Hong, G., Corter, C., Hong, Y., & Pelletier, J. (2011). Differential effects of literacy instruction time and homogeneous grouping in kindergarten: Who will benefit? Who will suffer? To appear in *Educational Evaluation and Policy Analysis*.

Online Supplementary Materials

Appendix

Marginal Mean Weighting through Stratification for Estimating Differential Effects

Here we explain the procedure of estimating the marginal mean weight for each child through propensity score stratification. This part of the analysis involves five major steps.

Step 1. Estimating propensity scores

After identifying all the observed class-level and school-level pretreatment covariates for each of the six treatments, we created missing indicators to capture different missing patterns among categorical covariates, and then imputed missing data in the continuous covariates via maximum likelihood estimation. We then analyzed a multinomial logistic regression model at the class level to estimate a vector of six propensity scores for each kindergarten class denoted by $\theta_{L0}, \theta_{L1}, \theta_{L2}, \theta_{H0}, \theta_{H1}$, and θ_{H2} corresponding to the six possible treatments. The six estimated propensity scores summed up to 1.0 for every kindergarten class. Each propensity score summarizes the observed pretreatment information including class composition, teacher characteristics, and school characteristics predicting the probability that the kindergarten class would adopt the corresponding treatment. As proven by Rosenbaum and Rubin (1983, 1984), the experimental units and the control units that have the same propensity score for a certain treatment are not systematically different on average in how they would respond to that treatment. Due to the richness of the observed pretreatment information that we used to estimate each propensity score, it seemed reasonable to assume, for example, that the kindergarten classes in the *L*0 group and the classes not in the *L*0 group that have the same estimated propensity score θ_{L0} are not systematically different in how their students would have achieved in literacy growth on average had they all been assigned to the *L*0 treatment. This is the so-called "weak ignorability assumption" for evaluating multiple or multi-valued treatments (Imbens, 2000).

Step 2. Identifying the analytic sample

Next, we compared each treatment group and the rest of the sample on the distribution of the logit of the estimated propensity of being assigned to that particular treatment. This procedure enabled us to empirically identify and exclude classes that did not have counterfactual information in the data.

Step 3. Stratifying the sample on each propensity score

For each of the six treatments, we then divided the analytic sample into either five or six strata on the basis of the corresponding propensity score. According to Cochran (1968), stratifying a sample into five subclasses typically removes at least 90% of the bias associated with a pretreatment covariate.

4. Computing marginal mean weight

Here we use the *L*0 group to illustrate. Once we have stratified the whole sample on the basis of the logit of the estimated θ_{L0}, within each stratum, the kindergarten classes in the *L*0 group and the classes not in the *L*0 group have the same distribution of the logit of θ_{L0}. When the weak ignorability assumption holds, we expect that within each stratum the observed mean outcome of children attending the kindergarten classes in the *L*0 group is an unbiased estimate of the population average potential outcome associated with *L*0 for all the children in that stratum regardless of the actual treatment assignment of their classes. The same result holds for all three subpopulations of children. This is because child prior ability indicated by child relative standing in class is independent of the class-level treatments. For example, the observed mean outcome of high-ability children in the *L*0 classes within each stratum provides an unbiased estimate of the subpopulation average potential outcome associated with *L*0 for all the high-ability children in that stratum. Hence, we can estimate the marginal mean potential outcome associated with *L*0 for the entire subpopulation of high-ability children through computing a weighted mean of the observed outcome of high-ability children in the *L*0 group.

Let a = 1, 2, 3 denote the low-ability, medium-ability, and high-ability subpopulations, respectively. As derived by Author (2010), in general, for a child from subpopulation a whose kindergarten class adopted treatment z and was found in stratum s_z, the marginal mean weight is

$$\text{MMW-S} = n_{s_z,a} \times pr(Z = z \mid a) / n_{Z=z,s_z,a}\,, \tag{a1}$$

where $n_{s_z,a}$ is the number of subpopulation a children in stratum s_z under the stratification on θ_Z; $n_{Z=z,s_z,a}$ is the number of sampled children from subpopulation a whose classes in stratum s_z actually adopted treatment z; $pr(Z = z \mid a)$ is the proportion of children in subpopulation a who attended kindergarten classes in treatment group z. Intuitively speaking, for each subpopulation of children, we assign weight to those in a certain treatment group such that the weighted composition of this treatment group approximates the pretreatment composition of the entire subpopulation.

Table A1 illustrates the construction of the weighted sample of high-ability children in the *L*0 group. For example, because the kindergarten classes in stratum 1 had a relatively low propensity of adopting *L*0, the high-ability children attending *L*0 classes in this stratum had a relatively low representation in the *L*0 group (15 out of 207) when compared with their representation in the whole sample (400 out of 1,203). The estimated marginal mean weight for these 15 children in stratum 1 was $(400 \times 207)/(15 \times 1{,}203) = 4.59$. Hence, the weighted *L*0 group would have $15 \times 4.59 = 68.85$ high-ability children in stratum 1. High-ability children attending kindergarten classes in a higher stratum had a relatively higher representation in the $L0$ group and thus would receive a relatively lower weight. As a result, the composition of high-ability children in the weighted $L0$ group would resemble that of the entire analytic sample as if their classes had been assigned at random to $L0$.

Applying Equation (a1) to the child-level data, we computed a marginal mean weight for each child as a function of the child's prior ability level, treatment group membership, and stratum membership. We applied the same strategy to each of the six treatments for each subpopulation of children. Table A2 displays the computed marginal mean weights for the six treatment groups within each subpopulation of children.

Step 5. Checking balance in pretreatment composition among treatment groups in the weighted sample. We examined the difference in each pretreatment covariate among the six weighted treatment groups for children at each prior ability level. Adopting a significance level of .05, we expected to see about 5% of the covariates showing significant differences under the null hypotheses. Indeed, no more than 5% of the hypotheses testing showed a statistically significant difference. We therefore concluded that, under the weak ignorability assumption, all the six treatment groups became comparable for children at the same prior ability level.

Table 3

Weighted Analysis of Differential Treatment Effects on Literacy Scale Score

Fixed Effects	Coefficient	*SE*	*t*
Literacy Pretest			
High Ability			
Intercept, γ_{010}	30.44	0.90	33.81***
$L0$, γ_{011}	0.60	1.32	0.45
$L1$, γ_{012}	0.02	1.12	0.02
$H0$, γ_{013}	0.17	1.32	0.13
$H1$, γ_{014}	0.49	1.03	0.48
$H2$, γ_{015}	2.77	2.01	1.38
Medium Ability			
Intercept, γ_{020}	17.68	0.38	47.09***
$L0$, γ_{021}	0.57	0.50	1.14
$L1$, γ_{022}	0.52	0.51	1.02
$H0$, γ_{023}	0.69	0.51	1.34
$H1$, γ_{024}	0.48	0.45	1.08
$H2$, γ_{025}	0.66	0.57	1.16
Low Ability			
Intercept, γ_{030}	12.08	0.58	20.74***
$L0$, γ_{031}	0.25	0.81	0.31
$L1$, γ_{032}	0.67	0.75	0.89
$H0$, γ_{033}	0.51	0.76	0.67
$H1$, γ_{034}	0.60	0.67	0.89
$H2$, γ_{035}	0.48	0.85	0.56
General Knowledge (γ_{040})	0.27	0.01	19.18***
SES (γ_{050})	1.01	0.13	8.01***
Literacy Growth			
High Ability			
Intercept, γ_{110}	12.53	1.07	11.68***
$L0$, γ_{111}	0.36	1.45	0.25
$L1$, γ_{112}	2.66	1.51	1.76
$H0$, γ_{113}	0.84	1.24	0.67
$H1$, γ_{114}	0.62	1.26	0.50

*H*2, γ_{115}	1.39	1.53	0.90
Medium Ability			
Intercept, γ_{120}	13.99	0.51	27.53***
*L*0, γ_{121}	0.01	0.66	0.01
*L*1, γ_{122}	0.39	0.73	0.54
*H*0, γ_{123}	0.73	0.69	1.06
*H*1, γ_{124}	1.39	0.61	2.27*
*H*2, γ_{125}	2.26	0.86	2.62**
Low Ability			
Intercept, γ_{130}	13.43	0.83	16.18***
*L*0, γ_{131}	2.05	0.99	2.07*
*L*1, γ_{132}	1.85	1.14	1.63
*H*0, γ_{133}	1.12	1.05	1.07
*H*1, γ_{134}	2.51	1.01	2.48*
*H*2, γ_{135}	1.60	1.59	1.01
General Knowledge (γ_{140})	0.13	0.02	6.21***
SES (γ_{150})	1.04	0.20	5.12***

Random Effects	Variance Component	*df*	χ^2
Level 2			
Student literacy pretest, r_{0ij}	15.62	6,966	18,038.41***
Student literacy growth rate, r_{1ij}	38.22	6,966	18,891.35***
Level 3			
Class literacy pretest u_{00j}	10.22	1,697	5,055.70***
Class literacy growth rate, u_{10j}	11.36	1,697	3,403.54***

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4

Estimated End-of Year Proficiency Probability in Literacy Subdomains by Instructional Treatment and Prior Ability

	Instructional Treatment					
Literacy Subdomains	*L*0	*L*1	*L*2	*H*0	*H*1	*H*2
Letter Recognition						
High ability	1.00	1.00	1.00	1.00	1.00	1.00
Medium ability	.99	.99	.99	1.00	1.00	1.00
Low ability	.98	.97	.96	.97	.98	.98
Beginning Sounds						
High ability	.99	.99	.99	.99	.99	.99
Medium ability	.84	.83	.82	.86	.87	.88
Low ability	.59	.58	.44	.51	.60	.60
Ending Sounds						
High ability	.95	.96	.94	.96	.94	.96
Medium ability	.48	.51	.45	.51	.51	.56
Low ability	.19	.22	.10	.16	.21	.20
Sight Words						
High ability	.47	.55	.29	.50	.42	.51
Medium ability	.01	.01	.01	.02	.02	.02
Low ability	.00	.00	.00	.00	.00	.00
Comprehension						
High ability	.11	.14	.08	.12	.11	.13
Medium ability	.00	.01	.00	.01	.01	.01
Low ability	.00	.00	.00	.00	.00	.00

Table 5

Weighted Analysis of Differential Treatment Effects in Literacy Subdomains

	Letter Recognition		Beginning Sounds		Ending Sounds		Sight Words		Words in Context	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
High Ability										
Intercept	8.40***	0.30	4.52***	0.274	2.70***	0.26	-0.88**	0.33	-2.38***	0.23
*L*0	-0.25	0.43	0.27	0.36	0.28	0.35	0.75	0.45	0.40	0.31
*L*1	0.26	0.41	0.58	0.38	0.58	0.34	1.08**	0.41	0.63*	0.29
*H*0	-0.11	0.36	0.34	0.33	0.42	0.32	0.89*	0.36	0.53*	0.25
*H*1	0.22	0.36	0.06	0.32	0.02	0.30	0.57	0.38	0.40	0.27
*H*2	0.27	0.48	0.72*	0.35	0.46	0.34	0.92*	0.42	0.57*	0.29
χ^2 (*df*)	4.82 (5)		10.90 (5)		10.35 (5)		14.40* (5)		7.12 (5)	
Medium Ability										
Intercept	5.04***	0.15	1.49***	0.14	-0.20	0.12	-4.58***	0.20	-5.42***	0.19
*L*0	0.12	0.20	0.16	0.17	0.12	0.16	0.35	0.25	0.11	0.24
*L*1	0.25	0.21	0.12	0.19	0.23	0.16	0.32	0.27	0.15	0.25
*H*0	0.53**	0.21	0.31	0.18	0.24	0.16	0.51*	0.25	0.38	0.24
*H*1	0.61**	0.19	0.38*	0.17	0.23	0.15	0.62*	0.24	0.49*	0.22
*H*2	0.71**	0.27	0.53*	0.22	0.43*	0.17	0.75**	0.26	0.47	0.24
χ^2 (*df*)	32.80*** (5)		16.88** (5)		9.31 (5)		17.98** (5)		18.29** (5)	
Low Ability										
Intercept	3.10***	0.22	-0.25	0.24	-2.18***	0.21	-7.57***	0.55	-7.69***	0.29
*L*0	0.84**	0.30	0.60*	0.28	0.73**	0.28	1.62***	0.42	1.40***	0.36
*L*1	0.54	0.28	0.57	0.30	0.89**	0.28	1.45***	0.41	1.17**	0.36
*H*0	0.50	0.28	0.31	0.30	0.53	0.31	1.21**	0.43	0.91*	0.37
*H*1	0.71**	0.26	0.65*	0.27	0.86***	0.26	1.57***	0.37	1.19***	0.34
*H*2	0.95*	0.39	0.65	0.37	0.82*	0.34	1.41**	0.51	0.88*	0.41
χ^2 (*df*)	15.51** (5)		12.26* (5)		19.53** (5)		34.35*** (5)		32.47*** (5)	

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 6

Weighted Analysis of Differential Treatment Effects on General Learning Behaviors

Fixed Effects	Coefficient	*SE*	*t*
High Ability			
Intercept	3.47	0.03	113.07***
*L*0	-0.03	0.05	-0.63
*H*0	0.02	0.04	0.37
*H*1	-0.04	0.04	-0.96
*H*2	-0.03	0.06	-0.49
Medium Ability			
Intercept	3.12	0.02	156.69***
*L*0	0.06	0.03	1.82
*H*0	0.02	0.03	0.85
*H*1	0.02	0.03	0.86
*H*2	0.08	0.04	2.21*
Low Ability			
Intercept	2.74	0.03	79.43***
*L*0	0.10	0.05	1.92
*H*0	0.06	0.06	1.11
*H*1	0.11	0.04	2.51*
*H*2	0.16	0.07	2.46*

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Table A1

Marginal Mean Weight for High-Ability Children Attending Kindergarten Classes with Low Reading Time and No Grouping (L0)

Stratum	Unweighted Sample			MMW-S	Weighted Sample
	*L*0 = 1	*L*0 = 0	Total		*L*0 = 1
1	15	385	400	4.59	68.85
2	19	216	235	2.13	40.47
3	60	241	301	0.86	51.60
4	78	145	223	0.49	38.22
5	35	9	44	0.22	7.7
Total	207	996	1,203	---	207

Table A2

Marginal Mean Weight for High-, Medium-, and Low-Ability Children

High Ability						
Stratum	*L*0	*L*1	*L*2	*H*0	*H*1	*H*2
1	4.59	4.02	2.45	4.05	2.53	3.17
2	2.13	1.52	0.98	1.96	2.27	1.35
3	0.86	1.09	0.96	0.82	0.95	0.48
4	0.49	0.47	0.29	0.53	0.81	0.31
5	0.22	0.27	0.23	0.39	0.60	0.13
6	---	---	---	0.42	0.40	---
Weighted *n*	207	233	136	221	300	106
Medium Ability						
Stratum	*L*0	*L*1	*L*2	*H*0	*H*1	*H*2
1	3.71	3.27	2.96	3.82	3.39	3.09
2	1.65	1.26	1.05	1.68	1.73	1.39
3	0.93	0.95	0.70	0.86	1.23	0.45
4	0.51	0.44	0.31	0.48	0.86	0.29
5	0.24	0.30	0.22	0.42	0.59	0.14
6	---	---	---	0.29	0.42	---
Weighted *n*	981	934	626	916	1,376	493
Low Ability						
Stratum	*L*0	*L*1	*L*2	*H*0	*H*1	*H*2
1	2.95	5.08	2.48	2.98	3.60	3.34
2	2.00	1.32	1.10	1.89	1.76	0.87
3	1.06	0.85	0.81	0.73	1.26	0.56
4	0.56	0.43	0.31	0.61	0.73	0.32
5	0.23	0.27	0.20	0.52	0.54	0.15
6	---	---	---	0.37	0.38	---
Weighted *n*	414	378	213	383	538	213