Scand. J. of Economics 107(2), 375–394, 2005
DOI: 10.1111/j.1467-9442.2005.00413.x

Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement

*Mikael Lindahl**
Stockholm University, SE-106 91 Stockholm, Sweden
mikael.lindahl@sofi.su.se

Abstract

The effects of class size on scholastic achievement are estimated using a seasonal feature of the school system. The fact that schools are in session during the school year and out of session during the summer makes it possible to control for non-school influences on both the level of and changes in achievement. Using Swedish data, smaller classes are found to generate higher test scores and this effect is larger for immigrants. The results are also compared with those from applying the same data to the widely used value-added model.

Keywords: Class size; summer and school-year learning; value-added models; difference-in-differences

JEL classification: *H*52; *I*21; *I*28

I. Introduction

Given the amount of resources used for education in most developed countries, it is not surprising that the efficiency of various school policies is subject to such debate. One of the school policy instruments that has received most attention is the size of classes. There are still too many conflicting results in the literature for any consensus to have been reached regarding whether decreasing class size has any significant effect on achievement.[1]

* I have benefited from the comments of two anonymous referees, Anders Björklund, Per-Anders Edin, Daniel Hallberg, Martin Hörnqvist, Alan Krueger, Maria Melkersson, Lena Nekby, Roope Uusitalo, Diane Whitmore, Olof Åslund, and seminar participants at the Swedish Institute for Social Research, Stockholm; Research Institute of Industrial Economics, Stockholm; Uppsala University; EALE/SOLE 2000; EEA 2000; IZA, Bonn; Tinbergen Institute, Amsterdam; and CEE, London. I also thank Ossian Wennström and Peter Björklund for excellent help with data collection, and the pupils and teachers who participated in this study. Financial support from the Humanities and Social Sciences Research Council and Handelsbanken is gratefully acknowledged.

[1] Angrist and Lavy (1999) for Israel, Case and Deaton (1999) for South Africa, and Krueger (1999) and Krueger and Whitmore (2001) for the US, find significant positive effects of smaller classes. Hoxby (2000) for the US and Dobbelsteen, Levin and Oosterbeek (2002) for

There are at least two reasons why simply regressing test scores on class sizes is likely to give a biased estimate of the class-size effect. First, resources are often distributed to school areas with many low-achieving pupils and to low-achieving pupils within schools. Second, families choose (or locate close to) those schools which have relatively more resources and smaller classes. Parents with the resources to make these choices more actively are likely to be those with high-achieving pupils. These two issues bias the class-size estimate in different directions. If the first issue dominates, the estimate will be biased away from finding positive achievement effects of smaller classes.

Ideally, for research purposes, pupils and teachers should be randomly assigned to classes of different sizes. However, large-scale randomized experiments are expensive and difficult to implement. Most often, researchers have to rely on non-experimental data. The standard way of estimating class-size effects has been to estimate so-called value-added models, which regress the test-score level in one grade on current school and family background characteristics and the test-score level in the previous grade. Hanushek (1997) summarizes the US literature on school resources and pupil performance, drawing on results from 90 published studies, the overwhelming majority of which use either level or value-added techniques. Only 15 percent of the estimates show statistically significant positive effects of smaller classes, and of the estimates using the value-added model (78 out of 277 estimates), only 12 percent show positive effects of smaller class sizes.

This study has three purposes. First, I investigate a potential weakness of the value-added model; if variables affecting learning are omitted and also correlated with class size, the class-size estimate will be biased. Second, I develop a new approach to estimate class-size effects; it controls for these unobservable learning effects and, in its most general form, has the value-added model as a special case. The new approach is built around the fact that schools are closed during summer vacation, but open during the school year. This seasonal feature makes it possible to separate the effects of school and non-school factors on learning. The idea of using summer learning as a counterfactual to school-year learning has a long tradition in educational sociology; see e.g. Heyns (1978) and Entwisle, Alexander and Olson (1997). Here, I extend their approach so as to estimate the achievement effects of specific school inputs, such as class size. Third, I estimate the effects of class

the Netherlands, find insignificant small effects. Rivkin, Hanushek and Kain (2005) for the US, find significant small effects of smaller classes. For Sweden, few academic studies have been conducted. Marklund (1962) uses data from 1955–1956 and 1959 and finds no effects. Lindsey and Cherkaoui (1975) use data from 1962 and find negative effects of smaller classes.

size on achievement in Swedish schools.[2] I apply the estimation strategies to a new sample of pupils from Stockholm, Sweden. The sample contains scores on a math test for the same pupils on three occasions: at the end of the fifth grade and at the beginning and end of the sixth grade. It also contains measures of school factors (class sizes for both grades and teacher variables) and non-school factors (pupils' demographic and social background).

The paper is organized as follows. Section II outlines the estimation strategies and Section III describes the new data set used in this study. In Section IV, test scores are related to class sizes. Some sensitivity analysis is also conducted. Section V concludes.

II. Educational Production Functions and Estimation Techniques

Value-added Models

A typical value-added model can be expressed as:

$$A_{it} - A_{it-1} = \theta + \phi F_{it} + \beta S_{it} + v_i + u_{it}, \tag{1}$$

where $A_{it} - A_{it-1}$ is learning (i.e., changes in achievement) between the end of grades $t-1$ and t for pupil i; F_{it} denotes a vector of non-school or family (i.e., demographic, background and neighborhood) characteristics; S_{it} denotes a vector of schooling variables such as class size and teacher quality, and θ is an intercept. The error term has two parts: the unobservable fixed learning effect v_i, capturing the student's ability to learn and everything else with a constant influence on learning, and the random error term u_{it} which is assumed to be orthogonal to F_{it}, S_{it} and v_i. If the lagged achievement level, A_{it-1}, affects learning across grades, equation (1) becomes:

$$A_{it} = \theta + \phi F_{it} + \beta S_{it} + \lambda A_{it-1} + v_i + u_{it}. \tag{2}$$

In (1) and (2), A_{it-1} captures the entire history (until $t-1$) of observed and unobserved school and non-school characteristics. Note that a classical measurement error in test scores will bias all estimates

[2] In this study, as well as in other non-experimental studies of class-size effects, the class-size estimate also captures effects of peers and unobservable teacher quality if these are correlated with class size and have an independent effect on student achievement. For instance, if good teachers are rewarded by assignment to smaller classes, an estimate showing a positive effect of smaller classes will be an overestimate. See Boozer and Rouse (2001), Hanushek, Kain and Rivkin (2003) for the US, and Bonesrønning, Falch and Strøm (2004) for Norway, for studies of teacher sorting.

in (2), but not in (1).[3] Equations (1) and (2) are referred to as the *VA (value-added)* specifications; see Hanushek (1979).

In general, the popular VA specification has generated small and insignificant class-size estimates; see Hanushek (1997). A reason for this could be that the VA model fails to eliminate the fixed learning effect, v_i. If unobservable time-constant factors have an effect on change in achievement through the fixed learning effect, in addition to a one-time effect on the achievement level, and the fixed learning effect is correlated with class size, then the estimate of the class-size effect will be biased. This type of bias will also appear due to reverse causality, if schools allocate pupils to classes of different sizes based on the students' ability to learn. If resources are compensatory within schools, slow-learning pupils will be assigned to smaller classes, which would underestimate the effect of smaller classes on achievement using the VA model. The approach outlined below attempts to eliminate biases due to fixed achievement level *and* fixed learning effects.

A New Approach to Estimating Educational Production Functions

It was assumed above that achievement-level measures were only available at the end of grade levels $t - 1$ and t. Suppose a measure is also available at the start of the school year in grade level t. For expository purposes, assume that each grade level t consists of two seasons of equal length: the summer vacation, when schools are closed, and the school period, when schools are in session. The first part is denoted $j = 1$ and the second part $j = 2$. For the time being, also assume unobserved ability to have the same effect on learning in both periods.[4] Equation (1) at grade t (for $j = 1, 2$) can then be expressed as:

$$\Delta A_{it,1} = \kappa_1 + \alpha_1 F_{it} + \delta_i + \varepsilon_{it,1}, \tag{3}$$

$$\Delta A_{it,2} = \kappa_2 + \alpha_2 F_{it} + \beta S_{it} + \delta_i + \varepsilon_{it,2}, \tag{4}$$

where $\Delta A_{it,1} = A_{it,1} - A_{it-1,2}$ is learning during the summer period; $\Delta A_{it,2} = A_{it,2} - A_{it,1}$ learning during the school period; $A_{it-1,2}$, $A_{it,1}$ and

[3] The estimation of (2) in the presence of a classical measurement error in test scores generates the following β- and λ-estimates (where we ignore F and assume class size to be the only S-variable and v_i to be uncorrelated with the explanatory variables):

$$p \lim \beta_{OLS} = \beta + \lambda \eta_{AS}(1 - R)/(1 - \rho_{AS}^2) \quad \text{and} \quad p \lim \lambda_{OLS} = \lambda[1 - (1 - R)/(1 - \rho_{AS}^2)],$$

where R is the estimated reliability ratio, η_{AS} the estimate from a regression of the test score level in the spring of the sixth grade on class size in the sixth grade, and ρ_{AS} the estimated correlation coefficient between these two variables.

[4] This is further discussed in Section IV. It is also assumed that F_{it} has the same value in both periods.

$A_{it,2}$ are the achievement levels at the end of the school period in grade $t-1$, at the start of the school period in grade t, and at the end of the school period in grade t, respectively; F_{it} and S_{it} are defined as before; α_1 and α_2 are coefficients allowing observable background variables to have different impacts during the school and summer periods; and κ_1 and κ_2 are intercepts allowing the change in average achievement to be different during the school and summer periods. The error terms have two parts: the fixed learning effect δ_i and the random error terms $\varepsilon_{it,j}$, where the latter are assumed to be uncorrelated with F_{it}, S_{it} and δ_i.

Equation (3) expresses summer vacation learning as a function of family characteristics and the fixed learning effect, and (4) expresses school period learning as a function of family and school characteristics and the fixed learning effect. What is important about (3) and (4) is that the school characteristics only affect learning when schools are in session, whereas family characteristics can influence learning both when schools are in session and when they are not.

We can eliminate the fixed learning effect by taking the difference between (4) and (3):

$$\Delta A_{it,2} - \Delta A_{it,1} = \kappa' + \alpha' F_{it} + \beta S_{it} + \Delta\varepsilon_{it,2}, \tag{5}$$

where the dependent variable is the difference between learning during the school and summer periods; $\alpha' = \alpha_2 - \alpha_1$; $\kappa' = \kappa_2 - \kappa_1$; and $\Delta\varepsilon_{it,2} = \varepsilon_{it,2} - \varepsilon_{it,1}$. If lagged achievement does not affect learning, the estimation of (5) produces consistent estimates of class-size effects, even in the presence of fixed achievement level *and* fixed learning effects.

The identification strategy is more complicated if a lagged achievement level affects learning. Equations (3) and (4) are then expressed as:

$$A_{it,1} = \kappa_1 + \alpha_1 F_{it} + \gamma_1 A_{it-1,2} + \delta_i + \varepsilon_{it,1}, \tag{6}$$

$$A_{it,2} = \kappa_2 + \alpha_2 F_{it} + \beta S_{it} + \gamma_2 A_{it,1} + \delta_i + \varepsilon_{it,2}, \tag{7}$$

both generalizations of (2).[5] Taking the difference between (7) and (6), we get:

$$\Delta A_{it,2} = \kappa' + \alpha' F_{it} + \beta S_{it} + \gamma_2 \Delta A_{it,1} + \Delta\gamma A_{it-1,2} + \Delta\varepsilon_{it,2}, \tag{8}$$

where $\Delta\gamma = \gamma_2 - \gamma_1$. It is not possible to consistently estimate the class-size parameter in (8) unless some restriction is imposed (or some instrument is available). Assuming $\gamma_1 = \gamma_2 = \gamma$, that is, the previous achievement level has the same effect on learning during the summer and during the school year, (8) can be rewritten as:

[5] Inserting (6) into (7), we get $A_{it,2} = \theta + \phi F_{it} + \beta S_{it} + \lambda A_{it-1,2} + v_i + u_{it}$, where $\theta = \kappa_2 + \gamma_2\kappa_1$, $\phi = \alpha_2 + \alpha_1\gamma_2$, $\lambda = \gamma_1\gamma_2$, $v_i = (1+\gamma_2)\delta_i$ and $u_{it} = \gamma_2\varepsilon_{it,1} + \varepsilon_{it,2}$, which is equivalent to (2).

$$\Delta A_{it,2} = c' + \alpha' F_{it} + \beta S_{it} + \gamma \Delta A_{it,1} + \Delta\varepsilon_{it,2}. \tag{9}$$

Due to the correlation between $\Delta A_{it,1}$ and $\Delta\varepsilon_{it,2}$ (since cov($A_{it,1}$, $\varepsilon_{it,1}$) $\neq 0$), OLS estimates of (9) will be biased. Therefore, I estimated (9) using $A_{it-1,2}$ as an instrument for $\Delta A_{it,1}$; see Anderson and Hsiao (1981). Still, estimates of the parameters in (9) could be inconsistent for at least four reasons. First, if the error terms are serially correlated, i.e., cov($\varepsilon_{it,2}$, $\varepsilon_{it,1}$) $\neq 0$; second, if lagged test-score levels have different effects on changes in test scores during the summer and the school year, respectively, i.e., $\gamma_1 \neq \gamma_2$; third, if $A_{it-1,2}$ has no statistically significant effect on $\Delta A_{it,1}$; and fourth, if the test scores are error-ridden measures of scholastic achievement.[6]

The main difference between equations (1) and (2) and equations (5), (8) and (9) is that the last three specifications eliminate the unobservable fixed learning effect, whereas the first two do not. Equations (5), (8) and (9) are referred to as *DD (difference-in-differences)* models. As shown above, the DD model (9) is a special case of the least restrictive DD model (8). Note also that the VA models (1) and (2) and the DD model (5) are all special cases of (8). We end up with the VA models (1) if $\gamma = -1$, and (2) if $\gamma_1 \neq \gamma_2 = -1$, and the DD model (5) if $\gamma = 1$.[7]

Further, note that the β-estimate from (5) will not be biased due to measurement error in the test scores or serial correlation in the error terms. The intuition for equations (5) and (9) is that by regressing school-year learning as a function of school characteristics, controlling for the most recent summer period learning, it is possible to isolate the effect of school characteristics, such as class size, on test scores. Hence, I used the experience during the summer months as a way of adjusting for non-school influences on the level and change in test scores.

III. Data[8]

The National Agency for Education distributes a math test to all schools in Stockholm, Sweden, to be taken by pupils during the spring semester of the fifth grade. I contacted schools at the start of the fall semester in 1998. I selected four parts of this test, which I then distributed to the pupils at the start and end

[6] Estimation of (9), with the classical measurement error in test scores, generates the following β- and λ-estimates (assuming one S-variable, ignoring F_{it}): $p \lim \beta_{IV} = \beta - \gamma(1 - R)(k/d)$ and $p \lim \gamma_{IV} = \gamma[(1 - R + d)/d]$, where R is the estimated reliability ratio, k the estimate from a regression of $\Delta A_{it,1}$ on S_{it}, and d the estimate from a regression of $\Delta A_{it,1}$ on $A_{it-1,2}$ minus the product of an estimate of S_{it} on $A_{it-1,2}$ and the estimate from $\Delta A_{it,1}$ on S_{it}. Solving for β and γ generates the measurement-error-corrected estimates.

[7] Note that (1) and (5), but not (2), are special cases of (9).

[8] For a detailed description of the data and the sampling design, see Lindahl (2000).

of the sixth grade. This test was also given during the period February–June in 1998 in the fifth grade, with the four parts of the test conducted on separate occasions. The test in the fall of the sixth grade was given between the last week in September and the first week in November 1998, and the test in the spring of the sixth grade was given during the last four weeks of the term (May–June) in 1999.[9] Summer vacation in Sweden lasts 10 weeks, from early June to late August. I, and not the teachers, graded the tests on all three occasions.

In total, 556 pupils took the test on all three occasions, and they did so under similar conditions regarding time allowed and teacher help.[10] The same test was used on all three occasions. Four parts of the test were used, each of a different kind, with questions ranging from simple counting exercises to more advanced problems. The average percentile rank over the four test parts (on each test occasion) was then used as the measure of each pupil's math achievement on each test occasion.[11] Table 1 shows summary statistics for the test scores on the three test occasions (spring of the fifth grade, fall and spring of the sixth grade), as well as for changes in test scores.[12] The correlations between test scores on the three test occasions are between 0.72 and 0.77.[13]

Data on school, class and teacher characteristics were gathered through a questionnaire distributed to the teachers at the time of the fall sixth-grade test. Teachers were asked to answer questions about themselves (their teaching experience in total and with their current class, and their education) and their students (pupils' genders and whether the pupil is an immigrant child) as well as to provide information about their class sizes (during teaching of both math and regular subjects).[14] To obtain information on pupils' social background,

[9] On average, the spring fifth-grade test was taken 9.7 weeks before the summer break, the fall sixth-grade test, 7.2 weeks after the summer break, and the spring sixth-grade test, 1.2 weeks before the next summer break.

[10] Even though all pupils took at least one part of the test in the spring of the fifth grade, the fall of the sixth grade and the spring of the sixth grade, not all pupils took all four parts of the test on each occasion. Of the $4 \times 3 = 12$ test parts conducted, 222 pupils did all 12 parts, 225 pupils 10–11 parts, 97 pupils 8–9 parts, and 12 pupils 5–7 parts. The average number of test parts done was 10.57.

[11] I scaled each test part on each test occasion in percentile ranks, ignoring any missing test-part score. Then, I took the average of these percentile rank scores on each occasion.

[12] In Lindahl (2000), I show that the absolute increase in the test score over the school period is almost four times higher as compared to the summer period. This suggests that even though the testing dates are far from ideal, the summer vacation really does affect the observed change in the test score.

[13] Assuming classical measurement error, we can consistently estimate the (alpha) reliability of the average test score on each test occasion using the formula $\alpha = Nr/[1 + (N - 1)r]$, where N is the number of test parts and r is the average of all test-part correlations; see Cronbach (1951). Using the test-score parts from the spring of the fifth grade, when all parts were taken on separate occasions, the estimate is 0.79 ($N = 4$ and $r = 0.48$).

[14] Controlling for teacher education should not be necessary since all teachers but one were certified and had a Bachelor's degree as their highest scholastic credential.

Table 1. *Descriptive statistics*

	Mean	Std. dev.	Min.	Max.
Test scores (percentile ranks)				
Fifth grade, spring: $A_{it-1,2}$	47.65	23.07	1	99.5
Sixth grade, fall: $A_{it,1}$	47.93	22.98	1	96.5
Sixth grade, spring: $A_{it,2}$	46.80	22.70	1.5	92
Dependent variable in (1): Change, fifth grade spring, to sixth grade spring: $A_{it} - A_{it-1}$	−0.85	17.24	−48.5	68.5
Dependent variable in (5): Change, sixth grade fall, to sixth grade spring minus, change, fifth grade spring, to sixth grade fall: $\Delta A_{it,2} - \Delta A_{it,1}$	−1.41	27.94	−98.5	85
Dependent variable in (9): Change, sixth grade fall, to sixth grade spring: $\Delta A_{it,2}$	−1.13	15.63	−48.8	49.5
Class sizes				
Class size, sixth grade (math)	19.90	4.40	5.5	25
Class size, sixth grade (regular)	23.11	4.16	13	28.5
Teacher variables, sixth grade				
Teacher experience in years	16.17	10.82	0.2	33
Teacher experience (years in the class)	1.62	1.04	0	5
Background variables				
Gender (Girl = 1)	0.50	0.50	0	1
Immigrant's child = 1	0.23	0.42	0	1
Parents' education	12.36	1.96	7.53	19.67
Log(family income)	12.60	0.54	11.19	14.75

Notes: The number of observations is 556. A pupil is counted as an immigrant child if he/she lacks parents with Swedish as their native language. In the cases where information on the teacher experience variables for the actual teacher in the math class is missing, the experience of the teacher responsible for the regular class is used. The data used for education and family income are from 1996. Parents' education is the average years of schooling in the block where the pupil lives. Years of schooling are calculated on the basis of information on the fraction of people in a block in each of eight educational categories and weighted by the average years of schooling for each category (estimated from a representative data set for Sweden (SLLS) for 1991). The logarithm of family income is calculated as the logarithm of the average family income in each block. For eight pupils with missing address information, the average of the education and family income of the pupils in the class, respectively, were assigned to those pupils. This was also done for an additional five pupils with missing family income data.

the pupils' addresses (from the class lists) were matched with block data on education and family income from Statistics Sweden. As shown in Table 1, on average, a pupil has a teacher with 16 years of teaching experience, but less than two years in the current class. One-fourth of the sample consists of immigrant pupils. The average pupil has parents with 12 years of education and a family income of SEK 300,000 (approximately $30,000 at this time).

This study focuses on the size of the class measured as the actual number of pupils present during math instruction (*math classes*), since the test used in this study is designed to capture math skills. The sample of pupils is from 38 math classes in 16 schools. The distribution of math class sizes is left skewed, with some pupils taught in very small classes. I also collected a measure of the number of pupils present during teaching in a typical subject (*regular classes*).

Table 2. *Correlation matrix for background and class-size variables*

	Girl	Immigrant's child	Parents' education	Log (family income)	Class size, sixth grade (math)	Class size, sixth grade (regular)
Girl	1.00					
Immigrant's child	−0.02	1.00				
	(0.55)					
Parents' education	0.03	−0.50	1.00			
	(0.55)	(0.00)				
Log (family income)	0.01	−0.55	0.72	1.00		
	(0.77)	(0.00)	(0.00)			
Class size, sixth grade (math)	−0.01	−0.47	0.51	0.47	1.00	
	(0.86)	(0.00)	(0.00)	(0.00)		
Class size, sixth grade (regular)	−0.09	−0.24	0.19	0.13	0.42	1.00
	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	

Note: The number of observations is 556. p-Values for the test of no correlation are in parentheses.

There are 32 regular classes. Most previous work on class-size effects has used aggregated data on class size, measured as pupil–teacher ratios at the school/district level.[15] The average math (regular) class size is 20 (23) pupils.

In Table 2, class sizes are correlated with background variables. Pupils with parents who are less educated, have lower family incomes and are immigrants, are found in smaller math classes. This pattern is due to sorting both between and within schools.[16] The former is expected, since educational resources in Sweden are redistributed towards schools with a high fraction of pupils with disadvantaged backgrounds.[17] The dispersion of class size is high as compared to many other data sets, partly due to the fact that both of these class-size measures are relatively more disaggregated. However, two-thirds of the variance in math class sizes stem from between-school variation in these data, so the redistributive school resource policy in Sweden is probably also important. Since the most accurate measure of math class size is the actual number of pupils who are taught in math, I focus on this measure throughout the rest of the paper.

[15] The 28 percent of the estimates summarized in Hanushek (1997), who uses classroom data, are relatively less likely to generate a positive effect of smaller classes, perhaps because compensating distribution of resources is mainly carried out within schools. If so, aggregated data can produce a more accurate estimate of the true class-size effect. However, if the distribution of school resources is mainly compensated between areas, disaggregated data might instead be more accurate.

[16] This can be seen by regressing math class size on each of the non-school variables separately, while controlling for school dummies (within-school sorting), and separately regressing school mean math class size on the school mean of each of the non-school variables (between school sorting). The estimates on the non-school variables always indicate sorting of disadvantaged students into smaller classes, both within and between schools. The signs of the estimates are: for education (+), log family income (+), immigrant status (−).

[17] See National Agency for Education (1999).

IV. Results

The results reported here include estimated class-size effects without controls, with demographic and family background controls (pupils' gender, whether parents are immigrants, parents' years of schooling and the logarithm of family income), and with additional controls for teacher characteristics (teacher experience, teacher experience squared and teacher experience with current class) or for school fixed effects.[18] Since classical measurement errors in test scores will bias the estimates of models (2) and (9), when lagged test scores are included as controls, I correct these estimates by assuming a reliability ratio in test scores of 0.79, which is the estimated alpha reliability from Section III. Since measurement error has different effects on the estimates in the VA and DD models, this is also necessary in order to see how much of any bias in the class-size estimates of the VA model is due to the fixed learning effect. All estimated standard errors in this section allow for correlated regression errors among pupils within the same school.

Level and Value-added Regressions

First, I show results from a simple-level model. I regressed the spring of the sixth-grade test scores on class size in the same year, controlling for background and teacher variables. The estimate in column 1 of Table 3 indicates a positive relation between math class size and test scores. Interpreting this estimate as a causal effect of class size, an increase in class size by one pupil would, on average, give rise to a 0.8 percentile rank higher test score. The obvious drawback of such level regressions is that class size is unlikely to be exogenous. Instead, it is likely to be correlated with other school, family or pupil characteristics from the present and/or previous time periods, and also to be partly determined by authorities with the aim of compensating weak pupils with more school resources and smaller classes. Therefore, the estimate is very likely a biased estimate of the causal class-size effect.

In columns 2–9 of Table 3, we turn to VA models (1) and (2). The dependent variable is the change in test scores between the spring of the fifth and sixth grades. Notwithstanding whether the test score at the beginning of the period is controlled for, the class-size estimates are insignificantly different from zero or positive and significant.[19] This result is consistent with the pattern in the literature using variations of the VA model. Additional family background variables, teacher variables or school

[18] In some of the later estimations, I controlled for school fixed effects, thereby controlling for any between-school sorting of relevant school inputs (such as teacher quality).

[19] Using model (2), the measurement error adjusted class-size estimate in column 6 of Table 3 is 0.17, whereas the unadjusted estimate is 0.48.

Table 3. *Level and value-added regressions*

Dependent variables:	Test score, spring, sixth grade	The difference between test scores in the spring of the sixth grade and the spring of the fifth grade							
		Unconditional on initial test				Conditional on initial test: corrected for measurement error in test scores			
	OLS	OLS	OLS	OLS	School-FE	OLS	OLS	OLS	School-FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Class size, sixth grade	0.84	−0.01	0.11	0.16	−0.03	0.17	0.19	0.23	0.01
	(0.35)	(0.30)	(0.30)	(0.28)	(0.41)	(0.14)	(0.16)	(0.16)	(0.21)
Teacher experience	0.32			0.36				0.36	
	(0.25)			(0.38)				(0.23)	
Teacher experience squared	−0.010			−0.014				−0.013	
	(0.008)			(0.011)				(0.007)	
Teacher experience in current class	−0.92			−0.55				−0.59	
	(0.96)			(1.01)				(0.56)	
Test score fifth grade, spring						−0.12	−0.11	−0.11	−0.04
						(0.03)	(0.04)	(0.04)	(0.04)
School dummies	No	No	No	No	Yes	No	No	No	Yes
p-Value: test of no joint effect of background variables	0.01	–	0.29	0.28	0.42	–	0.26	0.30	0.22
R^2	0.127	0.000	0.009	0.019	0.120	–	–	–	–

Notes: The number of observations is 556. Since there is no within-school variation in class size for three schools, only 445 observations are used in identifying the class-size estimates in columns 5 and 9. In columns 1–5, the standard errors allow for regression errors correlated among pupils in the same school. Test scores are measured in percentile ranks. The estimates and standard errors in columns 6–9 assume a true reliability ratio of 0.79 in the fifth-grade test-score percentile ranks. Columns 1, 3–5 and 7–9 include controls for the background variables: girl, immigrant's child, parents' education and the logarithm of family income. The dependent variable in column 1 is the percentile test-score rank in the spring of the sixth grade. The dependent variable in columns 2–9 is the difference between the percentile test-score rank in the spring of the sixth grade and the percentile test-score rank in the spring of the fifth grade. The estimate on lagged test scores from equation (2) is reported in columns 6–9 as $\lambda - 1$ to facilitate the comparison with equation (1).

fixed effects do not significantly affect the class-size estimates. If the fixed learning effect is poorly proxied for by the background variables, estimated class-size effects are likely to be biased using the VA model. I also note that if schools compensate slow-learning pupils by assigning them to smaller classes, the class-size effect estimates in Table 3 will be upward biased.

Difference-in-differences Regressions

Table 4 shows the estimates from the DD models (5) and (9). The dependent variable is the difference between changes in test score over the school period (fall and spring of the sixth grade) and the summer period (spring of the fifth grade and fall of the sixth grade). The first-stage effects of the spring fifth grade test scores on the change in summer test score are always significant in columns 5–7.[20] The class-size effect is now estimated to be significantly negative, and this result is robust to adding family background variables, lagged changes in test scores, teacher variables or fixed school effects as controls.[21] It appears that eliminating the fixed learning effect has a substantial effect on the results (as is clear from comparing results using the VA and DD models). A decrease in class size by one pupil is estimated to give about 0.4 to 1.0 percentile ranks higher test scores, on average.

Rivkin *et al*. (2005) use the difference in changes in achievement between subsequent grades as a dependent variable, which also eliminates the fixed learning effect. Since they only find a small effect of smaller classes, it could be argued that unobserved individual-specific factors do not bias estimates using the VA model. However, their approach could lead to class-size estimates biased toward zero, since it requires using changes in class sizes between grades as the independent variable, and the reliability ratio for the change in class size in subsequent grades is likely to be much lower than the reliability ratio for the class size within a specific grade level, which is used here. Hence, their results do not necessarily contradict the argument in this paper.

Although estimations of (5) and (9) give rise to positive achievement effects of smaller classes, the class-size estimates from these two models differ sufficiently to be statistically significant. Why is this so? Estimations of both these specifications require assumptions regarding the effect of lagged test scores on changes in test score. Equation (5) assumes no such

[20] If school fixed effects are added to the specification used in column 8 of Table 4, the first-stage effect of the spring fifth-grade test score is insignificant. Therefore, these estimates are not shown.

[21] When measurement error in test scores is not corrected for, we obtain a very similar β-estimate. The intuition for this is that using the lagged test-score level as an instrument for changes in test-score already in the first stage, to a large extent corrects for measurement error.

Table 4. *Difference-in-differences regressions*

	Dependent variable: difference between the school period and the summer period changes in test scores						
	Unconditional on initial test				Conditional on initial test: corrected for endogeneity and measurement error in test scores		
	OLS	OLS	OLS	School-FE	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Class size, sixth grade	−0.77	−0.95	−0.98	−1.38	−0.36	−0.38	−0.37
	(0.23)	(0.32)	(0.26)	(0.66)	(0.14)	(0.21)	(0.18)
Teacher experience			0.77				0.55
			(0.81)				(0.32)
Teacher experience squared			−0.019				−0.016
			(0.025)				(0.010)
Teacher experience in current class			−1.04				−0.77
			(1.11)				(0.50)
Lagged test score change ($\gamma - 1$)					−1.08	−1.07	−1.08
					(0.31)	(0.35)	(0.36)
School dummies	No	No	No	Yes	No	No	No
First-stage effect of instrument					−0.13	−0.12	−0.11
					(0.03)	(0.03)	(0.03)
p-Value: test of no joint effect of background variables	–	0.74	0.61	0.85	–	0.71	0.66
R^2	0.015	0.020	0.027	0.102	–	–	–

Notes: The number of observations is 556. Since there is no within-school variation in class size for three schools, only 445 observations are used in identifying the class size estimates in column 4. The standard errors, in parentheses, allow for regression errors correlated among pupils in the same school. Test scores are measured in percentile ranks. Columns 2–4 and 6–7 include controls for the background variables: girl, immigrant's child, parents' education and the logarithm of family income. The estimates and standard errors in columns 5–7 assume a true reliability ratio of 0.79 for test-score levels. Lagged test-score change is the change in test-score percentile rank between the fifth grade, spring, and the sixth grade, fall. Columns 5–7 use the test-score percentile rank in the fifth grade, spring, as an instrument for the lagged test-score change. The dependent variable is the difference between percentile test scores in the spring of the sixth grade and percentile test scores in the fall of the sixth grade, minus the difference between percentile test scores in the fall of the sixth grade and percentile test scores in the spring of the fifth grade. The estimate on lagged test-score change from equation (9) is reported in columns 5–7 as $\gamma - 1$ to facilitate the comparison with equation (5).

effect, whereas (9) allows for an effect, but assumes it to be the same in both seasons. Estimations of the seasonal specifications (6) and (7) suggest that these assumptions are too strong. The estimates on test scores at the beginning of the periods are zero and insignificant in the school-period regression, but negative and significant in the summer regression, and the difference between these estimates is statistically significant.

It would therefore have been expedient to estimate the most general specification (8). However, this requires an additional assumption about the value of the difference between lagged test scores between seasons, $\Delta\gamma$ in (8). Since the fixed learning effect is not eliminated in the seasonal equations (6) and (7), I did not know if lagged achievement estimates from these equations would be consistent. However, if the estimates on initial test scores are equally inconsistent for both seasons, an estimate of the difference between them will be consistent. Here, the best estimate of this difference equals 0.09.[22] Since this estimate is positive, the estimates of the lagged change in test scores in columns 5–7 of Table 4 would be too negative and the class-size estimates too positive.

To illustrate this, I estimated equation (8) with pre-summer test scores as an instrument for the change in summer test scores at the first stage, also controlling for class size and background variables (as in column 6 of Table 4). I then estimated the change in school-year test scores as a function of the predicted change in first-stage summer test scores, class size and the background variables, *as well as* the pre-summer test scores with a coefficient restricted to 0.09, in the second stage. The estimate on the lagged change in test scores is then -0.75 (instead of -1.07) and the estimate on class size -0.61 (instead of -0.38). Thus, the restriction imposed on (9) partly explains the difference between the class-size estimates in columns 1–3 and columns 5–7 in Table 4. The class-size estimates in columns 5–7 of Table 4 are therefore likely to understate the positive effect of smaller classes on achievement.

There is no evidence of a quadratic class-size effect. If class size squared is added to the specification estimated in column 3 of Table 4, the estimate (standard error) for class size is -2.21 (2.86), and for class size squared 0.04 (0.09).[23] The class-size estimates are basically unaffected when a median regression is run or if the pupils with the very lowest class sizes (less than nine pupils in a class) are excluded. If the analysis had been carried out using

[22] Estimation of (6), controlling for background variables, gives an estimate (standard error) of $\gamma_1 = 0.90$ (0.03). Estimation of (7), controlling for background variables and class size, gives an estimate (standard error) of $\gamma_2 = 0.99$ (0.03). These estimates are corrected for measurement error in test-score levels. Hence, an estimate of $\Delta\gamma = \gamma_2 - \gamma_1$ equals 0.09, with a standard error (assuming independence between estimates) equal to 0.05.

[23] Even though the quadratic in class size is insignificant, the sign is in line with Lazear's (2001) theoretical model, which predicts class-size effects to be more pronounced in smaller classes.

the measure of regular class size, the positive effect of smaller classes on achievement would have been somewhat larger as compared to using the math class-size measure. According to the sixth-grade class lists, 701 pupils were available for testing in the sampled schools, but only 556 took all tests. Since sample selection can bias the estimates and since I have information on background and school variables for all pupils, I implemented a so-called Heckman correction; see Heckman (1976). The sample-selection corrected class-size estimates were very similar to the uncorrected estimates in Table 4.[24]

To assess whether class-size effects differ among Swedish and immigrant children, I ran separate DD regressions for these two sub-samples. The results are shown in Table 5. Using model (5), columns 1 and 2 show that immigrant children benefit more from smaller classes than Swedish children. For immigrant children, a decrease in class size by one pupil generates almost 2 percentile ranks higher test scores, on average. In columns 3 and 4, estimates of model (9) are reported. Now the class-size effect almost disappears for Swedish pupils. Since the estimate of the first-stage effect is insignificant for immigrant children, the class-size estimate in column 4 is not identified with any acceptable degree of certainty. From Table 5, I conclude that the negative effect of class size is mainly due to its effect on the group of immigrant children.[25]

Sensitivity Analysis

So far, the analysis has relied on some strong assumptions calling for sensitivity analysis. First, I have assumed that the fixed learning effect has the same effect on changes in test scores during the summer and school-year periods. If this is not the case, the DD specification (5) becomes:

$$\Delta A_{it,2} - \Delta A_{it,1} = \kappa' + \alpha' F_{it} + \beta S_{it} + (\phi_2 - \phi_1)\delta_i + \Delta\varepsilon_{it,2}, \qquad (10)$$

where ϕ_1 and ϕ_2 are the effects of unobserved learning ability on changes in achievement during the summer and school-year periods, respectively. Since δ_i is unobserved, there is no direct way of testing $\phi_1 = \phi_2$. However, if the background variables are jointly insignificant in estimations of (5) and (9),

[24] The Heckman correction relies on normality of the error term, so I used the standardized values of the raw scores as the dependent variable. The class-size estimate (standard error) in column 3 of Table 4, for instance, is -0.041 (0.009) in standard deviation units. Correcting for sample selection, this estimate is -0.037 (0.019).

[25] The regular class-size estimates are similar for immigrant and Swedish pupils. The positive math class-size effect varies negatively with parents' education. However, this is mainly because immigrants have a lower education and their children benefit relatively more from smaller classes.

Table 5. *Difference-in-differences regressions for Swedish and immigrant children, separately*

	Dependent variable: difference between the school period and the summer period changes in test scores			
	No lag: OLS		Lag, corrected for endogeneity and measurement error in test scores: IV	
	Swedish children	Immigrant children	Swedish children	Immigrant children
	(1)	(2)	(3)	(4)
Class size in sixth grade (CS6)	−0.43 (0.37)	−1.86 (0.42)	−0.14 (0.23)	−1.19 (0.33)
Lagged test score change ($\gamma - 1$)			−1.09 (0.30)	−0.81 (2.10)
First-stage effect of instrument			0.15 (0.04)	−0.04 (0.07)
p-Value: test of no joint effect of background variables	0.519	0.361	0.402	0.598
R^2	0.028	0.127	–	–

Notes: The number of observations is 429 in columns 1 and 3, and 127 in columns 2 and 4. The standard errors, in parentheses, allow for regression errors correlated among pupils in the same school. Test scores measured in percentile ranks. The estimates and standard errors in columns 3–4 assume a true reliability ratio of 0.79 for test-score levels. All columns include controls for the background variables: girl, parents' education, the logarithm of family income, and the school variables: a quadratic in teacher experience and a linear for teacher experience in current class. Lagged test-score change is the change in percentile test-score ranks between the spring of the fifth grade and the fall of the sixth grade. Columns 3 and 4 use the percentile test-score rank in the spring of the fifth grade as an instrument for the lagged test-score change. The dependent variable is the difference between percentile test scores in the spring of the sixth grade and percentile test scores in the fall of the sixth grade, minus the difference between percentile test scores in the fall of the sixth grade and percentile test scores in the spring of the fifth grade. This variable has the mean (std. dev.) −2.14 (28.86) in columns 1 and 3, and 1.03 (24.52) in columns 2 and 4. The estimates and standard errors in columns 3–4 assume a true reliability ratio of 0.79 for test-score levels. The class-size mean (std. dev.) is 21.02 (3.85) in columns 1 and 3, and 16.09 (4.03) in columns 2 and 4, respectively.

this would at least indicate that this assumption is reasonable (if the observable background variables are good proxies for the fixed learning effect). Indeed, when I added the background variables to the DD model in Section IV, these variables were jointly (and individually) insignificant in all column specifications (see row 8 of Table 4). Hence, I conclude that it cannot be rejected that the unobserved learning effects are the same in both seasons.

What is the likely direction of bias in the class-size estimate, if the observable background variables are poor proxies for the fixed learning effect? The most likely scenario is that children are less keen to make use of their ability or willingness to learn during summer vacation (i.e., children do not attempt to learn as intensively as during the school year), since, for instance, children do homework during the school year but not during the summer. Given that high-ability children more easily make achievement gains from homework, the fixed learning effect will be more important for learning during the school year as compared to summer vacation, so that $\phi_2 > \phi_1$. The class-size estimate in (10) would then be biased away from turning out negative, since the fixed learning effect should be positively correlated with advantageous family background and the test-score level, and we know (from Tables 2 and 3) that these variables are positively correlated with class size. This suggests that the estimated positive effect of *smaller* classes on achievement, found earlier, is an underestimate of the true effect.[26]

Second, I have assumed that the timing of the tests, i.e., the tests were not taken immediately after and before the summer break, has no effect on the results. If school does not contribute to learning at the beginning and end of the school year, it can be assumed that learning during the summer is well captured by our observed changes in test scores between the spring of the fifth grade and the fall of the sixth grade. Another related fact is that summer vacation lasts only 10 weeks and is much shorter than the school period, which is something for which I have not adjusted. The simplest way of controlling for this fact is to add two controls, for the number of weeks

[26] A result in the literature that might generate a bias in the other direction is that children from different family backgrounds make similar achievement gains during the school period, but that children from disadvantageous family backgrounds lose skills during the summer period, whereas children from advantageous backgrounds gain or keep their level; see Entwisle *et al.* (1997). This would mean that pupils with advantageous backgrounds have lower $\Delta A_{it,2} - \Delta A_{it,1}$, the dependent variable in (10). However, these results on differential gains during the summer and school periods for children with different family backgrounds are unconditional on schooling variables. In Lindahl (2000), I found evidence that immigrant pupils gain more than Swedish pupils during the school year, but that both groups preserve their skill level during the summer period, whereas no differences in seasonal gains are found between girls and boys and for pupils of different socioeconomic backgrounds. Hence, the dependent variable in (10) is larger for immigrants than for Swedish pupils. However, if controls for schooling variables are included, there is no significant difference across immigrants and Swedish pupils.

between the summer and the school test period, respectively, to the estimations. After this is done for the specification in column 2 of Table 3, the estimate (standard error) changes from −0.81 (0.33) to −0.89 (0.40).[27]

Three other issues have also been examined in an earlier version of this paper; see Lindahl (2002). These issues are whether the class-size estimates here are biased because the same test was distributed to pupils on all test occasions (re-test bias), the assumption that class size in the fifth grade only affects learning in the fifth grade and not learning in the following summer period, and the assumption that class size in the sixth grade is uncorrelated with the error terms in equation (5). I concluded that the first issue is unlikely to affect the estimates, and that the other two issues, if anything, lead to underestimates of the positive effect of smaller classes on achievement.

V. Conclusions

I have introduced a new way of estimating the effects of class size on scholastic achievement, where I used the fact that schools are only in session during the school-year period, and out of session during the summer. By taking the difference between school- and summer-period changes in test scores, I was able to isolate the effect of school characteristics on achievement. An advantage of this identification strategy is that it can be applied to most countries or regions, as long as the break in schooling is sufficiently long. The data required are also easy to collect. Thus, this identification strategy has important advantages over those relying on some country-specific exogenous variation in class sizes.

I found that estimations using the popular value-added model yield class-size estimates insignificantly different from zero. But when the same data are applied to a difference-in-differences specification, eliminating unobservable learning fixed effects then, on the contrary, significant positive achievement effects of smaller classes are obtained. The basic difference-in-differences estimations gave class-size estimates between −0.4 and −1.0 percentile ranks, and the sensitivity analysis suggested the −0.4 percentile

[27] I also predicted test scores in the last week of school in the spring of the fifth grade and the sixth grade, and in the first week of school in the fall of the sixth grade, by assuming learning to be linear during the school year. Test scores at the end of the fifth grade were predicted by assuming that the individual learning rate in the fifth grade can be approximated by the estimated individual learning rate in the sixth grade, net class-size effects. I also scaled up the summer change in learning so that it would be comparable to the length of a school year. Estimating the DD specifications using these predicted percentile ranks test scores strengthens the positive achievement effect of smaller classes found earlier. For instance, using predicted scores and the specification in column 2 of Table 4, the class-size estimate (standard error) changes from −0.95 (0.32) to −2.46 (0.85).

rank to be a lower-bound estimate of the positive effect of smaller classes. Thus, reducing class size by one pupil in Swedish schools gives rise to at least 0.4 percentile rank higher test scores, on average. I also found that immigrants' children benefit more from smaller math classes. The magnitude of the class-size effect and the result that some disadvantaged groups benefit more from smaller classes are in line with the results in Angrist and Lavy (1999) for Israel and in Krueger (1999) for the US.

This study has also contributed to the methodology of estimating the effects of school inputs, by showing how sensitive the widely used value-added model is to an implicit assumption. The VA model does not appear to accurately capture the class-size effect, due to its failure to eliminate unobservable factors with an independent effect on learning.

References

Anderson, T. W. and Hsiao, C. (1981), Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association LXXVI*, 598–606.

Angrist, J. D. and Lavy, V. (1999), Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement, *Quarterly Journal of Economics CXIV*, 533–575.

Bonesrønning, H., Falch, T. and Strøm, B. (2004), Teacher Sorting, Teacher Quality, and Student Composition, forthcoming in *European Economic Review*.

Boozer, M. and Rouse, C. (2001), Intraschool Variation in Class Size: Patterns and Implications, *Journal of Urban Economics 50*, 163–189.

Case, A. and Deaton, A. (1999), School Quality and Educational Outcomes in South Africa, *Quarterly Journal of Economics CXIV*, 1047–1084.

Cronbach, L. J. (1951), Coefficient Alpha and the Internal Structure of Tests, *Psychometrika XVI*, 297–334.

Dobbelsteen, S., Levin, J. and Oosterbeek, H. (2002), The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition, *Oxford Bulletin of Economics and Statistics 64*, 17–38.

Entwisle, D. R., Alexander, K. L. and Olson, L. S. (1997), *Children, Schools and Inequality*, Westview Press, Boulder, CO.

Hanushek, E. A. (1979), Conceptual and Empirical Issues in the Estimation of Educational Production Functions, *Journal of Human Resources XIV*, 19–41.

Hanushek, E. A. (1997), Assessing the Effects of School Resources on Student Performance: An Update, *Educational Evaluation and Policy Analysis 19*, 141–164.

Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2003), Why Public Schools Lose Teachers, *Journal of Human Resources 39*, 326–354.

Heckman, J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement 5*, 475–492.

Heyns, B. (1978), *Summer Learning and the Effects of Schooling*, Academic Press, New York.

Hoxby, C. M. (2000), The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation, *Quarterly Journal of Economics CXV*, 1239–1285.

Krueger, A. B. (1999), Experimental Estimates of Education Production Functions, *Quarterly Journal of Economics CXIV*, 497–532.

Krueger, A. B. and Whitmore, D. M. (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.

Lazear, E. P. (2001), Educational Production, *Quarterly Journal of Economics 116*, 777–803.

Lindahl, M. (2000), Summer Learning and the Effects of Schooling: Evidence from Sweden, in *Studies of Causal Effects in Empirical Labor Economics*, Ph.D. thesis, SOFI, Stockholm University, Akademitryck.

Lindahl, M. (2002), Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement, mimeo, University of Amsterdam.

Lindsey, J. K. and Cherkaoui, M. (1975), Some Aspects of Social Class Differences in Achievements among 13-year-olds, *Comparative Education 11*, 247–260.

Marklund, S. (1962), *Skolklassens storlek och struktur: Studier rörande elevernas kunskaper i relation till klassens storlek och homogenitet* (Size and Structure of the School Class: Studies on Pupils' Knowledge in Relation to Class Size and Homogeneity), Almqvist & Wiksell, Stockholm.

National Agency for Education (1999), Samband mellan resurser och resultat: En studie av landets grundskolor med elever i årskurs 9 (The Relation between Resources and Results: A Study of Comprehensive Schools with Students in 9th Grade), Report no. 170, Stockholm.

Rivkin S. G., Hanushek, E. A. and Kain, J. F. (2005), Teachers, Schools and Academic Achievement, *Econometrica 73*, 417–458.

First version submitted August 2002;
final version received July 2004.