Economics of Education Review 31 (2012) 1058–1066

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Combination classes and educational achievement

Jaime L. Thomas *

Mathematica Policy Research, 505 14th Street, Suite 800 Oakland, CA 94612, United States

ARTICLE INFO

Article history:
Received 16 October 2009
Received in revised form 20 July 2012
Accepted 29 July 2012

JEL classification:
I21
I22

Keywords:
Costs
Educational economics
Educational finance
Human capital
Input output analysis

ABSTRACT

Using the ECLS-K and considering first graders in single-grade and K–1 and 1–2 combination classes, I discuss the mechanisms underlying the combination-class effect and address the systematic school-, teacher-, and student-level differences that confound estimates of this effect. I find evidence for positive selection into 1–2 classes, but using a rich set of control variables, find no relationship between class type and student achievement in first grade within schools, and no difference in overall first-grade achievement between single-grade and combination schools in a matched school sample. The results I present suggest that first graders are not harmed by being in a combination class or by their schools offering combination classes. As long as other stakeholders such as parents, teachers, and students in other grades are not made worse off, these results suggest that offering combination classes may be a Pareto-improving option for school administrators.

1. Introduction

The combination class, in which students from two adjacent grades are grouped within one classroom under one teacher, is a tool that school administrators can use to manage uneven class sizes and conserve scarce facility and personnel resources. Sims (2008) gives evidence that schools in California responded to the 1996 adoption of the California Class Size Reduction Program by increasing their use of combination classes. Even when class-size-reduction incentives weaken, as they have in California, combination classes continue to be a cost-saving option since they allow schools to use fewer teachers and classrooms. Cost-saving considerations become particularly important when state budgets are tight and education funding cuts loom. If combination-class membership has a nonnegative effect on student outcomes, offering such classes is an attractive strategy for schools looking to save money without sacrificing educational quality.

I present evidence for nonrandom assignment to combination classes, a concern in isolating a causal effect. However, after controlling for a rich set of variables that influence assignment and accounting for cross-school differences, I find strong evidence that first graders in combination classes do no better and no worse than their single-grade counterparts on reading, math, and general knowledge tests given at the end of the school year. In addition, I find no evidence that first graders in schools offering combination classes perform worse than first graders in matched schools that do not offer such classes, indicating that offering combination classes may be a viable cost-saving option for school administrators.

This paper proceeds as follows: Section 2 reviews the literature on combination classes. Section 3 describes the rich data set used in the analysis. In Section 4, I discuss the causal mechanisms by which combination classes affect learning. Section 5 focuses on confounders of this effect, describing how combination students, teachers, and schools differ from their single-grade counterparts due to nonrandom selection. Section 6 presents the main results of the paper and discusses some robustness checks, and Section 7 concludes.

* Tel.: +1 510 830 3717; fax: +1 510 830 3701.
E-mail address: jthomas@mathematica-mpr.com

0272-7757/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.econedurev.2012.07.013

2. A review of the literature

In discussing the literature on alternatives to single-grade classrooms, it is necessary to distinguish between multi-age and multi-grade classes. Multi-age classes are often implemented for pedagogical reasons, while multi-grade classes tend to be implemented for administrative reasons such as having fewer teachers than grade levels or uneven student enrollments (see Mariano and Kirby, 2009 for a more detailed discussion). I focus on two-grade combination classes, a type of multi-grade class.

The evidence on the effect of combination classes is mixed. In a large meta-analysis, Veenman (1995) concludes that students in multi-grade classes learn as much as their single-grade counterparts. Mason and Burns (1997) argue that conclusions about combination-class effects are confounded by the merging of combination-class and multi-age studies and by the assignment of certain types of students and teachers to combination classes. Indeed, Burns and Mason (2002) find that principals and teachers assign higher-ability and more independent students to combination classes, and that this nonrandom assignment affects class distributional properties.

Russell, Rowe, and Hill (1998) report negative associations between multi-grade membership and reading and math scores, though results differ by year. Wilkinson and Hamilton (2003) do not find differences in the ranges of abilities in single- and multi-grade classrooms. In contrast, using a doubly robust, within-district (but across-school) estimation strategy that takes into account students' relative grade levels, Mariano and Kirby (2009) find small but significantly negative effects of combination-class membership in third and fourth grade on English and math assessments.

Like Mariano and Kirby (2009), I carefully consider students' relative grade level in my analysis. In addition, I give evidence that previous findings of negative combination-class effects could be due to inadequate controls for differences across schools. I obtain results similar to those of Mariano and Kirby (2009) using a cross-school, doubly robust estimation strategy. Using two different methods that account for cross-school differences (school fixed effects and propensity score matching), I find that combination-class students do not score significantly better or worse than their single-grade counterparts.

3. Data and sample

The Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 (ECLS-K) Restricted-Use Data Set is an ongoing study focusing on children's early school experiences. It contains a rich set of student-, classroom-, and school-level variables, permitting analysis of the factors that influence a school's decision to offer combination classes and the teacher and student characteristics that influence assignment, as well as the relationship between combination-class membership and test scores.

I use data collected in the spring of the children's kindergarten and first-grade years. Spring first grade reading, math, and general knowledge standardized test scores are the outcome variables of interest. Spring kindergarten scores are prior test score controls. I consider a variety of additional child-, classroom-, and school-level variables: child characteristics (gender, ethnicity, age, and academic and disability histories), family background variables (SES, home language, and family structure), teacher characteristics (ethnicity, education, experience, paid and unpaid prep time, and job satisfaction), classroom characteristics (demographics, classroom activities, age distribution, and class size), and some school characteristics (location and calendar type). In addition, the ECLS-K contains behavior measures measured by the students' kindergarten teachers that are typically unobservable to the econometrician. Finally, I match schools to the National Center for Education Statistics (NCES) Common Core of Data (CCD) to obtain data on a number of other school characteristics: enrollment by grade level, demographics, free lunch eligibility, staffing, and district finances.

Table 1
Students, teachers, and schools by combination type.

Combination type	Number	Percentage
Students		
K–1 combination	170	2%
Single-grade first	9410	95%
1–2 combination	300	3%
Total	9880	100%
Teachers		
K–1 combination	70	2%
Single-grade first	2750	94%
1–2 combination	120	4%
Total	2940	100%
Schools		
Offers K–1 and 1	30	3%
Offers 1 only	890	88%
Offers 1 and 1–2	70	7%
Offers K–1, 1, and 1–2	20	2%
Total	1010	100%

Note: All unweighted sample sizes are rounded to the nearest ten to comply with the Institute for Education Sciences (IES) restricted-use data policy.

I restrict the sample to public-school first graders. Of these 12,870 children within 1540 schools, I further restrict the sample to those who are in a single-grade first-grade class or a K–1 or 1–2 combination, and attend schools offering at least one single-grade first-grade class. The final sample includes 9880 children with 2940 teachers in 1010 schools.[1]

Table 1 shows the number of students and teachers in and the number of schools offering each type of class. A large majority of students and teachers are in single-grade first-grade classes (94% and 95%, respectively), and a large majority of schools (approximately 90%) offer single-grade first-grade classes only. The most common type of combination class groups first and second graders. 63% of children in combination classes are in a 1–2

[1] In the results from the preferred regression model discussed in Section 6, the final analytic sample includes 6840 children with no missing data on the test score, child and family background, and teacher characteristic variables used in the model. All unweighted sample sizes are rounded to the nearest ten to comply with the Institute for Education Sciences (IES) restricted-use data policy.

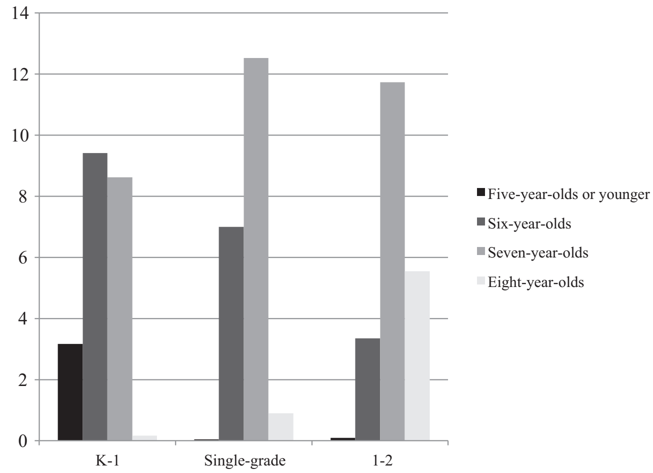

Fig. 1. Teacher-reported age distributions within combination and single-grade classes.

combination. Fifty-seven percent of schools offering at least one combination offer first and 1–2 only.

4. Mechanisms underlying the combination-class effect

In this paper, I am interested in the relationship between test scores and classroom organization in first grade. Combination classes differ from single-grade classes on several dimensions. Some of these differences are inherent to combination classes and would exist even if schools randomly decided to offer combination classes and if teachers and students were randomly assigned to these classes.

First, combination classes have a wider age span. K–1 classes have significantly more five- and six-year-olds than either single-grade or 1–2 classes, and significantly fewer seven- and eight-year-olds. 1–2 classes have significantly more eight-year-olds than K–1 and single-grade classes. Fig. 1 gives more information about the age distributions in the different class types.

Evidence on the effect of a wider age span within a classroom on student outcomes is inconclusive,[2] but there is evidence that students who are relatively older in their first year of school continue to outperform relatively younger students even after the age differences themselves become meaningless (Bedard & Dhuey, 2006).[3] First graders in K–1 combination classes are the relatively older students and may benefit from reviewing material they first encountered in kindergarten, as well as from the confidence they feel stemming from their status as the older children. Conversely, first graders in 1–2 combination classes, as the relatively younger students, may benefit from exposure to more advanced coursework.

In addition to their relative age differences, first-graders in K–1 and 1–2 combination classes experience different teaching methods and curricula. In a survey of 35 combination-class teachers in California, Mason, Burns, and Armesto (1993) find that teachers tend to use a mixed approach in combination classes, in which the teacher separates students by grade level for certain subjects such as math and reading and uses large-group instruction for subjects such as science and social studies. One of the criticisms of combination classes is that children are under-exposed to curriculum content because of the increased time spent preparing and teaching two curricula (Mason et al., 1993). Another potential criticism is that non-core topics such as social studies are neglected since teaching core topics like reading and math take more time.

In terms of teaching approach, ECLS-K data bear this out: children in combination classes receive less whole-class and more small-group instruction than single-grade children, as summarized in Table 2. In terms of subject matter, ECLS-K data show differences by grade level but do not indicate that non-core subjects are neglected in combination classes. K–1 students are taught reading and language arts, math, social studies, and science significantly less often than single-grade students (and all but math significantly less than 1–2 students), but are taught music and art more often. 1–2 students are taught math less often than single-grade students, but otherwise differences between subjects taught in 1–2 and single-grade classes are not significant.

In the next section, I discuss ways that combinations differ from single-grade classes in the context of systematic decision-making on the part of school administrators to offer combination classes and nonrandom assignment of students and teachers within schools to combination classes.

5. School-, teacher-, and student-level differences

5.1. How are combination schools different from single-grade schools?

Schools that decide to offer combination classes may be systematically different from those that do not. For example, multi-track year-round schools may have a small

[2] See, for example, Miller (1995) and Hattie (2002).

[3] Deming and Dynarski (2008) find that being relatively older is not necessarily beneficial, especially when lifetime outcomes are considered. They show that entering school later reduces educational attainment and depresses lifetime earnings.

Table 2
Duration of teaching methods and frequency of subjects taught in combination and single-grade classes.

	K–1	Single-grade	1–2
Whole-class instruction	3.915†	4.161	3.902†
	(0.070)	(0.006)	(0.048)
Small-group instruction	3.532†	3.347	3.499†
	(0.074)	(0.007)	(0.051)
Teach reading/language arts	4.927†,‡	4.983	4.978‡
	(0.017)	(0.002)	(0.011)
Teach math	4.867†	4.939	4.853†
	(0.033)	(0.003)	(0.023)
Teach social studies	3.411†	3.565	3.494
	(0.068)	(0.006)	(0.049)
Teach science	3.210†,‡	3.541	3.474‡
	(0.068)	(0.006)	(0.048)
Teach music	3.250†,‡	3.055	2.969‡
	(0.066)	(0.006)	(0.045)
Teach art	3.305†,‡	3.075	3.030‡
	(0.061)	(0.006)	(0.042)

Note: The table contains means with standard errors given in parentheses. Teachers answered the following question about whole-class and small-group activities: "In a typical day, how much time do the children spend on the following activities?" Answers range from 1 (no time) to 5 (3 h or more). In terms of subject matter, teachers answer the question, "How often do children in your class usually work on lessons or projects in the following general topic areas?" Answers range from 1 (never) to 5 (daily). The table contains the results of regressions of each variable on dummies for K–1 and 1–2 membership and school fixed effects. The constant in each regression is the within-school mean for single-grade classes. The combination-class means are computed by adding each coefficient to the constant. Data are unweighted.

† Indicates that the K–1 or 1–2 mean is significantly different from the single-grade mean at the 5% level.

‡ Indicates that the K–1 and the 1–2 means are significantly different at the 5% level.

Table 3
Comparing combination to single-grade schools.

	Single-grade schools	Combination schools
Percent minority	40.892***	49.978***
	(1.164)	(3.077)
Percent free-lunch eligible	37.165	41.390
	(0.949)	(2.472)
Pupil–teacher ratio	15.789***	17.511***
	(0.181)	(0.374)
Total enrollment	552.589	575.590
	(8.558)	(27.210)
Kindergarten enrollment	82.766	88.958
	(1.524)	(4.798)
First-grade enrollment	91.254	94.730
	(1.539)	(4.769)
Second-grade enrollment	89.381	93.314
	(1.465)	(4.814)
Local revenue as a percentage of total revenue	42.993***	38.381***
	(0.589)	(1.602)
Instructional expenses as a percentage of total expenses	52.331	52.440
	(0.237)	(0.653)
Year-round calendar	0.034***	0.110***
	(0.008)	(0.031)
Northeast	0.192	0.131
	(0.013)	(0.031)
Midwest	0.242*	0.164*
	(0.014)	(0.034)
South	0.370**	0.279**
	(0.016)	(0.041)
West	0.195***	0.426***
	(0.013)	(0.045)
City	0.386*	0.475*
	(0.016)	(0.045)
Suburb	0.406**	0.295**
	(0.017)	(0.041)
Town	0.088	0.115
	(0.010)	(0.029)
Rural	0.120	0.115
	(0.011)	(0.029)
K reading score for current first graders	50.233	50.113
	(0.246)	(0.510)
K math score for current first graders	50.256	49.500
	(0.243)	(0.508)
K general knowledge score for current first graders	49.597	49.583
	(0.260)	(0.531)

Note: Table contains means with standard errors given in parentheses. Statistical test performed is a two-sample t-test, assuming equal variances. Data are unweighted.

* Indicates that means are significantly different at the 10% level.

** Indicates that means are significantly different at the 5% level.

*** Indicates that means are significantly different at the 1% level.

number of students per grade level and choose to offer combination classes in order to use fewer classrooms. Calendar type may have an effect on student achievement independent of its association with combination classes. If a year-round calendar has a negative effect on student achievement, as shown in Graves (2010, 2011), the combination-class effect would be biased downwards in the absence of controls for calendar type.

Table 3 compares combination to single-grade schools. Combination schools are quite different on several measures and appear to be more disadvantaged than single-grade schools. They have a significantly higher percentage of minority students and a significantly higher pupil–teacher ratio. Local revenue as a percentage of total revenue is significantly lower for combination schools, though they spend about the same as single-grade schools on instructional expenses as a percentage of total expenses.

Schools differ on institutional characteristics and location as well. Combination schools are much more likely to have a year-round calendar. Burns, Mason, and Demiranda (1993) find that year-round principals, particularly those in multi-track schools, are constrained in their assignment of students to different types of classes since there are relatively low numbers of students in each grade level per track. Thus principals may have little choice but to combine adjacent grades into a combination class. Combination schools are also much more likely to be located in the West and much less likely to be located in suburban areas.

These differences highlight the importance of carefully controlling for school-level differences when estimating combination-class effects. In my main outcome regressions, I focus on within-school differences between combination- and single-grade students using school fixed effects. As a robustness check, I also match schools on their propensity to offer combination classes and consider associations between class type and student achievement in this matched-school sample.

5.2. How are combination teachers different from single-grade teachers?

In their survey of 72 school principals in California, Burns et al. (1993) find that the most common reasons

Table 4
Comparing combination to single-grade teachers.

	K–1	Single-grade	1–2
White	0.703	0.791	0.830
	(0.060)	(0.007)	(0.042)
Black	0.113	0.072	0.094
	(0.040)	(0.004)	(0.028)
Hispanic	0.143‡	0.091	0.016†,‡
	(0.044)	(0.005)	(0.031)
Asian	0.058	0.028	0.058
	(0.027)	(0.003)	(0.019)
Other	−0.017	0.019	0.003
	(0.022)	(0.002)	(0.015)
Male	0.002	0.030	0.001
	(0.030)	(0.003)	(0.021)
B.A. or less	0.280	0.277	0.310
	(0.076)	(0.008)	(0.052)
Some graduate school	0.260	0.328	0.326
	(0.081)	(0.009)	(0.056)
Graduate degree	0.460	0.395	0.363
	(0.082)	(0.009)	(0.057)
Years teaching	18.682†,‡	13.497	13.944‡
	(1.746)	(0.192)	(1.210)
Paid prep hours per week	3.588	4.122	3.925
	(0.385)	(0.042)	(0.270)
Unpaid prep hours per week	8.513	9.194	9.781
	(0.663)	(0.073)	(0.460)
Enjoys teaching	4.652†	4.398	4.439
	(0.128)	(0.014)	(0.088)
Thinks s/he makes a difference	4.771	4.517	4.641
	(0.104)	(0.011)	(0.072)
Would choose teaching again	4.310	4.284	4.458
	(0.166)	(0.018)	(0.115)

Note: The table contains means with standard errors given in parentheses. Teachers answered the following questions related to preparation time: "How many hours do you have designated as paid preparation periods per week?" and "Other than time spent during the work day, how many hours a week on average do you spend preparing for the class you teach?" Answers range from 1 (2 h or less per week) to 5 (15 or more hours per week). Teachers responded to the following statements with answers from 1(strongly disagree) to 5 (strongly agree): "I really enjoy my present teaching job," "I am certain I am making a difference in the lives of the children I teach," and "If I could start over, I would choose teaching again as my career." The table contains the results of regressions of each variable on dummies for K–1 and 1–2 class and school fixed effects. The constant in each regression is the within-school mean for single-grade teachers. The combination-class means are computed by adding each coefficient to the constant. Data are unweighted.

† Indicates that the K–1 or 1–2 mean is significantly different from the single-grade mean at the 5% level.

‡ Indicates that the K–1 and the 1–2 means are significantly different at the 5% level.

for assigning a teacher to a combination class are that a teacher volunteers, that a teacher is experienced, that the assignment is part of a teacher rotation process, and that the principal believes that the teacher is talented. If more talented teachers are assigned to combination classes, the positive effect of the teachers' skill will bias estimates of the combination-class effect upwards if it is not accounted for in the model. If, however, teachers are assigned on a rotating basis, the bias would be less severe.

Table 4 illustrates differences between combination and single-grade teachers. On most observable variables, the two types of teachers look similar. 1–2 teachers are significantly less likely than K–1 and single-grade teachers to be Hispanic. K–1 teachers have significantly more teaching experience than the other types of teachers. Rockoff (2004) finds that teaching experience significantly raises test scores, particularly in reading, highlighting the importance of controlling for observable teacher characteristics when estimating combination class effects.

K–1 teachers also enjoy teaching more than single-grade teachers. Principals may have selected them to teach K–1 classes on account of this trait. Alternatively, since enjoyment of teaching was measured during the first-grade year, this could be a result of teaching a K–1 class. Concerns about endogeneity preclude using enjoyment of teaching as a control variable, though the latter scenario is unlikely: 27 of the 35 teachers surveyed in Mason and Burns (1995) indicated that they preferred not to teach combination classes, many citing the extra work combination classes require.

It is possible, if not likely, that teachers differ on unobservable characteristics as well. However, to the extent that those unobservable characteristics are correlated with included variables such as teacher experience, concerns about bias are mitigated. Still, since sample size limitations preclude the use of school and teacher fixed effects in my outcome regressions, the possibility of nonrandom teacher assignment must be kept in mind when interpreting the results in Section 6.

5.3. *How are combination students different from single-grade students?*

In this section, I analyze student-level variables to determine if there is positive or negative selection into combination classes. According to a survey of principals, the main reason for nonrandom assignment is to make these classes more attractive to teachers (Burns et al., 1993). Generally, the goal is to make class ability more homogeneous than it would be under random assignment, or to populate the class with independent workers.

Combination and single-grade students look fairly similar in terms of most background characteristics, but there are notable differences indicative of positive selection into combination classes, particularly into the 1–2 option. Table 5 contains within-school means of student background characteristics by class type. Both K–1 and 1–2 students are older, on average, than single-grade students, though only the 1–2 difference is significant. This provides some evidence that more mature students are assigned to combination classes. Combination students are less likely to be Black. K–1 students are more likely than both 1–2 and single-grade students to have been diagnosed with a disability, indicating that they may have been placed there to catch up to their peers.[4] 1–2 students have significantly higher SES than the other two types, evidence of positive selection. Finally, K–1 students have more siblings than single-grade and 1–2 students.

[4] For a similar reason, one might think that students who repeated kindergarten may be more likely to be placed in a combination class. However, I find no significant differences in the percentage of first-time kindergarteners across class type. Because it might proxy for other characteristics, this variable is included as a covariate in the outcome regressions discussed in Section 6.

Table 5
Comparing combination to single-grade students: background characteristics.

	K–1	Single-grade	1–2
Male	0.402	0.512	0.492
	(0.507)	(0.005)	(0.039)
Age in months (spring first grade)	87.468	87.016	87.737†
	(0.480)	(0.044)	(0.324)
Speaks a language other than English at home	0.159	0.149	0.128
	(0.033)	(0.003)	(0.022)
White	0.539	0.541	0.578
	(0.041)	(0.004)	(0.028)
Black	0.089†	0.153	0.102†
	(0.029)	(0.003)	(0.020)
Hispanic	0.244	0.184	0.181
	(0.034)	(0.003)	(0.023)
Asian	0.091	0.069	0.082
	(0.026)	(0.002)	(0.017)
Other	0.037	0.054	0.057
	(0.022)	(0.002)	(0.015)
First-time kindergartener in 1998–1999	0.958	0.961	0.949
	(0.023)	(0.002)	(0.015)
Diagnosed with a disability	0.274†,‡	0.162	0.148‡
	(0.045)	(0.004)	(0.030)
SES	−0.080‡	−0.107	0.014†,‡
	(0.073)	(0.007)	(0.049)
Mom and dad married	0.735	0.654	0.712
	(0.055)	(0.005)	(0.037)
Number of siblings	1.921†,‡	1.532	1.607‡
	(0.137)	(0.012)	(0.092)

Note: The table contains means with standard errors given in parentheses from results of regressions of each variable on dummies for K–1 and 1–2 class and school fixed effects. The constant in each regression is the within-school mean for single-grade students. The combination-class means are computed by adding each coefficient to the constant. Data are unweighted.
† Indicates that the K–1 or 1–2 mean is significantly different from the single-grade mean at the 5% level.
‡ Indicates that the K–1 and the 1–2 means are significantly different at the 5% level.

A major advantage of using the ECLS-K in this analysis is that it contains several behavior measures that are typically unobservable to the econometrician. Students' kindergarten teachers rated their behavior along five dimensions. The Approaches to Learning Scale measures behaviors that affect the ease with which children can benefit from the learning environment. The Self-Control Scale has four items that indicate the child's ability to control behavior. The five Interpersonal Skills items rate the child's ability to get along with others. The Externalizing Problem Behaviors scale rates the frequency with which a child acts out, and the Internalizing Problem Behavior Scale asks about the apparent presence of anxiety, loneliness, low self-esteem, and sadness. Table 6 shows that combination students exhibit more positive and fewer negative behaviors than single-grade students. Only the 1–2 differences are significant, however.

In addition, Table 6 presents evidence that 1–2 classes contain more academically advanced students than other classes: 1–2 students have higher kindergarten reading, math, and general knowledge test scores than the other types of students. K–1 students have the lowest scores, but the differences between K–1 and single-grade scores are not significant.

Comparing combination to single-grade classes using answers from the teacher survey about the class as a whole,

Table 6
Comparing combination to single-grade students: kindergarten achievement and behavior.

	K–1	Single-grade	1–2
Kindergarten reading score	50.143‡	50.448	53.997†,‡
	(1.019)	(0.091)	(0.693)
Kindergarten math score	50.054‡	50.388	53.310†,‡
	(0.989)	(0.090)	(0.666)
Kindergarten general knowledge score	49.793‡	50.300	52.405†,‡
	(1.013)	(0.091)	(0.691)
Approaches to learning	3.211	3.142	3.304†
	(0.075)	(0.007)	(0.051)
Self-control	3.319	3.194	3.299†
	(0.069)	(0.006)	(0.047)
Interpersonal behavior	3.207	3.133	3.285†
	(0.071)	(0.006)	(0.048)
Externalizing problem behaviors	1.424†	1.645	1.517†
	(0.071)	(0.007)	(0.049)
Internalizing problem behaviors	1.526	1.552	1.514
	(0.058)	(0.005)	(0.039)

Note: The table contains means with standard errors given in parentheses from results of regressions of each variable on dummies for K–1 and 1–2 class and school fixed effects. The constant in each regression is the within-school mean for single-grade students. The combination-class means are computed by adding each coefficient to the constant. Data are unweighted.
† Indicates that the K–1 or 1–2 mean is significantly different from the single-grade mean at the 5% level.
‡ Indicates that the K–1 and the 1–2 means are significantly different at the 5% level.

I find that combination classes are larger, yet rated as better behaved as a group.[5] 1–2 classes are the largest with an average of 21.6 students, followed by K–1 classes with 21.5 students and single-grade classes with 20.5 students. K–1 and 1–2 classes are rated as 3.9 on a scale from 1, meaning, "Group misbehaves very frequently and is almost always difficult to handle" to 5, meaning "Group behaves exceptionally well." 1–2 classes earn a score of 3.4, on average.

Overall, I find that combination students are positively selected in terms of behavior, and that 1–2 students are also positively selected on academic ability. This will bias estimates of the combination-class effect unless I can control for the variables influencing class assignment. In the next section, I discuss my results as well as efforts to account for the systematic decision to offer combination classes at the school level and nonrandom assignment of students and teachers to combination classes.

6. Results

6.1. Student-level regressions

First, I discuss the results from five outcome regression models. The dependent variables are first-grade reading, math, and general knowledge test scores, standardized within the sample to have a mean of zero and a standard deviation of one. I run one regression per test score for a total of three regressions per model. The independent variables differ by model, but all include school fixed effects. Model 1 contains only dummies for class type, with

[5] The literature on the effect of class size on student achievement is extensive and results are mixed. See, for example, Cho, Glewwe, and Whitler (2012), Hoxby (2000) or Mishel and Rothstein (2002).

Table 7
Results of five outcome regression models.

	K–1 combination Coefficient (standard error)	1–2 combination Coefficient (standard error)	Adjusted R-squared
Dependent variable: first-grade reading test score			
Model 1	−0.030	0.261**	0.212
	(0.155)	(0.103)	
Model 2	0.003	0.025	0.656
	(0.114)	(0.067)	
Model 3	−0.026	0.028	0.661
	(0.121)	(0.068)	
Model 4	0.028	0.017	0.673
	(0.124)	(0.064)	
Model 5	0.013	0.011	0.674
	(0.128)	(0.068)	
Dependent variable: first-grade math test score			
Model 1	−0.050	0.324***	0.192
	(0.170)	(0.101)	
Model 2	0.008	0.091	0.652
	(0.138)	(0.061)	
Model 3	0.003	0.077	0.657
	(0.144)	(0.063)	
Model 4	0.033	0.075	0.666
	(0.168)	(0.065)	
Model 5	0.012	0.066	0.666
	(0.171)	(0.069)	
Dependent variable: first-grade general knowledge test score			
Model 1	−0.013	0.182**	0.306
	(0.120)	(0.086)	
Model 2	−0.029	−0.048	0.730
	(0.091)	(0.046)	
Model 3	−0.008	−0.048	0.727
	(0.093)	(0.048)	
Model 4	0.005	−0.039	0.726
	(0.115)	(0.046)	
Model 5	−0.007	−0.042	0.726
	(0.116)	(0.047)	

Note: The table contains regression results from five models. Model 1 contains only dummies for class type, with single-grade classes being the omitted category. Model 2 contains class-type dummies as well as kindergarten test scores. Model 3 contains dummies, scores, and kindergarten behavior measures. Model 4 contains dummies, kindergarten test scores, kindergarten behavior measures, and student background characteristics. Model 5 adds teacher characteristics: gender, race/ethnicity, education, and experience. All models contain school fixed effects. Standard errors are clustered at the school level. Data are unweighted.
** Indicates that means are significantly different at the 5% level.
*** Indicates that means are significantly different at the 1% level.

single-grade classes being the omitted category. Model 2 contains class-type dummies as well as kindergarten test scores. Model 3 contains dummies, kindergarten test scores, and kindergarten behavior measures. Model 4 contains dummies, kindergarten test scores, kindergarten behavior measures, and student background characteristics. Finally, Model 5 adds teacher characteristics: gender, race/ethnicity, education, and experience.[6]

Table 7 contains the coefficients on the K–1 and 1–2 dummies from each of the three regressions in each of the five models. The coefficient on the 1–2 dummy is highly significant in Model 1 for all of the outcomes, but becomes statistically indistinguishable from zero as more controls are added. The coefficient on the K–1 dummy is never significantly different from zero. These results illustrate the importance of including a detailed set of controls when trying to measure a combination-class effect, even when school fixed effects are included. Kindergarten test scores are particularly important. Results obtained without such controls could lead researchers and policymakers to draw incorrect conclusions about the effect of combination classes.

Note that the coefficients on both dummies in the reading and math test scores remain positive (though not significantly different from zero) when all controls are used. These results contradict those of Mariano and Kirby (2009), who find small but significant negative effects on language arts and math scores of combination-class membership in third grade in a 2–3 combination and fourth grade in a 4–5 combination in models that do not include school fixed effects or consider a matched sample of schools. As a robustness check and also to highlight the importance of accounting for cross-school differences, I estimate K–1 and 1–2 coefficients in separate regressions using a doubly robust technique following Mariano and Kirby (2009).

Doubly robust estimation is a two-stage process. In the first stage, propensity scoring techniques are used to weight the "control" (i.e., single-grade) students so that the distribution of their characteristics matches that of the "treatment" (i.e., combination-class) students' characteristics. In the second stage, these weights are used to estimate a "treatment" effect in a weighted multiple linear regression model (Mariano & Kirby, 2009). The estimator is doubly robust because it remains consistent when either the model of selection into treatment or the model for the distribution of the counterfactual data is correct (Bang & Robins, 2005).

Table 8 compares results using the doubly robust technique in two samples, using the control variables from Model 5: all schools, and a matched sample of 90 combination-class schools and 70 control schools.[7] I construct the matched sample in order to approximate an experiment in which schools were randomly assigned to one of two groups: a treatment group of schools offering combination classes and a comparison group of schools offering single-grade classes only. Doubly robust estimation on the unmatched sample only controls for observable student differences. Matching at the school level allows me to create two groups of schools that are similar in terms of observable characteristics but that differ in their single-grade or combination class offerings.

I estimated each school's propensity to offer combination classes based on percent minority in the school, percent free-lunch eligible, pupil–teacher ratio, total enrollment, enrollments in grades K, 1, and 2, district finances, calendar type, region and urbanicity, and kindergarten test scores. I then matched each treated (combination) school

[6] The Model 5 analysis sample contains 6840 students with no missing data. On average, the covariates have four percent missing data (range: zero to 13 percent). Results from Models 1 through 4 are similar whether or not I restrict each sample to the 6840 children in the Model 5 analysis sample.

[7] School fixed effects are not included in these estimations.

Table 8
Doubly robust estimation.

	First grade reading test score	First grade math test score	First grade general knowledge test score
Doubly robust estimate, all schools			
K–1	−0.111**	−0.147**	−0.076
	(0.053)	(0.061)	(0.049)
1–2	−0.043	−0.001	−0.029
	(0.042)	(0.042)	(0.041)
Doubly robust estimate, matched school sample			
K–1	0.029	0.146*	−0.033
	(0.058)	(0.079)	(0.062)
1–2	0.009	0.080*	0.001
	(0.043)	(0.046)	(0.048)

Note: This table contains doubly robust regression results for two samples: all schools, and a matched sample of 90 combination-class schools and 70 control schools. To obtain the matched sample, I estimated each school's propensity to offer combination classes and matched each treated (combination) school to its nearest control (single-grade) school, based on the absolute value of the difference in propensity scores. Control variables are those from Model 5, above, excluding school fixed effects. Standard errors are clustered at the school level.

* Indicates that means are significantly different at the 10% level.

** Indicates that means are significantly different at the 5% level.

to its nearest control (single-grade) school, based on the absolute value of the difference in propensity scores.

In the upper panel of the table, which includes results using all schools in the analysis sample, all coefficient estimates are negative and the coefficients on the K–1 dummies in the reading and math regressions are significant, similar to Mariano and Kirby's (2009) result. In the lower panel, which considers only the matched sample of schools, the coefficients in the reading and math regressions are positive. This is similar to the results from Model 5 in Table 7 (though the doubly robust estimates in the math regressions are significant at the ten percent level, whereas the Model 5 estimates are not significant). The results from Tables 7 and 8 show that it is important to take school-level differences as well as students' relative grade level into account when estimating combination-class effects.

6.2. First grade achievement in a matched sample of schools

Sims (2008) finds that children in schools with a higher percentage of students in combination classes perform worse than children in schools with fewer combination-class students. This could be because, once combination classes are implemented, single-grade students do worse than they would have if the school had not implemented combination classes, perhaps because resources are diverted to the combination classes and away from single-grade classes. This reasoning could explain why the estimates in Table 7, both significant and insignificant, were predominately positive.

As a check that, overall, first-graders are not harmed by a school's decision to offer combination classes, I compare student achievement in treated and control schools in my matched sample of schools. Table 9 shows that average first grade reading, math, and general knowledge test scores are statistically indistinguishable across treated and control schools. The implication is that, when similar schools are considered, offering combination classes in first grade has no relationship with first-grade achievement overall.[8]

Table 9
Treatment vs. control schools.

	Single-grade schools	Combination schools
First grade reading test score	50.457	49.970
	(0.506)	(0.497)
First grade math test score	50.406	50.147
	(0.500)	(0.470)
First grade general knowledge test score	48.790	49.286
	(0.757)	(0.572)

Note: This table contains means with standard errors given in parentheses. Statistical test performed is a two-sample *t*-test, assuming equal variances.

7. Conclusion

In this paper, I address the selection issues involved in all three levels of implementing a combination class. To address school-level selection, I use school fixed effects in the outcome regressions, and, as a robustness check, re-estimate my full model without school fixed effects on a sample of matched schools using a doubly robust estimation technique. I document observable teacher differences across class types with the caveat that unobservable teacher differences are likely to exist as well. I find evidence for positive selection into 1–2 classes, but using a rich set of control variables, find no relationship between class type and student achievement in first grade within schools, and no difference in overall first-grade achievement between single-grade and combination schools.

Of course, any observational technique, including propensity score matching and doubly robust estimation, is subject to omitted variable bias if all relevant variables are not included in the estimation. The results presented here suggest that combination classes in first grade have no effect on reading, math, or general knowledge test scores, but I cannot rule out the existence of bias due to unobservable differences between combination and single-grade students (or between combination and single-grade teachers).

One potential omitted variable is parental involvement. The direction of bias is unclear. On one hand, more motivated parents may resist placing their children in a combination class due to fears of watered-down curriculum or teacher distraction. These parents may also work more with their children to enhance their performance in school. This would inflate estimates of the effectiveness of single-grade classes relative to combination classes. On the other hand, more motivated parents may lobby for their children's placement in combination classes, biasing estimates of the combination class effect upwards. As Mariano and Kirby (2009) note, parents may want their children to be in the lower of the two grades in a combination class and could work with them to master the higher-grade material.

8 It is important to note that my matched sample of schools is small and these results may not generalize to a larger population of schools.

Parents who feel their children need review of concepts taught in previous years may lobby for their children to be placed in the higher of the two grades and also work with them to enhance school performance. Insofar as parental involvement is not captured by covariates such as pretest scores, behavior measures, and demographic characteristics, it remains a potential source of bias.

Combination classes are a cost-saving option that allows schools to use fewer teachers and classrooms. The results I present suggest that first graders are not harmed by being in a combination class or by their schools offering combination classes. As long as other stakeholders such as parents, teachers, and students in other grades are not made worse off, these results suggest that offering combination classes may be a Pareto-improving option for school administrators.

Extending this analysis to combination classes in later grades is an interesting avenue for future research. Such an analysis would inform the relative age literature by shedding light on how the effect of relative age changes as children get older.

Acknowledgements

I would like to thank my colleagues at Mathematica Policy Research and my faculty advisors at the University of California, San Diego (Drs. Julie Cullen, Nora Gordon, Julian Betts, and Kate Antonovics) for their insightful comments and suggestions.

References

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *69*, 962–972.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, *121*(4), 1437–1472.

Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. *American Education Research Journal*, *39*(1), 207–233.

Burns, R. B., Mason, D. A., & Demiranda, M. A. (1993). *How elementary principals assign teachers and students to combination classes*. Riverside: California Educational Research Cooperative, University of California.

Cho, H., Glewwe, P., & Whitler, M. (2012). Do reductions in class size raise students' test scores? Evidence from population variation in Minnesota's elementary schools. *Economics of Education Review*, *31*(3), 77–95.

Deming, D., & Dynarski, S. (2008). The lengthening of childhood. *Journal of Economic Perspectives*, *22*(3), 71–92.

Graves, J. (2010). The academic impact of multi-track year-round school calendars: A response to school overcrowding. *Journal of Urban Economics*, *67*(3), 378–391.

Graves, J. (2011). Effects of year-round schooling on disadvantaged students and the distribution of standardized test performance. *Economics of Education Review*, *30*(6), 1281–1305.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*(9), 449–481.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, *115*, 1239–1285.

Mariano, L. T., & Kirby, S. N. (2009). *Achievement of students in multigrade classrooms: Evidence from the Los Angeles Unified School District*. RAND Education Working Paper WR-685-IES.

Mason, D. A., & Burns, R. B. (1995). Teachers' views of combination classes. *Journal of Educational Research*, *89*(1), 36–45.

Mason, D. A., & Burns, R. B. (1997). Reassessing the effects of combination classes. *Educational Research and Evaluation*, *3*(1), 1–53.

Mason, D. A., Burns, R., & Armesto, J. (1993). *Teachers' views about combination classes*. Riverside: California Educational Research Cooperative, University of California.

Miller, W. (1995). *Are multi-age grouping practices a missing link in the educational reform debate? NAASP bulletin 1995*. Reston, VA: National Association of Secondary School Principals.

Mishel, L., & Rothstein, R. (Eds.). (2002). *The class size debate*. Washington, DC: Economic Policy Institute.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247–252.

Russell, V. J., Rowe, K. J., & Hill, P. W. (1998). *Effects of multigrade classes on student progress in literacy and numeracy: Quantitative evidence and perceptions of teachers and school leaders*. Paper presented at the 1998 annual conference of the Australian association for research in education.

Sims, D. (2008). A strategic response to class size reduction: Combination classes and student achievement in California. *Journal of Policy Analysis and Management*, *27*, 457–478.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, *65*(4), 319–381.

Wilkinson, I. A. G., & Hamilton, R. J. (2003). Learning to read in composite (multigrade) classes in New Zealand: Teachers make the difference. *Teaching and Teacher Education*, *19*(2), 221–235.