

# ZEW

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research



Rechts-, Wirtschafts- und  
Verwaltungswissenschaftliche  
Sektion  
Fachbereich  
Wirtschaftswissenschaften

Diskussionspapiere der DFG-  
Forschergruppe (Nr.: 3468269275):

Heterogene Arbeit: Positive und Normative  
Aspekte der Qualifikationsstruktur der Arbeit

---

Christian Wiermann

**Class Size, Instruction Time and Central Exit  
Examinations: Disentangling the Relative  
Contributions to Scholastic Achievement**

April 2005

---

Diskussionspapier Nr. 05/04

<http://www.wiwi.uni-konstanz.de/forschergruppewiwi/>

Nr. 05/04, April 2005

---

## Class Size, Instruction Time and Central Exit Examinations: Disentangling the Relative Contributions to Scholastic Achievement

Christian Wiermann

Universität Konstanz  
Fachbereich Wirtschaftswissenschaften  
Fach D138  
78457 Konstanz  
Germany

mail: [christian.wiermann@uni-konstanz.de](mailto:christian.wiermann@uni-konstanz.de)

phone: +49-7531-88-2137

fax +49-7531-88-3130

---

### Zusammenfassung:

We analyze the effects of instruction time and class size on student achievement in Germany. Using econometric evaluation techniques we are able to deal with unobserved heterogeneity in the student body. Specifically we apply first-difference methods and matching across subjects to control for overall student ability and specific skills. We find that an increase in class size reduces the performance of the students, while additional lessons improve the test score achievement. Both effects differ across federal states and individuals: Additional instruction time in states with central exit examinations enhances performance, while there is no effect in states without central exit examinations and more able students profit more from additional lessons as compared to their less able fellow students. We furthermore show that that reductions in class size accompanied by a decrease in instruction time can be a cost neutral instrument to raise student performance in states with central exit examinations.

JEL Klassifikation : I21

Schlüsselwörter : Educational Economics, PISA, Instruction Time, Class Size

Download/Reference : <http://www.wiwi.uni-konstanz.de/forschergruppewiwi/>

# Class Size, Instruction Time and Central Exit Examinations: Disentangling the Relative Contributions to Scholastic Achievement

Christian Wiermann  
University of Konstanz \*†

April 20, 2005

## Abstract

We analyze the effects of instruction time and class size on student achievement in Germany. Using econometric evaluation techniques we are able to deal with unobserved heterogeneity in the student body. Specifically we apply first-difference methods and matching across subjects to control for overall student ability and specific skills. We find that an increase in class size reduces the performance of the students, while additional lessons improve the test score achievement. Both effects differ across federal states and individuals: Additional instruction time in states with central exit examinations enhances performance, while there is no effect in states without central exit examinations and more able students profit more from additional lessons as compared to their less able fellow students. We furthermore show that that reductions in class size accompanied by a decrease in instruction time can be a cost neutral instrument to raise student performance in states with central exit examinations.

**Keywords:** Educational Economics, PISA, Instruction Time, Class Size

**JEL Classification:** I21

---

\***Correspondence:** Box D-138, University of Konstanz, 78457 Konstanz, Germany, Christian.Wiermann@uni-konstanz.de

†I would like to thank Heinrich W. Ursprung, Winfried Pohlmeier, Anton Flossman and Michael Rauber for helpful comments and discussions. Financial support of the Deutsche Forschungsgemeinschaft (DFG) through the research project “Heterogeneous Labor: Positive and Normative Aspects of the Skill Structure of Labor” is gratefully acknowledged. Comments are very welcome. I’m fully responsible for all the remaining errors.

## 1 Introduction

Class size and the weekly teaching time per subject are two of the most important institutional characteristics of a school system, that politicians can change. As decreasing class size and increasing teaching time are not only generally believed to improve scholastic achievement, but also very costly, a clear understanding of their effects is essential.

Taking a look at the relevant literature reveals that the understanding of the consequences of changes in both instruments is far from being clear cut. This study circumvents many of the econometric difficulties usually encountered when evaluating the effects of changes in class size and instruction time. We do this by applying econometric evaluation techniques which can handle the estimation biases typically induced by unobservable student characteristics, in particular abilities. Specifically we use first-difference and skill-specific matching methods across subjects to neutralize the unobserved heterogeneity in our dataset, namely the German PISA and PISA-E 2000 study. We estimate specific effects for states with and without CEEs and across skill groups. In a final step we use our results to show that reductions in class size accompanied by a decrease in instruction time can be a cost neutral instrument to raise student performance in states with CEEs.

The estimation of the consequences of additional instruction time very complex. First, instruction time can differ along three dimensions. Lessons (hours) per day, schooling days per week and schooling weeks per year create three policy dimensions, which politicians can alter when deciding upon the regular schooling time of a pupil. We focus in our study on lessons per week, as this policy dimension is the measure reported in the PISA study. Second, one needs to distinguish between the time the teacher spends with the students as a body and the time spend with an individual student.<sup>1</sup> Also there is time lost in the course of administration, enforcement of discipline and absenteeism. Finally, the time of a lesson might differ across subjects and institutional settings.

The literature, which explicitly focusses on the allocation of teaching time is not very large. The reasons for this are manifold. First, there do not exist many datasets with a large variation in the teaching time per week. Second, one very often has to deal with endogenous selection. Academically weaker students are assigned extra lessons, which gives rise to biased estimates in standard OLS approaches. Third, due to displacement effects and inefficiencies in the production of education it is very hard to find significant effects of additional lessons at all. Wössmann (2003), for example, cannot find any overall significant effect of instruction time on student performance for a cross-section of East Asian countries in an education production function estimation approach. On the other hand, Fuchs and Wössmann (2004) find positive and significant, but small effects of additional instruction time per year on achievement in mathematics and sciences, but not in reading in a different cross-section of 32 countries. Coates (2003) reports in a study on elementary school students in Illinois that instruction times does matter in determining test scores. He furthermore finds that complementarities and

---

<sup>1</sup>See for example Rossmiller (1983) on this issue, who finds for the US that during a typical school year of 1080 hours, students are actually only taught for 364 hours.

complementarities between subjects exist. English instruction, for example, raises the mathematics score and vice versa, while on the other hand more instruction in social sciences harms for a given amount of teaching the mathematics performance.

There is of course an extensive literature on the effects of class size. Hanushek (1999a) surveys this literature and reports that studies with positive significant effects and studies with negative significant effects of class size reductions have the same weight in the literature, leading him to the conclusion that the actual effect would be no different from zero.<sup>2</sup> In a U.S. House of Representatives hearing he specified, that “existing evidence indicates that achievement for the *typical student* will be unaffected by instituting the types of class size reductions that have been recently proposed or undertaken. The most noticeable feature of policies to reduce overall class sizes will be a dramatic increase in the costs of schooling, an increase unaccompanied by achievement gains.”<sup>3</sup> Hoxby (2003) cannot find any positive effects of class size reductions either. Her study is very relevant, because it benefits from having a purely exogenous, i.e. performance independent variation in class size.

However, the literature clearly shows that there are small effects for - as Hanushek puts it - “non-typical students”, e.g. Afro-American, academically weak or very young students. Hanushek (1999b) himself reports on the basis of the Tennessee Project STAR - which was the largest ever conducted experiment on class size effects during the mid 1980s - that support in terms of better performance is found in the data for very large class size reductions at young age. Using the same data set, Krueger (1999) comes to a different conclusion. He finds positive effects of smaller classes across all ages, which are somewhat larger for minorities than for the other students. Krueger (2003) furthermore argues that even large reductions in class size are beneficial from a cost-benefit-analysis point of view. Boozer and Rouse (1995) also find support for class size reductions. In a study on interschool variation in class size they discover that when taking teacher-pupil ratios as instruments for class size one can identify positive effects on the achievement. Using the same IV-estimation technique to account for the non-random assignment, Akerhielm (1995) finds that smaller classes matter, though not across the whole student population. A similar result is reported for Sweden by Lindahl (2001). An international comparison of class size effects can be found in Wössmann and West (2002). They find that the effects vary across countries and account this variety of results to differences in teacher quality. Their results confirm that class size sometimes does matter, while under different institutional arrangements it might not.

The literature on the effects of central exit examinations is very small. Using international TIMMS microdata, Wössmann (2002) estimates the achievement gain to be equivalent to about one school year. Juerges et al. (2003) employ German TIMMS data and come up with one third of a school year equivalents. In a follow up paper Juerges et al. (2004) analyze the influence of CEEs on teacher quality. They show that the existence of this external standard significantly raises teaching quality.

---

<sup>2</sup>Altogether Hanushek looked at 377 education production function regressions in 90 scientific publications.

<sup>3</sup>See Hanushek (1998)

Our plan of the paper is as follows. In the next section, we develop our model and introduce the econometric specification to identify class size and instruction time effects. In section 3 we describe the dataset and in section 4 we present our estimation results. The final section concludes.

## 2 The Model

Consider the following system of education production functions for five subjects mathematics [M], physics [P], reading [R], chemistry [C] and biology [B].<sup>4</sup>

$$TS_i^M = \alpha_1^M + \alpha_2 A_i + \alpha_3^M A_i^M + \beta_1^M X_i + \beta_2^M F_i + \beta_3^M T_i + \gamma_1^M C_i^M + \gamma_2^M L_i^M + \varepsilon_i^M \quad (1)$$

$$TS_i^P = \alpha_1^P + \alpha_2 A_i + \alpha_3^P A_i^P + \beta_1^P X_i + \beta_2^P F_i + \beta_3^P T_i + \gamma_1^P C_i^P + \gamma_2^P L_i^P + \varepsilon_i^P \quad (2)$$

$$TS_i^R = \alpha_1^R + \alpha_2 A_i + \alpha_3^R A_i^R + \beta_1^R X_i + \beta_2^R F_i + \beta_3^R T_i + \gamma_1^R C_i^R + \gamma_2^R L_i^R + \varepsilon_i^R \quad (3)$$

$$TS_i^B = \alpha_1^B + \alpha_2 A_i + \alpha_3^B A_i^B + \beta_1^B X_i + \beta_2^B F_i + \beta_3^B T_i + \gamma_1^B C_i^B + \gamma_2^B L_i^B + \varepsilon_i^B \quad (4)$$

$$TS_i^C = \alpha_1^C + \alpha_2 A_i + \alpha_3^C A_i^C + \beta_1^C X_i + \beta_2^C F_i + \beta_3^C T_i + \gamma_1^C C_i^C + \gamma_2^C L_i^C + \varepsilon_i^C \quad (5)$$

The left-hand-side variable  $TS_i^X$  are the PISA test scores of student  $i$  in subject  $X$ . The coefficients and explanatory variables are the following:  $\alpha_1^X$  are the subject specific constants,  $\alpha_2$  picks up ability  $A_i$  in its most *general* -i.e. not subject specific - form,  $\alpha_3^X$  is the estimate for subject *specific* ability,  $\beta_1^X$  denotes a vector of individual background factors,  $F$  is the federal state and  $T$  denotes the school track in the hierarchical German school system. Our variables of interest are class size  $C$  and instruction time per week  $L$ . The last expression is a subject specific error-term.

The problem in estimating these equations separately is that some serious omitted variable bias might occur, since we cannot control for individual ability. We circumvent this problem by applying a difference estimation technique. Subtracting each single equation (2) to (5) from equation (1) yields the following system:

---

<sup>4</sup>Reading, as it is in the PISA study, denotes the subject, in which the domestic language is taught.

$$\begin{aligned}
TS_i^M - TS_i^P &= (\alpha_1^M - \alpha_1^P) + \alpha_3^M A_i^M - \alpha_3^P A_i^P + (\beta_1^M - \beta_1^P)X_i + (\beta_2^M - \beta_2^P)F_i \\
&+ (\beta_3^M - \beta_3^P)T_i\gamma_1^M C_i^M - \gamma_1^P C_i^P + \gamma_2^M L_i^M - \gamma_2^P L_i^P + (\varepsilon_i^M - \varepsilon_i^P)
\end{aligned} \tag{6}$$

$$\begin{aligned}
TS_i^M - TS_i^R &= (\alpha_1^M - \alpha_1^R) + \alpha_3^M A_i^M - \alpha_3^R A_i^R + (\beta_1^M - \beta_1^R)X_i + (\beta_2^M - \beta_2^R)F_i \\
&+ (\beta_3^M - \beta_3^R)T_i\gamma_1^M C_i^M - \gamma_1^R C_i^R + \gamma_2^M L_i^M - \gamma_2^R L_i^R + (\varepsilon_i^M - \varepsilon_i^R)
\end{aligned} \tag{7}$$

$$\begin{aligned}
TS_i^M - TS_i^C &= (\alpha_1^M - \alpha_1^B) + \alpha_3^M A_i^M - \alpha_3^P A_i^C + (\beta_1^M - \beta_1^B)X_i + (\beta_2^M - \beta_2^B)F_i \\
&+ (\beta_3^M - \beta_3^B)T_i\gamma_1^M C_i^M - \gamma_1^B C_i^B + \gamma_2^M L_i^M - \gamma_2^B L_i^B + (\varepsilon_i^M - \varepsilon_i^B)
\end{aligned} \tag{8}$$

$$\begin{aligned}
TS_i^M - TS_i^B &= (\alpha_1^M - \alpha_1^C) + \alpha_3^M A_i^M - \alpha_3^P A_i^B + (\beta_1^M - \beta_1^C)X_i + (\beta_2^M - \beta_2^C)F_i \\
&+ (\beta_3^M - \beta_3^C)T_i\gamma_1^M C_i^M - \gamma_1^C C_i^C + \gamma_2^M L_i^M - \gamma_2^C L_i^C + (\varepsilon_i^M - \varepsilon_i^C)
\end{aligned} \tag{9}$$

By using this specification we cancel out the bias-inducing general ability. Nevertheless, we still face the problem that we cannot observe the differences in the subject specific ability. Simply neglecting the respective influence leads to a severe bias in our estimates. However we can solve this problem by using the fact that we can always match one specific ability. It is generally acknowledged for example that competence in mathematics is highly correlated with competence in physics, we thus speculate that  $\alpha_3^M A_i^M - \alpha_3^P A_i^P = 0$ .

Moreover, we set  $\gamma_1^M = \gamma_1^P$  and  $\gamma_2^M = \gamma_2^P$ , as mathematics and physics are very similar subjects with respect to common teaching techniques. However we need to allow the other  $\gamma$ -coefficients to differ from each other, i.e.  $\gamma_j^M, \gamma_j^R, \gamma_j^C$  and  $\gamma_j^B$  with  $j = 1, 2$  need not necessarily be the same. Second, we can get rid of nearly all of the explanatory individual background variables by setting all the  $\beta_1$ 's equal. We do this because there is usually not a large heterogeneity in the influence of the background variables on the test scores in the different subjects. The impact of a single room for the student at home - not shared with any siblings - should usually exert a similar positive influence on his achievement in all subjects. The only variable we allow to differ is the gender dummy (Female=1, Male=0) for the student, because there is a well established distinct relationship between the gender and the achievement in different subjects. Furthermore we restrict the coefficients  $\beta_2$  to be equal across all equations. Institutional characteristics of federal states apply to all subjects taught, thus the estimated  $\beta_2$  coefficients should be equal. Finally, we keep the school track dummy in our estimation, because usually different school tracks place different weights and importance on the subjects.<sup>5</sup>

---

<sup>5</sup>We actually checked all our assumptions concerning the size of the coefficients across all subjects by using hypothesis tests after standard OLS education production function regressions using a large set of background factors. Furthermore we estimated the whole system relaxing various restrictions, which did not change our results significantly.

All these assumptions reduce our system to

$$\begin{aligned}\Delta TS^{MP} &= \mu^{MP} + (\beta_1^M - \beta_1^P)Female_i + (\beta_3^M - \beta_3^P)T_i \\ &+ \gamma_1^M(C_i^M - C_i^P) + \gamma_2^M(L_i^M - L_i^P) + \nu^{MP}\end{aligned}\quad (10)$$

$$\begin{aligned}\Delta TS^{MR} &= \mu^{MR} + \alpha_3^M A_i^M - \alpha_3^R A_i^R + (\beta_1^M - \beta_1^R)Female_i + (\beta_3^M - \beta_3^R)T_i \\ &+ \gamma_1^M C_i^M - \gamma_1^R C_i^R + \gamma_2^M L_i^M - \gamma_2^R L_i^R + \nu^{MR}\end{aligned}\quad (11)$$

$$\begin{aligned}\Delta TS^{MC} &= \mu^{MC} + \alpha_3^M A_i^M - \alpha_3^C A_i^C + (\beta_1^M - \beta_1^C)Female_i + (\beta_3^M - \beta_3^C)T_i \\ &+ \gamma_1^M C_i^M - \gamma_1^C C_i^C + \gamma_2^M L_i^M - \gamma_2^C L_i^C + \nu^{MC}\end{aligned}\quad (12)$$

$$\begin{aligned}\Delta TS^{MB} &= \mu^{MB} + \alpha_3^M A_i^M - \alpha_3^B A_i^B + (\beta_1^M - \beta_1^B)Female_i + (\beta_3^M - \beta_3^B)T_i \\ &+ \gamma_1^M C_i^M - \gamma_1^B C_i^B + \gamma_2^M L_i^M - \gamma_2^B L_i^B + \nu^{MB}\end{aligned}\quad (13)$$

where  $\nu^{MP} = (\varepsilon_i^M - \varepsilon_i^P)$  and  $\mu^{MP} = (\alpha_1^M - \alpha_1^P)$  in equation (10) [and the respective expressions for the other subject differences in equations (11)-(13)].

The coefficients of particular interest in this simultaneous equation system are  $\gamma_1^M$  and  $\gamma_2^M$ . The  $\gamma_1^M$  measures the influence of an additional student in the Math class on the test score performance in mathematics. The  $\gamma_2^M$  likewise measures the effect of an additional hour of instruction in mathematics on the mathematics achievement.

We estimate equations (10) to (13) with a seemingly unrelated regression (SURE), in which we impose the restriction that the  $\gamma_1^M$  and the  $\gamma_2^M$  need to be the same across all equations.<sup>6</sup> This restriction enables us make use of the additional variation inherent in the other other equations. Note that we can only interpret the coefficients of equation (10), because we cannot control for the specific ability differentials in the other equations, which thus leads to biased coefficients in these equations due to the missing subject specific abilities  $A_i^X$ 's. Also note that the SURE method allows for a nonzero covariance between the error terms across the different equations for one individual and is therefore the correct estimation method.

So far we have described a way to estimate class size and instruction time effects in mathematics. Subtracting equation (1), (3), (4), (5) from equation (2), setting  $\gamma^P = \gamma^M$  in equation (2) and restricting all the  $\gamma^P$ s to be the same across all five equations, we can estimate the effects on the physics performance. Likewise one can combine the other equations, always imposing one restriction on specific abilities and therefore on the  $\gamma$ s. We decided that the combinations mathematics / physics, physics / mathematics, chemistry / biology and biology / chemistry are plausible. When looking at

<sup>6</sup>The SURE regression is going back to Zellner (1962).



the reading score things become more complicated with respect to the proper matching. Nevertheless we estimate with the same SURE method reading / biology, though the matching of specific abilities might not be completely convincing. We are now set to apply our model framework to the German PISA dataset, which we describe in the next section, before then going on to present the estimation results.

### 3 PISA - The Data

#### 3.1 The PISA 2000 and PISA-E 2000 Student Achievement Test

The PISA 2000 and PISA-E(xtention) 2000 student achievement tests were conducted in May and June 2000. While PISA 2000 is part of an international study, in which 5.037 Germans students at the age of fifteen were assessed in reading, mathematical and scientific literacy, the PISA-E 2000 study was a German national extension. In the PISA-E study 45.899 students were assessed. This group of students includes the original 5.037 students from the international PISA study. The PISA-E study consists of two overlapping samples. The first sample is made up of 33.766 students in 9th grade. The second sample includes 33.809 students at the age of fifteen.

PISA-E is not only an extension with respect to the number of pupils, but also with respect to the tests and accompanying questionnaires. The national study allows for a more detailed analysis of the scientific literacy, as disaggregated tests in physics, biology and chemistry were conducted. Furthermore it is possible to look at different schooling tracks.

The dataset we use in this paper is a combination of the dataset made up of the 15 year old students and the students from 9th-grade. As we want to compare instruction time and class size effects across subjects, we altered the dataset in the following way. First, we dropped all students, who were not assessed in all five subjects. Then we deleted all the students, who attend vocational schools. Finally we dropped a small number of observations due to missing variables. all this leaves us with a dataset consisting of 6535 observations.

#### 3.2 Descriptive Statistics

Table 1 shows the descriptive statistics of our dataset. We briefly discuss these, before proceeding to our estimation results in the next section.

The average test scores are slightly higher than those officially reported. The reason might be that we dropped the rather weakly performing pupils from vocational schools; moreover there might exist possible bias due to dropping observations with missing data. The German students scored highest in biology with around 517 average test score points. The subject with the weakest performance is physics with approximately 500 average test score points.

VARIABLE	DESCRIPTION	MEAN	STD.ERR.
Test Scores			
Reading	Diff. adj. score in reading literacy	504.718	85.795
Mathematics	Diff. adj. score in mathematics	505.656	83.364
Biology	Diff. adj. score in biology	516.918	90.708
Chemistry	Diff. adj. score in chemistry	504.013	92.492
Physics	Diff. adj. score in physics	499.553	76.516
School Data			
L-Read	Lessons in reading per week	3.7018	0.8892
L-Math	Lessons in mathematics per week	3.7883	0.8475
L-Bio	Lessons in biology per week	1.5645	0.8187
L-Che	Lessons in chemistry per week	1.9066	0.7563
L-Phy	Lessons in physics per week	1.7221	0.8247
S-Read	Number of students in reading class	23.6505	4.7040
S-Math	Number of students in mathematics class	23.3970	4.8802
S-Bio	Number of students in biology class	23.7110	4.1919
S-Che	Number of students in chemistry class	23.0528	4.9275
S-Phy	Number of students in physics class	22.9448	4.9023
School Stratification Type			
Gymn	1 if Grammar School	0.3426	-
Real	1 if Intermediate Schools	0.2506	-
Haupt	1 if Secondary General Schools	0.1216	-
Gesamt	1 if Comprehensive School	0.1034	-
Middle	1 if Middle School	0.1818	-
Individual Data			
Female	1 if female, 0 if male	0.5163	-

Table 1: Descriptive Statistics: 6535 observations

The average number of lessons in reading and mathematics per week is around 3.7, while the average number of lessons in the sciences is between 1.5 in biology and 1.9 in chemistry. This reflects the legislation of the federal states pretty well.<sup>7</sup> Class size in all subjects is more or less 23 students. However note that there exists sufficient variation in the data on both measures.

Around 34% of the students in the sample go to grammar schools [Gymnasium], 25% attend intermediate schools [Realschule], 18% go to the middle schools [Combination of Haupt- and Realschule only found in some federal states ], 12% are enrolled in secondary general schools [Hauptschule] and around 10% attend comprehensive schools [Gesamtschule]. Altogether there are 52% girls in the sample.

<sup>7</sup>Standard syllabuses in Germany specify between 3 and 4 lessons in mathematics and reading per week and 0-3 lessons in sciences, which are often taught alternately.

Dependent Variable: $\Delta TS$	CLASS SIZE: $\gamma_1$	LESSONS PER WEEK: $\gamma_2$
Math - Physics: $\Delta TS^{MP}$ $R^2 = 0.03$	-0.183* (0.10)	1.267** (0.54)
Physics - Math: $\Delta TS^{PM}$ $R^2 = 0.03$	-0.182* (0.10)	1.055** (0.52)
Chemistry - Biology: $\Delta TS^{CB}$ $R^2 = 0.04$	-0.274*** (0.09)	1.287** (0.57)
Biology - Chemistry: $\Delta TS^{BC}$ $R^2 = 0.04$	-0.447*** (0.11)	0.645 (0.49)
Reading - Biology: $\Delta TS^{RB}$ $R^2 = 0.10$	0.724*** (0.12)	0.812* (0.49)

6535 observations. Standard errors are reported in brackets. Coefficients are significant on the 10%(\*), 5%(\*\*) and 1%(\*\*\*) significance level. Controlling variables in the above regressions are all variables reported in the descriptive statistics.

Table 2: Full Sample Specification

## 4 Results

Table 2 shows our estimation results of the model described in section 2.

An increase in class size lowers the test score in mathematics and the three sciences. The effect is significant in all of the four regressions though on different levels. The quantitative effects range from -0.182 average test score points in physics up to -0.447 average test score points in biology, i.e. a reduction by 5 students increases performance on the average by 1 point in physics and by 2.2 points in biology. Surprisingly the effect of one additional student in the reading class is positive and significant. One reason for this counter intuitive results might be that it is due to the imperfect match of specific abilities between reading and biology.

Turning to the instruction time per week we find that additional lessons have the expected positive effect on individual performance in all, but the positive impact is not significant in biology. The size of the coefficient varies from 0.812 average test score points in reading to 1.267 average test score points in chemistry.

### 4.1 Instruction Time, Class Size and Central Exit Examinations

Taking a closer look at our results reveals some interesting insights. We divided our sample into two subsamples. One subsample contains all the observations from students, who live in federal states that have CEEs. The other sample consists of the students, who do not have to take a CEE at the end of their mandatory school career. CEE are part of the institutional design of the school system and are a mechanism to create incentives on the part of the pupils and teachers. Thus, CEE should be reflected in a more effective use of instruction time. Seven of the sixteen German federal states (Meck-

Class Size	CEE STATES: $\gamma_1$	NON-CEE STATES: $\gamma_1$
Math - Physics: $\Delta TS^{MP}$	-0.413*** (0.16)	-0.068 (0.14)
Physics - Math: $\Delta TS^{PM}$	-0.421*** (0.16)	-0.007 (0.14)
Chemistry - Biology: $\Delta TS^{CB}$	-0.496*** (0.15)	-0.093 (0.13)
Biology - Chemistry: $\Delta TS^{BC}$	-0.962*** (0.23)	-0.212 (0.14)
Reading - Biology: $\Delta TS^{RB}$	0.746*** (0.17)	0.364*** (0.18)

Instruction Time	CEE STATES: $\gamma_2$	NON-CEE STATES: $\gamma_2$
Math - Physics: $\Delta TS^{MP}$	2.189*** (0.74)	-0.364 (0.80)
Physics - Math: $\Delta TS^{PM}$	1.818** (0.72)	-0.024 (0.76)
Chemistry - Biology: $\Delta TS^{CB}$	1.816** (0.83)	-0.060 (0.57)
Biology - Chemistry: $\Delta TS^{BC}$	1.251* (0.83)	-0.033 (0.80)
Reading - Biology: $\Delta TS^{RB}$	1.337** (0.12)	0.741 (0.71)

3630 observations in the CEE-sample, 2905 observations in the Non-CEE sample. Standard errors are reported in brackets. Coefficients are significant on the 10%(\*), 5%(\*\*) and 1%(\*\*\*) significance level. Controlling variables in the above regressions are all variables reported in the descriptive statistics.

Table 3: Effects of Additional Instruction Time and Class Size by CEE / Non-CEE states

lenburg Western Pommerania, Saxony, Saxony-Anhalt, Thuringia, Bavaria, Baden-Württemberg and the Saarland) do have CEEs.<sup>8</sup> Our CEE subsample consists of 3630 observations, the other one accordingly consists of 2905 data points. The results summarized in Table 3 confirm our intuition: Instruction time is used more efficiently in CEE states.

The upper part of Table 3 reports the effects of a reduction in class size by one student for the CEE and the Non-CEE states. In all of the five regressions, the coefficient in the CEE states is highly significant. Except the reading coefficient, all estimates have the expected negative sign, ranging from -0.413 points in mathematics up to -0.962 points in biology. As in our aggregated regression,

<sup>8</sup>Note that not all student in Saarland, Saxony-Anhalt and in Mecklenburg are effected by CEE's. In the latter ones they do not have to be taken in the Hauptschule, in Saarland they only have to be taken in the Gymnasium. Nevertheless we do divide the sample with respect to the federals states having CEEs or not, but not with respect to the individual pupil having a CEE or not.

the reading coefficient has an unexpected sign. Turning to the Non-CEE states, we see that class size has no significant effect on schooling achievement apart from reading, where the coefficient is again positive. In none of the five subjects additional instruction time improves the performance of students in the Non-CEE states. All the coefficients are statistically not significantly different from zero. The opposite is true for the CEE states. Additional lessons seem to increase the test scores of the students in a range between 1.251 average points in biology and 2.189 average points in mathematics. All the coefficients are significant.

Our results point towards the conclusion that CEEs are a mechanism to raise the test score results of students. Not only the amount of time spend in class in absolute value, but also the reduction of the teaching time per student (via the class size effect) exert some positive influence on the scholastic achievement in CEE states. The channels through which this mechanism works are unclear. Two interpretation are possible. First, students and teachers might have a higher motivation, which is reflected in a more efficient use of time spend in class. Second, one can interpret our result in a neoclassical fashion. The marginal product of an additional lessons is obviously higher in CEE states than in Non-CEE states. But as the average number of lessons taught in CEE and Non-CEE states is in our dataset more or less the same, this must imply a better education production technology in CEE states.<sup>9</sup>

## 4.2 Instruction Time, Class Size and Heterogeneous Skills

A similar but less extreme picture emerges when we investigate returns across skill levels. We address the question whether additional lessons and class size reductions have different effects on high and low achievement students. Again we spilt our dataset into two subsamples in each of the five regressions. The dataset is divided according to whether the student is a better performer than the median in the respective subject or not.

The results are displayed in Table 4. Again, the upper part of the table reports the effects of a reduction in class size when the test score of the respective student is above (left) or below (right) the median. Class size reductions exert a positive and statistically significant effect on the more able students in all three sciences, the effect is statistically not significantly different from zero in mathematics and reading. The effect is highest in chemistry with -0.493 test score points and lowest in physics with -0.124 test score points. Looking at the performance quantile below the median we see that the significant biology coefficient does have the expected sign (-0.386 points), whereas the reading coefficient is significant and positive. A increase in the reading class size by one student vis-a-vis the biology class leads to an increase in the test scores by 0.745 points. Turning to additional instructional time we can see the following: If the test score of the student is above the median, then additional lessons have a positive effect on the performance of the student in mathematics, physics

---

<sup>9</sup>Note that we abstain from the conclusion that the Non-CEE states are in their production optimum and thus the marginal effects are therefore zero. This explanation does not reflect the fact that the Non-CEE states are performing worse compared to the CEE states.

Class Size	TEST SCORE > MEDIAN: $\gamma_1$	TEST SCORE < MEDIAN: $\gamma_1$
Math - Physics: $\Delta TS^{MP}$	-0.124 (0.15)	-0.178 (0.14)
Physics - Math: $\Delta TS^{PM}$	-0.271* (0.15)	-0.180 (0.14)
Chemistry - Biology: $\Delta TS^{CB}$	-0.493*** (0.14)	-0.092 (0.13)
Biology - Chemistry: $\Delta TS^{BC}$	-0.484*** (0.19)	-0.386** (0.78)
Reading - Biology: $\Delta TS^{RB}$	0.154 (0.18)	0.745*** (0.16)

Instruction Time	TEST SCORE > MEDIAN: $\gamma_2$	TEST SCORE < MEDIAN: $\gamma_2$
Math - Physics: $\Delta TS^{MP}$	1.468** (0.72)	0.654 (0.74)
Physics - Math: $\Delta TS^{PM}$	1.376* (0.71)	0.726 (0.72)
Chemistry - Biology: $\Delta TS^{CB}$	1.719** (0.75)	0.968 (0.81)
Biology - Chemistry: $\Delta TS^{BC}$	-1.111 (0.75)	0.798 (0.78)
Reading - Biology: $\Delta TS^{RB}$	0.716 (0.67)	0.683 (0.68)

3267 observations below the median, 3268 observations above the median. Standard errors are reported in brackets. Coefficients are significant on the 10%(\*), 5%(\*\*) and 1%(\*\*\*) significance level. Controlling variables in the above regressions are all variables reported in the descriptive statistics.

Table 4: Effects of Additional Instruction Time and Class Size conditional on Test Scores

and chemistry. The effect is statistically not significantly different from zero in the remaining two subjects biology and reading. When looking at students, whose performance is below the median in the respective subject, we see that additional instruction time has no effect at all. None of the coefficients is significant. This is consistent with the repeatedly expressed fear that weak students are often lost in class and thus not motivated to profit from additional lessons.

### 4.3 The Trade-Off between Class Size and Instruction Time

If one assumes that the size of our estimates in the preceding sections are linear within sizeable intervals of class size and instruction time around the status quo, one can easily perform a cost-benefit analysis on the optimal number of students in class and lessons per week, while at the same time keeping the incurred costs constant. We calculate the effects of identical percentage changes in class size and instruction time, and sum up over both effects, since a say 10% reduction in class size and

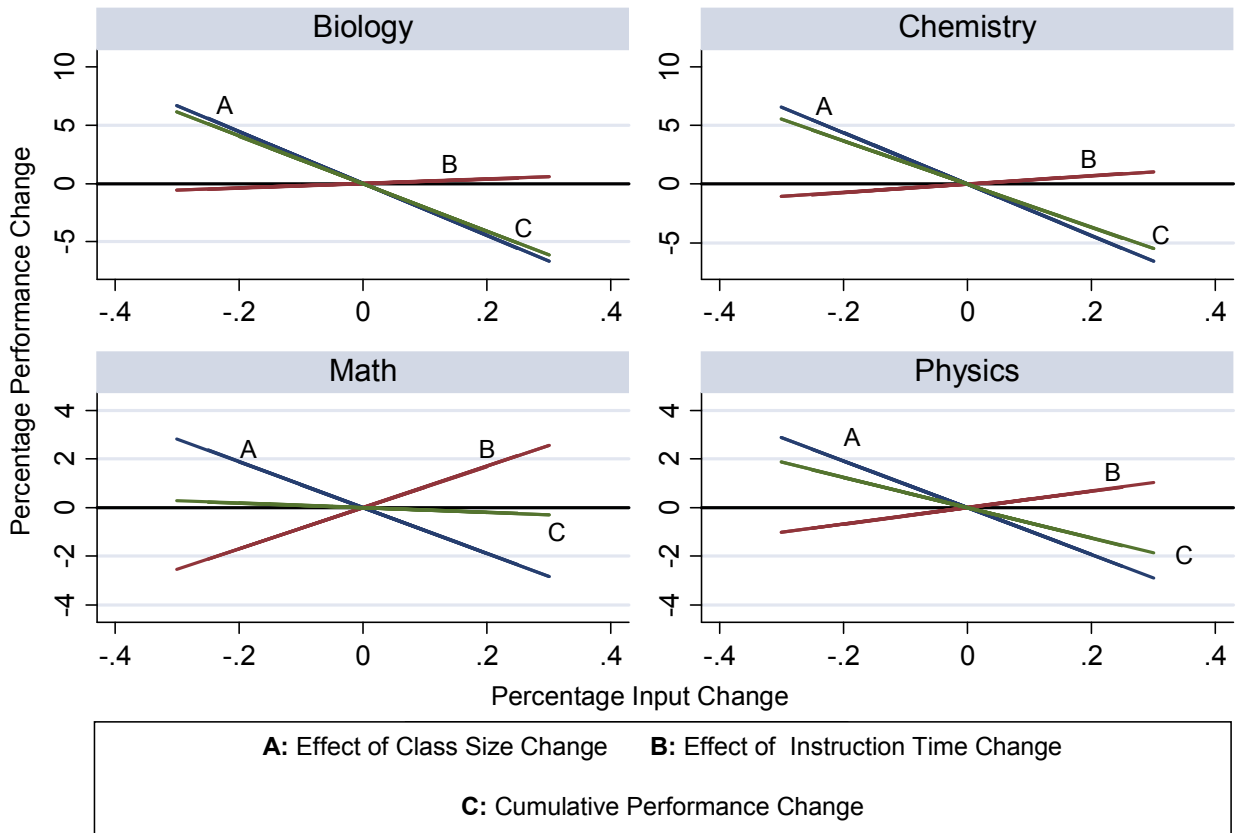


Figure 1: Trade-Off Analysis

a 10% decrease in the number of lessons taught does not alter the required amount of teachers. We performed this analysis for a change up to 30%.<sup>10</sup>

We perform our analysis only for the CEE states and for chemistry, biology, physics and mathematics, as we have for all of these subjects significant and plausible estimates. In each subject we start of with the average number of student in class and the average instruction time. We then alter both input variables under the constraint that the overall schooling costs are not allowed to vary.<sup>11</sup> Under our linearity assumption we can calculate the overall performance effect of the change in the composition of the two input factors, using our estimates from Table 3. Our results are displayed in Figure 1.

On the abscissas we plot for each subjects the percentage changes in the input variables class size and instruction time, ranging from -30% to 30%. The ordinates measure the induced performance change for the CEE states taken from Table 3. The A-lines depict the effects of changes in class size,

<sup>10</sup>A large change would definitely lead to different effects on the performance, as the education production function is presumably not linear beyond our specified interval, i.e. a reduction in class size down from 10 to 5 student should have different consequences than a reduction from 20 to 15 students. The same applies to an extension of instruction time.

<sup>11</sup>We believe that this assumption is justifiable if we abstract from capacity effects due to lack of rooms etc.

the B-lines depict the effects of changes in instruction time. The cumulative performance changes are depicted by the C-lines, which represents the sum of the class size effect (A) and the instruction time effect (B).

For all four subjects the qualitative effects are similar. In biology and chemistry a 30% decrease in class size accompanied by a 30% decrease in instruction time pays off with a substantial performance gain of namely 6.1 and 5.5 test score points, respectively. This would imply that German students close up to 50% of the gap to the OECD mean. In mathematics and physics these effects are substantially smaller i.e. only 0.28 and 1.8 test score points, respectively. Thus one can conclude that under our linearity assumptions at least in biology and chemistry a cost neutral recombination of the educational production factors “class size” and “instruction time” may generate significant performance gains.

## 5 Conclusion

In this paper we have investigated the effects of additional instruction time and class size in an econometric evaluation model, where we used fist-differences across subjects in order to control for specific abilities. Using the German PISA 2000 and PISA-E dataset we first found out that class size has got a negative effect on the test score results of the students in mathematics, biology, physics and chemistry. We then analyzed the relationship between instruction time and the test score results. Our analysis consisted of three steps. First, we investigated whether one can find a positive effect of additional teaching at all. We found out that except in biology there is a performance enhancing effect of more instruction time.

Then we proceeded to analyze whether there is a difference in the results between German federal states with and without central exit examinations. Surprisingly our regression results are very clear cut and robust across all five subjects. In CEE states instruction time in all subjects seems to matter as the performance of the students rises. The opposite is true for the Non-CEE states. There we do not find any significant relationship at all. Our results point towards more motivated teachers and students and a more effective use of instruction time. The incentives set by CEE’s seem to function quite well.

In a next step we compared instruction time effects across skill groups. If a student is more able in a certain subject, he seems to grasp more out of an additional lesson, although this result is only valid for mathematics, physics and chemistry. On the other hand, the lower quantile of the ability distribution seems to get nothing out of additional lessons. All our estimates are insignificant.

In a final step we checked using our estimates from the preceding sections, whether a possible cost neutral change in both input factors would have a performance enhancing effect for the students. For the CEE-states we got the result that a decrease in the class size combined with a reduction of the



teaching time is a possible way raise test score results. This effect is larger in biology and chemistry compared to mathematics and physics.

Our results highlight two further research questions. First, it remains unclear, whether additional lessons in CEE states have a positive impact due to incentives or whether teaching methods, syllabuses or further class and school characteristics are different in these states, making a lesson more productive. Second, it is important to find out why weaker students are not even able to extract small gains from additional lessons. Both research topics require to take a closer look at what happens inside the class and to carefully investigate the institutional framework in which students are taught.

## References

- AKERHIELM, K. (1995): “Does Class Size Matter?,” *Economics of Education Review*, 14(2).
- BOOZER, M., AND C. ROUSE (1995): “Interschool Variation in Class Size: Patterns and Implications,” *NBER Working Paper No. 5144*.
- COATES, D. (2003): “Education Production Functions Using Instructional Time as Input,” *Education Economics*, 11(3).
- DEUTSCHES PISA KONSORTIUM (2000): *Ergänzungen zum Skalenhandbuch*, Berlin.
- FUCHS, T., AND L. WÖSSMANN (2004): “What accounts for international differences in student performance? A re-examination using PISA data,” *IZA Discussion Paper No. 1287*.
- HANUSHEK, E. A. (1998): “The Evidence on Class Size. Summary of Testimony,” Hearing of: Subcommittee on Early Childhood, Youth, and Families Committee on Education and the Workforce U.S. House of Representatives.
- (1999a): “The Evidence on Class Size,” in *Earning and Learning: How Schools Matter*, ed. by S. E. Mayer, and P. Peterson. Brookings Institution, 1999.
- (1999b): “Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from other Investigations of Class Size Effects,” *Educational Evaluation and Policy Analysis*, 21(2).
- HOBY, C. M. (2000): “The effects of class size on Student achievement: New evidence from population variation,” *Quarterly Journal of Economics*, 115(3).
- JUERGES, H., W. RICHTER, AND K. SCHNEIDER (2004): “Teacher Quality and Incentives. Theoretical and Empirical Standards on Teacher Quality,” *CESifo Working Paper No. 1296*.
- JUERGES, H., K. SCHNEIDER, AND F. BUECHEL (2003): “The Effect of Central Exit Examinations on Student Achievement: Quasi-experimental Evidence from TIMSS Germany,” *CESifo Working Paper No. 939*.

- KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114(2).
- (2003): "Economic Considerations and Class Size," *The Economic Journal*, 113(485).
- LINDAHL, M. (2001): "Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement," *IZA Discussion Paper 261*.
- MAX-PLANCK-INSTITUTE FOR HUMAN DEVELOPMENT (2000): *PISA 2000: Dokumentation der Erhebungsinstrumente*, Berlin.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2000a): *Manual for the PISA 2000 Database*, Paris.
- (2000b): *PISA 2000 Technical Report*, Paris.
- ROSSMILLER, R. (183): "Time on task: A look at what erodes time for instruction," *NASSP Bulletin*.
- WÖSSMANN, L. (2003): "Educational Production in East Asia: The Impact of Family Background and Schooling Policies on Student Performance," *IZA Discussion Paper No. 745*.
- WÖSSMANN, L., AND M. R. WEST (2002): "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS," *IZA Discussion Paper 485*.
- WÖSSMANN, L. (2002): "Central Exams Improve Educational Performance: International Evidence," *Kiel Discussion Papers 397*.
- ZELLNER, A. (1962): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias," *Journal of the American Statistical Society*, 57.