IWQW

Institut für Wirtschaftspolitik und Quantitative
Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 07/2011

Teacher Experience and the Class Size Effect - Experimental Evidence

Steffen Mueller
University of Erlangen-Nuremberg

ISSN 1867-6707

Friedrich-Alexander-Universität
IWQW
Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Teacher Experience and the Class Size Effect - Experimental Evidence

Steffen Mueller *

University of Erlangen-Nuremberg

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 most pronounced for higher and average-performing students. We further show that senior teachers outperform rookies only in small classes. The results have straightforward policy implications. Interestingly, the class size effect is most likely due to a higher quality of instruction in small classes and not due to less disruptions.

Keywords: class size reduction, teacher experience, student achievement

JEL Classification: I2, H4, J4

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

*I thank A. Colin Cameron and Regina T. Riphahn for valuable comments. Of course, any remaining errors are my own.

with random assignment of students and teachers to classes of different size. In particular, Krueger (1999) drew 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, p. 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 suggests that student achievement can likely be influenced by class size reduction but little by observed teacher characteristics. The fact that unobserved teacher characteristics are important is of limited help for optimal resource allocation because the policy maker is then required to rank teachers according to some criteria that cannot be observed and has to be estimated first. In the absence of random matching of students, teachers, and schools, such rankings are inherently prone to criticism.[1]

As pointed out by Rice (2002), 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 increase the positive 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 moderate class size effects and are both observable and controllable by the policy maker. Teacher experience is such a possibly important moderating factor.

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

[1] Typically, teacher quality is estimated using value added models. Rothstein (2010) gives a good treatment of this method.

STAR experiment. One main empirical result is that assigning an inexperienced teacher to a small class almost fully offsets the beneficial effect of class size reductions. On the other hand, the rookie is as effective as a senior teacher in regular size classes. Obviously, both findings combined generate the policy advice to assign senior teachers to small classes and inexperienced teachers to regular size classes in order to maximize student achievement with a given number of senior and rookie teachers. 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 assess whether teacher experience and class size reductions have different effects on higher or lower performing students, we extend our analysis and allow for differing interaction effects of class size and teacher experience along the unconditional student achievement distribution using RIF regressions as proposed by Firpo et al. (2009).

2 Literature

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, p. 148) states that "there is no strong or consistent relationship between school resources and student performance." A theoretical model of Lazear (2001) helps understand 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 would overcome problems of sorting and omitted variables and allow causal inference. The only large scale data for the United States that is collected under random assignment is the Tennessee's Student/Teacher Achievement Ratio. 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).[2]

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 literature and analyzes the joint effect of teacher experience

[2] Some of those studies look at modifiers of teacher experience effects, e.g., with respect to subject taught (Nye et al. 2004; Clotfelter et al. 2006).

and class size reductions on student achievement.[3] 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 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, 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.[4]

In Equation 1, learning is influenced via two channels: the disruption channel p^n and the quality-of-instruction channel $q(n, E)$. Per definition, the disruption channel induces negative class size effects as long as $0 < p < 1$. Supporting this specification, Rice (1999) and Blatchford et al. (2002) find that in small classes more time is devoted

[3] In a recent discussion paper and without presenting the results, McKee et al. (2010) state that teacher experience does not interact with class size. They use the same data we do and, similar to our definition, define teacher experience as an indicator variable that is equal to one if the teacher has less than 3 years of experience. However, the authors use only test scores for one grade, namely kindergarten. With this definition, their estimates for teacher experience in small classes are based on about 20 inexperienced teachers. Finding no significant effect does therefore not necessarily mean that class size effects do not differ with teacher experience in general.

[4] We borrow from Lazear (2001) the distinction between the time available for instruction (resulting from p^n) and the quality of this time. In this framework, p does not depend on teacher experience. We will relax this assumption below.

to instruction. To structure the discussion below, we 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 student. 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 primarily improved by on-the-job experience constituting an advantage for senior teachers. 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[5]

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

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.

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

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

[5] 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.

The negative class-size effect will become more (less) pronounced with higher teaching experience if the cross derivative $\frac{\partial^2 L}{\partial n \partial E}$ is negative (positive). Contrarily, the class size effect will become less negative or even positive with higher teacher experience if the cross derivative is positive. With $\frac{\partial q(n,E)}{\partial E} \geq 0$ 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 effect 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 (a minimum of) experience is necessary for the effective use of more instructional time per student.

If $\frac{\partial^2 q(n,E)}{\partial n \partial E} < 0$, then $\frac{\partial^2 L}{\partial n \partial E}$ is also below zero and, therefore, the class size effect is amplified with higher teacher experience via an increase in the quality of instruction. It would then be optimal for the policy maker to assign senior teachers to small classes and rookies to regular ones because this would be the resource allocation that maximizes overall student achievement with a given number of experienced and unexperienced teachers.

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

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

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 $\{n \cdot \ln p(E) + 1\} > 0$. For values of

$p \geq 0.97$, $\{n \cdot \ln p(E) + 1\}$ is positive up to a class size of 32. As a result, Equation 4 as a whole may become positive even if Equation 3 was negative. Hence, the class-size effect doesn't 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 counterweights any potential advantage of seniors with respect to the class-size effect on teaching quality (i.e. $\frac{\partial^2 q(n,E)}{\partial n \partial E}$).[6]

The existence of the disruption channel and 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. We then compare the outcome difference between seniors and rookies by class size. If p does not depend on experience, any changes in the outcome difference can be attributed to the quality-of-instruction channel. If we can't rule out the existence of a disruption channel in the first step, changes in the outcome difference between teacher types can't unambiguously be 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.[7]

[6] For values of $p \leq 0.95$, $\{n \cdot \ln p(E) + 1\}$ is negative for reasonable class sizes and the sign of the additive term in Equation 4 as well as the sign of Equation 4 as a whole is unclear.

[7] The latter two class types will be pooled in our analysis as most studies found no sizeable differences in student performance and because also the regular classes without full-time aide were supported by part-time aides at the time.

Achievement in reading and math was measured via the Stanford Achievement Tests (SAT) that provides test scores that can be compared across grades.[8]

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, particularly, whether random assignment was actually done properly. While Hanushek (1999) and Hoxby (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 are unrelated to class type assignment and teacher characteristics, attrition will not affect our results. Nye et al. (1999) p. 137 found 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

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

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 subset of 45 students that move 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 some 21,500 observations.[9]

Third, because of student mobility, some overlap occurred in the actual class size 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 first compare the achievement difference between rookie and senior teachers in regular classes. If no difference shows up there, seniors have no advantage with respect to disruptive behavior. This is because we plausibly assumed (and presented extensive empirical evidence) that the quality of instruction of seniors teachers is at least as high as the rookies' quality. In the absence of the disruption channel, 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 can't rule out the disruption channel, we have no chance to disentangle

[9] Clearly, excluding a selective group does not solve all the problems. We rather argue that the potentially problematic group of 17 students is too small to drive our results.

the two channels.

5.1 Achievement Levels

We begin by estimating the following regression:

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

where i denotes individual student, c class, g grade, and s school. Y_{icgs} is the SAT test score standardized to mean zero and variance one. The vector S contains 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 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

[10] Summary statistics are presented in Table 1.

[11] Our own experimentations show that average student achievement does not further increase when teacher experience exceeds three years. However, there is a slight decline in teacher effectiveness for very experienced teachers with more than 25 years of service. We have also 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

	Grade				
Variables	K	1	2	3	All
	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.

small class rookies in third grade would then be 13. In the following analysis, students are pooled over all grades with the grades controlled by a set of dummies γ_g in Equation 5. 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 classes (i.e. teachers) and within students over time. 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 case, the two-way cluster-robust standard errors are very close to those obtained by simply clustering at the class level.

OLS estimates the effects of small classes and inexperienced teachers on the mean of the student achievement distribution. However, it is also interesting to know whether those 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.

Conditional quantile regression (CQR) as proposed by Koenker and Basset (1978) provides information about the effect of a covariate (class size) on the within group dispersion. A "group" consists of students who have the same covariates excluding class size. However, CQR does not consider the effects of a covariate on the between group dispersion. Unconditional quantile regression (UCQR) as introduced by Firpo et al. (2009) tells us whether the overall dispersion changes 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 on individuals located at a certain quantile of the *unconditional* achievement distribution. Hence, UCQR allows assessing whether class size reductions increase or decrease the achievement differences between good and bad students while CQR does not. Because we focus on distinguishing the class size effects on good and bad students rather than on the (weighted) within group dispersion, we apply the technique of Firpo et al. (2009).

The results from the basic specification in Equation 5 are presented in Table 2. The reference category are students in regular classes that have a senior teacher. Hence, β_2 measures the difference in student achievement between senior and rookie teachers in regular classes and β_3 identifies the difference within small classes. As β_2 is generally insignificant and close to zero, we find no support for the existence of the disruption channel. The finding is consistent with the basic theoretical model that sets $\frac{\partial p(E)^n)}{\partial E} = 0$ and we conclude that the probability of disruptive behavior is not affected by teacher experience in our data. Additionally, this is a very interesting finding because it uncovers an important heterogeneity in the widespread view that teacher experience increases

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***	(8.68)	-.005	(0.21)	-.125***	(3.21)
0.1	.084***	(6.83)	-.010	(0.43)	-.092**	(2.25)
0.2	.114***	(7.65)	-.040*	(1.65)	-.111***	(2.62)
0.3	.143***	(7.85)	-.081**	(2.27)	-.127*	(1.87)
0.4	.172***	(8.80)	-.024	(0.65)	-.178***	(2.87)
0.5	.144***	(8.09)	-.044	(1.45)	-.144***	(2.92)
0.6	.152***	(10.59)	.016	(0.72)	-.116***	(2.90)
0.7	.163***	(10.83)	.044*	(1.88)	-.146***	(3.90)
0.8	.175***	(9.93)	.047*	(1.88)	-.142***	(3.69)
0.9	.156***	(7.86)	.020	(0.88)	-.158***	(3.80)
	21,443 Observations					
	Standardized SAT Score on Math					
OLS	.162***	(7.82)	.036	(1.09)	-.143***	(2.75)
0.1	.115***	(5.70)	-.016	(0.48)	-.143**	(2.30)
0.2	.165***	(8.73)	-.000	(0.01)	-.188***	(3.59)
0.3	.191***	(9.97)	-.007	(0.20)	-.107**	(2.01)
0.4	.181***	(10.76)	-.018	(0.64)	-.102**	(2.03)
0.5	.162***	(9.27)	.007	(0.27)	-.077*	(1.72)
0.6	.182***	(10.37)	.055**	(2.11)	-.142***	(3.21)
0.7	.169***	(9.45)	.056***	(2.25)	-.126***	(3.05)
0.8	.167***	(9.34)	.101***	(3.68)	-.146***	(3.23)
0.9	.164***	(6.68)	.107***	(3.29)	-.176***	(3.23)
	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). For quantile regression (OLS), absolute t-values (z-values) 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.

student achievement (see Krueger 1999 or Clotfelter et al. 2006).

The first column in Table 2 presents the small class effect for experienced teachers. The OLS estimate for reading (math) indicates that students in such a class perform on average 0.14 (0.16) test score standard deviations better than those in the regular classes with senior teachers. 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.[12] As we haven't found 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.[13] Given $\frac{\partial q(n,E)}{\partial n} \leq 0$, this finding is only consistent with the view that neither the quality of instruction nor the available time for instruction (via the disruption channel) increases for rookie teachers as class size decreases.[14]

The main results are that only seniors generate class size effects and that the class size effect comes through an increase in teaching quality per unit of instructional time. Hence, our results are not in line with theories that explain class size effects 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.

The unconditional quantile regression results allow a deeper look into what exactly happens to good and bad students. Students at the lowest deciles of the achievement distribution gain less from small classes with senior teachers than better performing students. Hence, the introduction of small classes with senior teachers increases overall achievement inequality due to a larger inequality at the bottom 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

[12] The picture does not change if we use actual class size instead of class type as regressor.

[13] 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.

[14] We assumed throughout the paper that $p(E)$ (not $p(E)^n$) is independent of class size. Although we think that this standard assumption is a plausible one, relaxing it allows teacher experience differences in the change of p when class size is reduced (i.e. the cross derivative of p with respect to n and E). This would make a direct test of the quality-of-instruction channel impossible. However, the fact that p does not depend on class size for inexperienced teachers supports the standard assumption of p being generally independent of class size.

(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 good students, this means that rookies outperform seniors in regular classes while the opposite is true in small classes for both subjects. Both results again support our prior findings: the seniors' 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 ($\frac{\partial p(E)^n)}{\partial E} = 0$).[15]

5.2 A Value Added Specification

Comparison of achievement levels may be inappropriate because differences in levels cloud all initial differences different students bring into a certain grade level. Such differences will bias 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 result, e.g., by including previous year's test score as an additional regressor.[16] 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

[15] Note that the conditional quantile regression gives qualitatively similar results in our study. However, conditional quantile regression is estimating a more steady increase in the small class effect over the distribution. As the corresponding effect of the interaction term steadily decreases, the rookie small class effect is essentially zero at any point of the conditional distribution. The conditional quantile regression also hides the beneficial effects of rookies for high achievers in regular classes.

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

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 the reference category. 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) p. 176 argues that "... the necessary exclusion restriction is that teacher assignments are orthogonal to all other determinants of the so-called gain score." 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.[17] In our empirical implementation of the VAM, we therefore run the following regression separately by class type

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

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 in kindergarten, independent of the teacher type they are assigned to. As they start from the same

[17] 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.

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***	(4.64)	-.012	(0.43)
0.1	-.122***	(2.95)	-.005	(0.21)
0.2	-.117***	(2.93)	-.006	(0.27)
0.3	-.117***	(2.86)	-.003	(0.12)
0.4	-.317***	(4.56)	-.007	(0.20)
0.5	-.282***	(4.17)	-.031	(0.53)
0.6	-.221***	(4.66)	-.059	(1.35)
0.7	-.220***	(5.15)	.001	(0.03)
0.8	-.177***	(4.23)	.019	(0.70)
0.9	-.140***	(3.41)	-.014	(0.46)
Observations	4,648		9,337	
	Standardized SAT Score on Math			
OLS	-.131**	(2.43)	-.034	(0.82)
0.1	-.231***	(3.24)	.043	(1.12)
0.2	-.225***	(3.57)	-.010	(0.29)
0.3	-.122**	(2.02)	.030	(0.81)
0.4	-.164***	(3.12)	.009	(0.23)
0.5	-.203***	(3.53)	-.036	(0.91)
0.6	-.198***	(3.80)	.007	(0.19)
0.7	-.100**	(2.29)	.032	(0.84)
0.8	-.049	(0.95)	.063*	(1.90)
0.9	-.010	(0.16)	.116**	(2.47)
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). For RIF regression (OLS), absolute t-values (z-values) 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.

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.[18]

The results for the VAM are presented in Table 3 and corroborate our main findings from the estimation in levels as shown in Table 2. 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 the policy maker, the most important results are

1. only senior teachers generate a beneficial class size effect
2. this effect is lower for the lowest performing students
3. senior and rookie teachers perform similar in regular size classes.

It is clear from these findings that only senior teachers should be assigned to classes of reduced size. If class size is reduced, then additional classes have to be installed and, hence, there will be demand for additional teachers. If there are not enough teachers, new teachers have to be trained. As stated in the third result, 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]

[18] This is true for both OLS and RIF regression.

[19] This may not be true if the additional demand for teachers decreases average teacher quality. Hanushek (1999) emphasizes that a decrease in average teacher quality due to the additional demand for teachers may offset beneficial class size effects and Rivkin et al. (2005) find teacher quality to

The second finding indicates that overall student achievement is maximized if only good students are assigned to small classes (with senior teachers). 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.[20] However, as the effect for bad students 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 is able to do so by assigning bad performing students to small classes with senior teachers and good students to classes of regular size.

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.[21] 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 students' 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 additionally assess the number of 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 real wages and the number of grades in which class size reduc-

be at least as important as class size reductions. Furthermore, Jepsen and Rivkin (2002) argue in their analysis of the 1996 California class size reduction program that the massive influx of new teachers decreased average teacher quality because inexperienced and less skilled teachers were hired. Our results clearly confirm the view of Jepsen and Rivkin (2002) with respect to the assignment of inexperienced teachers to small classes and, hence, no contradiction arises here. Nevertheless, one might ask whether the additional influx of inexperienced teachers into regular classes, as proposed in our paper, may lead to a decrease in the average quality of rookies. 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 is necessarily the case when attracting additional young people to become teachers. Even if there is a limited pool of capable teacher candidates, we argue that a potential decrease in the quality of rookies is a second order effect that plays a minor role for overall student achievement.

[20] See Table 2.

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

Tab. 4: Present value of costs and benefits as well as the internal rate of return for reducing class size from 22 to 15 for several discount rates, wage growth rates, and different numbers grades 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 %
		1st grade			1st and 2nd grade	
0.02	2,937	11,045	16,164	5,816	15,571	22,769
0.04	2,880	5,510	7,814	5,649	7,761	11,007
0.06	2,826	2,943	4,057	5,492	4,145	5,714
0.08	2,666	1,670	2,245	5,181	2,352	3,162
Internal Rate of Return		0.061	0.073		0.050	0.061
		1st to 3rd grade			1st to 4th grade	
0.02	8,638	20,088	29,374	11,405	24,605	35,978
0.04	8,312	10,013	14,200	10,873	12,264	17,393
0.06	8,006	5,348	7,372	10,379	6,550	9,029
0.08	7,553	3,034	4,079	9,921	3,717	4,997
Internal Rate of Return		0.046	0.057		0.044	0.055

Assumptions: a 1 standard deviation increase of math or reading test scores translates into 8 percent higher income; test score advantages for different durations in small classes with senior teachers are taken from Table 5; "Cost" are additional costs per pupil a class size reduction causes in terms of the salaries of teachers and other instructing staff.

tions are performed. The table compares the present value of costs with the present value of future real earnings advantages in US Dollars 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 4.4 (first four grades) and 6.1 percent (only first grade), depending on the number of grades to be considered. Given a moderate increase in real wages of 1 percent per year, the internal rate of return rises by at least 1.1 percentage points 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 with the number of grades included. Remembering that the highest class size effects have been found in the initial grades (see Table 5 in the appendix 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 the class size reductions from the initial grade to later grades. From the second and the fifth column of Table 4 we see that the additional costs per pupil of extending the investment to the second year 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 below 4 percent. From a simple cost-benefit point of view, it may therefore 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. 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 disruption. 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 the achievement of higher and average-performing students in small classes.

Hanushek (1999) p. 153 and Hoxby (2000) p. 1241 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 can't be expected in non-experimental settings. Even if this is true, our results show that senior teachers were able to fulfill expectations while rookies were not. Hence, the important finding of this paper is that, given the right incentives, senior teachers are able to use small classes efficiently while rookies are not.

Our results have straightforward 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 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 4.4 to 7.3 percent, depending on the discount rate and future real wage growth. It is highest for the first year of school attendance.

8 Appendix

We want to approximate the internal rate of return of 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

$$\begin{aligned} Y_{icgs} &= \beta_0 + \beta_1 iSMALL_{icgs} + \beta_2 YSSENIOR_{icgs} + \beta_3 YSROOKIE_{icgs} + \\ &\quad \beta_4 YRROOKIE_{icgs} + \beta_4 S_{icgs} + \beta_5 T_{cgs} + \alpha_f + \alpha_s + \gamma_g + \epsilon_{icgs}. \end{aligned} \tag{7}$$

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 thus the number of years in regular classes with senior teachers serves as the reference category.

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

Subject	iSMALL		YSSENIOR		YSROOKIE		YSROOKIE	
Reading	.040*	(1.86)	.041***	(4.40)	-.006	(0.24)	.005	(0.29)
Math	.076**	(2.39)	.035***	(2.66)	.012	(0.38)	.029	(1.35)

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 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). Z-values in parentheses. ***,**,* 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 for some subjects.

The results are presented in Table 5. As expected from our previous results, we find insignificant coefficients on $YSROOKIE$ and $YRROOKIE$. The initial assignment to a small class as well as having attended small classes with senior teachers has significant beneficial effects. Having been assigned to small classes with senior teachers for four years cumulates in an advantage over having been 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.

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 duration. 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[22] and therefore the additional cost per pupil of increasing the number of classes by $7/15 = 47$ percent would be 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:

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

$$\sum_{t=18-4}^{65-4} \frac{E_t \cdot \beta \left(\delta_M + \delta_R\right)}{\left(1+r\right)^t} \tag{9}$$

where E_t is the yearly wage at time t, β is the percentage wage increase associated with a one standard deviation higher test score, and δ_M and δ_R being the class size effects in math and reading, respectively, 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. As in Krueger (2003), we refer to the empirical studies of Currie and Thomas (1999), Murnane et al. (1995), and Neal and Johnson (1996) and assume that β is .08. It is further assumed that people start working at age 18 and retire at age 65 and that there is a stable age-earnings profile.

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

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.

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, Socioeconomic Status and Future Outcomes," Working Paper 6943, National Bureau of Economic Research 1999.

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.

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.

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

Koenker, Roger and Gilbert Basset, "Regression Quantiles," *Econometrica*, 1978, *46* (1), 33–50.

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.

McKee, Graham, Steven Rivkin, and Katherine Sims, *Disruption, Skill and Heterogeneous Benefits of Smaller Classes*, http://economics.uconn.edu/Seminar Series/rivkin10.pdf, mimeo, 2010.

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

Murnane, Richard J., John B. Willett, and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determinatio," *The Review of Economics and Statistics*, 1995, *77* (2), 251–66.

Neal, Derek A. and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 1996, *104* (5), 869–95.

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.

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*, Washington, D.C.: Economic Policy Institute, 2002, pp. 89–94.

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.

Diskussionspapiere 2011
Discussion Papers 2011

01/2011 **Klein, Ingo, Fischer, Matthias, Pleier, Thomas:** Weighted Power Mean Copulas: Theory and Application

02/2011 **Kiss, David:** The Impact of Peer Ability and Heterogeneity on Student Achievement: Evidence from a Natural Experiment

03/2011 **Zibrowius, Michael:** Convergence or divergence? Immigrant wage assimilation patterns in Germany

04/2011 **Klein, Ingo, Christa, Florian:** Families of Copulas closed under the Construction of Generalized Linear Means

05/2011 **Schnitzlein, Daniel**: How important is the family? Evidence from sibling correlations in permanent earnings in the US, Germany and Denmark

06/2011 **Schnitzlein, Daniel**: How important is cultural background for the level of intergenerational mobility?

Diskussionspapiere 2010
Discussion Papers 2010

01/2010 **Mosthaf, Alexander, Schnabel, Claus and Stephani, Jens:** Low-wage careers: Are there dead-end firms and dead-end jobs?

02/2010 **Schlüter, Stephan and Matt Davison:** Pricing an European Gas Storage Facility using a Continuous-Time Spot Price Model with GARCH Diffusion

03/2010 **Fischer, Matthias, Gao, Yang and Herrmann, Klaus:** Volatility Models with Innovations from New Maximum Entropy Densities at Work

04/2010 **Schlüter, Stephan, Deuschle, Carola:** Using Wavelets for Time Series Forecasting – Does it Pay Off?

05/2010 **Feicht, Robert, Stummer, Wolfgang:** Complete closed-form solution to a stochastic growth model and corresponding speed of economic recovery.

06/2010 **Hirsch, Boris, Schnabel, Claus:** Women Move Differently: Job Separations and Gender.

07/2010 **Gartner, Hermann, Schank, Thorsten, Schnabel, Claus:** Wage cyclicality under different regimes of industrial relations.

08/2010 **Tinkl, Fabian**: A note on Hadamard differentiability and differentiability in quadratic mean.

Diskussionspapiere 2009
Discussion Papers 2009

01/2009 **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009 **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009 **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009 **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009 **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009 **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009 **Herrmann, Klaus:** Non-Extensitivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009 **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009 **Dewenter**, Ralf, **Haucap**, Justus and **Wenzel**, Tobias: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009 **Stuehmeier**, Torben and **Wenzel**, Tobias: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009 **Klein**, Ingo, **Köck**, Christian and **Tinkl**, Fabian: Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

12/2009 **Schlüter**, Stephan: Constructing a Quasilinear Moving Average Using the Scaling Function

13/2009 **Blien,** Uwe, **Dauth,** Wolfgang**, Schank,** Thorsten and **Schnabel,** Claus: The institutional context of an "empirical law": The wage curve under different regimes of collective bargaining

14/2009 **Mosthaf**, Alexander, **Schank**, Thorsten and **Schnabel**, Claus: Low-wage employment versus unemployment: Which one provides better prospects for women?

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl**: Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl**: Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm**: Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner**: The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias**: On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus**: Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation