



Is identification with school the key component in the ‘Black Box’ of education outcomes? Evidence from a randomized experiment

Jason M. Fletcher*

School of Public Health, Yale University, 60 College Street, New Haven, CT 06510, United States

ARTICLE INFO

Article history:

Received 16 July 2008

Accepted 23 January 2009

JEL classification:

I2

I21

Keywords:

Human capital

Expenditures

Demand for schooling

ABSTRACT

In this paper, we follow up the important class size reduction randomized experiment in Tennessee in the mid 1980s (Project STAR) to attempt to further understand the long-lasting influences of early education interventions. While STAR led to large test score benefits during the intervention, these benefits quickly faded at its conclusion. However, research has recently shown that the STAR experiment led to long term benefits, including increases in college entrance exams participation (ACT/SAT), especially for minority students. We collect new follow up data on high school participation in extracurricular activities to examine whether (1) STAR increased participation in high school activities and (2) whether this increase in participation in high school is the explanation behind the long term benefits of the intervention. We find suggestive evidence that STAR did indeed increase some aspects of high school participation, including scholastic honors and participation in sports, especially for minority students. In contrast, we find little evidence that this increase in participation is the mechanism that has conferred higher rates of college-going to the STAR students.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Whether class size reduction programs represent effective and cost-effective school reforms continues to be one of the most controversial issues in education research. A new twist in the debate has recently been introduced by examinations of potential long term benefits to class size reductions in early elementary school. Using experimental data from Tennessee (Project STAR), researchers have consistently found clear initial achievement gains for students who were randomly assigned to small classes in early elementary school. However, these gains largely disappeared after the students returned to normal-size classes. Then in high school, 8 or 9 years later, students who were randomly assigned to small classes in early elementary school were more likely to take college entrance exams and attend college (Krueger & Whitmore, 2001). An essential question for researchers is why this effect exists. We examine

whether small class size in the early elementary years had an important (and yet unmeasured) enduring effect on these students' identification with school that then led to the long term benefits of the intervention.

In particular, we test two related hypotheses. The first hypothesis is that individuals in small classes in elementary school will, because of their increased identification with the schooling process, be more likely to participate in school activities. We chose to examine participation in school activities because the data are relatively easy to collect and the link between participation in school activities and identification with school is well established in the education literature (Finn, 1989; Holland & Andre, 1987; Ma, 2003). In order to test this hypothesis, we gather and combine new data on participation in high school from students who were involved in a randomized class size reduction experiment in early elementary school (Project STAR). The measures of participation that we use include sports, clubs, and other activities during high school. As it is well-documented that minority students lose substantial ground in achievement in the first years of schooling (e.g. Fryer & Levitt, 2004, 2006), we also examine whether there

* Tel.: +1 203 785 5760; fax: +1 203 785 6287.

E-mail address: jason.fletcher@yale.edu.

are differential effects of small classes on school identification for minority students.

For our second hypothesis, we test whether increased identification with school had an effect on whether a student took a college entrance exam. We test this hypothesis by examining the probability that an individual took a college entrance exam. Including our measure of identification with school should increase the likelihood of taking a college entrance exam and should reduce the relationship between small class size and the college outcome documented in previous research (Krueger & Whitmore, 2001). Importantly, the random assignment of students to classrooms in the Project STAR experiment should allow the relationships in both of our analyses to be interpreted as causal. Finding identification with school to be a critical element that makes schools work and educational interventions effective would imply that efforts to increase student identification with school should be a top priority in future education policy suggestions.

To preview our results, we find evidence that small classes during grades K-3 increased some measures of high school participation, especially for minority students. The results are strongest for participation in scholastic clubs and sports but weak for participation in non-scholastic clubs.

While we find causal evidence of effects of small class size on participation in high school activities, we are unable to link these increases in high school participation to an increase in the likelihood of the students taking a college entrance examination. Using additional measures of identification with school administered in 8th grade that have recently been provided in the public access STAR database, we also find that the increases in identification with school provided through the class size reduction seem to be orthogonal to the college exam participation. Although we are unable to provide the mechanism linking class size reductions with college exam participation, we extend the literature by providing new evidence of additional benefits from class size reduction in early elementary school—identification with school in 8th grade and participation in high school activities.

2. Background literature

There is a large literature documenting the long term benefits of many types of early investments in children, particularly disadvantaged children (see Heckman, 2006 for a recent summary; also see Reynolds, Wang, & Walberg, 2003). For example, the Perry Preschool Program, an early intervention on Black children, is estimated to have a rate of return of approximately 15% (Rolnick & Grunewald, 2003) and a benefit-cost ratio of over 8:1 due to its long term benefits in decreasing arrests, cash assistance, out-of wedlock births, etc. Importantly, this intervention showed small immediate benefits, and researchers hypothesize that the benefits accrued in part because the pre-schoolers “bonded” with their school (Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984; Finn, 1989). Another example of targeted early interventions with long term benefits is the Abecedarian program, which started when participants were 4 months old. This

expensive and intensive program has been shown to permanently raise IQ and non-cognitive skills. A broader set of investments societies make in children include a large range of educational inputs. In this paper, we focus on one such input: class size.

The debate over the benefits of reducing class size has raged over several decades. In fact, Hanushek (1999) states, “No topic in education has received such public and professional attention as class size”. The Tennessee Student/Teacher Achievement Ratio (STAR) program has been at the center of the debate because it was a large scale randomized experiment designed specifically to examine the efficacy of class size reductions. During the STAR experiment over 11,000 students in 79 schools were randomly assigned to small classes (13–17 students per teacher), regular classes (22–25 students per teacher, and regular-with-aide classes (22–25 students with a full-time teacher’s aide).

While there is much debate regarding the robustness of some of the findings of the STAR experiment, several research findings appear uncontroversial. Achievement for students in small classes increased more than their counterparts during the experiment (Kindergarten through 3rd grade). Additionally, the achievement-test gains for students in small classes continued through 4th grade (Finn, Fulton, Zaharias, & Nye, 1989). The consensus for the effects of class size on student outcomes stops at the 4th grade. Finn, Gerber, Achilles, and Boyd-Zaharias (2001) present evidence that achievement gains continue through 8th grade for students who were in small classes between Kindergarten and 3rd grade. Hanushek (1999) analyzes the same data and is not persuaded of long-lasting achievement gains. Additionally, Hanushek stresses two reasons to be skeptical of the effectiveness of class size reductions: there have been long term decreases in class size in the US with no apparent effect on achievement and class size reductions appear to have a one-time effect on achievement rather than a cumulative effect. Krueger and Whitmore (2001) similarly show that the initial test score gains seem to fade after the children return to regular class sizes. However, all researchers emphasize the potential for unmeasured benefits from class size that could be both important and long term.

Longer term effects of small classes have recently been examined. In particular, Krueger and Whitmore (2001) present a perplexing finding. They document that students who were randomized to small classes in Project STAR in grades K-3 were more likely to take college entrance exams and attend college 8 or 9 years after the class size program ended. This increase in college entrance exams was particularly strong for Black students and students on free/reduced lunch. Resolving the finding that small class size had initial effects on achievement that faded with time, and then the effects reemerged at the end of high school is of fundamental importance to improving education. The uncertainty in whether class size reductions are effective (and if so, why?) leaves researchers, practitioners, and parents with more questions than answers. In this paper we seek to answer this question by examining whether class size reductions increased students’ identification with school and whether this increase in identification with school is the mecha-

nism linking class size reductions to college entrance exam participation.

For our conceptual framework, we draw on recent work in economics by Akerlof and Kranton (2000, 2002) as well as other social scientists (e.g. Voelkl, 1997) on the importance of student identification with school ideals for educational outcomes. This framework predicts that exposure to small classes better acclimates students to being in school and develops a sense of belonging, with the result of better education outcomes. These benefits are unlikely to be documented by solely relying on standardized test scores (see also Finn, Pannozzo, & Achilles, 2003; Sizer, 1984).

For our measures of identification with school we both collect new data on participation in high school activities as well as investigate largely-unexamined measures of identification with school collected in 8th grade. Participation in high school activities has been shown to be strongly linked with measures of identification with school by educational researchers. Ma (2003) suggests that participation in school activities may be the key to students' sense of belonging in school. Finn (1989) states that the ability to increase participation in school activities may provide a straightforward way of increasing identification with school. Holland and Andre (1987) show that participation in school activities is highly correlated with self-esteem, feelings of self-control, and educational aspirations. These and other educational researchers have strongly suggested that participation in school activities is an important component and necessary condition for students' identification with school (Rutter, Maughan, Mortimore, Ouston, & Smith, 1979).

3. Empirical methods

Our methodology is relatively straightforward. The random assignment of students and teachers to classrooms in the Project STAR experiment allows us to utilize relatively simple research methods in testing our hypotheses.¹ Our analysis has two parts. We first estimate whether there is a causal relationship between small class size and participation, which is our measure of identification with school. In particular, we estimate the following empirical model:

$$A_{is} = \beta X_{is} + \delta \text{Classtype}_{is} + \theta_s + \varepsilon_{is} \quad (1)$$

where student i in school s chooses activity A depending on individual characteristics (e.g. gender, race, poverty status), class type (small class, regular class, regular class with teacher's aide), school fixed effects and an idiosyncratic shock. In some specifications, we will also include an interaction term between race and small class size, as there is substantial evidence that small class size benefits Black students more than White students. Additionally, as we have rich participation data, we examine the differences in the number of activities that students are involved in and

the types of activities to further examine the causal link between small class size and participation behaviors.

For the second part of our analysis we will test whether identification with school is the missing piece in the puzzle of why small class sizes led to increases in college test-taking. We test this hypothesis by predicting college test-taking with school participation and small class size as primary control variables. If our hypothesis is correct, using school participation in the regression should substantially shrink (and may eliminate) the direct effect of small class size on college test-taking found in Krueger and Whitmore (2001). This finding would be consistent with our hypothesis and implies that it was not small class size *per se* that led to increases in college entrance exam participation. Rather, class size reductions increased identification with school, which led to the increase in college entrance exam participation. Importantly, the random assignment of students to classrooms in the Project STAR experiment allows the statistical relationships found in our analyses to have a causal interpretation.

4. Data

For the data collection aspect of our work, we linked currently available information for students in the Project STAR class size experiment in Tennessee² with school yearbook data for these students during their high school years. We accomplished this by collecting school yearbooks from high schools the STAR students attended between 1995 and 1998 (the on-time year of graduation for STAR students was 1998). We matched names of STAR students with yearbook photographs in order to create counts of high school participation for each student in our sample. Since not all yearbooks were available for all years and not all schools were cooperative with our data collection activities, we have incomplete participation information for the STAR students (see the data appendix for further details of the creation of our measures). Importantly, we use within-school estimators in our analysis to control for between-school differences in access to information as well as the type of information contained in each high school's yearbook. Overall, we collected information on over 1200 STAR students in over 20 high schools. Summary statistics for the full sample of STAR students as well as our follow up sample are presented in Tables 1 and 2. We follow Krueger and Whitmore (2001) and combine the White and (small number of) Asian students. All other students are considered "non-White."

Our main categories of high school activities include sports, non-scholastic club, scholastic honors/clubs, music, and superlatives.³ Thirty-one percent of our sample was involved in some high school activity. The largest single

² The public use version of this dataset is available online: <http://www.heros-inc.org/data.htm>.

³ Examples of each category include: sports (varsity and junior varsity football, soccer, baseball, basketball), clubs (student government, fellowship of Christian athletes, language clubs, DECA, Future Farmers of America), honors (beta club, academic awards), music (choir, marching band, ensemble), superlatives (Mr./Ms. High school, Homecoming court, Class officer).

¹ The integrity of the initial (Kindergarten) random assignment of students has been verified statistically by several researchers, including Krueger and Whitmore (2001). Some researchers have questioned the random assignment of children in grades 1–3 (Ding & Lehrer, in press). We estimate two sets of results: (1) using only Kindergarten assignment and (2) using initial assignment in all grades as done in Krueger and Whitmore.

Table 1
Summary statistics. Project STAR yearbook follow up.

Variable	Year	Obs	Mean	Std. dev	Min	Max
Male	All	1213	0.49	0.50	0	1
Non-White	All	1212	0.29	0.45	0	1
Small class	Entry	1213	0.26	0.44	0	1
Small class	Kindergarten	732	0.30	0.46	0	1
Small class	1st grade	880	0.30	0.46	0	1
Reduced/free lunch	Entry	1169	0.41	0.49	0	1
Reduced/free lunch	Kindergarten	730	0.36	0.48	0	1
Reduced/free lunch	1st grade	855	0.39	0.49	0	1
Special education	Entry	952	0.01	0.09	0	1
Special education	Kindergarten	732	0.05	0.22	0	1
Special education	1st grade	880	0.01	0.08	0	1
Reading test score	Kindergarten	684	442.19	32.09	380	627
Math test score	Kindergarten	692	494.01	45.03	354	626
Average test score	Kindergarten	684	55.35	25.17	0.60	100
Average test score	8th grade	1013	53.03	25.46	0	99
ACT/SAT test	High school	1213	0.57	0.49	0	1
All activities	High school	1213	0.65	1.29	0	9
Any activity	High school	1213	0.31	0.46	0	1
Sports activities	High school	1213	0.17	0.46	0	3
Any sports	High school	1213	0.14	0.35	0	1
Club activities	High school	1213	0.29	0.72	0	5
Any clubs	High school	1213	0.18	0.39	0	1
Superlatives	High school	1213	0.08	0.35	0	4
Any superlative	High school	1213	0.06	0.23	0	1
Honors activities	High school	1213	0.08	0.31	0	3
Any honors	High school	1213	0.07	0.25	0	1
Music activities	High school	1213	0.04	0.20	0	2
Any music	High school	1213	0.04	0.18	0	1
Identification scale	8th grade	727	35.84	6.45	14	67

category of activity was clubs, where 18% of the students participated, followed by sports (14%), honors (7%), superlatives (6%), and music (4%). Table A1 stratifies the sample by race. The non-White students are more likely to receive free/reduced lunch, have lower test scores and are less likely to participate in all of the participation categories we measure—the participation rates for any activities are 35% for White and 20% for non-Whites.

Table A2 shows that the characteristics of our follow up sample match the full sample in some dimensions, though

Table 2
Probability of appearing in high school follow up sample.

	HS sample	HS sample
Small K	−0.004 (0.008)	−0.003 (0.008)
Aide K	−0.006 (0.006)	−0.007 (0.006)
Free lunch	−0.050*** (0.013)	−0.030*** (0.010)
White	−0.004 (0.010)	0.001 (0.010)
Female	0.021*** (0.006)	0.009 (0.006)
In 8th grade sample		0.114*** (0.017)
Constant	0.145*** (0.011)	0.074*** (0.011)
Observations	11029	11029
School × entry wave fixed effects	Yes	Yes
R-squared	0.01	0.04
Number of schools	79	79

Robust standard errors in parentheses.

*** $p < 0.01$.

our follow up sample is more likely to be White, less likely to receive free/reduced lunch, and our sample has higher test scores. Since the original random assignment was done within-schools, all our analyses use school fixed effects or school by entry wave fixed effects. Table 2 predicts the correlates of being in our sample. Since our data collection was done at the school level and our analysis is within-schools, the most likely reason that our sample may not reflect the distribution of characteristics of the full sample is attrition. Students enter the sample between grades Kindergarten and 3rd grade. We collect information for students in grades 9–12. Students who change schools are less likely to be captured in our sample. In column 1, we show that free/reduced lunch status is negatively related with being followed in our sample. Females are more likely to be in our sample than males. When we control for whether the student was still followed through 8th grade, both associations shrink considerably.

Next, in Table 3, we show that the measured effects of attending a small class are similar in our sample as the full sample. Following Krueger and Whitmore (2001), we estimate the following set of empirical models:

$$Y_{isg} = \beta_{0g} + \beta_{1g}Small_{is} + \beta_{2g}X_{is} + \theta_{sw} + \varepsilon_{isg}$$

where Y_{isg} represents the test score percentile rank for student i in grade g ($g = K, \dots, 8$) who initially attended school s ($s = 1, \dots, 79$) in wave w . $Small$ is a dummy variable that equals one if the student was initially assigned to a small class. X_{is} is a vector of covariates such as sex, race, and free/reduced price lunch status at school entry. θ_{sw} is a set of school by entry wave dummy variables. To ensure that the coefficient estimates on our class-type variables are not

Table 3

Replicating test score gains from small class size. Full sample vs. high school follow up sample.

Grade	K	K	1	1	2	2	3	3	4	4
Small K class	6.094*** (1.576)	5.397 (3.333)	7.710*** (1.389)	8.799** (3.323)	6.850*** (1.471)	5.993* (3.450)	5.578*** (1.228)	7.773** (2.932)	3.983*** (1.291)	2.908 (3.299)
Aide K class	0.460 (1.471)	-3.955 (2.855)	1.255 (0.811)	-0.580 (2.341)	1.298 (0.967)	-1.505 (2.258)	0.350 (0.879)	-2.544 (1.901)	-0.525 (0.892)	-3.251* (1.891)
Non-White	-10.426*** (1.646)	-10.382** (4.318)	-8.056*** (1.426)	-11.733*** (3.269)	-8.331*** (1.522)	-9.652** (3.729)	-7.300*** (1.616)	-7.149 (5.147)	-6.080*** (1.981)	-5.989 (4.955)
Female	5.134*** (0.638)	3.958** (1.829)	3.300*** (0.699)	3.043* (1.644)	3.560*** (0.679)	1.828 (1.507)	3.338*** (0.771)	3.433** (1.427)	5.540*** (0.935)	5.936*** (1.828)
Free lunch status	-14.476*** (1.013)	-12.860*** (2.993)	-14.586*** (0.943)	-12.230*** (1.579)	-14.824*** (0.756)	-12.472*** (1.937)	-13.338*** (0.924)	-14.637*** (2.061)	-12.418*** (0.899)	-13.792*** (2.136)
Constant	54.309*** (1.154)	57.794*** (2.529)	54.841*** (0.940)	56.489*** (1.770)	55.504*** (0.991)	58.755*** (2.031)	54.976*** (0.879)	57.559*** (1.750)	53.103*** (0.776)	54.776*** (1.362)
Observations	5619	788	6148	904	5814	919	5781	907	3861	699
School × entry wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.09	0.08	0.09	0.09	0.09	0.06	0.07	0.08	0.07	0.08
Number of schools	79	68	77	63	76	50	75	51	70	37
Sample	5 Full	5 HS	6 Full	6 HS	7 Full	7 HS	8 Full	8 HS	8 Full	8 HS
Small K class	2.256 (1.546)	1.945 (2.410)	2.496*** (0.935)	2.336 (2.160)	3.573*** (1.078)	1.944 (2.491)	2.385*** (0.894)	1.730 (1.886)		
Aide K class	-2.521** (1.224)	-3.396 (2.533)	-0.207 (0.671)	0.203 (1.811)	-0.668 (0.894)	-2.639 (2.394)	-1.036 (0.758)	-0.248 (1.866)		
Non-White	-7.840*** (1.950)	-9.135 (6.002)	-9.260*** (1.480)	-6.140 (4.151)	-12.090*** (2.193)	-9.522* (4.866)	-12.094*** (1.161)	-10.896*** (3.937)		
Female	3.968*** (1.077)	3.849 (2.329)	2.230*** (0.702)	1.592 (1.498)	3.552*** (0.737)	3.309** (1.339)	3.049*** (0.690)	1.661 (1.436)		
Free lunch status	-13.802*** (1.125)	-13.405*** (2.624)	-14.993*** (0.759)	-14.312*** (1.655)	-14.609*** (0.806)	-12.445*** (2.572)	-14.655*** (0.691)	-13.127*** (1.832)		
Constant	54.298*** (1.082)	55.286*** (1.687)	58.124*** (0.709)	58.837*** (1.777)	58.967*** (0.999)	59.019*** (2.222)	58.879*** (0.576)	60.099*** (1.523)		
Observations	2499	573	6158	1065	4685	887	5930	1071		
School × entry wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
R-squared	0.07	0.06	0.08	0.06	0.09	0.06	0.09	0.07		
Number of schools	58	33	80	63	79	61	79	62		

Robust standard errors in parentheses.

* $p < 0.1$.** $p < 0.05$.*** $p < 0.01$.

subject to bias because of potential non-random sorting after the initial class assignments, these dummy variables are based only on the initial assignment and are fixed.

Our results are similar to those presented elsewhere (e.g. Krueger & Whitmore, 2001). In particular, we show that the large test score differences remain throughout the intervention (K-3) but they are largely reduced following the conclusion of the intervention. Importantly, the estimates for our follow up sample are quite similar to those for the main sample, though because of our smaller sample size are often less precisely estimated.

5. Main results

We now turn to our main results. First, we examine the causal effects of small class size on participation in a variety of high school activities. Then, we examine whether high school activities, our measure of identification with school, are the link between class size reduction and long term benefits, specifically college test-taking.

Table 4 presents results for our five categories of high school participation as well as a combined category. In these results, all outcomes measures the participate/no

participate decision. Results that examine the number of activities are similar and are available upon request. For each category, we examine the main effect of small class size and then the interaction effect between small class size and being non-White. For the superlatives and sports categories, we find some suggestive evidence that small class size in early elementary school increases these activities. However, for music, honors, and clubs, the results are negative. Interestingly, the interaction between small class size and non-White status suggests a differential benefit for non-White students in all categories except clubs. Other results include that students in poverty (during elementary school) are far less likely to be involved in all high school activities. Overall, we find mixed evidence of the causal increases in high school participation in activities stemming from early class size interventions.

As a supplementary analysis, we use pre-existing data collected on a sub-sample of the STAR students during 8th grade that sought to measure identification with school using self-reports. The scale was computed as the sum of 16 items rated by the student, such as "School is one of my favorite places to be". This measure has been shown to be correlated with 4th grade teacher-ratings of partic-

Table 4
Effects of small Kindergarten class size on high school participation.

	Superlatives	Superlatives	Sports	Sports	Music	Music	Music	Honors	Honors	Clubs	Clubs	Total	Total
Small K class	0.032 (0.020)	0.018 (0.016)	0.001 (0.028)	-0.004 (0.030)	-0.031* (0.017)	-0.043** (0.019)	-0.022 (0.031)	-0.033 (0.035)	-0.010 (0.030)	-0.004 (0.036)	-0.055 (0.036)	-0.072*** (0.035)	-0.072*** (0.035)
Small K class × Non-White		0.069 (0.058)		0.023 (0.056)		0.056*** (0.019)		0.055 (0.049)		-0.033 (0.043)		0.083 (0.064)	0.083 (0.064)
Aide K class	0.034 (0.023)	0.034 (0.023)	-0.002 (0.029)	-0.001 (0.029)	0.009 (0.022)	0.010 (0.022)	-0.023 (0.025)	-0.022 (0.025)	0.009 (0.036)	0.008 (0.036)	-0.002 (0.046)	-0.001 (0.046)	-0.001 (0.046)
Non-White	0.115* (0.061)	0.097 (0.059)	0.229* (0.122)	0.223* (0.123)	-0.040** (0.018)	-0.054*** (0.018)	0.092 (0.073)	0.078 (0.078)	0.016 (0.107)	0.025 (0.107)	0.106 (0.095)	0.084 (0.096)	0.084 (0.096)
Female	0.013 (0.014)	0.012 (0.014)	-0.029 (0.021)	-0.029 (0.021)	0.016 (0.016)	0.015 (0.016)	0.051 (0.025)	0.050 (0.025)	0.115*** (0.033)	0.116*** (0.033)	0.080** (0.035)	0.079** (0.035)	0.079** (0.035)
Free lunch K	-0.087*** (0.026)	-0.087*** (0.026)	-0.137*** (0.033)	-0.137*** (0.033)	-0.007 (0.019)	-0.008 (0.019)	-0.116*** (0.030)	-0.117*** (0.030)	-0.085 (0.044)	-0.085 (0.044)	-0.179*** (0.030)	-0.180*** (0.030)	-0.180*** (0.030)
Constant	0.046** (0.020)	0.050** (0.019)	0.175*** (0.037)	0.177*** (0.037)	0.056** (0.012)	0.059*** (0.012)	0.127*** (0.025)	0.130*** (0.025)	0.152*** (0.036)	0.150*** (0.037)	0.365*** (0.036)	0.369*** (0.035)	0.369*** (0.035)
Observations	730	730	730	730	730	730	730	730	730	730	730	730	730
R-squared	0.02	0.03	0.03	0.03	0.01	0.01	0.03	0.03	0.03	0.03	0.03	0.04	0.04
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of schools	55	55	55	55	55	55	55	55	55	55	55	55	55

Robust standard errors in parentheses. Clustered standard errors (Kindergarten school level).

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

Table 5

Effects of small Kindergarten class size on 8th grade identification with school total score.

Sample	Identification score			
	Full	Full	HS	HS
Small K class	0.599* (0.314)	0.661** (0.326)	0.109 (0.688)	0.094 (0.707)
Aide K class	0.228 (0.362)	0.226 (0.362)	-0.336 (0.536)	-0.334 (0.539)
Non-White	0.055 (0.779)	0.133 (0.774)	-3.525 (2.180)	-3.554 (2.141)
Female	-2.260*** (0.311)	-2.261*** (0.311)	-2.324*** (0.696)	-2.323*** (0.694)
Free lunch	0.478 (0.388)	0.471 (0.391)	1.125 (0.876)	1.125 (0.876)
Small K class × Non-White		-0.270 (0.624)		0.142 (2.549)
Constant	37.019*** (0.371)	37.005*** (0.366)	37.439*** (0.588)	37.440*** (0.588)
School fixed effects	Yes	Yes	Yes	Yes
Observations	2136	2136	501	501
Number of schools	78	78	51	51
R-squared	0.04	0.04	0.05	0.05

Robust standard errors in parentheses. Clustered standard errors (Kindergarten school level).

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

ipation as well as test scores (see Voelkl, 1997 for further details for this measure). Table 5 shows that in the full sample, small class size does seem to increase the identification with school scale. However, this increase is not more beneficial to non-White students, and the estimated sign of the interaction term is negative. This result suggests that this measure of identification with school will not be able to provide the link between class size and college test-taking, particularly because there is no positive differential impact on non-White students. The results using our follow up sample are also mixed. The estimated magnitudes of the effects are much smaller and not statistically significant.

While we find only mixed evidence for a strong causal link between small class size and high school participation or 8th grade identification with school, we proceed to examine whether we find any evidence that our measures of identification with school could be an explanation behind the long term benefit of small class size on college test-taking. In Table 6, we predict whether students take either the ACT or SAT college entrance test. We follow Krueger and Whitmore (2001) and utilize the full sample, comparing students who entered the experiment by controlling for school by entry wave fixed effects in all specifications. In column 1, our estimate of the causal effect of small class size on college test-taking is a 2.3 percentage point increase compared with estimates between 2.7 and 1.7 by Krueger and Whitmore.⁴ In column 2 of Table 6 we estimate the same specification using only our follow up

⁴ These coefficients are from Table 4, column 3 and Table 5, column 3, respectively in Krueger and Whitmore (2001). Krueger and Whitmore use a somewhat smaller sample as they restricted their sample to individuals who were on grade level for the first result. The second result uses the full sample.

Table 6

Replication of Krueger/Whitmore, comparison with high school sample, and inclusion of high school activities.

	KW replication Full	KW replication HS	Activities HS	KW replication Black	KW replication HS/Black	Activities HS/Black
Small K class	0.023* (0.013)	0.027 (0.035)	0.032 (0.035)	0.045* (0.023)	0.116** (0.046)	0.110** (0.046)
Non-White	0.051*** (0.019)	0.142** (0.063)	0.117** (0.056)			
Female	0.149*** (0.010)	0.150*** (0.021)	0.129*** (0.022)	0.140*** (0.018)	0.113** (0.044)	0.105** (0.044)
Aide K class	0.008 (0.010)	0.061** (0.029)	0.070*** (0.025)	0.024 (0.016)	0.122** (0.056)	0.133** (0.054)
Free/reduced lunch	−0.306*** (0.012)	−0.302*** (0.033)	−0.259*** (0.032)	−0.249*** (0.021)	−0.229*** (0.074)	−0.202*** (0.066)
Activities			0.096*** (0.010)			0.094*** (0.023)
Constant	0.411*** (0.012)	0.531*** (0.026)	0.462*** (0.027)	0.391*** (0.019)	0.579*** (0.069)	0.517*** (0.065)
Observations	11,027	1340	1340	4101	391	391
School × entry wave FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.10	0.09	0.14	0.07	0.05	0.09
Number of schools	80	73	73	68	45	45

Robust standard errors in parentheses. Clustered standard errors (School entry by wave).

* $p < 0.1$.** $p < 0.05$.*** $p < 0.01$.

sample. Importantly, the causal effect of small class size on college test-taking is quite similar—2.7 percentage points. On the other hand, we do find some other differences, such as the correlation between non-White and college test-taking is much higher in our sample than the full sample. In contrast, the coefficients for female and free/reduced lunch are nearly identical. In column 3, we control for whether the student participated in any activity during high school. The correlation between our measure of participation and college test-taking suggests an increase of 9.6 percentage points. However, controlling for high school participation does not seem to substantially change the relationship between small class size and college test-taking. This is evidence that the channel linking small class size and college test-taking may not be high school participation, which is evidence against our second hypothesis.

In columns 4–6, we repeat our analysis on the sample of Black students. Our replication of Krueger and Whitmore suggests a causal effect of small class size on college test-taking of 4.5 percentage points for Black students—Krueger and Whitmore estimate effects between 5.7 and 4.1, depending on the sample. When we estimate the same specification in column 4 on our high school follow up sample, we find much larger causal effects of small class size on college test-taking. Our estimate increases from 4.5 points to 11.6 points. This suggests that we (unintentionally) collected follow up data from schools where students received larger class size benefits. Nonetheless, our results are similar to those in columns 1–3: the magnitude of the relationship between participation in high school activities and college test-taking is 9.6 points and controlling for participation in high school does not substantively alter the relationship between small class size and college test-taking. Again, the results suggest that identification with school is not the mechanism creating the long term benefits of small class size.

6. Conclusions

In this paper, we follow up the important class size reduction randomized experiment in Tennessee in the mid 1980s (Project STAR) to attempt to further understand the long-lasting influences of early education interventions. An empirical puzzle in the economics of education literature is *why* some early educational interventions have long term benefits and other do not. In fact, some long term benefits appear unexpectedly, as intermediate (measurable) benefits are small or non-existent. This is the case for the timing of benefits for the randomized class size reduction experiment, Project STAR. During the intervention, small class size increased test scores considerably, but these test score gains largely diminished following the conclusion of the experiment. Several papers have then shown long term benefits accruing more than 10 years after the experiment, including an increase in college test-taking. These long term benefits also seem to have disproportionately accrued to minority students. In this paper, we examine whether identification with school is the missing link tying together the intervention and the long term outcomes.

In particular, we pursue two complementary research goals. First, we examine whether small class size causally increased participation in high school activities. We find moderate evidence of these effects, which is stronger for superlatives and sports but not participation in music activities. Second, we examine whether this increase in participation in high school is the explanation behind the long term benefits of the intervention. We find little evidence that this increase in participation is the mechanism that has conferred higher rates of college-going to the STAR students. Therefore, our results suggest that although increasing high school participation is another likely long term benefit from class size reductions in early elementary

school, the mechanism linking other long term benefits, such as college test-taking, remains unclear.

Acknowledgements

The author is especially grateful to Helen Pate-Bain and Jayne Zaharias for helpful discussions, advice, and providing the original Project STAR data. The author also thanks George Akerlof and Barbara Wolfe for helpful discussions and advice throughout the project. Rachel Croson, Rachel Kranton, and participants at the American Economic Association and American Education Finance

Association Annual meetings provided useful comments. The author thanks Peter Dale, Alicia Dumont, Laura Kelley, Don Mitchell, Paula Tapner, and Cliff Tonsberg for excellent research assistance in collecting the yearbooks and Darryn Beckstrom and Angeline Liu for excellent research assistance in coding the yearbook data. This research has been supported by the Russell Sage Foundation Small Grants in Behavioral Economics Program.

Appendix A.

See Tables A1–A3.

Table A1
Summary statistics. Comparison of full sample vs. high school follow up.

Variable	Year	Obs	Mean	Std. dev	Obs	Mean	Std. dev
Male	All	11,405	0.53	0.50	1213	0.49	0.50
Non-White	All	11,291	0.37	0.48	1212	0.29	0.45
Small class	Entry	11,425	0.26	0.44	1213	0.26	0.44
Small class	Kindergarten	6,189	0.30	0.46	732	0.30	0.46
Small class	1st grade	6,732	0.28	0.45	880	0.30	0.46
Reduced/free lunch	Entry	10,977	0.55	0.50	1169	0.41	0.49
Reduced/free lunch	Kindergarten	6,164	0.48	0.50	730	0.36	0.48
Reduced/free lunch	1st grade	6,557	0.51	0.50	855	0.39	0.49
Special education	Entry	8,465	0.02	0.14	952	0.01	0.09
Special education	Kindergarten	6,188	0.03	0.18	732	0.05	0.22
Special education	1st grade	6,729	0.01	0.11	880	0.01	0.08
Reading test score	Kindergarten	5,662	436.91	31.78	684	442.19	32.09
Math test score	Kindergarten	5,740	485.67	47.69	692	494.01	45.03
Average test score	Kindergarten	5,659	48.58	26.79	684	55.35	25.17
Average test score	8th grade	6,075	49.66	26.51	1013	53.03	25.46
Identification scale	8th grade	3,601	36.35	6.50	727	35.84	6.45

Table A2
Summary statistics of high school follow up sample. Black and White students.

Variable	Year	Obs	Mean	Std. dev	Obs	Mean	Std. dev
Male	All	347	0.44	0.50	865	0.51	0.50
Non-White	All	347	1.00	0.00	865	0.00	0.00
Small class	Entry	347	0.24	0.43	865	0.27	0.45
Small class	Kindergarten	170	0.31	0.46	562	0.30	0.46
Small class	1st grade	207	0.29	0.46	672	0.30	0.46
Reduced/free lunch	Entry	334	0.81	0.39	834	0.25	0.43
Reduced/free lunch	Kindergarten	170	0.84	0.37	560	0.22	0.41
Reduced/free lunch	1st grade	206	0.84	0.37	648	0.24	0.43
Special education	Entry	231	0.01	0.09	720	0.01	0.09
Special education	Kindergarten	170	0.02	0.13	562	0.06	0.25
Special education	1st grade	207	0.00	0.07	672	0.01	0.08
Reading test score	Kindergarten	155	436.85	29.68	529	443.75	32.62
Math test score	Kindergarten	160	487.94	49.93	532	495.84	43.34
Average test score	Kindergarten	155	56.39	26.05	529	55.05	24.92
Average test score	8th grade	275	51.03	25.70	737	53.72	25.32
ACT/SAT test	High school	347	0.55	0.50	865	0.58	0.49
All activities	High school	347	0.40	1.02	865	0.76	1.37
Any activity	High school	347	0.20	0.40	865	0.35	0.48
Sports activities	High school	347	0.13	0.38	865	0.19	0.48
Any sports	High school	347	0.11	0.31	865	0.15	0.36
Club activities	High school	347	0.15	0.51	865	0.35	0.79
Any clubs	High school	347	0.10	0.30	865	0.22	0.41
Superlatives	High school	347	0.07	0.36	865	0.08	0.35
Any superlative	High school	347	0.05	0.22	865	0.06	0.23
Honors activities	High school	347	0.03	0.18	865	0.10	0.35
Any honors	High school	347	0.03	0.16	865	0.08	0.28
Music activities	High school	347	0.02	0.14	865	0.05	0.22
Any music	High school	347	0.02	0.14	865	0.04	0.20
Identification scale	8th grade	122	33.46	5.46	605	36.32	6.53

Table A3

Effects of small Kindergarten on high school activities. Results for Black students.

	Superlatives Black students	Sports Black students	Music Black students	Honors Black students	Clubs Black students	All activities Black students
Small K class	0.069 (0.064)	−0.001 (0.067)	−0.020 (0.030)	0.027 (0.040)	0.041 (0.026)	0.019 (0.081)
Aide K class	−0.018 (0.023)	−0.037 (0.070)	−0.041 (0.051)	−0.027 (0.033)	0.106** (0.050)	0.008 (0.092)
Female	0.010 (0.031)	−0.048 (0.039)	0.035 (0.037)	0.028 (0.029)	−0.003 (0.044)	−0.038 (0.049)
Free lunch	−0.227** (0.099)	−0.213** (0.094)	−0.021 (0.054)	−0.170* (0.084)	0.036 (0.068)	−0.238** (0.086)
Constant	0.243*** (0.078)	0.366*** (0.090)	0.054 (0.060)	0.194* (0.077)	0.030 (0.082)	0.458*** (0.088)
Observations	169	169	169	169	169	169
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.12	0.06	0.02	0.06	0.03	0.04
Number of schools	30	30	30	30	30	30

* $p < 0.1$.** $p < 0.05$.*** $p < 0.01$.

Appendix B. Data collection appendix

Collecting the data on participation in high school activities was accomplished in several steps. First, the HEROS corporation, which maintains an archive of all Project STAR data, distributed a file to us that contained the names and high school of each student in the dataset. Although there were originally 79 elementary schools in the STAR experiment, there are approximately 150 high schools in Tennessee that received at least one Project STAR student. From the list of 150 high schools, we selected approximately 30 high schools from which we would attempt to gather yearbook information. We would then sample all students in the selected schools with the goal of gathering data for over 3000 students.

In selecting the 30 candidate high schools, we decided to over-sample schools with large numbers of STAR students and also include schools with smaller numbers that are geographically close to the “large” schools. We made this choice in order to maximize sample size and minimize expenses in collecting the data. We then contacted the selected schools to verify the existence of a catalog of high school yearbooks for the years 1994–1999 and to determine the proper protocol for visitors to use the high school library.

Several factors created the need for us to deviate from our some aspects of our original plan. First, unfortunately, the Memphis City School System rejected our application to visit any schools within its system. We also had difficulties gaining access to several schools outside of Memphis, but often in urban areas (Nashville and Knoxville). As Memphis-area schools represent an important number of STAR students, we attempted to collect yearbooks by alternative methods. We contacted individuals in the relevant high school class-years through several websites devoted to high school social networks, such as Classmates.com and Reunion.com. We also relied on personal networks of individuals in the Memphis area to collect yearbooks through friends and friends-of-friends.

These activities results in our collection of yearbooks from over 20 high schools, which we then coded and cre-

ated a database of individual names and schools linked with their activities during high school. We then sent this file to HEROS in order for them to link our names/activities file with their names/STAR data file, de-identify the data, and send the data back to us to analyze. Unfortunately, the HEROS names data was not as clean as we expected and was missing first names from several hundred of the students we collected data on. Thus, we did not attain a perfect match rate, and this decreased our sample from which to base our analysis. Nonetheless, we attained a sample size of over 1300 STAR students matched with their high school activities. We then performed the analysis we outlined in our proposal to examine the effects of small class size in Kindergarten through 3rd grade on high school activities, including clubs, sports, honors, music, superlatives, and a combined category.

References

- Akerlof, G., & Kranton, R. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3), 715–753.
- Akerlof, G., & Kranton, R. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 50, 1167–1201.
- Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: The High/Scope Press.
- Ding, W., & Lehrer, S. (in press). Estimating treatment effects from contaminated multi-period education experiments: The dynamic impact of class size reductions. *Review of Economics and Statistics*.
- Finn, J. (1989). Withdrawing from school. *Review of Educational Research*, 59(2), 117–142.
- Finn, J., Fulton, D., Zaharias, J., & Nye, B. (1989). Carry-over effects of small classes. *Peabody Journal of Education*, 67(1), 75–84.
- Finn, J., Gerber, S., Achilles, C., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2), 145–183.
- Finn, J., Pannozzo, G., & Achilles, C. (2003). The ‘Why’s’ of class size: Student behavior in small classes. *Review of Educational Research*, 73(3), 321–368.
- Fryer, R., & Levitt, S. (2004). Understanding the Black–White test score gap in the first two years of school. *The Review of Economics and Statistics*, 86(2), 447–464.
- Fryer, R., & Levitt, S. (2006). The Black–White test score gap through third grade. *American Law and Economics Review*, 8(2), 249–281.

- Hanushek, E. (1999). The evidence on class size. In S. Mayer, & P. Peterson (Eds.), *Earning and learning: How schools matter*. Brookings Institute Press.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900–1902.
- Holland, A., & Andre, T. (1987). Participation in extracurricular activities in secondary school—What is known, what needs to be known. *Review of Educational Research*, 57(4), 437–466.
- Krueger, A., & Whitmore, D. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, 111, 1–28.
- Ma, X. (2003). Sense of belonging to school: Can schools make a difference? *Journal of Educational Research*, 96(6), 340–349.
- Reynolds, A. J., Wang, M. C., & Walberg, H. J. (2003). *Early childhood programs for a new century*. Washington, DC: Child Welfare League of America Press.
- Rolnick, A., & Grunewald, R. (2003). *Early childhood development: Economic development with a high public return*. Federal Reserve Bank of Minneapolis, Minneapolis, MN. Technical Report.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Sizer, T. (1984). *Horace's compromise: The dilemma of the American high school*. Boston: Houghton Mifflin Company.
- Voelkl, K. (1997). Identification with school. *American Journal of Education*, 105(3), 294–318.