mild depressive symptoms (score of 5–9); (3) moderate depressive symptoms (score of 10–14); and (4) severe depressive symptoms (score of 15–36). Children were coded "1" if their mother was diagnosed as having moderate or severe depressive symptoms, and "0" if their mother had no symptoms or mild symptoms. Over 20% of sample children had mothers who exhibited such symptoms at baseline.

Low maternal literacy skills. Study mothers were also administered at baseline the reading subtest of the Kaufman Functional Academic Skills Test (KFAST; Kaufman & Kaufman, 1994). The KFAST is a brief, norm-referenced (M = 100, SD = 15) test of functional reading abilities for individuals (α = .90). Similar to the direct child assessments, we defined a low KFAST score as more than 1 *SD* below national norms (below an 85). Children of mothers who scored below an 85 at baseline were given a "1" for this risk factor, and children of mothers who scored at or above an 85 on the KFAST were coded "0." Almost 50% of sample children had mothers with this risk factor at baseline.

Economic difficulty. Although nearly all Head Start-eligible children are low-income as determined by the federal poverty line, additional information on household monetary resources was collected from the primary caregiver at baseline. They were asked if the family had experienced any difficulty in the past three months in paying rent, paying electric and heating bills, buying food for the family, and buying clothes for the children. Children of mothers who responded yes to any of these questions were coded "1" for economic difficulty. Children of mothers who indicated they had not experienced any difficulty were coded "0." Over a third of sample children resided in households that had experienced recent economic difficulty.

Covariates. In order to adjust for departures from randomization and increase the precision of our point estimates, we used the same set of baseline covariates as was used in the Final Report of the HSIS (U.S. DHHS, 2010a, 2010b). Child covariates included: gender (1 = male, 0 = female); whether the child's race was black (1 = yes, 0 = no); whether the child was of Hispanic or Latino ethnicity (1 = yes, 0 = no); whether the child was classified as having a disability at baseline (1 = yes, 0 = no); the language of baseline testing (1 = Spanish baseline testing, 0 = English baseline testing); number of elapsed weeks from September 1, 2002 until the spring 2003 assessment; and age in weeks at the spring 2003 assessment. Family covariates included: caregiver age in years; highest level of maternal education; whether the mother was a recent immigrant to the United States; primary language used at home (1 = Spanish, 0 = English); and three family structure variables including whether both biological parents lived with the child, whether the child's mother was married, and whether the mother was a teenage mother at the child's birth. A lagged dependent variable was also included in the model to control for the child's baseline level of pre-academic skills or behavior problems, respectively. For purposes of analyses, all the covariates were centered at their mean. Descriptive statistics for all covariates are displayed in Table 1.

Non-response. As with any longitudinal dataset, there was non-response in the HSIS. In particular, child assessment and parent interview response rates were correlated with treatment status as well as child gender and race. To control for this potential bias, we weighted all our analyses using the study-supplied weights, which included an adjustment for probability of selection into the sample as well as for non-response (U.S. DHHS, Technical Report, 2010a). The exact weights that were used depended on the outcome of interest and wave of the study. Therefore, explicitly listed at the bottom of each table are the specific weights used for each analysis. As a robustness check, we also ran multiple imputation for missing data and found that the results of our analyses did not change significantly. Hence below we report the weighted case results (N = 3540).

Analysis plan

Residualized growth models were used to estimate the effects of each risk factor on pre-academic skills and behavior problems above and beyond what can be explained by the lagged dependent variable and what can be predicted based on covariates. Thus, our principal regression model was:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2(FACTOR)_{t-1} + \beta_3 TX + \beta_4(FACTOR_{t-1} \times TX) + \gamma COVARIATES + e,$$

where Y_t was the outcome variable of interest; Y_{t-1} was the lagged outcome variable at baseline; $FACTOR_{\ t-1}$ was the individual risk factor at baseline; TX was the dummy variable for random assignment to Head Start; $FACTOR_{t-1} \times TX$ was an interaction of the individual risk factor at baseline with assignment to Head Start; *COVARIATES* was the vector of child and family covariates described above used in the model; and e_t was an error term. The primary parameter of interest, β_4, tested whether the relation between each individual risk factor and child outcomes varied whether the child was randomly assigned to Head Start.

We then used similar growth models to estimate the effects of cumulative risk on pre-academic skills and behavior problems. Here our principal model was:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2(RISK)_{t-1} + \beta_3 TX + \beta_4(RISK_{t-1} \times TX) + \gamma COVARIATES + e_t,$$

where the primary parameter of interest, β_4, tested whether the relation between each additional risk factor and child outcomes varied whether the child was randomly assignment to Head Start. Given the strong potential for Head Start center-level variation, all models included Head Start center-level fixed effects with the standard errors properly adjusted for weighting and clustering.

Results

Descriptive analyses

Table 1 displays the descriptive statistics for the two primary outcomes of pre-academic skills and total child behavior problems, the main independent variable of cumulative risk, and the covariates. There were no significant differences between treatment and control group children on either pre-academic skills or total behavior problems at baseline. By the end of one academic year, Head Start children had significantly higher pre-academic skills (p < .001) and significantly fewer behavior problems (p < .05) compared with control group children. Consistent with the Final Report of the HSIS (U.S. DHHS, 2010a, 2010b), these cognitive and behavioral advantages disappeared by the end of kindergarten.

Additionally, as would be expected from random assignment, there were no significant differences between Head Start and control group children on the risk index or single risk factors at baseline. The mean number of risk factors was about 2.5 on a scale from 0 to 10, indicating the distribution was positively skewed. About 12% of the sample had no or one risk factors, 68% of the sample had between two and five risk factors, and 10% had six or more risk factors. Despite this skewedness, our modeling techniques are appropriate because OLS regression provides the best linear unbiased estimator of the regression coefficients even when these variables are not normally distributed (Wooldridge, 2009).

Balance was achieved between treatment and control on all covariates except child disability. The imbalance may be due to

Table 2
Relationship between risk factors, assignment to Head Start, and their interaction.

	Outcomes									
	(1) Pre-academic skills: direct assessments end of HS year		(2) Total behavior problems: maternal-report end of HS year		(3) Pre-academic skills: direct assessments end of K year		(4) Total behavior problems: maternal-report end of K year		(5) Total behavior problems: teacher-report end of K year	
	Single interaction	Full set of interactions	Single interaction	Full set of interactions	Single interaction	Full set of interactions	Single interaction	Full set of interactions	Single interaction	Full set of interactions
Interactions of Head Start treatment with baseline:										
Child pre-academic skills										
Low receptive vocabulary skills	0.02	0.02	0.16*	0.18	0.15	0.23	0.07	0.06	−0.21***	−0.18
	(0.09)	(0.10)	(0.08)	(0.11)	(0.16)	(0.16)	(0.11)	(0.18)	(0.12)	(0.23)
Low reading ID skills	0.14***	0.13	0.13***	0.05	0.12	0.15	0.04	−0.10	−0.18	−0.26
	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	(0.11)	(0.11)	(0.14)	(0.12)	(0.16)
Low early math skills	0.08	0.04	−0.12	−0.09	0.13	0.15	0.20	0.18	0.02	0.09
	(0.09)	(0.08)	(0.10)	(0.10)	(0.12)	(0.11)	(0.15)	(0.15)	(0.16)	(0.16)
Child behavior problems										
High hyperactivity	0.01	0.02	0.24*	0.54**	0.09	0.03	0.49**	0.39***	0.00	0.18
	(0.12)	(0.13)	(0.11)	(0.18)	(0.19)	(0.15)	(0.16)	(0.22)	(0.15)	(0.24)
High aggressiveness	0.04	0.01	0.29*	0.21	0.07	0.14	0.55**	0.42	−0.10	−0.39
	(0.16)	(0.08)	(0.15)	(0.19)	(0.14)	(0.20)	(0.20)	(0.27)	(0.23)	(0.30)
Child health										
Never seen the dentist	0.15	0.08	−0.05	0.01	0.06	0.09	−0.01	−0.07	−0.09	−0.14
	(0.10)	(0.10)	(0.08)	(0.10)	(0.12)	(0.12)	(0.12)	(0.15)	(0.11)	(0.14)
Suboptimal overall health	0.07	0.04	−0.08	−0.17	0.04	0.07	0.10	0.18	−0.09	−0.06
	(0.11)	(0.11)	(0.10)	(0.13)	(0.13)	(0.13)	(0.14)	(0.18)	(0.14)	(0.23)
Family factors										
Maternal depression	0.07	0.09	0.07	0.02	0.09	0.10	0.19	0.16	−0.08	−0.01
	(0.09)	(0.10)	(0.09)	(0.11)	(0.12)	(0.12)	(0.15)	(0.16)	(0.16)	(0.19)
Low maternal literacy skills	0.05	0.04	0.03	−0.00	0.06	0.07	−0.07	−0.08	−0.22*	−0.01
	(0.08)	(0.08)	(0.09)	(0.12)	(0.10)	(0.11)	(0.11)	(0.15)	(0.10)	(0.15)
Economic difficulty	0.05	0.06	−0.04	−0.07	0.14	0.15	0.10	0.00	0.02	0.01
	(0.08)	(0.08)	(0.08)	(0.11)	(0.12)	(0.11)	(0.10)	(0.13)	(0.11)	(0.15)

Note. Standard errors in parentheses calculated using jackknife replicate weights. Standardized regression coefficients. * $p<0.05$. ** $p<0.01$. *** $p<0.10$.
PPVT = Peabody Picture Vocabulary Test. WJ = Woodcock–Johnson. Pre-academic skills are the average scores of 3 direct standardized assessments: PPVT, WJ Letter–Word test, & WJ Applied Problems test. Head Start center-level fixed effects included in the model. Covariates (centered at mean): lagged DV, child cohort, child race, child gender, child disability status, child age at spring 2003 assessment, number of weeks elapsed between 09/01/02 and spring 2003 child assessment, language of baseline child assessment, maternal education, maternal marital status, caregiver depression, teenage mother status, caregiver age, maternal immigration status, and home language. Weight used = CHSPR2003WTCA for pre-academic skills end of HS; CHSPR2003WTPI for maternal-report behavior problems end of HS; CHSPR2005WTCA for pre-academic skills end of K; CHSPR2005WTPI for maternal-report behavior problems end of K; S05TRWTCA_PI_TS for teacher-report behavior problems end of K.

Head Start's stated commitment to serve children with disabilities and its purposeful reservation of at least 10% of enrollment slots for these children (U.S. DHHS, Final Report, 2010b), Head Start was more likely to serve children with disabilities as opposed to the control condition ($p<.01$). A simultaneous joint test of all baseline covariates between treatment and control was marginally significant owing mainly to this difference in serving children with disabilities ($F(16, 3468) = 0.98$, $p<.10$).

Principal regression analyses

In order to provide the most flexible and transparent look at the relation between risk, random assignment to Head Start, and child outcomes, we first tested each individual risk factor in our models separately. We did these analyses in two ways. First, we ran regressions for our five outcomes of interest (two at the end of one academic year in Head Start or the control condition and three at the end of kindergarten) with the main effects of each risk factor, the main effect of Head Start, their interaction, and covariates, one by one for each individual risk factor. We then ran the same five regressions with the main effects of each risk factor, the main effect of Head Start, their interaction, and covariates, all while controlling for all the other individual risk factors and their interactions with Head Start. This latter set of regression runs helped isolate the effect of each particular risk factor holding constant all the others. For parsimonious presentation, only the interaction results from these models are found in Table 2.

Results from these analyses indicate that there were no statistically significant interactions with the individual risk factors and assignment to Head Start on pre-academic skills at either time point of the end of the Head Start year or the end of kindergarten. However, on behavior problems, there was some indication that two individual risk factors were particularly associated with maternal-, though not teacher-, report of behavior. A child with high hyperactive behavior ($\beta = .24$, $p<.05$) or high aggressive behavior ($\beta = .29$, $p<.05$) at baseline had higher maternal-reported behavior problems at the end of the Head Start year if they were assigned to Head Start compared with the control condition. By the end of kindergarten, these detrimental interactions increased in magnitude. A child with high hyperactive behavior ($\beta = .49$, $p<.01$) or high aggressive behavior ($\beta = .55$, $p<.01$) at baseline had higher maternal-reported behavior problems at the end of kindergarten if they were assigned to Head Start compared with the control condition. These results primarily held in the models that contained all the individual risk factors and their interactions with Head Start. Interestingly, for teacher-report of behavior, none of the interactions between the individual risk factors and Head Start was statistically significant.

We next estimated models using our cumulative risk index displayed in Table 3.

Models 1 and 2 display the results from the end of one academic year in Head Start or the control condition. Because interaction variables are included in all the models, the main effects of cumulative risk are for control group children. Consequently, for each outcome,

Table 3
Relationship between cumulative risk, assignment to Head Start, and their interaction.

	Outcomes				
	(1) Pre-academic skills: direct assessments end of HS year	(2) Total behavior problems: maternal-report end of HS year	(3) Pre-academic skills: direct assessments end of K year	(4) Total behavior problems: maternal-report end of K year	(5) Total behavior problems: teacher-report end of K year
Cumulative risk (continuous index)	−0.06*	0.08***	−0.04	0.15***	0.12**
	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)
Treatment (dummy assignment to HS)	0.14**	−0.06∼	0.08	−0.04	−0.07
	(0.05)	(0.03)	(0.06)	(0.05)	(0.05)
Cumulative risk × treatment	0.05	0.08∼	0.03	0.15*	−0.10∼
	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)
N	3014	3014	2054	2054	2054
R^2	0.61	0.49	0.57	0.44	0.29

Note. Standard errors in parentheses calculated using jackknife replicate weights. Standardized regression coefficients. ∼ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.
Head Start center-level fixed effects included in model. Covariates (centered at mean): lagged DV, child cohort, child race, child gender, child disability status, child age at spring 2003 assessment, number of weeks elapsed between 09/01/02 and spring 2003 child assessment, language of baseline child assessment, maternal education, maternal marital status, caregiver depression, teenage mother status, caregiver age, maternal immigration status, and home language. Weight used = CHSPR2003WTCA for pre-academic skills end of HS; CHSPR2003WTPI for maternal-report behavior problems end of HS; CHSPR2005WTCA for pre-academic skills end of K; CHSPR2005WTPI for maternal-report behavior problems end of K; S05TRWTCA_PI_TS for teacher-report behavior problems end of K.

we first discuss the results for the control group and then for the Head Start group.

Model 1 indicates that the more risks control group children had at baseline, the lower their pre-academic skills at the end of one academic year. On average, the pre-academic skills of control group children decreased for each additional risk factor they had ($\beta = -.06$, $p < .05$), compared with control group children with no extra risk factors. We compare the scores for Head Start children with cumulative risk to those for Head Start children with no additional risks by adding the point estimates for the main and interaction effects of risk. Similar to control group children, the pre-academic skills of Head Start children decreased, on average, for each additional risk factor they had ($\beta = -.01$), compared with Head Start children with no extra risk factors. Because the interaction effect was not statistically significant, there were essentially no differences between Head Start and control group children in how much each additional risk factor decreased their pre-academic skills at the end of the program year.

Model 2 shows that the more risks control group children had at baseline, the higher their maternal-reported behavior problems at the end of one academic year. On average, the maternal-reported behavior problems of control group children increased for each additional risk factor they had ($\beta = .08$, $p < .001$), compared with control group children with no extra risk factors. Using the same calculation as above in adding the main and interaction effects, similar to control group children, the maternal-reported behavior problems of Head Start children also increased, on average, for each additional risk factor they had ($\beta = .16$), compared with Head Start children with no extra risk factors. The interaction effect for risk and Head Start was marginally statistically significant, providing suggestive evidence that a child assigned to Head Start, on average, had higher maternal-reported behavior problems at the end of the program year for each additional risk factor they had ($\beta = .08$, $p < .10$), compared with control children.

Models 3, 4, and 5 display the results for the end of kindergarten. Model 3 demonstrates that by the end of the kindergarten, the statistically significant detrimental effect of baseline cumulative risk for control group children on pre-academic skills had faded, though it should be noted the point estimate was in the expected negative direction ($\beta = -.04$). Similarly, the statistically significant detrimental effect of cumulative risk for Head Start children on pre-academic skills had also faded, though again it should be noted the point estimate was in the expected negative direction ($\beta = -.01$). The interaction term was also not significant. Thus, there were essentially no differences between treatment and control group children in how much each additional risk factor decreased their pre-academic skills at the end of kindergarten.

Model 4 indicates that the more risks control group children had at baseline, the higher their maternal-reported behavior problems at the end of kindergarten. On average, the maternal-reported behavior problems of control group children increased for each additional risk factor they had ($\beta = .15$, $p < .001$), compared with control group children with no extra risk factors. Using the same calculation in adding the main and interaction effects, similar to control group children, the maternal-reported behavior problems of Head Start children also increased, on average, for each additional risk factor they had ($\beta = .30$), compared with Head Start children with no extra risk factors. The interaction effect for risk and Head Start was statistically significant. A child assigned to Head Start, on average, had higher maternal-reported behavior problems at the end of kindergarten for each additional risk factor they had ($\beta = .15$, $p < .05$), compared with control children.

To better understand this interaction, Fig. 1 graphically displays the predicted maternal-report of behavior at the end of kindergarten for a child with one risk factor, the mean number of total risk factors (about 2.5), and 10 total risk factors. Head Start children are indicated by the light gray line while control group children are indicated by the dark gray line. Head Start children have a steeper slope than control group children as the number of total risk factors linearly increases. This interaction was more robust than the one at the end of the program year, achieving formal significance.

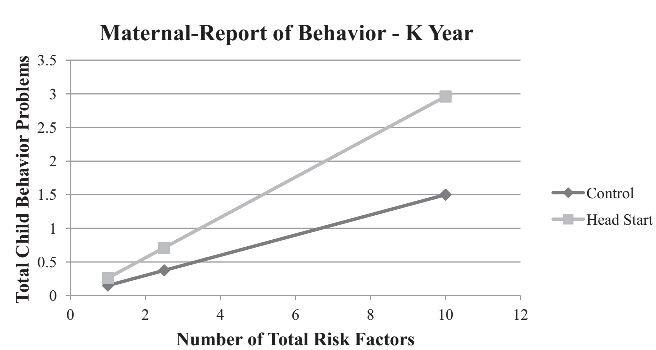

Fig. 1. Predicted maternal-report of behavior problems at the end of kindergarten for a child with 1 total risk factor, the mean number of total risk factors (about 2.5), and 10 total risk factors. Interaction significant at the $p < .05$ level.

Lastly, Model 5 shows that the more risks control group children had at baseline, the higher their teacher-reported behavior problems at the end of kindergarten. On average, the teacher-reported behavior problems of control group children increased for each additional risk factor they had ($\beta = .12$, $p < .01$), compared with control group children with no extra risk factors. Similarly, adding together the main and interaction effects, the teacher-reported behavior problems of Head Start children also increased, on average, for each additional risk factor they had ($\beta = .02$), compared with Head Start children with no extra risk factors. The interaction effect for risk and Head Start was marginally statistically significant providing suggestive evidence that a child assigned to Head Start, on average, had *fewer* teacher-reported behavior problems at the end of kindergarten for each additional risk factor they had ($\beta = -.10$, $p < .10$), compared with control children. This finding is particularly interesting given that the point estimate is in the opposite direction than the point estimate for maternal-report of behavior.

In sum, from the regressions with individual risk factors, we found evidence that high hyperactive or aggressive activity at baseline was associated with higher maternal-report of problem behavior at both time points. From the regressions with the cumulative risk index, we found evidence of a marginally significant detrimental interaction between cumulative risk and Head Start on maternal-report of behavior problems at the end of the program year, a significant detrimental interaction between cumulative risk and Head Start on this same outcome at the end of kindergarten, and a marginally significant beneficial interaction between cumulative risk and Head Start on teacher-report of behavior problems at the end of kindergarten. We found no significant interactions between risk and Head Start on pre-academic skills at either time point in models with the individual risk factors or using cumulative risk.

Supplemental analyses

Given the indication that the two behavior risk factors might be driving the interaction results on behavior between cumulative risk and Head Start, we wanted to test this empirically. We, therefore, ran regressions omitting these two risk factors from the cumulative index. Results are presented in the online supplemental material and indicate that for maternal-report of problem behavior, these two behavior risk factors were entirely responsible for the negative interaction between cumulative risk and Head Start as this interaction was no longer significant when they were omitted. Interestingly, for teacher-report of behavior, the interaction between cumulative risk and Head Start remained marginally significant even with the omission of the behavior risk factors from the index. This is consistent with our initial findings that none of the interactions between the individual risk factors and Head Start were statistically significant on this outcome.

Further robustness checks. Our primary analysis sample included both 3- and 4-year-old study children together, with a covariate for child cohort, because we had no theoretical reason for separating the two age cohorts and to increase statistical power. In the absence of hypotheses of age differences therefore, as a further robustness check, we reran our principal analyses separately for each age cohort to test whether the key findings were consistent across cohorts. These findings are presented in the online supplemental material and indicate the full sample results held independently for both age cohorts.

Additionally, about five centers had unusually large concentrations of at-risk children. Our principal results still held even when we excluded these centers from our analysis.

Discussion

This paper conducted an empirical test of the stated conceptual goals of the Head Start program using the HSIS: does the Head Start program differentially benefit children at greatest risk through its comprehensive services of preschool education; medical, dental, and mental health care; nutrition assistance; and efforts to help parents foster their children's development. It sought to determine whether these multiple services provided by the program helped children at risk in cognitive development, socio-emotional development, health, and family functioning.

Despite positive Head Start treatment main effects, our interaction results were largely negative, suggesting that Head Start was less effective for higher-risk children compared with the control condition. In other words, we found no evidence that Head Start was particularly effective in benefitting high-risk children's pre-academic skills at either time point using either the individual risk factors or the cumulative risk index. Therefore, although Head Start increased children's pre-academic skills, on average, compared with control group children (at least after the program year), Head Start did not differentially increase the pre-academic skills of children whose risk profiles matched those targeted by the Head Start service model. Given Head Start's ambitious goal of improving a wide array of developmental outcomes for low-income children as well as the difficulty of increasing young children's academic achievement (U.S. DHHS, Final Report, 2010b), especially for high-risk children's literacy outcomes (DiPrete & Eirich, 2006; Stanovich, 1986; Walberg & Tsai, 1983), these results are not surprising.

On total child behavior problems, our results varied depending on whether the respondent was the child's mother or teacher. Maternal- and kindergarten teacher-report of behavior were only modestly correlated at .28, consistent with other prior literature (Achenbach, McConaughy, & Howell, 1987; NICHD ECCRN, 2004; Stanger & Lewis, 1993; Verhulst & Akkerhuis, 1989) and indicating they represented two distinct measures of a child's behavior. Using the maternal-report, children with higher aggressive or hyperactive behavior at baseline had more behavior problems when randomly assigned to Head Start than the control condition. In fact, these two behavior risk factors were entirely responsible for the marginally significant detrimental interaction between Head Start and the cumulative risk index at the end of the program year and the significant detrimental interaction at the end of kindergarten. Thus, in the cumulative interactions, a child assigned to Head Start, on average, had higher maternal-reported behavior problems for each additional risk factor they had, compared with controls. These findings somewhat converge with previous results from the NICHD Study of Early Child Care and Youth Development that more time children spent in center-based care was related to less harmonious mother-child interactions and less maternal sensitivity in the first three years of life (NICHD ECCRN, 1998, 1999).

Conversely, using the kindergarten teacher-report of behavior problems, in the regressions with the individual risk factors, none of the interactions between the individual risk factors and Head Start was statistically significant. However, the interaction between the cumulative risk index and Head Start was marginally significant and beneficial. A child assigned to Head Start, on average, had marginally *fewer* behavior problems for each additional risk factor they had, compared with control group children. This marginally significant interaction held with further empirical testing when we removed the behavior risk factors from the cumulative index.

These mixed results on behavior may be at least partially explained by the amount of time children spend in each care setting and the situational specificity of what may constitute problem behavior for each of the respondents (Achenbach, Edelbrock, et al., 1987; Achenbach, McConaughy, et al., 1987; Verhulst & Akkerhuis, 1989). For example, because children spend several hours in child care or in kindergarten class, they may come home tired and cranky from a full day and act out accordingly. Therefore, the maternal report of increased problems may have represented a limited snapshot of behavior during the day, based largely on such undesirable conduct, particularly if the child was prone to problem behavior in the first place.

In contrast, the teacher-report of behavior may have represented a more stable longer-term picture of how a child behaved throughout the day, allowing teachers more opportunity to see positive conduct. It appears, therefore, the Head Start program differentially benefitted high-risk children by helping them to behave better in classroom settings. Thus, Head Start decreased children's teacher-reported behavior problems, on average, compared with control group children and marginally more so for children whose risk profiles matched those targeted by the Head Start service model. Similarly, Cooper and Lanza (2014) also found some evidence of mixed results on behavior using the HSIS that tended to be more positive for teacher-report and more negative for maternal-report. Future studies should continue to examine whether the effects of Head Start on children's behavior differ depending on the respondent.

Limitations and future directions

Some study limitations should be noted. First, we used a cumulative risk model comprised of items based on Head Start's theory of change model because the program expects that the synergy among the services offered and the benefits received will help even children at greatest risk with the transition to kindergarten. Ideally, however, a multifactorial experiment, in which different combinations of the Head Start treatment were offered, would have helped us ascertain the most cost-effective of Head Start's "whole child" services for high-risk children. Since the HSIS was not designed in such a multifactorial way, we consequently view our analytical approach as an appropriate, albeit limited, alternative to understand whether Head Start's comprehensive services were beneficial for children with differing risk profiles. Moreover, since the Head Start impacts in the HSIS are evaluated relative to counterfactual conditions, which include other center-based services, and because our study is an intent-to-treat (ITT) analysis as fitting the nature of our conceptual test of Head Start, our results are likely conservative estimates of the causal impact of this interaction for high-risk children who actually participated in the program. Future studies should make use of such a multifactorial experiment to more definitively determine which of the program services are most effective and for which children, as well as to address the Head Start versus no center-based care comparison.

Second, more discussion is warranted on whether it is better to conceptualize the risk measure as a set of individual items or as a summary index, and which most accurately represented children's experiences prior to Head Start. Prior research indicates that cumulative risk models provide a more comprehensive and precise representation of a child's disadvantage than examining each factor individually (Garmezy et al., 1984; Rutter, 1987; Sameroff et al., 1987, 1993). Consistent with these previous studies, the majority of the children in the HSIS sample had between two and five risk factors, and interactions between the cumulative risk index and Head Start were at least marginally significant in three out of our five models. Thus, given the multiple services approach of Head Start to provide preschool education; medical, dental, and mental health care; nutrition assistance; and efforts to help parents foster their children's development, the cumulative risk index is the most conceptually and arguably empirically matched to program goals, with child-level risk factors in this instance more important than family-level factors for differential program benefits.

Nonetheless, particular components of the index such as high aggressive behavior and high hyperactive behavior played a primary role in determining these interactions between cumulative risk and Head Start, at least with regard to maternal-report of problem behavior. Underlying our stated research question on the extent to which Head Start particularly benefitted children at greatest risk in the domains targeted by the program's services to improve school readiness, was the fundamental premise that an approach that classifies risk consistently with how it is classified by Head Start's theory of change would best determine if the program is meeting its goals. Therefore, the concept of a risk index matched to the domains targeted by Head Start has more utility for policy than a set of individual items given the comprehensive approach of the program. However, the evidence supporting the use of such an index for determining differential program benefits for high-risk children was modest at best. Though we see considerable value in keeping the risk factors together in a highly conceptually based index, rather than separating them out individually, this is still an empirical question requiring further study. We look forward to future work further confirming this hypothesis.

Lastly, although the risks for this study were aligned with Head Start goals, the outcomes were limited to widely accepted, standard academic achievement and behavioral outcomes. We chose these outcomes precisely because they are so established and reliable. However, future studies could see how these risk profiles affect children's development in a more comprehensive outcome framework. The findings we presented here do not preclude possible positive outcomes in other areas such as increased access to dental or mental health care.

In sum, this study tested how well Head Start's multiple services differentially benefitted children with differing risk profiles. It is the first explicit empirical test, based on experimental data, of the stated conceptual risk-related goals of the Head Start program. Bioecological theory (Bronfenbrenner & Morris, 2006) expects that the interactions between a child and their environmental context may vary depending on the quality of the match of services to needs as well as the developmental outcome of interest. Consistent with this stipulation, the mixed interaction findings provided at least suggestive evidence that Head Start was better able to effectively individualize the program to improve teacher-reported behavior rather than pre-academic school readiness. Prior research has shown that the types of risk factors that Head Start targets to be more related to behavior than to test scores (Klebanov & Brooks-Gunn, 2006; Sameroff, 2006; U.S. DHHS, Final Report, 2010b), and the results from our study largely converge with this finding. Although attempts to individualize the Head Start program occur and vary at the center level, there are no systematic or formal guidelines for doing so at the national level (Zigler & Styfco, 2010). Therefore, although the differential benefits of Head Start for high-risk children are limited, nonetheless children enrolled should continue to receive services as, on average, the program boosts pre-academic skills and reduces problem behavior. Future research should continue to examine ways to help Head Start be more responsive to the individual needs of high-risk children and families, particularly in the cognitive domain, providing them with higher levels of support than they currently receive.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecresq.2015.08.001.

References

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology*, *15*(4), 629–650. http://dx.doi.org/10.1007/BF00917246

Achenbach, T. M., McConaughy, S. H., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-information correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213–232. http://dx.doi.org/10.1037/0033-2909.101.2.213

Barton, L. R., Spiker, D., & Williamson, C. (2012). Characterizing disabilities in Head Start programs: Not so clearcut. In M. Burchinal, & A. Winsler (Eds.), *Supporting optimal child development through Early Head Start and Head Start programs: Secondary data analyses of FACES and EHSREP* (pp. 596–612) (*Early Childhood Research Quarterly*, *27*(4). doi:10.1016/j.ecresq.2012.04.002).

Bloom, H., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. (Submitted for publication).

Bollen, K. A. (2002). Analysis of latent variables. *Annual Review of Psychology*, *53*, 605–634. http://dx.doi.org/10.1146/annurev.psych.53.100901.135239

Bradley, R. H. (2004). Chaos, culture, and covariance structures: A dynamic systems view of children's experiences at home. *Parenting: Science and Practice*, *4*(2–3), 243–257. http://dx.doi.org/10.1080/15295192.2004.9681272

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (vol. 1) (6th ed., vol. 1, pp. 793–828). New York, NY: Wiley.

Bulotsky-Shearer, R. J., Wen, X., Faria, A.-M., Hahs-Vaughn, D. L., & Korfmacher, J. (2012). National profiles of classroom quality and family involvement: A multilevel examination of proximal influences on Head Start children's school readiness. In M. Burchinal, & A. Winsler (Eds.), *Supporting optimal child development through Early Head Start and Head Start programs: Secondary data analyses of FACES and EHSREP* (pp. 627–639) (*Early Childhood Research Quarterly*, *27*(4). doi:10.1016/j.ecresq.2012.02.001).

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: A comparison of statistical risk models. *Developmental Psychology*, *36*(6), 793–807. http://dx.doi.org/10.1037/0012-1649.36.6.793

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist*, *53*(2), 221–241. http://dx.doi.org/10.1037/0003-066X.53.2.221

Cooper, B. R., & Lanza, S. T. (2014). Who benefits most from Head Start? Using latent class moderation to examine differential treatment effects. *Child Development*, *85*(6), 2317–2338. http://dx.doi.org/10.1111/cdev.12278

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, *32*, 271–297. http://dx.doi.org/10.1146/annurev.soc.32.061604.123127

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test, third ed. (PPVT)*. Circle Pines, MN: American Guidance Services.

Egeland, B., Carison, L., & Sroufe, L. A. (1993). Resilience as process. Special issue: Milestones in the development of resilience. *Development and Psychopathology*, *5*(4), 517–528. http://dx.doi.org/10.1017/S0954579400006143

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, *139*(6), 1342–1396. http://dx.doi.org/10.1037/a0031808

Farber, E. A., & Egeland, B. (1987). Invulnerability among abused and neglected children. In E. J. Anthony, & B. J. Cohler (Eds.), *The invulnerable child* (pp. 253–288). New York, NY: Guilford Press.

Furstenberg, F. F., Brooks-Gunn, J., & Morgan, S. P. (1987). *Adolescent mothers in later life*. New York, NY: Cambridge University Press.

Garmezy, N., Masten, A. S., & Tellegen, A. (1984). The study of stress and competence in children: A building block for developmental psychopathology. *Child Development*, *55*(1,), 97–111.

Halle, T. G., Hair, E. C., Wandner, L. D., & Chien, N. C. (2012). Profiles of school readiness among four-year-old Head Start children. In M. Burchinal, & A. Winsler (Eds.), *Supporting optimal child development through Early Head Start and Head Start programs: Secondary data analyses of FACES and EHSREP* (pp. 613–626) (*Early Childhood Research Quarterly*, *27*(4). doi:10.1016/j.ecresq.2012.04.001).

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467–475.

Kaufman, A. S., & Kaufman, N. L. (1994). *Kaufman Functional Academic Skills Test (K-FAST)*. Bloomington, MN: Pearson Assessments.

Klebanov, P., & Brooks-Gunn, J. (2006). Cumulative, human capital, and psychological risk in the context of early intervention: Links with IQ at ages 3, 5, and 8. *Annals of the New York Academy of Sciences*, *1094*, 63–82. http://dx.doi.org/10.1196/annals.1376.007

Lutz, M. N., Fantuzzo, J., & McDermott, P. (2002). *Adjustment Scales for Preschool Intervention (ASPI)*. Philadelphia, PA: University of Pennsylvania.

Masten, A. S., Coatsworth, J. D., Neemann, J., Gest, S. D., Tellegen, A., & Garmezy, N. (1995). The structure and coherence of competence from childhood through adolescence. *Child Development*, *66*(6), 1635–1659. http://dx.doi.org/10.1111/j.1467-8624.1995.tb00956.x

McCoy, D. C., Morris, P. A., Connors, M. C., Yoshikawa, H., & Gomez, C. (2015). *Differential effectiveness of Head Start in urban and rural communities*. (Submitted for publication).

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, *53*(2), 185–204. http://dx.doi.org/10.1037/0003-066X.53.2.185

Miller, E. B., Farkas, G., Vandell, D. L., & Duncan, G. J. (2014). Do the effects of Head Start vary by parental pre-academic stimulation? *Child Development*, *85*(4), 1385–1400. http://dx.doi.org/10.1111/cdev.12233

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, *69*(4), 1145–1170. http://dx.doi.org/10.1111/j.1467-8624.1998.tb06165.x

NICHD Early Child Care Research Network. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology*, *35*(6), 1399–1413. http://dx.doi.org/10.1037/0012-1649.35.6.1399

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development*, *69*(4), 1–146. http://dx.doi.org/10.1111/j.1540-5834.2004.00312.x

Radloff, L. D. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. http://dx.doi.org/10.1177/014662167700100306

Raikes, H. H., Vogel, C., & Love, J. M. (2013). Family subgroups and impacts at ages 2, 3, and 5: Variability by race/ethnicity and demographic risk. In J. M. Love, R. Chazan-Cohen, H. Raikes, & J. Brooks-Gunn (Eds.), *What makes a difference: Early Head Start Evaluation findings in a developmental context* (pp. 64–92) (*Monographs of the Society for Research in Child Development*, *78*(1). doi:10.1111/j.1540-5834.2012.00703.x).

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*(2), 302–316. http://dx.doi.org/10.1037/a0015302

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, *57*(3), 316–331. http://dx.doi.org/10.1111/j.1939-0025.1987.tb03541.x

Sameroff, A. J. (2006). Identifying risk and protective factors for healthy child development. In A. Clarke-Stewart, & J. Dunn (Eds.), *Families count: Effects on child and adolescent development* (pp. 53–76). http://dx.doi.org/10.1017/CBO9780511616259.004

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development*, *64*(1), 80–97. http://dx.doi.org/10.1111/j.1467-8624.1993.tb02896.x

Sameroff, A. J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987). Intelligence quotient scores of 4-year-old children: Social-environmental risk factors. *Pediatrics*, *79*(3), 343–350.

Seifer, R., Sameroff, A. J., Baldwin, C. P., & Baldwin, A. (1992). Child and family factors that ameliorate risk between 4 and 13 years of age. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*(5), 893–903. http://dx.doi.org/10.1097/00004583-199209000-00018

Shager, H. M., Schindler, H. S., Magnuson, K. A., Duncan, G. J., Yoshikawa, H., & Hart, C. M. D. (2013). Can research design explain variation in Head Start research results? A meta-analysis of cognitive and achievement outcomes. *Educational Evaluation and Policy Analysis*, *35*(1), 76–95. http://dx.doi.org/10.3102/0162373712462453

Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology*, *22*(1), 107–116. http://dx.doi.org/10.1207/s15374424jccp2201_11

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407. http://dx.doi.org/10.1598/RRQ.21.4.1

U.S. Department of Health and Human Services. Administration for Children and Families. Office of Planning, Research and Evaluation. Head Start Impact Study (HSIS). (2002–2006). ICPSR 29462-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-06-15. doi:10.3886/ICPSR 29462.v4.

U.S. Department of Health and Human Services, Administration for Children and Families. (2011). *Head Start approach to school readiness – Overview*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/sr/approach

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

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

Verhulst, F. C., & Akkerhuis, G. W. (1989). Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4–12. *Journal of Child Psychology and Psychiatry*, *30*(1), 123–136. http://dx.doi.org/10.1111/j.1469-7610.1989.tb00772.x

Walberg, H. J., & Tsai, S.-L. (1983). Matthew effects in education. *American Educational Research Journal*, *20*(3), 359–373. http://dx.doi.org/10.3102/00028312020003359

Werner, E. E. (1997). The value of applied research for Head Start: A cross-cultural and longitudinal perspective. *NHSA Research Quarterly: A Research to Practice Journal for the Early Childhood Field*, *1*(1), 15–24. http://dx.doi.org/10.1207/s19309325nhsa0101_2

Werner, E. E. (2001). Protective factors and individual resilience. In J. P. Shonkoff, & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 115–132). Cambridge, UK: Cambridge University Press.

Werner, E. E. (2004). The role of caring adults and in the lives of children of alcoholics. *Substance Use and Misuse*, *39*(5), 699–720. http://dx.doi.org/10.1081/JA-120034012

Whitehurst, G. J., & Lonigan, C. J. (2003). Emergent literacy: Development from prereaders to readers. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (vol. 1) (pp. 11–29). New York, NY: Guilford.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of achievement*. Itasca, IL: Riverside.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Mason, OH: Cengage Learning.

Wright, M. O., Masten, A. S., & Narayan, A. J. (2013). Resilience processes in development: Four waves of research on positive adaptation in the context of adversity. In S. Goldstein, & R. B. Brooks (Eds.), *Handbook of resilience in children* (pp. 15–37). New York, NY: Springer.

Yesil-Dagli, U. (2011). Predicting ELL students' beginning first grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills. *Early Childhood Research Quarterly*, *26*(1), 15–29. http://dx.doi.org/10.1016/j.ecresq.2010.06.001

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