Abstract Title Page

Title

Which Combination of High Quality Infant-Toddler and Preschool Care Best Promotes School Readiness?

Authors

Weilin Li[1], George Farkas[1], Greg J. Duncan[1], Margaret R. Burchinal[1], Deborah L. Vandell[1], Erik A. Ruzek[1], Tran T. Dang[1]

Weilin Li (Corresponding Author)
Department of Education
University of California, Irvine
2075 Education
Irvine, CA 92697-5500
weilinl@uci.edu

[1]Department of Education
University of California – Irvine

Abstract Body

Background / Context

Research has demonstrated that much of the achievement gap observed in economically disadvantaged children is already present when formal schooling begins at age 5 (Fryer & Levitt, 2004), and can be largely attributed to child care experiences from birth to age 5 (Duncan & Brooks-Gunn, 1997). This fact has stimulated the attention of policy-makers and researchers on education programs for children before age 5. Evidence suggests that these programs do improve school readiness skills, but the effects are not large enough to fully eliminate the achievement gap experienced by disadvantaged children (Ludwig & Phillips, 2007; Magnuson, Ruhm, & Waldfogel, 2007). Moreover, when the effect of high quality preschool is estimated controlling baseline cognitive performance, the effect size is quite modest – less than .10 increase in 54-month achievement or cognitive ability associated with a 1-point (roughly 2 standard deviations) increase in the study's quality measure (NICHD-ECCRN & Duncan, 2003).

In an effort to create programs that can further reduce the achievement gap, attention has turned to high quality child care prior to preschool, i.e. child care for infants and toddlers. Research has demonstrated that high quality infant-toddler care is associated with better cognitive development and language comprehension (Burchinal, et al., 1996, 2000; Clarke-Stewart, et al., 2000; McCartney, et al., 1985; Vandell & Wolfe, 2000). Early enrichment programs such as Early Head Start and the Infant Health and Development Program (IHDP) has been shown to have positive effect for 3-year-old children in cognitive and language development (Love, et al., 2001, 2005; McCormick, et al., 2006).

However from the literature, there is no evidence about which childcare combination for the two stages most strongly increases school readiness. In order to make better policy decisions that are appropriately tailored to each developmental stage, we need to understand the dynamic mechanism of skill formation and educational investment at different stages.

Purpose / Objective / Research Question / Focus of Study

Following the literature, this paper aims to test the following hypotheses:

Hypothesis 1 (H1): Everything else the same, high quality infant-toddler care will increase children's cognitive scores immediately (i.e. at 24 months of age). However, without subsequent high quality preschool, children with high quality infant-toddler care will not have higher cognitive and achievement scores when formal school begins (i.e. at 54 months of age) than children with low quality infant-toddler care.

Hypothesis 2 (H2): Everything else the same, high quality preschool will positively affect children's cognitive and academic scores immediately (i.e. at 54 months of age). In addition, the combination of high quality infant-toddler and preschool care will produce higher cognitive and achievement scores at 54 months of age than the combination of high quality preschool but low quality infant-toddler care.

Hypothesis 3 (H3): Everything else the same, high quality child care in both infancy-toddlerhood and preschool stages will produce higher children's cognitive and academic scores at school entry than any other child care quality combinations.

Setting

Data analyzed in this paper are from the NICHD Study of Early Child Care and Youth Development (henceforward SECCYD). In 1991, a socio-economically diverse sample of children and their families were recruited from designated hospitals at 10 data collection sites:

Little Rock, Arkansas; Irvine, California; Lawrence, Kansas; Boston, Massachusetts; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Charlottesville, Virginia; Morgantown, North Carolina; Seattle, Washington; and Madison, Wisconsin.

Population / Participants / Subjects

SECCYD enrolled 1,364 families with full-term healthy newborns in accordance with a conditionally random sampling plan, which was designed to ensure that the recruited families reflected the diversity in terms of socio-economic status, race, and ethnicity. Response rates at the point of the six-month interview were around 50% (Duncan & Gibson-Davis, 2006).

Intervention / Program / Practice

Child care quality. This paper aims to test effects of combinations of high quality child care during infancy-toddlerhood and preschool. The child care quality is measured by the Observational Record of the Caregiving Environment (henceforward ORCE), which was developed to assess the quality of caregiving for individual children at 6, 15, 24, 36, and 54 months. The ORCE can be used in different settings such as home care and center-based care to assess different types of caregivers such as relative, nanny, and teacher. Each assessment consists of four 44-minute child-focused observations across 2 days. By means of observation, each aspect of caregivers' and children's behavioral frequencies and qualitative ratings of interactions between caregivers and children contributed to a 4-point subscale. The final quality rating score is the mean of the subscales. Higher scores indicate that caregivers were more sensitive to children's behaviors, more cognitive stimulating, more warm and positive, more exploration fostering, and less emotionally detached. Cronbach's alphas for the composite score ranged from .83 to .89 and reliabilities are greater than .80 at all ages.

Child outcomes. Our outcome measure at 24 month is the Bayley Mental Developmental Index based on a 1993 revision of the test (Bayley, 1993). Two child outcomes at 54 month are considered, cognitive development and academic achievement. The cognitive development outcome is the mean of the Woodcock-Johnson Picture Vocabulary and memory for Sentences tests and the Preschool Language Scale Expressive and Receptive tests. The academic achievement is the mean of the Woodcock-Johnson Applied Problems, Letter-Word Identification, and Incomplete words Scales, assessing mathematical skill, reading skill, and phonological knowledge respectively (Neisser et al., 1996).

Controls. In this paper, time-insensitive demographic controls are measured at one month after birth. These control covariates are study sites, child race (white and non-white), child gender, child's birth order, child's temperament (maternal rating), maternal attitudes for raising kids, maternal age, maternal education level (in years), and paternal education level (in years). Time-sensitive control covariates are measured at both 1 month and 24 months of age. These controls are child's heath, maternal separation anxiety, maternal employment status, whether mother's partner presents in the household, number of adult and children in household, and family income-to-needs ratio.

Research Design

Since the SECCYD is an observational study, the treatment and control groups exhibit imbalance on covariates. The imbalance may lead to *selection bias*, i.e. potential bias from imbalanced treatment assignment conditional on certain covariates. The normal OLS could help to reduce the selection bias by accounting for many potential confounding factors; however, the

OLS estimates may not solve the bias problem when there is not a sufficient overlap between treatment and control groups (Cochran, 1965). Therefore, we use propensity score matching approach to reduce biases from "selection on observables" (Ravallion, 2001) and to avoid results that could be extrapolated beyond the region of the data in normal OLS estimates.

Specifically, the propensity score matching in this analysis involves four steps. In the first step, we divide the SECCYD sample into six groups according to child care quality in the two stages: (1) the group with low quality infant-toddler care and low quality preschool (henceforward low-low), (2) the group with high quality infant-toddler care and low quality preschool (henceforward high-low), (3) the group with low quality infant-toddler care and high quality preschool (henceforward low-high), (4) the group with high quality infant-toddler care and high quality preschool (henceforward high-high), (5) the group with high quality infant-toddler care no matter how preschool is (henceforward early-high), (6) the group with low quality infant-toddler care no matter how preschool is (henceforward early-low), (7) the group with high quality preschool no matter how infant-toddler care was before preschool (henceforward late-high), and (8) the group with low quality preschool no matter how infant-toddler care was before preschool (henceforward late-low). We use 3.0 point of the averaged ORCE scores at 6, 15, and 24 months as the cut point for low and high quality infant-toddler care, and 3.0 point of the averaged ORCE scores at 36 and 54 months as the cut point for low and high quality preschool. In the second step, we built up contrasts to test the three hypotheses, which are demonstrated in Table 1. The first column shows which hypothesis to test. The second and third columns show the targeted and comparison groups in each contrast. Observations in comparison groups will be selected to match observations in the corresponding targeted groups in the following steps. The fourth column shows the outcome measure(s) timing.

[Insert Table 1 here]

In the third step, we use a logistic model for each contrast to predict a propensity score for each individual, defined as the conditional probability of being selected into the targeted group given a full set of covariates. Since we have six contrasts here (note that the second contrast for H2 is the same as the third contrast for H3), a series of six binomial logistic regressions are used to generate propensity scores. Variables measured at one month after birth are used in logistic models to generate propensity scores of being selected into low-low, high-low, low-high, or high-high group. Time-insensitive variables measured at one month after birth, updated demographic variables and cognitive scores measured at 24 months of age are used in logistic models to generate propensity scores of being selected into late-high or late-low group.

In the fourth step, for each contrast, observations in the comparison group are selected to match observations in the corresponding targeted group. Take the first contrast for H2 as an example, "late high" is the targeted group and "late low" is the comparison group; therefore, observations in "late low" group are selected to match those in "late high" group. The match is conducted within sites so as to eliminate the unobserved demographic differences (Cook et al., 2008) and within caliper width of 0.01 to ensure sufficient overlap in propensity scores between targeted and comparison groups (Rosenbaum, 2002). The standardized differences after matching are within the interval between -.1 and .1, which ensures the balance between the matched targeted and comparison groups conditional on observed covariates (Rubin, 2001).

Data Analysis

After propensity score matching, we conduct two approaches to test hypotheses. The first approach is the most traditional one in propensity score matching. We calculate the standardized

mean difference for each pair of balanced targeted and comparison groups and use that standardized mean difference to obtain inferences for the hypotheses.

The second approach is propensity score adjusted regression. In the process of generating propensity scores, a series of binomial models for six different contrasts are estimated. This method is more practical than conducting multinomial models, which is computationally more burdensome and statistically less robust since a mis-specification in one of the series will compromise all others (Lechner, 2001). However, separate estimation of the six binomial models may cause bias for between-contrast comparison (Bryson et al., 2002). Hence propensity score adjusted regressions with all observations in the comparison groups are conducted to correct the possible within-contrast comparison bias.

The propensity score adjusted regression is described as follows. For children in targeted groups, we generate weights as inverse propensity scores ($w = 1/p$). For children in comparison groups, we generate weights by formula $w = 1/(1-p)$. These weights ensures that the children in the comparison group that are most like the children in the targeted group can be weighted more, and those less like the children in targeted group can be weighted less (Rubin, 2001).

Findings / Results:

Table 2 presents descriptive statistics for the outcome and control variables in the analysis. The column panels in which the mean and standard deviation are listed correspond to groups with different child care quality during infancy-toddlerhood and preschool. As might be expected, there is severe imbalance between different groups.

[Insert Table 2 here]

Cognitive development

The first two row panels in Table 3 present standardized mean differences in cognitive scores between targeted and comparison groups before and after propensity score matching. The third row panel shows comparison results from propensity score adjusted regressions.

The first panel implies the significant difference between treatment and control groups in cognitive scores measured at both 24 and 54 months across all of the original contrasts.

The second and the third row panel show standardized mean differences after selection bias is reduced by propensity score matching and adjusted regressions. It implies that the high quality infant-toddler care itself raises cognitive scores by .16 to .25 standard deviation (henceforward SD) at 24 months of age, but this effect fades out to -.04 to .06 without subsequent high quality preschool. High quality preschool raises the group's average cognitive score by .13 to .22 SD at 54 months regardless of the infant-toddler care the child received. The impact of high quality preschool on cognitive score at 54 month can be augmented by .18 SD for children with high quality infant-toddler care. Consistent high quality child care during both infant-toddler and preschool stage increases cognitive scores by .25 to .28 SD at 54 month compared to low quality child care in both stages, by .13 to .17 SD compared to high infant-toddler care only, and by .18 SD compared to high preschool only.

[Insert Table 3 here]

Academic achievement

Table 3 also presents standardized mean differences in academic achievement between targeted and comparison groups by propensity score matching and adjusted regressions.

As might be expected, the first panel implies that, before propensity score matching, there is a significant difference between treatment and control groups in academic achievement measured at both 24 and 54 months across all the six original contrasts.

The second and the third row panel show standardized mean differences after reducing selection bias conditional on observed covariates. It implies that without subsequent high quality preschool, high quality infant-toddler care itself does not significantly raise academic achievement at 54 month. High quality preschool raises academic achievement by .14 to .22 SD at 54 month regardless how infant-toddler care was before preschool. The impact of high quality preschool on academic achievement at 54 month can be augmented by .18 to .19 SD for children with high quality infant-toddler care and higher cognitive scores at 24 month. Consistent high quality child care during both infant-toddler and preschool stage increases academic achievement by .30 to .31 SD at 54 month compared to low quality child care in both stages, by .07 to .27 SD compared to high infant-toddler care only, and by .18 to .19 SD compared to high preschool only. Therefore, consistent high quality child care during both stages is the best combination that promotes academic achievement at school entry.

Robustness checking

Currently there are two methods of conducting propensity score matching for multi-level data. The first method is demonstrated in the former part of this paper, which is to match within each site to eliminate the unobserved site-specific confounding factors. The other method is to match across sites and then run propensity score adjusted regressions with site fixed-effects. Here we use the second method to check robustness of results in Table 3.

Results of site fixed-effect regressions are presented in Table 4. Similar to results in the after matching panel of Table 3, high quality infant-toddler care itself will increase cognitive scores by .27 SD immediately after infancy-toddlerhood; however, it does not impact cognitive and academic outcomes at 54 month. Regardless of infant-toddler care quality, high quality preschool will increase cognitive and academic scores significantly (.14 and .15 SD); and high quality infant-toddler care boosts effects of high quality preschool on both cognitive development (.15 SD) and academic achievement (.16 SD) at 54 month. Consistent high quality child care from birth to 54 month increases both cognitive and academic scores significantly, -- .24 SD compared to low quality child care in both stages, .15 and .22 SD compared to high infant-toddler care only, .15 and .16 SD compared to high preschool only.

Conclusions

The major finding of this paper is that high quality infant-toddler care itself does not affect child outcomes in the long run without subsequent high quality preschool. High quality child care in the very early period can raise the immediate cognitive outcomes by .16 to .19 SD. However, this impact is found to fade out within two or three years if there is no high quality child care following.

Regardless of the infant-toddler care quality, high quality preschool positively affects children's cognitive and academic scores at school entry. That positive effect is augmented for children with high quality infant-toddler care and higher cognitive scores at preschool entry. This implies a positive association between marginal productivity of preschool investment and the cognitive skills developed during infancy-toddlerhood.

Therefore, to invest only in high quality infant-toddler care without subsequent high quality preschool is not productive in the long term. High cognitive and academic scores at school entry require consistent high quality infant-toddler care and high quality preschool. Findings of this paper suggest the desirability of spreading investment across early childhood periods as opposed to front-loading investment on infant-toddler care.

Appendices

Appendix A. References

Bayley, N. (1993). *Bayley II Scales of Infant Development*. New York: Psychological Corporation.

Bryson, A., Dorsett, R. and Purdon, S. (2002) The use of propensity score matching in the evaluation of labour market policies. Working Paper No. 4, Department for Work and Pensions.

Burchinal, M. R., Roberts, J. E., Nabors, L. A., and Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development*, 67, 606-620.

Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S. A., Neebe, E., and Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, 71-2, 339-357.

Clarke-Stewart, K. A., D. L. Vandell, M. Burchinal, M. O'Brien, and K. McCartney. (2000). Do Features of Child Care Homes Affect Children's Development? Unpublished paper, University of California, Irvine.

Cook, T. D., W. J. Shadish, and V. C. Wong. 2008. Three conditions under which observational studies produce the same results as experiments. *Journal of Policy Analysis and Management* 274:724–50.

Duncan, G. J., & Brooks-Gunn, J. (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.

Duncan, G. J., Dowsett, C., Classens, A., Magnuson, K., Huston, A., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428-1446.

Duncan, G. J. and Gibson-Davis, C. (2006), Connecting Child Care Quality to Child Outcomes: Drawing Policy Lessons from Nonexperimental Data, *Evaluation Review*, 30(5), 1-20.

Fryer, R. and Levitt, S. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *Review of Economics and Statistics*, 86, 447-64.

Lechner, M. (2001) Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M. Lechner and F. Pfeiffer (eds), *Econometric Evaluation of Labour Market Policies* (pp. 1–18). Heidelberg: Physica.

Lechner, M. (2001) A note on the common support problem in applied evaluation studies. Discussion Paper No. 2001-01, University of St Gallen, SIAW.

Love, John M., Ellen Eliason Kisker, Christine Ross, Helen Raikes, Jill Constantine, Kimberly Boller, Jeanne Brooks-Gunn, Rachel Chazan-Cohen, Louisa Banks Tarullo, Christy Brady-Smith, Allison S. Fuligni, Peter Z. Schochet, Diane Paulsell, and Cheri Vogel (2005). "The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs." *Developmental Psychology*, 41, pp. 885-901.

Love, J. M., Harrison, L., Sagi-Schwartz, A., van Ijzendoorn, M. J., Ross, C., Ungerer, J. A., Raikes, J., Brady-Smith, C., Boller, K., Brooks-Gunn, J., Constantine, J., Kisker, E. E., Paulsell, K., and Chazan-Cohen, R. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, 74-4, 1021-1033.

Ludwig, J. and Phillips, D. (2007), "The Benefits and Costs of Head Start," NBER Working Paper 12973.

Magnuson, K. A., Ruhm, C., Waldfogel, J. (2007). Does prekindergarten improve school preparations and performance? *Economics of Education Review*, 26, 33-51.

McCartney, K., S. Scarr, D. Phillips, and S. Grajek. (1985). Day Care as Intervention: Comparison of Varying Quality Programs. *Journal of Applied Developmental Psychology* 6: 247–260.

McCormick, C., et al. (2006). Early intervention in low birth weight premature Infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics*. 117 (3). 771-780.

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

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W.,Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knownsand unknowns. *American Psychologist*, 51, 77 – 101.

Rosenbaum, P. R. (2002) *Observational Studies*. New York: Springer.

Rubin, D. B., (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology* 2:169–188.

Vandell, D. L. and Wolfe, B. (2000). Child care quality: Does it matter and does it need to be improved? Institute for Research on Poverty Special Report, University of Wisconsin at Madison, November, 20.

Appendix B. Tables and Figures

Table 1

Contrasts for testing hypotheses

hypothesis	Targeted group	Comparison group	Measured time
H1: Everything else the same, high quality infant-toddler care will increase children's cognitive scores at 24 months of age. However, without subsequent high quality preschool, children with high quality infant-toddler care will not have higher cognitive and achievement scores at 54 months of age than children with low quality infant-toddler care.	early high	early low	24 month
	high-low	low-low	54 month
H2: Everything else the same, high quality preschool will positively affect children's cognitive and academic scores at 54 months of age. In addition, the combination of high quality infant-toddler and preschool care will produce higher cognitive and achievement scores at 54 months of age than the combination of high quality preschool but low quality infant-toddler care.	late high	late low	54 month
	high-high	low-high	54 month
H3: Everything else the same, high quality child care in both infancy-toddlerhood and preschool stages will produce higher children's cognitive and academic scores at school entry than any other child care quality combinations.	high-high	low-low	54 month
	high-high	high-low	54 month
	high-high	low-high	54 month

Table 2

Description of Sample and Analysis Variables

	low-low		high-low		low-high		high-high		early-low		early-high		late-low		late-high	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Child outcomes																
Cognitive development at 24 month	-0.06	0.97	0.26	1.00	0.17	0.91	0.39	0.88	0.05	0.94	0.35	0.92	0.17	0.91	0.39	0.88
Cognitive development at 54 month	-0.11	0.97	0.13	0.90	0.26	0.91	0.44	0.80	--	--	--	--	0.26	0.91	0.44	0.80
Academic achievement at 54 month	-0.08	0.93	0.11	0.94	0.19	0.90	0.46	0.83	--	--	--	--	0.19	0.90	0.46	0.83
Family characteristics																
at one month after birth																
Maternal age (years)	28.84	5.54	29.20	4.63	29.69	5.54	29.81	4.85	29.34	5.49	29.55	4.76	29.11	5.22	29.78	5.22
Maternal education (years)	14.09	2.35	15.01	2.59	14.88	2.14	15.60	2.31	14.50	2.28	15.31	2.45	14.45	2.44	15.21	2.24
Maternal paid leave	46%	50%	66%	48%	36%	48%	67%	47%	43%	50%	68%	47%	54%	50%	50%	50%
Maternal child-rearing attitudes	75.87	15.91	72.48	14.48	70.72	15.07	67.71	14.83	73.02	15.30	69.80	14.96	74.19	15.23	69.07	14.82
Paternal/partner education (years)	14.18	2.40	14.71	2.61	15.24	2.51	15.53	2.63	14.74	2.51	15.16	2.66	14.45	2.48	15.41	2.58
Maternal separation anxiety	68.69	12.93	66.54	12.22	70.20	12.70	66.72	11.71	68.83	12.75	66.66	12.02	--	--	--	--
Mother has a job	61%	49%	84%	37%	53%	50%	81%	39%	58%	49%	83%	38%	--	--	--	--
Father/partner in household	95%	22%	97%	17%	96%	19%	99%	12%	96%	20%	98%	15%	--	--	--	--
Family income-to-needs ratio	2.58	2.17	3.24	2.31	3.37	2.76	3.97	3.03	2.96	2.48	3.65	2.76	--	--	--	--
Number of adult relatives present at home	4.15	1.30	3.77	0.88	4.07	0.99	3.69	0.97	4.09	1.11	3.72	0.92	--	--	--	--
at 24 months of age																
Maternal separation anxiety	--	--	--	--	--	--	--	--	--	--	--	--	62.10	13.23	61.31	12.84
Mother has a job	--	--	--	--	--	--	--	--	--	--	--	--	73%	44%	71%	45%
Father/partner in household	--	--	--	--	--	--	--	--	--	--	--	--	92%	28%	95%	21%
Family income-to-needs ratio	--	--	--	--	--	--	--	--	--	--	--	--	3.83	3.01	4.59	3.20
Number of adult relatives present at home	--	--	--	--	--	--	--	--	--	--	--	--	4.11	1.15	3.96	0.97

Table 2

Description of Sample and Analysis Variables

	low-low		high-low		low-high		high-high		early-low		early-high		late-low		late-high	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Child characteristics																
at one month after birth																
Gender (male = 1)	52%	50%	47%	50%	47%	50%	50%	50%	51%	50%	47%	50%	51%	50%	48%	50%
Ethnicity: White/non-Hispanic	13%	34%	12%	33%	13%	34%	10%	30%	13%	33%	11%	31%	12%	32%	12%	33%
Difficult temperament	3.28	0.64	3.29	0.70	3.32	0.65	3.38	0.63	3.28	0.64	3.33	0.66	3.27	0.65	3.34	0.64
child's birth order	1.93	0.97	1.69	0.78	1.93	0.91	1.53	0.75	1.91	0.92	1.61	0.77	1.84	0.89	1.75	0.87
Health of Baby	3.66	0.56	3.71	0.50	3.73	0.51	3.76	0.49	3.69	0.54	3.74	0.49	--	--	--	--
at 24 months of age																
Health of Baby	--	--	--	--	--	--	--	--	--	--	--	--	3.19	0.69	3.25	0.71
Sample size	275		132		202		144		407		346		477		276	

Table 3

Standardized mean difference before and after matching

				cognitive score		achievement score	
hypothesis	targeted group	comparison group	time of measure(s)	mean difference (se)		mean difference (se)	
Standardized mean difference before matching							
H1	early high	early low	24 month	0.30***	(0.07)	--	--
	high-low	low-low	54 month	0.19*	(0.10)	0.15	(0.10)
H2	late high	late low	54 month	0.35***	(0.07)	0.32***	(0.07)
	high-high	low-high	54 month	0.21*	(0.09)	0.28**	(0.09)
H3	high-high	low-low	54 month	0.53***	(0.09)	0.53***	(0.09)
	high-high	high-low	54 month	0.34***	(0.10)	0.38***	(0.11)
	high-high	low-high	54 month	0.21*	(0.09)	0.28**	(0.09)
Standardized mean difference after matching							
H1	early high	early low	24 month	0.25*	(0.06)	--	--
	high-low	low-low	54 month	-0.04	(0.12)	0.12	(0.10)
H2	late high	late low	54 month	0.22***	(0.06)	0.22***	(0.06)
	high-high	low-high	54 month	0.18*	(0.09)	0.18**	(0.07)
H3	high-high	low-low	54 month	0.25***	(0.09)	0.30***	(0.09)
	high-high	high-low	54 month	0.13	(0.12)	0.07	(0.12)
	high-high	low-high	54 month	0.18*	(0.09)	0.18**	(0.07)
Propensity score adjusted regression result							
H1	early high	early low	24 month	0.16*	(0.07)	--	--
	high-low	low-low	54 month	0.06	(0.08)	0.04	(0.09)
H2	late high	late low	54 month	0.13**	(0.05)	0.14*	(0.06)
	high-high	low-high	54 month	0.18**	(0.07)	0.19**	(0.07)
H3	high-high	low-low	54 month	0.28***	(0.08)	0.31***	(0.08)
	high-high	high-low	54 month	0.17*	(0.08)	0.27*	(0.10)
	high-high	low-high	54 month	0.18**	(0.07)	0.19**	(0.07)

Note. When calculating standardized mean difference before matching, the sample size of comparison group is 275. When calculating standardized mean difference after matching, the sample size of comparison group is equal to that of the corresponding treatment group because 1:1 matching is conducted.

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

Table 4

Propensity score adjusted regressions with site fixed-effect

				cognitive score		achievement score	
hypothesis	targeted group	comparison group	time of measure(s)	mean difference (se)		mean difference (se)	
H1	early high	early low	24 month	0.27***	(0.07)	--	--
	high-low	low-low	54 month	0.05	(0.09)	-0.03	(0.10)
H2	late high	late low	54 month	0.14*	(0.06)	0.15*	(0.06)
	high-high	low-high	54 month	0.15*	(0.07)	0.16**	(0.08)
H3	high-high	low-low	54 month	0.24***	(0.10)	0.24***	(0.10)
	high-high	high-low	54 month	0.15	(0.10)	0.22*	(0.11)
	high-high	low-high	54 month	0.15*	(0.07)	0.16**	(0.08)

Note. When calculating standardized mean difference before matching, the sample size of comparison group is 275. When calculating standardized mean difference after matching, the sample size of comparison group is equal to that of the corresponding treatment group because 1:1 matching is conducted.

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