Teacher Experience and the Class Size Effect - Experimental Evidence

Steffen Mueller *

University of Erlangen-Nuremberg and CESifo

Abstract

We analyze teacher experience as a moderating factor for the effect of class size reduction on student achievement in the early grades using data from the Tennessee STAR experiment with random assignment of teachers and students to classes of different size. The analysis is motivated by the high costs of class size reductions and the need to identify the circumstances under which this investment is most rewarding. We find a class size effect only for senior teachers. The effect is less pronounced at lower quantiles of the achievement distribution. We further show that senior teachers outperform rookies only in small classes. Interestingly, the class size effect is likely due to a higher quality of instruction in small classes and not due to less disruptions.

Keywords: class size, teacher experience, student achievement

JEL Classification: I2, H4, J4

---

*Friedrich-Alexander-University Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany, email: steffen.mueller@wiso.uni-erlangen.de, phone: +49 9115302344, fax: +49 9115302178

1 Introduction

The conflicting results of the early literature on the effect of school resources on student achievement as summarized by Hanushek (1986) led to a large experimental project with random assignment of students and teachers to classes of different size. Krueger (1999) draws two conclusions from the Tennessee Student/Teacher Achievement Ratio (STAR) experiment: First, class size matters for student achievement and second, "measured teacher characteristics explain relatively little of student achievement" (Krueger 1999:514). Utilizing (non-experimental) data from Texas, Rivkin et al. (2005) find large effects of unobserved teacher heterogeneity while they also conclude that the effects of observable teacher characteristics are generally small. Aaronson et al. (2007) arrive at similar conclusions using data from Chicago. From a policy maker's point of view, these findings suggest that student achievement can likely be influenced by class size reduction but little by observed teacher characteristics. The result that unobserved teacher characteristics seem to be important is of limited help for optimal teacher allocation because the policy maker would be required to rank teachers according to some criteria that cannot be observed directly. In the absence of random matching of students, teachers, and schools, such rankings are inherently prone to criticism.[1]

Ding and Lehrer (2010:41) conclude that "small classes do not work consistently and unconditionally". Also, Rice (2002) pointed out that it is of special interest for the policy maker to know the circumstances under which expensive class size reductions are most effective. By relating student test scores to subsequent earnings, Krueger (2003) estimated that the up-front investments necessary for reducing class size from 22 to 15 students has an internal rate of return of 5 to 7 percent. In that view, finding (controllable) moderating factors that amplify beneficial class size effects is equivalent to identifying circumstances where the investment in class size reductions is more rewarding. A natural starting point is to look at factors that influence class size effects and are observed by the policy maker. Teacher experience is such a possibly important moderating factor.

---

[1] Typically, teacher quality is estimated using value added models. Rothstein (2010) gives a good treatment of the assumptions that have to be made to arrive at reliable results with this method.

Therefore, we study the influence of teacher experience on the class size effect. We derive hypotheses from a theoretical model and test them using data from the Tennessee STAR experiment. One of our main empirical results is that only senior teachers generate class size effects. On the other hand, rookies are as effective as senior teachers in regular size classes. Our theoretical considerations point at improved instructional quality rather than improved class discipline as the most likely reason for class size effects. The policy advice is to assign senior teachers to small classes and inexperienced teachers to regular size classes in order to maximize student achievement. We also provide some back of the envelope calculations for the internal rate of return on investments in class size reductions.

Furthermore, a society may have preferences regarding the inequality of the achievement distribution. It may, e.g., pursue equality of opportunity goals and support the learning of weaker or disadvantaged students. Alternatively, a society may support the emergence of an elite that is clearly outperforming the median student. To explore this, we extend our analysis and allow for differing interaction effects of class size and teacher experience along the unconditional student achievement distribution using unconditional quantile regressions as proposed by Firpo et al. (2009). We find that senior teachers generate the most beneficial class size effects at middle and higher deciles of the achievement distribution.

2 Literature

The empirical literature on class size effects disagrees about class size reductions as a means for better student learning. In his summary of the literature, Hanushek (1997:148) states that "there is no strong or consistent relationship between school resources and student performance." A theoretical model of Lazear (2001) explains how this lack of evidence can nevertheless be consistent with the existence of beneficial effects of class size reductions. His model derives the optimal class size from student behavior and the costs of smaller classes. According to Lazear (2001), students learn more from a lecture of given length if they experience less disruptions within the classroom. As disruptions

are primarily caused by misbehaving students, so his argument goes, these students are frequently sorted into smaller classes in practice. This can explain why class size effects are not found using data that cannot account for student sorting that is based on misbehavior.

What is more, most studies surveyed in Hanushek (1997) cannot draw on an experimental design that ensures random assignment of students and teachers to small and regular classes and are therefore subject to this kind of criticism. Besides the sorting problem stressed by Lazear (2001), the usual problem of omitted variables may invalidate the results of these studies. In addition, Krueger (2003) shows that an alternative weighting of the studies surveyed in Hanushek (1997) leads to a systematic relationship between class size and student achievement.

Random assignment of teachers and students to classrooms of different size overcomes problems of sorting and omitted variables and allows causal inference. The Tennessee Student/Teacher Achievement Ratio experiment is the only large scale data set for the US that is collected under random assignment and allows for the analysis of class size effects. Studies based on this data (e.g. Finn and Achilles 1990; Mosteller 1995; Krueger 1999) find a positive effect of class size reductions that is both statistically and economically significant. However, as like many social experiments, the STAR project was not perfect in the sense of random assignment and I will briefly address some concerns below.

Similar to class size effects, teacher effects on student achievement have been an important field of academic research for decades. It seems to be accepted wisdom in the literature that unobserved teacher characteristics are more important than observed characteristics (see e.g. Rivkin et al. 2005). Among the observed characteristics, although not large in magnitude, the effect of teacher experience on student achievement is found to be positive by many studies (Goldhaber and Brewer 1997, Jepsen and Rivkin 2002, Nye et al. 2004, Rockoff 2004, Clotfelter et al. 2006).

Although Rivkin et al. (2005), for example, compare the effect sizes of teacher quality and class size reductions, to the best of our knowledge, there is no study that combines the two strands of the literature and systematically analyzes the joint effect of teacher experience and class size reductions on student achievement. What is more, no study

analyzes the effect of class size reductions and/or teacher experience on different quantiles of the unconditional achievement distribution. This study aims at filling both gaps in the literature.

3   The Interaction of Teacher Experience and Class Size

It is well recognized that any effect of class size reduction on student achievement must be transmitted via different learning and/or teaching processes in the classroom. It seems reasonable to assume that teacher experience is an important determinant of the functioning of such processes. As there exists no elaborate theory on how teacher experience influences knowledge transfer in small vs. regular classes, we structure our thoughts about this question in a simple model building on the work of Lazear (2001)

$$L_{ics} = p_{cs}^n \cdot q(n, E)_{cs} + X_{ics}, \tag{1}$$

where L_{ics} is the learning outcome of student i in class c of school s, p is the probability that a student is not disrupting his own or others' learning at any moment in time (with $p > 0$), n is the number of students in class c, q is the value of a unit of instructional time, E is teacher experience, and X are student, teacher, and school characteristics. We borrow from Lazear (2001) the distinction between the time available for instruction (resulting from p^n) and the quality of this time (q). In this framework, p does not depend on teacher experience and we will drop this restriction below.

In Equation 1, learning is influenced via disruptions p^n and the quality of instruction $q(n, E)$. The existence of disruptions ($p < 1$) induces beneficial class size effects. Supporting this specification, Rice (1999) and Blatchford et al. (2002) find that more time is devoted to instruction if the class is smaller.

To structure the discussion below, we now discuss the partial derivatives of q with respect to n and E. Studies from educational science (e.g. Blatchford et al. 2002) tell us that teachers use smaller classes for more individualized teaching and more task-oriented interactions between teacher and students. Teachers know their class much better and

can accommodate the needs of the individual student. Thus, we find it reasonable to assume that the quality of instruction per unit of instructional time does at least not decrease if class size is reduced, i.e., $\frac{\partial q(n,E)}{\partial n} \leq 0$. From the theoretical point of view, the sign of $\frac{\partial q(n,E)}{\partial E}$ is more controversial. One could argue that young teachers come with the most recent knowledge, a higher enthusiasm, or up-to-date teaching methods. Contrarily, teaching quality may be first and foremost improved by on-the-job experience constituting an advantage for senior teachers. For instance, being familiar with the content that has to be taught, experienced teachers might spend more effort on how to teach and thereby improve the quality of instruction. Empirical evidence on the effect of teacher experience on student achievement clearly points to a positive relationship (see e.g. the studies of Goldhaber and Brewer 1997, Jepsen and Rivkin 2002, Nye et al. 2004, Rockoff 2004, Clotfelter et al. 2006) and we therefore assume in the following that $\frac{\partial q(n,E)}{\partial E} \geq 0$.

The class-size effect is the first derivative of Equation 1 with respect to n and, dropping subscript cs, is given by

$$\frac{\partial L}{\partial n} = p^n \cdot \ln p \cdot q(n, E) + p^n \cdot \frac{\partial q(n, E)}{\partial n}. \tag{2}$$

With the above assumptions, the sign of the class size effect is negative and thus points to a higher amount of learning in smaller classes.[2]

To assess the optimal allocation of experienced and inexperienced teachers to classes of different size, we are interested in the effect of teacher experience on the class size effect and therefore take the first derivative of Equation 2 with respect to E

$$\frac{\partial^2 L}{\partial n \partial E} = p^n \left(\ln p \cdot \frac{\partial q(n, E)}{\partial E} + \frac{\partial^2 q(n, E)}{\partial n \partial E} \right). \tag{3}$$

The negative class size effect will become more pronounced with higher teaching experience if the cross derivative $\frac{\partial^2 L}{\partial n \partial E}$ is negative. Given $\frac{\partial q(n,E)}{\partial E} \geq 0$, the sign of the cross derivative depends on the sign of $\frac{\partial^2 q(n,E)}{\partial n \partial E}$ which indicates whether the class size ef-

---

[2] Due to random assignment of students and teachers into classes of different size in Project STAR, the X variables in Equation 1 do not depend on class size and, therefore, do not show up in the first derivative.

fect on teaching quality (i.e. $\frac{\partial q(n,E)}{\partial n}$) increases or decreases with teaching experience. As $\frac{\partial q(n,E)}{\partial n} \leq 0$, $\frac{\partial^2 q(n,E)}{\partial n \partial E} < 0$ would suggest that class size reductions are the more beneficial the more experienced the teacher is and vice versa. Intuitively, this would be consistent with the assertion that experience is necessary for the effective use of more instructional time per student.

However, one may wonder whether there is a second effect of teacher experience on learning that takes effect via a change in disruptive student behavior. Augmenting the model by allowing p to depend on E extends Equation 3 to

$$\frac{\partial^2 L}{\partial n \partial E} = \left(p(E)^n \cdot \frac{\partial p(E)^n}{\partial E} \right) \left(\{n \cdot \ln p(E) + 1\} \cdot q(n, E) + n \cdot \frac{\partial q(n, E)}{\partial E} \right)$$
$$+ p(E)^n \left(\ln p(E) \cdot \frac{\partial q(n, E)}{\partial E} + \frac{\partial^2 q(n, E)}{\partial n \partial E} \right). \tag{4}$$

Hence, the term in the first two sets of large parentheses is added to Equation 3. The most plausible assumption about the sign of $\frac{\partial p(E)^n}{\partial E}$ is that more experienced teachers have less disruptions within their class room. Rice (1999) indeed finds that senior teachers need less time to keep order. Assuming this, the overall sign of the two additional terms in Equation 4 is positive if $\{n \cdot \ln p(E) + 1\} > 0$, which is true for values of $p \geq .97$ and class sizes below 32.[3] As a result, Equation 4 as a whole may become positive even if Equation 3 was negative. Hence, the class-size effect does not necessarily increase with teacher experience even if $\frac{\partial^2 q(n,E)}{\partial n \partial E} < 0$. Intuitively, this makes sense because $\frac{\partial p(E)^n}{\partial E} > 0$ constitutes the highest advantage of senior teachers with respect to disruptions in the largest classes and this may counterweight any potential advantage of seniors with respect to the class size effect on teaching quality (i.e. $\frac{\partial^2 q(n,E)}{\partial n \partial E}$).[4]

Whether teacher experience influences the class size effect via the disruption channel

---

[3] Rice (1999) provides some guidance on plausible values of p. Her results point at no more than 10 percent of time spent on maintaining order. As her classes have on average 23 students, this implies $p = .995$ in my model. However, Rice (1999) analyzes 8th graders who might have higher discipline than younger students. Nevertheless, a value of $p = .97$ in classes of 22 students would be equivalent with half of the class time lost due to disruption, which seems implausibly high even for very young students.

[4] This trade off exists only for realistic values of p and n, say $p \geq .95$ and $n \geq 15$.

and/or the quality-of-instruction channel is tested in two steps. First, we test whether the disruption channel plays a role, i.e. whether p depends on teacher experience. Second, we compare the outcome difference between seniors and rookies by class size. If p does not depend on experience, any changes in the outcome difference that follow class size reductions can be attributed to the quality-of-instruction channel.[5] If we cannot rule out the existence of a disruption channel in the first step, changes in the outcome difference between teacher types cannot be unambiguously attributed to disruption or quality.

4   The STAR Data

The Tennessee Student/Teacher Achievement Ratio (STAR) experiment was legislated by the State of Tennessee and designed to assess the effect of class size on student achievement. The experiment took place in 79 public elementary schools and followed one cohort of about 6,500 students from kindergarten through third grade, beginning in the fall of 1985 and ending in 1989.

To allow causal inference, teachers and students were randomly assigned within schools to classes of different size. The three class types are small classes (13-17 students), regular classes (22-25 students), and regular classes with a full-time aide.[6] Achievement in reading and math was measured via Stanford Achievement Tests (SAT) that provide test scores that can be compared across grades.[7]

4.1   Validity of the Experiment

The proper implementation of random assignment was permanently supervised by university staff and was not under the control of school personnel. Nevertheless, there was some debate about the validity of the experiment. While Hanushek (1999) and Hoxby

---

[5] We will discuss later whether dropping the assumed independence of p and n would allow an alternative interpretation of our results.

[6] The latter two class types will be pooled in our analysis as we, like most other studies, find no sizeable differences in results and because also the regular classes without full-time aide were supported by part-time aides at the time, which would additionally complicate the interpretation of any differences.

[7] Additionally, the Basic Skills First (BSF) test was conducted. As the BSF scores cannot be meaningful compared across grades, we will not use them.

(2000) criticize the implementation of the experiment or have doubts with respect to the insights that can be gained from experiments at all, Krueger (1999) and Nye et al. (1999) show that some of the criticisms put forward do not seem to affect results. Three implementation problems and their consequences are briefly discussed below.

First, since kindergarten was not compulsory in Tennessee at the time, a number of students joined the project when they entered first grade. Additionally, ordinary student mobility into and out of Project STAR schools happened. To deal with this, new students were randomly assigned to class types regardless of the grade at which they entered STAR. Under the assumption that parental decisions leading to student attrition out of STAR schools are unrelated to class type assignment and teacher characteristics, attrition will not affect our results. Nye et al. (1999:137) find that "the students who dropped out of the small classes actually evidenced higher achievement than those who dropped out of the larger classes, suggesting that the observed differences in achievement between students who had been in small and larger classes were not due to attrition." Therefore, students who switch between STAR schools or leave the sample before third grade are not excluded from our analysis.

Second, although students were intended to stay in the class type they were originally assigned to, 250 students managed to switch from regular to small classes or vice-versa *within the same school*. Comparing their prior achievement, we generally find that students who moved into small classes had a slightly lower achievement in the prior grade than the non-switchers and, hence, they are not expected to amplify any beneficial class size effect. Contrarily, the 45 students that moved from small into regular classes were above average if they moved after first grade and below average if they moved after second grade (n=17). To deal with within-school switching as a potential source of self selection bias, we exclude all post-switching observations of the 250 students and we end up with 21,443 (21,748) observations on reading (math) scores.[8]

Third, because of student mobility, some overlap occurred in the actual class size

---

[8] Excluding a selective group may not solve all the problems. We rather argue that the potentially problematic group of 17 students is too small to drive our results.

between small and regular classes: i.e. some small classes may have had more students than regular classes. Therefore, we will check whether results qualitatively change with actual class size instead of class type as a regressor.

5   Empirical Model and Results

The aim of the paper is to assess whether the class size effect depends on teacher experience. If this is the case, the theoretical model provides the framework to additionally test whether any difference in the class size effect by teacher experience is due to differences in disruptive behavior, i.e. time available for instruction, or teaching quality per unit of time available for instruction.

The implementation of the test that distinguishes between the two channels is done in two steps. We will first compare the achievement difference between inexperienced and experienced teachers in regular classes. If no difference shows up there, seniors have no advantage with respect to disruptive behavior.[9] In that case, our model predicts that any change in the senior-rookie difference that occurs when class size is reduced must be due to differences in the change of the quality of instruction, i.e. $\frac{\partial^2 q(n,E)}{\partial n \partial E}$. However, if we cannot rule out the disruption channel, we have no chance to disentangle the two channels.

5.1   Achievement Levels

We begin by estimating the following regression:

$$
\begin{aligned}
Y_{icgs} \;=\;\; & \beta_0 + \beta_1 SMALL_{cgs} + \beta_2 ROOKIE_{cgs} + \beta_3(SMALL_{cgs} \cdot \\
& ROOKIE_{cgs}) + \beta_k S_{icgs} + \beta_j T_{cgs} + \alpha_s + \gamma_g + \epsilon_{icgs}
\end{aligned}
\tag{5}
$$

where i denotes individual students, c classes, g grades, and s schools. Y_{icgs} is the SAT test score standardized to mean zero and variance one. The vector S contains

---

[9] This conclusion is possible because we assumed that the quality of instruction of senior teachers is at least as high as the rookies' quality.

student characteristics like gender, race, and socioeconomic background while T includes teacher characteristics like gender, race, and highest degree achieved. The class type $SMALL$ indicates assignment to a small class, $ROOKIE$ measures teacher experience, and $SMALL \cdot ROOKIE$ is the interaction of both.[10]

In the definition of teacher experience, we follow the more recent economic literature (Jepsen and Rivkin 2002, Nye et al. 2004, Rockoff 2004, Rivkin et al. 2005) and collapse the information into a binary variable that is one if the teacher has less than three years of experience and zero otherwise. With this definition we have 162 rookies in the data, of whom 63 were assigned to small classes. Although a higher number of rookie teachers in small classes may allow more precise estimation of β_3, increasing the number of rookie teachers by defining inexperience as having less than, say, four or five years of experience will dilute the marked differences between seniors and rookies and is therefore not a promising alternative.[11]

Although the data could in principle be analyzed separately by grade, the number of rookie teachers in small classes would be too small to do so. For instance, the number of small class rookies in third grade is 13. In the following analysis, students are pooled over all grades with the grades controlled by a set of dummies γ_g. As random assignment took place within schools, Equation 5 contains school fixed effects by adding a dummy variable α_s for each school. If random assignment was effective, ϵ_{icgs} is uncorrelated with each of the regressors of Equation 5 and a simple OLS estimation will yield unbiased estimates of the average treatment effects.

Errors are correlated within students over time and within classes (i.e. teachers) in the cross section. Cameron et al. (2011) derive an estimator for standard errors that are robust to this sort of non-nested two-way cluster structure and we apply their method for our OLS estimations. In our study, the two-way cluster-robust standard errors are very close to those obtained by simply clustering at the class level.

---

[10] Summary statistics for the variables used are presented in Table 1.

[11] Our own experimentations show that average student achievement does not further increase when teacher experience exceeds three years. We have nevertheless checked our results with different definitions of a rookie. In line with prior expectations, the effect of being an inexperienced teacher gets smaller on average, the more teachers we define to be inexperienced by moving the cutoff to higher experience levels.

Tab. 1: Summary statistics of Regressors in Equation 5, Means

Variables	K	1	2	3	All
			Grade		
		Student Level			
small class	.304	.280	.274	.283	.285
rookie teacher	.135	.164	.119	.080	.126
small class and rookie teacher	.047	.038	.032	.026	.036
male student	.520	.522	.526	.522	.522
white student	.683	.676	.655	.679	.673
on free lunch	.475	.501	.485	.478	.485
Observations	5,366	5,934	5,228	5,220	21,748
		Teacher Level			
small class	.391	.366	.391	.416	.390
rookie teacher	.142	.155	.116	.084	.125
small class and rookie teacher	.062	.048	.044	.041	.048
male teacher	.000	.006	.009	.031	.012
white teacher	.837	.824	.788	.784	.809
lowest degree (i.e. bachelor)	.649	.646	.638	.572	.626
Teachers	325	336	320	320	1,301

Descriptive statistics for the 21,748 observations used in the OLS estimation on math achievement as reported in Table 2.

OLS quantifies the effects of small classes and inexperienced teachers at the mean of the student achievement distribution. However, it is also interesting to know whether the effects are higher for low achieving or high achieving students. If, for example, an equality of opportunity policy is pursued then greater equality in student achievement by helping weaker students is likely intended. Contrarily, if society favors the formation of a student elite, it will appreciate beneficial effects for the best students. Unconditional quantile regression as introduced by Firpo et al. (2009) estimates how quantiles of the unconditional achievement distribution change due to class size reductions. Importantly, "unconditional" does not mean that other covariates are not held constant. It means that we estimate ceteris paribus effects at certain quantiles of the unconditional achievement

distribution.[12] We will use the method of Firpo et al. (2009) to illustrate whether our results for the conditional mean carry over to all quantiles.

Tab. 2: OLS and Unconditional Quantile Regression Estimates of the Joint Effect of Class Size and Teacher Experience on Achievement

Quantile	SMALL		ROOKIE		SMALL*ROOKIE	
	Standardized SAT Score on Reading					
OLS	.136***	(.016)	-.005	(.024)	-.125***	(.039)
0.1	.084***	(.012)	-.010	(.022)	-.092**	(.041)
0.2	.114***	(.015)	-.040*	(.024)	-.111***	(.042)
0.3	.143***	(.018)	-.081**	(.035)	-.127*	(.068)
0.4	.172***	(.020)	-.024	(.037)	-.178***	(.062)
0.5	.144***	(.018)	-.044	(.030)	-.144***	(.049)
0.6	.152***	(.014)	.016	(.022)	-.116***	(.040)
0.7	.163***	(.015)	.044*	(.023)	-.146***	(.037)
0.8	.175***	(.018)	.047*	(.025)	-.142***	(.038)
0.9	.156***	(.020)	.020	(.023)	-.158***	(.042)
	21,443 Observations					
	Standardized SAT Score on Math					
OLS	.162***	(.021)	.036	(.033)	-.143***	(.052)
0.1	.115***	(.020)	-.016	(.033)	-.143**	(.062)
0.2	.165***	(.019)	-.000	(.034)	-.188***	(.052)
0.3	.191***	(.019)	-.007	(.035)	-.107**	(.053)
0.4	.181***	(.017)	-.018	(.028)	-.102**	(.050)
0.5	.162***	(.017)	.007	(.027)	-.077*	(.045)
0.6	.182***	(.018)	.055**	(.026)	-.142***	(.044)
0.7	.169***	(.018)	.056***	(.025)	-.126***	(.041)
0.8	.167***	(.018)	.101***	(.027)	-.146***	(.045)
0.9	.164***	(.025)	.107***	(.033)	-.176***	(.055)
	21,748 Observations					

Dependent variables are standardized to mean zero and variance one. For example, 0.136 means that achievement is 0.136 standard deviations higher. The effects on the unconditional quantiles are estimated via RIF regressions as proposed in Firpo et al. (2009). Standard errors in parentheses. ***,**,* denote significance at the 1, 5, or 10 percent level, respectively. OLS standard errors are robust to two-way clusters at the teacher level (i.e., class level) and at the student level (over time) applying the method of Cameron et al. (2011). For quantile regression, standard errors based on 200 bootstrap replications are reported. The differences by subject in the number of observations are due to missing test score information.

The results from the basic specification in Equation 5 are presented in Table 2 and we will start with a discussion of the OLS results first. The reference category are students in regular classes that have a senior teacher. Hence, β_2 measures the difference in student

---

[12] See Ding and Lehrer (2011) for a recent application of this method to estimate class size effects for kindergarten.

achievement between senior and rookie teachers in regular classes and β_3 identifies the difference within small classes. As β_2 is insignificant and close to zero for both subjects and because teaching quality is assumed to rise in teacher experience, the finding is consistent with the basic theoretical model that sets $\frac{\partial p(E)^n}{\partial E} = 0$. We therefore conclude that the representative student's probability of disruptive behavior is not affected by teacher experience in our data. Thus, we find no support for the notion that class size effects are moderated by teacher experience via the disruption channel.[13] Additionally, the similar performance of seniors and rookies in regular size classes uncovers an important heterogeneity in the widespread view that teacher experience increases student achievement (see Krueger 1999 or Clotfelter et al. 2006).

The first column of Table 2 presents the small class effect for experienced teachers. The OLS estimate for reading (math) shows that students in such a class perform on average 0.14 (0.16) test score standard deviations better than those in a regular class with senior teacher. However, the large negative coefficient β_3 in the third column indicates that the beneficial class size effect completely vanishes if a rookie teaches a small class.[14] As we were not able to detect effects on the class size effect via the disruption channel (because our estimate of β_2 was zero), this finding suggests an influence of teacher experience on the class size effect via the quality-of-instruction channel.

Finally, the results show that student achievement in classes of inexperienced teachers does not vary with class size.[15] Given $\frac{\partial q(n,E)}{\partial n} \leq 0$, this finding is only consistent with the explanation that neither the quality of instruction nor the available time for instruction

---

[13] Remember that we do not apply a direct measure for disruptive behavior and that we derive this conclusion from a theoretical model that ignores peer effects that may origin in other channels than disruptive behavior, e.g. in class composition. This simplified framework is motivated by randomization in Project STAR assuring that there are no systematic differences in class composition after controlling for class size and school affiliation.

[14] The picture does not change if we use actual class size instead of class type as regressor. Note also that rookies and seniors are very similar in gender, race, and their probability of teaching a small class but differ in their highest degree achieved: only 7 percent of the rookies have more than a bachelor's degree compared to 42 percent of the seniors. Having no more than a bachelor's degree is always controlled for in my regressions. Adding interaction terms between bachelor degree and small class assignment or instead between bachelor degree and rookie status changes nothing in our findings with the additional interaction terms being insignificant. Results are available upon request.

[15] As $\beta_1 + \beta_3 = \beta_2$ cannot be rejected by the data (p-value for reading = 0.75 and for math = 0.82), no class size effect exists within the group of inexperienced teachers.

increases for rookie teachers as class size decreases. Our theoretical framework can offer only one explanation for this finding: p is one or very close to it. A p close to unity challenges the view that class size effects are predominantly driven by a reduction in disruptive behavior.

Could our results be in line with a disruption effect if the theoretical model would allow p to depend on n? We assumed throughout the paper that the average individual probability of disciplined behavior in a class, p, does not depend on class size. While we think this is a reasonable assumption as disruptive behavior should primarily be driven by personal student characteristics (e.g. Carrell and Hoekstra 2010), a reduction in class size may students feel to be more under the teacher's supervision and therefore raise p. Looking only at senior teachers, a disruption story consistent with our findings could then be that p increases as class sizes is reduced. However, the only way to make this a plausible story in the light of our findings for both teacher types (i.e. in the presence of equal performance of seniors and rookies in regular classes and the lack of small class effects for rookies) is to drop the assumption $\frac{\partial p(E)^n}{\partial E} \geq 0$. The picture that emerges, then, is one of a p equal or close to unity for rookies and a lower p for seniors that increases as class size is reduced (at the same time, instructional quality of seniors must exceed the rookies', independent of class size). The assumption that rookies create a better class discipline than experienced teachers is necessary for this alternative interpretation but implausible given, e.g., the results summarized in the meta-analysis of Veenman (1984) pointing to class discipline as the most severe problem of beginning teachers.

Thus, the main results are that only seniors generate class size effects and that the class size effect likely comes through an increase in teaching quality per unit of instructional time.[16] Our results are not in line with theories that explain class size effects solely via assumed reductions in disruptive behavior. Instead, the results confirm scholars that argue on grounds of improvements in teaching quality that become possible for certain kinds of teachers in smaller classes.

---

[16] All results persist if regular classes with and without fulltime aide are not pooled in the regressions. Results are available upon request.

We will now discuss two alternative explanations for our findings. First, it could be that experienced teachers in small classes react to the experiment and provide more effort than in a non-experimental setting ("Hawthorne effect"). To test this we follow Krueger (1999:527) who argued that such experimental incentives should be absent within the control group, i.e. in regular size classes. Restricting the analysis to the control group and using actual class size instead of class type as regressor, we again find class size effects to be stronger for more experienced teachers and therefore add to Krueger (1999). This finding is also not in line with the concern that, for some reason, randomization might have failed in such a way that the worst experienced teachers have been assigned to regular classes and that this may drive our results. What is more, such a manipulation of the randomization process would have required that the experimenters know the ability of participating teachers, which seems very unlikely in the light of the study of Jacob and Lefgren (2008) showing that even school principals are limited in their ability to detect performance differences between teachers.

A second alternative explanation for our findings could be that low-ability teachers may drop out of the school system within the first years of teaching so that our measure of experience also contains the influence of selection into the group of stayers. The beneficial effect of having more experience would then be biased upwards because teacher experience would be positively correlated with teacher ability, which is part of the error term. We cannot test for selective attrition of teachers with STAR data because each teacher is observed only once. However, the literature suggests that higher-ability teachers leave the profession. For instance, Murnane and Olson (1990) find that teachers scoring higher at the National Teacher Exam are more likely to leave the teaching profession. Nevertheless, this is only part of the story as teachers with good exams may not be the most effective teachers in terms of student achievement. Addressing this question, Rivkin et al. (2005) show that teachers leaving the profession after one year had similar student outcomes as stayers. As a final argument, selective teacher attrition cannot explain the similar outcome of both teacher types in regular size classes found in our study. We therefore do not believe that selective teacher attrition is amplifying our positive experience effects.

However, further research on this question is necessary.

The unconditional quantile regression results in Table 2 allow a deeper look into what exactly happens to high and low quantiles. The lowest deciles of the achievement distribution gain less from small classes with senior teachers than higher quantiles. Hence, the introduction of small classes with senior teachers increases overall achievement inequality due to a larger inequality at the bottom half of the unconditional achievement distribution. From the third decile upwards, the coefficient on $SMALL$ is roughly stable in both subjects and no increase in inequality happens there. While for reading, rookie teachers do not generate a class size effect at any part of the distribution (i.e. $\beta_1 + \beta_3 = 0$), for math this is only true for the lower and upper deciles. Students located in the range between the third and the seventh decile perform better in small classes even if an inexperienced teacher instructs math. Interestingly, the coefficients on $ROOKIE$ increase along the achievement distributions in reading and math. For higher quantiles, this means that rookies outperform seniors in regular classes. It also contributes to the advantage of rookies in small classes over seniors in regular classes at most upper deciles. Explaining this heterogeneity might be an interesting topic for future research. The overall picture, however, supports our prior findings: the senior's advantage in teaching small classes ($\frac{\partial^2 q(n,E)}{\partial n \partial E} < 0$) and the absence of a general advantage of seniors with respect to class discipline ($\frac{\partial p(E)^n}{\partial E} = 0$).

5.2   A Value Added Specification

Comparison of achievement levels may be inappropriate because differences in levels cloud all initial differences different students may bring into a certain grade level. Such differences will bias our results if those with a starting advantage still have an advantage at the end of the year and if starting levels are systematically different for different class sizes or teacher experience levels. Differences in starting endowments may be due to family background or school experiences.

The standard tool for assessing teacher effectiveness that deals with this problem is

a value added model (VAM). It measures achievement gains between a student's current and past test score results, e.g., by including previous year's test score as an additional regressor.[17] The lagged dependent variable implicitly controls for school experiences, socioeconomic status, individual background factors, i.e., all of individual history that is related to achievement, as long as it is reflected in the previous year's test score.

There are two specific characteristics in the application of a VAM to data with random assignment of teachers and students that have to be addressed before presenting the empirical specification. First, as we are dealing with random assignment, a starting advantage in the first year of STAR is ruled out. Nevertheless, different starting endowments in the following grades may arise. Second, the VAM specification will give biased results of the value that is added by current class type or teacher experience if the student history also affects the rate of learning today (a point that was e.g. made by Ballou et al. 2004). It is typically assumed that past advantages increase the rate of learning today. If this is the case and students stay in their class type, the teacher that is assigned to a small class in grades following kindergarten will teach students having higher initial rates of learning than students in regular size classes. Hence, the class size effect could be biased in the VAM specification despite random assignment because random assignment took place in earlier periods.

In the context of VAM's, Rothstein (2010:176) argues that "... the necessary exclusion restriction is that teacher assignments are orthogonal to all other determinants of the so-called gain score." Hence, as long as random assignment of teachers holds, the difference between senior and rookie teachers *within a class type* is estimated correctly because both types of teachers face on average the same initial rate of learning within their classrooms, respectively. They face the same initial rate of learning because students of a certain class type have on average the same class type history.[18] In our empirical implementation of

---

[17] There are different types of VAM that are valid under different assumptions (see Rothstein 2010).

[18] To ensure this, we now restrict the sample to students who entered STAR in kindergarten. Remember that within-school class type switchers are excluded throughout the whole analysis.

the VAM, we therefore run the following regression separately by class type

$$Y_{icgs} = \beta_0 + \beta_1 ROOKIE_{cgs} + \beta_2 Y_{ics,g-1} + \beta_k S_{icgs} + \beta_j T_{cgs} + \alpha_s + \gamma_g + \epsilon_{icgs}. \qquad (6)$$

Tab. 3: OLS and Unconditional Quantile Regression Estimates of the Effect of Inexperienced Teachers in a Value Added Model by Class Type

Quantile	Small Class		Regular Class	
	Standardized SAT Score on Reading			
OLS	-.172***	(.037)	-.012	(.028)
0.1	-.122***	(.041)	-.005	(.024)
0.2	-.117***	(.040)	-.006	(.024)
0.3	-.117***	(.041)	-.003	(.025)
0.4	-.317***	(.070)	-.007	(.035)
0.5	-.282***	(.068)	-.031	(.058)
0.6	-.221***	(.048)	-.059	(.043)
0.7	-.220***	(.043)	.001	(.038)
0.8	-.177***	(.042)	.019	(.027)
0.9	-.140***	(.041)	-.014	(.030)
Observations	4,648		9,337	
	Standardized SAT Score on Math			
OLS	-.131**	(.054)	-.034	(.042)
0.1	-.231***	(.073)	.043	(.036)
0.2	-.225***	(.068)	-.010	(.036)
0.3	-.122**	(.058)	.030	(.036)
0.4	-.164***	(.054)	.009	(.039)
0.5	-.203***	(.058)	-.036	(.034)
0.6	-.198***	(.051)	.007	(.037)
0.7	-.100**	(.046)	.032	(.037)
0.8	-.049	(.049)	.063*	(.034)
0.9	-.010	(.062)	.116**	(.044)
Observations	4,702		9,489	

The table shows estimated coefficients on $ROOKIE$. Dependent variables are standardized to mean zero and variance one. For example, -0.172 means that achievement is 0.172 standard deviations lower. The effects on the unconditional quantiles are estimated via RIF regressions as proposed in Firpo et al. (2009). Standard errors in parentheses. ***,**,* denote significance at the 1, 5, or 10 percent level, respectively. OLS standard errors are robust to two-way clusters at the teacher level (i.e., class level) and at the student level (over time) applying the method of Cameron et al. (2011). For quantile regression, standard errors based on 200 bootstrap replications are reported. The differences by subject in the number of observations are due to missing test score information.

Estimating gains in achievement typically leads to the loss of the first observation for each student because $Y_{ics,g-1}$ is not available for the first year. However, note that random

assignment assures that all students entering the project in kindergarten have the same expected endowment level at the time of school enrollment. As they start from the same level, it is possible to replace $Y_{ics,g-1}$ for kindergarten with a constant, say zero, in order to keep the first year of the data. The value assigned to the constant will only affect the estimates for the intercept and the grade dummies in Equation 6, and has no consequence for the estimation of the parameters of interest.

The results for the VAM are presented in Table 3 and corroborate our main findings from the estimation in levels: In small classes, inexperienced teachers add significantly less to the average student's knowledge than seniors while there is no difference in regular classes. The small class difference between both types of teachers is largest at the middle of the student achievement distribution. For math, the senior's advantage is also large at the first two deciles but does not exist at the eighth and ninth decile of the small class distribution. Similar to the results of the levels specification shown in Table 2, rookies outperform seniors at the top deciles of the math distribution of regular size classes.

6   Policy Implications

For the policy maker, the most important results are

1. only senior teachers generate a beneficial class size effect

2. senior and rookie teachers perform similar in regular size classes

3. the seniors' class size effect is lower at the lowest quantiles of the achievement distribution.

In the short run, an obvious policy implication arising from the first two results is to assign seniors to small classes and rookies to large classes in order to maximize achievement. If class size is reduced, then additional classes have to be installed and additional teachers are needed. If there are not enough teachers available, new teachers have to be trained. These newly trained teachers can be expected to perform (on average) as well as senior teachers in classes of regular size and, hence, they can be assigned to regular classes

without loss of student achievement. Therefore, student achievement can be improved at the aggregate level without the need for additional experienced teachers.[19]

As our results point at a moderating effect of teacher experience on the effect small classes have on the quality of instruction, the most plausible interpretation is that seniors adapt to small classes and rookies do not. Whether the above short run policy advice also holds in the longer run, therefore, depends on whether utilizing the advantages of small classes has to be learned by actually teaching them. If learning by doing is necessary, the short run policy advice will break down in the longer run as the share of teachers without small class experience will grow each year. But can a learning-by-doing-story plausibly be the reason for our findings? As teachers are exposed to the experiment only for one year, a necessary assumption for this story to hold is that seniors made their small class experience before the STAR experiment started, e.g. because of fluctuations in class size across previous student cohorts. Seniors with, say, 15 years of experience would then have a higher probability of small classes experience than seniors with only, say, 5 years of teaching experience and should, thus, perform better than the latter. To check this empirically (applying a levels specification similar to Equation 5), we split up seniors into six categories each covering three years of teaching experience plus one category containing teachers with at least 21 years of experience. Interacting these seven dummies with the small class indicator, we find beneficial class size effects for each senior category (with rookies as the reference category). While all interaction terms point at effects of at least .10 standard deviations (in math), coefficients are somewhat higher for young seniors with no more than nine years of experience which is not in line with the predictions of the learning-by-doing story.

---

[19] This may not be true if the additional demand for rookie teachers decreases average rookie quality. Jepsen and Rivkin (2002) argue in their analysis of the California class size reduction program that the massive influx of new teachers decreased average teacher quality because inexperienced and low skilled teachers were hired. While it is convincing to assume that a massive hiring of unemployed experienced teachers as in California deteriorates average teacher quality, we do not see why this should be necessarily the case when attracting additional young people to become educated as teachers. In the light of our results, the Californian problem was rather that additional inexperienced teachers were assigned to small classes, which confirms the view of Jepsen and Rivkin (2002).

If all teachers possess the techniques to appropriately deal with small classes and learning by doing is thus not necessary: Why should only seniors apply small class techniques? Intuitively, this could be the case because teachers are more concerned with basic tasks in their first years, e.g. administrative tasks, maintaining order, and standard knowledge transfer. More formally, one might think that applying small class techniques is just another task to perform.[20] Each task imposes costs on the teacher in terms of effort or cognitive demand. Marginal costs are assumed to increase in the amount of tasks performed as individuals face boundaries to their physical and cognitive capacity. Irrespectively of class size and prior small class experience, seniors' general experience reduces their costs for performing basic tasks and thus reduces their marginal costs of applying small class techniques. This, in turn, increases their propensity to actually use small class techniques. As a case in point, Rice (1999) finds that experienced teachers need less time for basic tasks. Following this argumentation, the short run policy implication mentioned above is also valid in the long run. However, further research is needed to understand the moderating role teacher experience plays with respect to the effect of class size on teaching quality.

The third finding listed above suggests that overall student achievement is maximized if no weak students are assigned to small classes. For instance, the class size effect for senior teachers at the ninth decile of the student achievement distribution in reading is roughly twice the effect at the first decile.[21] As the effect for the lowest quantiles is still positive, these figures also allow a different interpretation: if the policy maker aims at reducing the gap between good and bad students, she might achieve this goal by assigning bad performing students to small classes with senior teachers and good students to classes of regular size. However, peer effects will change and this may lead to unpredictable outcomes (see Carrell et al. 2011).

---

[20] Pate-Bain et al. (1992) list techniques that STAR teachers reported to be beneficial in small classes.
[21] See Table 2.

Similarly to Krueger (2003), we now do some back of the envelope calculations to approximate the rate of return an investment in class size reduction yields.[22] Building on estimates from Project STAR but not considering teacher experience as a moderating factor of the class size effect, Krueger (2003) compared the costs of reducing class size from 22 to 15 students with future increases in student's earnings that are assumed to arise from this investment. He estimated an internal rate of return on the investment in class size in the range of 5 to 7 percent. Based on the results of our cumulative specification and additional calculations (both presented in the appendix), we conduct a similar analysis and additionally identify the grades in which class size reduction should be performed in order to maximize the internal rate of return.

The results for the internal rate of return as presented in Table 4 depend on the expected growth rate in US real wages[23] and the number of grades in which class size reductions are performed. The table compares the present value of additional costs per student arising from reducing class size from 22 to 15 students with the present value of future real earnings advantages in US Dollars per student for different discount rates and two conservative scenarios for future real wage growth. For instance, assuming stagnating real wages during the next decades, the investment in reducing class size yields internal rates of return between 5.1 (first four grades) and 7.0 percent (only first grade). Given a moderate increase in real wages of one percent per year, the internal rate of return rises by at least one percentage point in each specification.

Hence, the internal rate of return is highest if class size is reduced only for the grade at which students enter school and steadily decreases as further grades are included. Remembering that the highest class size effects have been found in the initial grade (see Table 5 in the appendix of this paper or Table IX in Krueger 1999), this pattern comes as no surprise. Although the internal rates of return are substantial throughout all durations presented in Table 4, the policy maker may ask whether it pays to extend class

---

[22] Krueger (2003) presents in detail the assumptions necessary to perform this kind of calculations. The criticisms that are valid with respect to his calculations also apply to ours.

[23] More precisely, it depends on the future annual percentage increase of the cross sectional age-earnings profile drawn from the 2007 Current Population Survey that serves as the basis for our calculations of the present of value benefits of class size reductions. See the appendix for further details.

Tab. 4: Present value of costs and benefits of reducing class size from 22 to 15 for several discount rates as well as the internal rate of return on investment for different wage growth scenarios and different numbers of years with senior teachers

| | | Increase in Income for Wage Growth of: | | | Increase in Income for Wage Growth of: | |
Discount Rate	Cost	0 %	1 %	Cost	0 %	1 %
		Kindergarten (KG)			KG and 1st grade	
0.02	2,937	13,818	20,205	5,816	19,464	28,461
0.04	2,880	6,887	9,768	5,649	9,702	13,759
0.06	2,826	3,678	5,071	5,492	5,181	7,143
0.08	2,773	2,078	2,806	5,341	2,940	3,953
Internal Rate of Return		0.070	0.080		0.058	0.069
		KG to 2nd grade			KG to 3rd grade	
0.02	8,638	25,110	36,717	11,405	30,756	44,973
0.04	8,312	12,516	17,750	10,873	15,330	21,741
0.06	8,006	6,684	9,215	10,379	8,187	11,287
0.08	7,719	3,793	5,099	9,921	4,646	6,246
Internal Rate of Return		0.054	0.065		0.051	0.063

Assumptions: a one standard deviation increase in test scores translates into 20 percent higher income; cumulated test score advantages for different durations in small classes with senior teachers are computed as the mean of the predicted reading and math advantages as can be obtained from Table 5, e.g. four years in such classes yield reading (math) scores that are .20 (.22) standard deviations higher than in the reference category - the mean test score advantage for the four year period is thus .21; "Cost" denotes additional costs per pupil a class size reduction from 22 to 15 pupils causes in terms of the salaries of teachers and other instructing staff. See the appendix for further details.

size reductions from the initial grade to later grades. From the second and fifth column of the upper panel of Table 4 we see that the additional costs per pupil of extending the investment to the second grade are about 2,800 dollars depending on the discount rate. By comparing columns three and six, we also see that the present value of benefits exceeds these additional costs only if the discount rate is not larger than 4 percent. If only future earnings advantages are considered (and other beneficial class size effects are neglected), it may thus not be advisable to reduce class size in other than the first grade students enter school.

7 Conclusions

This study analyzes teacher experience as a moderating factor for the effect of class size reduction on student achievement in the early grades. It is motivated by the high costs of class size reductions and the need to identify the circumstances under which this investment is most rewarding. We utilize data from a large experiment with random assignment of teachers and students to classrooms of different size, the Tennessee Student/Teacher Achievement Ratio (STAR).

The main finding is that only experienced teachers are able to generate a beneficial class size effect on average student achievement. Within the framework of our theoretical model, the results are consistent with the view that teacher experience amplifies class size effects via gains in the quality of instruction but not via less disruptions. What is more, teacher experience does not matter in larger classes. Therefore, at least in the STAR experiment, the positive effects of both teacher experience and class size reductions, which are repeatedly reported in the literature, are driven by senior teachers in small classes only. The results support scholars who emphasize the improvements in teaching quality that become possible for certain kinds of teachers in smaller classes. Using unconditional quantile regression, we further find that the class size effect stems mainly from the ability of senior teachers to improve middle and high quantiles of the achievement distribution.

With respect to the external validity of the STAR results, Hanushek (1999) and Hoxby (2000) object that teachers reacted to the experimental setting of Project STAR. Both authors suspect that small class teachers may have worked harder because they felt monitored and aimed at fulfilling the expectations that they thought arise from teaching the small class. In that view, small class effects emerged from the incentives the STAR experiment provided and cannot be expected in non-experimental settings. Even if one is willing to follow this argument (Krueger 1999 and we present evidence against it), our results show that senior teachers were able to fulfill expectations while rookies were not.

We also showed that learning by teaching is not a plausible explanation for the better small class performance of seniors. Rather, we argued that beginning teachers may

be too concerned with performing basic teaching tasks and therefore fail to exploit the opportunities a small class offers. With this argument, we arrive at the following policy implications. As senior teachers do better than rookies in small classes only, the highest returns on investments into class size reductions can be expected by assigning only experienced teachers to small classes and inexperienced teachers to classes of regular size. Although class size reductions induce additional demand for teachers, the proposed reallocation by experience and class size ensures that the additional demand can be met with inexperienced teachers and is therefore feasible in the short run. The internal rate of return on reducing class size from 22 to 15 students (and assigning a senior to the small class) ranges from 5.1 to 8.0 percent. It is highest if class size is reduced only in the first year of school attendance.

A   Appendix

We aim at approximating the internal rate of return on investments in class size reductions for different durations in small classes. To do so, we first estimate a cumulative specification of the learning production function to assess the cumulative effects of having been in a particular class type for a certain number of years by adjusting Equation 5 to

$$
\begin{aligned}
Y_{icgs} \quad = \quad & \beta_0 + \beta_1 iSMALL_{icgs} + \beta_2 YSSENIOR_{icgs} + \beta_3 YSROOKIE_{icgs} + \\
& \beta_4 YRROOKIE_{icgs} + \beta_k S_{icgs} + \beta_j T_{cgs} + \alpha_f + \alpha_s + \gamma_g + \epsilon_{icgs}.
\end{aligned} \tag{7}
$$

The regressor $YSSENIOR$ ($YSROOKIE$) counts the number of years student i visited small classes taught by a senior (rookie) teacher while $YRROOKIE$ summarizes the number of years in regular classes taught by a rookie. The current year is always included in the count. Being initially assigned to a small class is captured by the $iSMALL$ dummy. The grade when the individual student entered the STAR project is controlled for by the three dummies α_f and the number of years in regular classes with senior teachers serves as the reference category.

The results are presented in Table 5. As expected from our previous results, we find

Tab. 5: The Cumulative Effect of Small Class Attendance by Teacher Experience

Subject	iSMALL		YSSENIOR		YSROOKIE		YRROOKIE	
Reading	.040*	(.021)	.041***	(.092)	-.006	(.024)	.005	(.016)
Math	.076**	(.032)	.035***	(.013)	.012	(.031)	.029	(.022)

Method OLS. Dependent variables are standardized to mean zero and variance one. For example, 0.040 means that achievement is 0.040 standard deviations higher. Standard errors in parentheses are robust to two-way clusters at the teacher level (i.e., class level) and at the student level (over time) applying the method of Cameron et al. (2011). ***,**,* denote significance at the 1, 5, or 10 percent level, respectively. The estimations with the reading (math) SAT score use 21,441 (21,746) observations. The differences in the number of observations is due to missing test score information.

insignificant coefficients on $YSROOKIE$ and $YRROOKIE$. The initial assignment to a small class as well as having attended small classes with senior teachers have significant beneficial effects. Having been assigned to small classes with senior teachers for four years cumulates in an advantage over having been for the same time in the reference category of $\beta_1 + 4 \cdot \beta_2 = 0.20$ standard deviations in reading (0.22 for math). Correspondingly, initial assignment to a small class with senior teacher raises achievement by $\beta_1 + \beta_2 = 0.081$ standard deviations in reading and 0.111 in math. With the exception of initial assignment to a small class with rookie teacher (0.045 for reading, 0.105 for math), further years with rookie teachers do generally not accumulate into an advantage over the reference group.[24]

As the results of this paper suggest that investments in class size reductions will not translate into higher future earnings of students if inexperienced teachers are assigned to small classes, we compare the benefits of being in a small class with a senior teacher for a certain number of years (as compared to being in a regular class during that time) to the costs of reducing class size for the same time span. We start with the 4-year period and, in contrast to Krueger (2003), calculate the additional costs per student for 4 years. In 2007, the US average per pupil expenditures for instruction amounted to 6,373 dollars[25]

---

[24] Ding and Lehrer (2010:42) also report positive effects of initial assignment to small classes. The lack of evidence for dynamic effects in their analysis might, at least to some extent, be explained by the fact that they do not distinguish between seniors and rookies.

[25] See U.S. Department of Education (2009), Table 183. We take the per pupil expenditures for instruction and instructional staff and neglect investments in equipment or facilities. This is justified if the increased number of lectures can be given in the same rooms but following a different schedule or if the investment in equipments and buildings are seen as fix costs that are borne only once and are therefore not taken into account when evaluating permanent class size reductions in the early grades. Clearly, this

and, therefore, the additional cost per pupil of increasing the number of classes by 7/15 = 47 percent is 2,995 dollar per year. The present value of costs is

$$\sum_{t=1}^{4} \frac{C_t}{(1+r)^t} \tag{8}$$

with $C = 2,995$ and r as the discount rate. As in Krueger (2003), we use the wage information from the Current Population Survey to approximate the age-earnings profile that is necessary to compute the present value of benefits, i.e. the value of future earnings advantages due to class size reductions discounted back to kindergarten. The present value of benefits is:

$$\sum_{t=18-4}^{65-4} \frac{E_t\,(\beta \cdot \delta)}{(1+r)^t} \tag{9}$$

where E_t is the yearly wage at time t, approximated as the average wage for a certain age according to the 2007 Current Population Survey, β is the percentage wage increase associated with a one standard deviation higher test score, and δ is the class size effect on student achievement, also measured in test score standard deviations. Our results in Table 5 suggest that the effect of being four years in a small class with a senior teacher amounts to .20 test score standard deviations in reading and .22 in math, respectively. To obtain δ, we average the reading and math outcomes presented in Table 5, arriving, e.g., at $\delta = .21$ in the four year case.

Test scores are proxies for cognitive and non-cognitive skills that are expected to translate into future earnings. Chetty et al. (2011) estimate that a one standard deviation higher test score in kindergarten increases earnings at age 25-27 by 18 log points, i.e. 19.7 percent. A number of other studies (c.f. Currie and Thomas 2001, Mulligan 1999, Murnane et al. 2000) arrive at results in the range of 10 to 15 log points while Hanushek and Zhang (2009) estimate a 20 log points earnings increase from a one standard deviation increase in a literacy test score. The mentioned studies differ both in the grades at which the test scores are measured and in the age at which income is observed. As Chetty et al. (2011)

---

approach is the more plausible the less grades are subject to class size reductions.

use STAR data and directly observe income of former STAR students, we will use their result in the following and proxy β in (9) with .20. Their estimate can be viewed as a rather conservative estimate of lifetime earnings differences because they observe earnings at age 25-27 and the percentage earnings gap between high skilled and less skilled workers is expected to increase at later ages.

Based on these assumptions, Table 4 gives the comparison of the present value of costs and benefits for several discount rates as well as the internal rate of return for different durations of small class attendance and two conservative scenarios of future real wage growth. Future real wage growth shifts future age-earnings profiles and therefore amplifies the present value of benefits.

References

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 2007, *25* (1), 95–135.

Ballou, Dale, William Sanders, and Paul Wright, "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational and Behavioral Statistics*, 2004, *29* (1), 37–65.

Blatchford, Peter, Viv Moriarty, Suzanne Edmonds, and Clare Martin, "Relationships between Class Size and Teaching: A Multimethod Analysis of English Infant Schools," *American Educational Research Journal*, 2002, *39* (1), 101–132.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust Inference with Multi-way Clustering," *Journal of Business and Economic Statistics, forthcoming*, 2011.

Carrell, Scott E. and Mark L. Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–228.

__ , **Bruce I. Sacerdote, and James E. West**, "From Natural Variation to Optimal Policy? The Lucas Critique meets Peer Effects," Working Paper 16865, National Bureau of Economic Research 2011.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How does your Kindergarten Classroom affect your Earnings? Evidence from Project STAR," *The Quarterly Journal of Economics, forthcoming*, 2011.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Teacher-Student Matching and the Assessment of Teacher Effectiveness," *The Journal of Human Resources*, 2006, *41* (4), 778–820.

Currie, Janet and Duncan Thomas, "Early Test Scores, School Quality and SES: Longrun Effects of Wage and Employment Outcomes," *Worker Wellbeing in a Changing Labor Market*, 2001, *20*, 103–132.

Ding, Weili and Steven F. Lehrer, "Estimating Treatment Effects from Contaminated Multiperiod Education Experiments: The Dynamic Impacts of Class Size Reductions," *The Review of Economics and Statistics*, 2010, *92* (1), 31–42.

__ **and** __ , "Experimental estimates of the impacts of class size on test scores: robustness and heterogeneity," *Education Economics*, 2011, *19* (3), 229–252.

Finn, Jeremy D. and Charles M. Achilles, "Answers and Questions about Class Size: A Statewide Experiment," *American Educational Research Journal*, 1990, *27* (3), 557–577.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux, "Unconditional Quantile Regressions," *Econometrica*, 2009, *77* (3), 953–973.

Goldhaber, Dan D. and Dominic J. Brewer, "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity," *The Journal of Human Resources*, 1997, *32* (3), 505–523.

Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 1986, *24* (3), 1141–1177.

_ , "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, 1997, *19* (2), 141–164.

_ , "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis*, 1999, *21* (2), 143–163.

_ **and Lei Zhang**, "Quality-Consistent Estimates of International Schooling and Skill Gradients," *Journal of Human Capital*, 2009, *3* (2), 107–143.

Hoxby, Caroline M., "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *The Quarterly Journal of Economics*, 2000, *115* (4), 1239–1285.

Jacob, Brian A. and Lars Lefgren, "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, 2008, *26* (1), 101–136.

Jepsen, Christopher and Steven G. Rivkin, *What is the Tradeoff between smaller Classes and Teacher Quality?*, Cambridge, MA: National Bureau of Economic Research, Working Paper No. 9205, 2002.

Krueger, Alan B., "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 1999, *114* (2), 497–532.

_ , "Economic Considerations and Class Size," *Economic Journal*, 2003, *113* (485), 34–63.

Lazear, Edward P., "Educational Production," *The Quarterly Journal of Economics*, 2001, *116* (3), 777–803.

Mosteller, Frederick, "The Tennessee Study of Class Size in the Early School Grades," *The Future of Children*, 1995, *5* (2), 113–127.

Mulligan, Casey B., "Galton versus the Human Capital Approach to Inheritance," *The Journal of Political Economy*, 1999, *107* (S6), 184–224.

Murnane, Richard J. and Randall J. Olson, "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina," *The Journal*

of Human Resources, 1990, *25* (1), 106–124.

_ , **John B. Willett, Yves Duhaldeborde, and John H. Tyler**, "How important are the cognitive skills of teenagers in predicting subsequent earnings?," *Journal of Policy Analysis and Management*, 2000, *19* (4), 547–568.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos, "The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis*, 1999, *21* (2), 127–142.

_ , **Spyros Konstantopoulos, and Larry V. Hedges**, "How large are Teacher Effects?," *Educational Evaluation and Policy Analysis*, 2004, *26* (3), 237–257.

Pate-Bain, Helen, Charles. M. Achilles, Jayne Boyd-Zaharias, and Bernard McKenna, "Class Size does make a Difference," *The Phi Delpa Kappan*, 1992, *74* (3), 253–256.

Rice, Jennifer King, "The Impact of Class Size on Instructional Strategies and the Use of Time in High School," *Educational Evaluation and Policy Analysis*, 1999, *21* (2), 215–229.

_ , "Making the Evidence Matter," in Lawrence Mishel and Richard Rothstein, eds., *The Class Size Debate*, Economic Policy Institute, 2002.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 2005, *73* (2), 417–458.

Rockoff, Jonah E., "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *The American Economic Review*, 2004, *94* (2), 247–252.

Rothstein, Jesse, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, 2010, *125* (1), 175–214.

U.S. Department of Education, *Digest of Education Statistics 2009*, Washington DC: National Center for Education Statistics, 2009.

Veenman, Simon, "Perceived Problems of Beginning Teachers," *Review Of Educational Research*, 1984, *54* (2), 143–178.