Early Childhood Research Quarterly 27 (2012) 684–694

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's school performance in kindergarten☆

Xiaoli Wen[a,*], Christine Leow[b], Debbie L. Hahs-Vaughn[c], Jon Korfmacher[b], Sue M. Marcus[d]

[a] National-Louis University, Early Childhood Education Department, 122 South Michigan Avenue, Chicago, IL 60603, United States
[b] Erikson Institute, United States
[c] University of Central Florida, United States
[d] Columbia University, United States

ARTICLE INFO

Keywords:
Head Start
Program duration
School performance
Propensity scores

ABSTRACT

Using data from a nationally representative sample, this study examined Head Start children's school outcome differences by the end of Kindergarten between children who attended Head Start program for two years and the ones who attended for one year. Propensity scores were used to match children who experienced different durations of the program on a series of demographic characteristics in order to achieve a precise estimation of the effects of program duration. The results showed that in comparison to a demographically comparable group of children who attended the Head Start program for one year, the children who experienced two years of intervention services had statistically significantly higher performance on all six academic and social outcome measures by the end of Kindergarten, which included PPVT, Woodcock–Johnson Reading Skills, Woodcock–Johnson Math Reasoning Skills, teacher-reported composite academic skills, preschool learning behaviors, and social skills. Policy and practice implications are discussed.

Early educational interventions, such as Head Start, have been widely recognized as an effective way to mitigate the negative effects of poverty on early learning and development (Burger, 2010; Camilli, Vargas, Ryan, & Barnett, 2010). Participation in early education interventions has shown short- and long-term positive effects on low-income children's academic skills, language development, social competence, emotional adjustment, reduced grade retention, and reduced need for special education (e.g., Belfield, Nores, Barnett, & Schweinhart, 2006; Gory, 2001; Ludwig & Phillips, 2007; Magnuson, Ruhm, & Waldfogel, 2007; Ramey et al., 2000; Temple & Reynolds, 2007). But what we know about early education interventions has primarily focused on the overall effectiveness, and we know much less about the specific program and participant factors and mechanisms that lead to favorable program outcomes (Berlin, O'Neal, & Brooks-Gunn, 1998; Guralnick, 1997; Reynolds, 2004). In this study, we looked at Head Start programs, the nation's largest early educational intervention, and examined the impact of one program design factor, program duration (defined as the length of program enrollment), on child outcomes.

Program duration is one way of measuring dosage, or the amount of intervention services families received. Intervention dosage is a multi-dimensional construct, and has been measured in several different forms beyond duration, including the amount of program contact (e.g., number of activities attended), intensity of intervention (e.g., full-day vs. half-day), and percentage or ratio of completed to expected amount of program contact, as defined by program protocol (Korfmacher et al., 2008; Littell, Alexander, & Reynolds, 2001). This study focuses on program duration in center-based educational intervention programs for low-income children because of its significant policy implications for the field of early education and intervention. In the past decade, there has been a strong expansion of early childhood programming (including Head Start and state-funded prekindergarten programs), but recent economic uncertainty calls into question the extent to which this expansion can be maintained. A tension exists between serving as many children as possible and providing the most impact with limited economic resources (e.g., Barnett & Hustedt, 2011; Steuerle, Reynolds, & Carasso, 2007), making the study of program design factors, such as length of programming, critical to efforts to serve low-income or at-risk children in the most efficient fashion.

Does the amount of intervention services children receive have causal impact on the amount of gains they accrue from the program? Theoretically and hypothetically, the dosage of intervention has been highlighted as a key variable in predicting program effectiveness (Shonkoff & Phillips, 2000). Available research evidence

☆ The authors thank the Inter-University Consortium for Political and Social Research (ICPSR) for providing access to the FACES 2003 database. We are especially grateful to the children, parents, and teachers who participated in this study.
* Corresponding author. Tel.: +1 312 261 3160; fax: +1 312 261 3160.
E-mail address: Xiaoli.Wen@nl.edu (X. Wen).

0885-2006/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2011.07.006

suggests that the most effective early intervention programs that were able to maintain long-term impacts were those that begin during children's early years of life, continue for multiple years, and provide support to families (e.g., Bogard & Takanishi, 2005; Reynolds, Ou, & Topitzes, 2004; Zigler & Styfco, 1994). Therefore, the length of intervention might be at least one of the factors that could determine the magnitude of program impact. However, it is challenging to make a causal conclusion regarding whether children and families who experience a longer duration of intervention would perform better on measured program outcomes than those who are enrolled for a relatively shorter time, because participants who experienced different amounts of intervention may differ in other ways as well, including their demographic characteristics (Hill, Brooks-Gunn, & Waldfogel, 2003; Powell, 2005). Simply stratifying participants by intervention duration or estimating the impact of duration in a standard regression model will not typically yield unbiased estimates because selection bias might be operating.

In the current study, we used a rigorous statistical methodology to make a less biased estimation of the impact of Head Start participation duration on Head Start children's school performance by the end of kindergarten. This study utilized secondary data from the national evaluation of Head Start, known as the Family and Child Experiences Survey (FACES, 2003 cohort; U.S. DHHS, 2008), to address whether Head Start children who entered the program at three years of age and were eligible to receive two years of program services performed academically and socially better in kindergarten than those who entered the program at four years of age and were eligible to receive only one year of program services.

1. Program duration impact

Intervention duration, as a critical program design component, has not received much research attention (Reynolds, 2004). A meta-analysis of 123 comparative studies of preschool intervention programs concluded that most studies did not collect detailed information on program duration, and for those studies that did examine duration, no significant impact on child cognitive and social outcomes was found (Camilli et al., 2010). Most empirical evidence regarding the impact of program duration is from examinations of small-scale model programs that usually are initiated by researchers, operate on a single site, have relatively narrow program foci, and are subject to close quality monitoring.

As one example, studies of the Carolina Abecedarian Project provide strong evidence regarding the effects of intervention duration (Campbell & Ramey, 1994). The Carolina Abecedarian Project is an early educational intervention for impoverished children, designed to prevent mild retardation and school failure through the provision of a supportive learning environment, beginning in infancy. In an experimental study, children were randomly assigned to one of the four conditions: educational treatment from infancy through third grade in public school (i.e., up to age eight); infancy through preschool treatment only (i.e., infancy to age five); primary school treatment only (ages five to eight); or an untreated control group. The design permitted the investigators to estimate the relative efficacy of interventions with different durations and different entry ages. The study also allows the estimation of the importance of reinforcing children's gains from early childhood treatment during the transition into early elementary school. A follow-up study of the children of the Carolina Abecedarian Project four to seven years after the intervention showed that the length of duration predicted Verbal IQ, Reading, and achievement scores. Findings generally supported the hypothesis that child academic performance increased as duration of treatment increased (eight years > five years > three years > none) (Campbell & Ramey, 1994). A more extended follow-up study of the same group of children when they reached age 15 demonstrated a similar conclusion: the largest program effects accrued to children whose program participation continued the longest (Campbell & Ramey, 1995).

Demonstration programs such as Abecedarian are different from large public programs sponsored by federal or state governments, and one must be cautious in generalizing findings that emerge from these model programs to large-scale public programs such as Head Start. The link between program duration and outcomes is not well established in Head Start programs. One study with a small sample of Head Start children explicitly compared child and family outcomes between children who attended the program for one year and those who attended for two years, and found that program duration was positively associated with home environment and parents' frequency of reading to children, but not child outcomes (Ritblatt, Brassert, Johnson, & Gomez, 2001). This study of Head Start population is correlational. Although attempts were made to control statistically for confounding variables that might influence program outcomes, the evidence for the impact of program duration is suggestive, but not conclusive.

The issue of program duration was touched upon in the Head Start Impact Study, the most recent nationwide randomized study of Head Start. A nationally representative sample of nearly 5000 newly entering three- and four-year old Head Start applicants were randomly assigned either to a treatment group that had access to Head Start services or to a control group that could receive any other non-Head Start services available in the community, chosen by their parents (U.S. DHHS, 2010). The three-year old children were eligible for two years of Head Start, while the four-year old children were eligible for one year of services. The study examined program outcomes separately for these two age cohorts. However, due to ethical concerns about possible denial of services to eligible children, one year after the randomization, children in the control group were allowed to enter the Head Start program (although typically *not* at the study sample centers), and some treatment group children left Head Start. By design, this study did not attempt to control whether three-year old children were enrolled for two years of Head Start or four-year old children were enrolled for one year (U.S. DHHS, 2010). Therefore, regretfully the study could not make a precise estimation of program duration impact. In addition, the sample statistics showed that the two age cohorts varied in their demographic characteristics, such as their ethnic distribution.

In general, the study demonstrated program impact on three-year old children's social-emotional development (e.g., less hyperactive and problem behaviors, better social skills and positive approaches to learning, and more positive relationships with parents), and these positive effects were maintained through first grade (U.S. DHHS, 2010). In contrast, four-year old children did not show these benefits. The program had minimal impact on cognitive development for both three- and four-year old children. The benefits of access to Head Start largely disappeared by first grade for the program participants as a whole. However, for three-year old children, there were a few sustained benefits (e.g., more positive relationships with parents, authoritarian parenting, and parents' less use of spanking). These results are suggestive of the influence of Head Start program duration, but the absence of randomization to different lengths of program and the demographic differences between the two age cohorts make it difficult to draw more causal conclusions.

Very few studies have made meaningful comparisons between the three- and four-year old age groups enrolled in Head Start. A careful review of the federal reports generated from the FACES studies (http://www.acf.hhs.gov/programs/opre/hs/faces/) showed that many analyses were conducted with the overall combined sample. When the program outcomes were compared and contrasted between the two age groups, usually a direct comparison was made without carefully controlling for confounding