

THE EFFECT OF CENTRAL EXIT EXAMINATIONS ON STUDENT ACHIEVEMENT: QUASI-EXPERIMENTAL EVIDENCE FROM TIMSS GERMANY

HENDRIK JÜRGES
KERSTIN SCHNEIDER
FELIX BÜCHEL

CESIFO WORKING PAPER NO. 939
CATEGORY 1: PUBLIC FINANCE
MAY 2003

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the CESifo website:* www.CESifo.de

THE EFFECT OF CENTRAL EXIT EXAMINATIONS ON STUDENT ACHIEVEMENT: QUASI-EXPERIMENTAL EVIDENCE FROM TIMSS GERMANY

Abstract

This paper makes use of the regional variation in schooling legislation within the German secondary education system to estimate the causal effect of central exit examinations on student performance. We propose a difference-in-differences framework that exploits the quasi-experimental nature of the German TIMSS middle-school sample. The estimates show that students in federal states with central exit examinations clearly outperform students in other federal states, but that only part of the difference can be attributed to central exit examinations. Our results suggest that central examinations increase student achievement by about one third school year equivalent.

JEL Code: I28.

Keywords: education, central examinations, difference-in-differences, quasi-experiment.

Hendrik Jürges
MEA
University of Mannheim
L 13, 17
68131 Mannheim
Germany
juerges@mea.uni-mannheim.de

Kerstin Schneider
Department of Economics
University of Dortmund
44221 Dortmund
Germany
kerstin.schneider@wiso.uni-dortmund.de

Felix Büchel
Max Planck Institute for Human Development
Lentzeallee 94
14195 Berlin
Germany
Buechel@mpib-berlin.mpg.de

We are grateful to Rainer Watermann, Max Planck Institute for Human Development, for editing the regional information used in this study, and to Jürgen Baumert for granting special permission to use this data.

1. Introduction

The poor performance of German students in TIMSS and PISA gave rise to an intense political discussion about the need to reform the German school system. From the perspective of an economist, reforms can be targeted at the allocation of financial resources or at changing institutions. Increasing financial resources alone appears to be no successful approach, as the discussion of the class size effect on achievement has demonstrated (Hanushek, 1999, Hoxby, 2000a, Jürges and Schneider, 2003). Even if small positive effects of financial resources can be identified, they are very costly.

Institutions of the schooling system, however, seem to explain more of the student performance. Changing the characteristics of the school system, thereby creating the right incentives appears to be more promising than simply allocating more financial resources to the education sector.¹ Competition between schools is one topic that has received attention (Hoxby, 2000b). Another issue, and that is the topic of this paper, are central exit examinations (CEEs).

The economic literature almost unanimously shows that CEEs and hence centralised standards improve student performance and might even raise welfare (Costrell, 1997; Effinger and Polborn, 1999). It is argued that central exit examinations do better at setting the right incentives to students, teachers, and schools than decentralised examinations (e.g. Bishop, 1997, 1999). Students for example benefit because results from CEEs are more valuable as signals on the job market than results from non-central examinations, simply because results are comparable. Furthermore, if an external standard is to be met at the end of the school career, students have no incentives to establish a low achievement cartel in class, possibly with the tacit consent of the teachers. Student test results can be used to monitor teacher and teaching quality on a regular basis. Whether incentives to improve teaching quality, an

¹ See Jürges and Schneider (2003) for a discussion in the context of the international TIMSS database.

arguably important factor in the education production function, should come solely from reputation effects or even be in the form of better pay for better teachers is open to discussion (Hanushek et al, 1998). Finally, the reputation of entire schools can be based on the achievement of its students, with good schools attracting good students (provided that aggregate CEE results are publicly available).

The empirical studies of the effect of CEEs on the academic performance of students claims that in cross-country as well as in single-country studies, the existence of central exit examinations significantly improves student performance (Bishop, 1997, 1999). The cross country results are obtained from using results from the international TIMSS database. However, as Jürges and Schneider (2003) show, the positive effect of CEEs on achievement in cross-country analysis based on TIMSS is not robust. Besides the international evidence, Bishop (1997, 1999) also presents results from Canadian micro-data. In 1990-1991 Canada had, just like Germany, a mixed system. Some provinces administered central exit exams at the final year of high school, whereas other provinces did not. Bishop estimates the effect to be between three fifth of a US grade-level equivalent for science and four fifth of the US grade level in math.

However, it is important to separate simple correlation from causation. The possibility that countries or federal states with CEE place in general higher priority on education and achievement has to be accounted for. In that case, high average student achievement and CEEs only reflect the electorates preferences for good education. Earlier papers have tried to deal with this issue by asking whether CEE states also differ along other dimensions than achievement, e.g. student discipline and absenteeism (Bishop, 1997). However, the data did not allow a convincing identification strategy and leaves unresolved issues.

We use the German federalised education system as a unique source of exogenous variation to identify the causal effect of CEEs on student achievement. Earlier studies

(Baumert and Watermann, 2000) have found differences at the upper secondary level for students in non-specialised mathematics courses only. However, as will be argued below, the German system of upper secondary education is not well suited to analyse the CEE effect. Instead we focus on the effect of exit exams at the end of lower secondary education.

In the remainder of the paper we give a brief account of German secondary education. Section 3 describes the data, and Section 4 discusses issues of identification and estimation. The results are presented in Section 5 and Section 6 concludes.

2. German secondary education in perspective

In this section we will give a concise description of the German school system, in which we try to emphasise those aspects that are most relevant for the understanding of central exit examinations in the German context.² Figure 1 presents a stylised overview of primary and secondary education.

Primary school – attended by all children in Germany – covers grades 1 to 4, or in some federal states grades 1 to 6. There is no formal exit examination at the end of primary school. The transition from primary school to one of the three secondary school types is generally guided by the students' abilities and performance. Admission to secondary school usually occurs on the basis of recommendations from the primary school including an evaluation of the student's suitability for secondary schools. If the primary school considers a student suitable for a certain type of school, the student will be admitted without any special admission procedure. In cases of conflict between the primary school's recommendations and the parents' wishes, the final decision about the future course of education is either with the parents, the receiving school or the school supervisory authority, depending on the laws of the federal state.

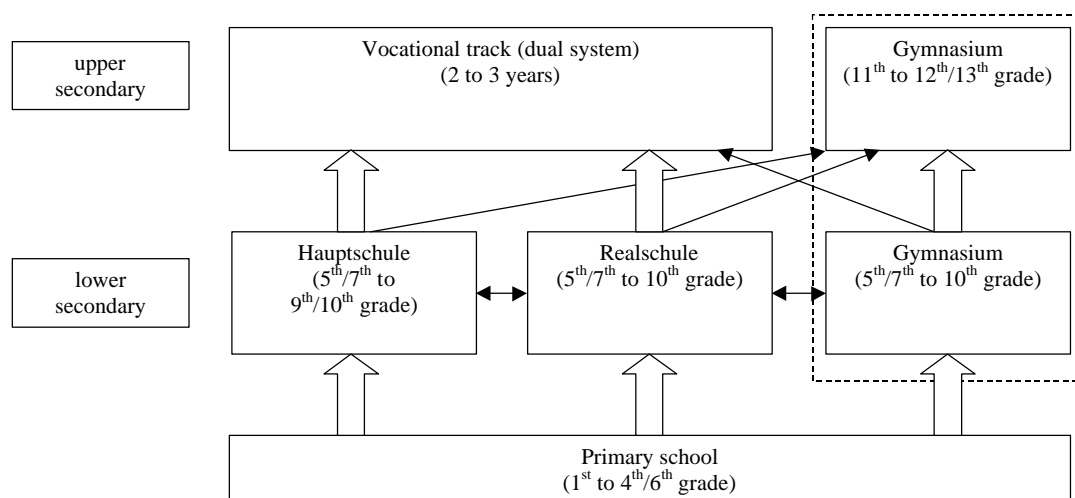


Figure 1: A stylised model of the German school system

Hauptschule, *Realschule* and *Gymnasium* are secondary schools offering one single course of education, leading to a school specific leaving certificate. The *Hauptschule* provides its students with basic general education. It normally covers grades 5 to 9 (or 10 in some states). The *Realschule* provides a more extensive general education, usually covering grades 5 to 10. The *Gymnasium* provides an intensified general education. The course of education in the standard *Gymnasium* comprises both the lower and upper secondary level and usually covers grades 5 to 13 (or 12 in some former GDR states). Depending on academic performance, students can switch between school types.³

On completion of the lower secondary level, students in *Hauptschule* or *Realschule* receive a leaving certificate, provided that they have successfully completed grade 9 or 10. Only in some federal states students are required to pass central exit examinations (Table 1 describes the status quo in 1995, the year the TIMSS data was collected). Students at the *Gymnasium* are not issued leaving certificates at the end of the lower secondary level, but a

² A detailed description of the German school system can be found in Jonen and Boene (2001).

³ A fourth type of school, *Gesamtschule* (comprehensive school), does not appear in our figure. This type of secondary school offers all lower secondary level school-leaving certificates as well as the entitlement to enter upper secondary school. Comprehensive schools also offer upper-secondary education. It only plays a minor role in most federal states. Less than 10 percent of all students in grade 8 are in a comprehensive school.

qualification to attend the upper level of the *Gymnasium*. Students leaving *Hauptschule* and *Realschule* usually undergo vocational training in the so-called dual system. The system is called dual because students are trained in parallel at the workplace (a private or public sector employer) and in a vocational school.

Table 1: CEE by federal state and type of degree; proportion of students by school/degree type (as of 1995)

	<i>Hauptschule</i> degree	<i>Realschule</i> degree	High school diploma (Abitur)	Students in 8 th grade <i>Haupt-/</i> <i>Realschule</i>	Students in non-Abitur tracks**
Baden-Württemberg (BW)	G/M/F/O	G/M/F	A	71.8	71.6
Bavaria (BY)	G/M/F/S*/O	G/M/F/S*/O	A	73.2	73.1
Berlin (BE)				67.2	57.2
Brandenburg (BB)				71.7	54.8
Bremen (HB)				69.1	63.9
Hamburg (HH)				65.6	56.7
Hesse (HE)				67.2	61.4
Mecklenburg-W. Pomerania (MV)		G/M/F	A	72.2	70.7
Lower Saxony (NI)				73.2	72.0
North Rhine-Westphalia (NW)				69.6	64.7
Rhineland-Palatinate (RP)				72.0	70.5
Saarland (SA)			A	71.8	66.0
Saxony (SN)	G/M/F	G/M/S	A	69.9	69.9
Saxony-Anhalt (ST)		G/M	A	67.6	67.2
Schleswig-Holstein (SH)				73.7	71.5
Thuringia (TH)	G/M	G/M/F	A	67.7	67.4

G = German; M = Mathematics; F = Foreign Language (mostly English); S = Science; O = Other; A = Any subject chosen for the written exams; *subject to student choice; ** Students in 8th grade *Haupt-* and *Realschule* plus two thirds of all students in 8th grade comprehensive school.

Central exit examinations are most common at the end of upper-secondary education. In 1995, seven out of the sixteen German federal states had a central *Abitur* (high-school diploma) on the state level. These states are concentrated in the south (Baden-Württemberg, Bavaria, Saarland) and east (Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia). The other states had decentralised systems, where teachers design problems for exit examinations individually subject to the approval of the school supervisory authority. Six states have central exit examinations at the end of *Realschule* and only four have them at the end of *Hauptschule*.

German exit examinations never cover all subjects taught at school. At the *Abitur* level, students can choose three or four subjects (within certain limits that vary by federal

state). This leads to difficult self-selection problems, which are unlikely to be solved convincingly with the available TIMSS data. At *Realschule* and *Hauptschule*, German and mathematics are always part of the exit examinations, i.e., mathematics is compulsory for all students passing exit examinations in these two school types. In order to assess the effect of CEEs on student achievement, we will thus concentrate on mathematics performance in *Hauptschule* and *Realschule* as the main outcome variable thought to be affected by CEEs. Other subjects in central exit examinations are languages (mostly English) or – less common – science.

3. Data description

The international data set of TIMSS Germany contains data on a total of 5763 students from 7th and 8th grade in 137 schools, collected in the 1994/95 school year. Students are from 14 out of 16 German federal states (Baden-Württemberg and Bremen did not participate), and we have data from all major types of secondary schools: *Hauptschule*, *Realschule*, *Gymnasium*, and *Gesamtschule*. However, for reasons explained below we only use data from students in *Haupt-* and *Realschule*. In addition to the actual test results in mathematics and science, TIMSS data contains a wide range of background variables on student background and attitudes, and teacher and school background. Despite the richness of the available data, we will follow a rather parsimonious approach and select a limited number of control variables for student and school background that have proven to have sizeable explanatory power for student achievement.

Table 2 contains variable definitions and descriptive statistics by type of exit examination. Note that we use sampling weights throughout the entire paper. In contrast to publications with focus on international comparisons of student achievement, we do not take the international standardised math (and in some analyses) science scores as dependent

variables. For sake of intra-German comparability we chose the national Rasch scores, standardised to have mean zero and variance one. The size of our regression parameters are thus directly interpretable in terms of standard deviations.

Table 2: Descriptive Statistics (weighted)

Variable	non-CEE		CEE	
	Mean	StdDev	Mean	StdDev
Mathematics score	-0.134	0.943	0.213	0.971
Science score	-0.114	0.966	0.216	0.969
Sex (1 = girl)	0.486		0.459	
Books at home: 0-10	0.183		0.188	
Books at home: 11-25	0.291		0.333	
Books at home: 26-100	0.168		0.175	
Books at home: 101-200	0.234		0.209	
Books at home: 200+	0.123		0.096	
Immigrant child (both parents born abroad)	0.191		0.077	
School type (1 = <i>Realschule</i>)	0.522		0.541	
Grade (1 = 8 th grade)	0.498		0.478	
Repeated class at least once	0.359		0.246	
Science at home=0 hours/day	0.174		0.203	
Science at home<1 hours/day	0.639		0.626	
Science at home=1-2 hours/day*	0.173		0.157	
Science at home=3-5 hours/day	0.010		0.012	
Science at home>5 hours/day	0.005		0.022	
East Germany	0.042		0.455	
Cumulative math lessons (in 1000s)	1.198	0.121	1.292	0.105
N obs.	1834		1363	

* response categories did not cover 2-3 hours

The most notable difference between students with and without central exit examinations is in terms of achievement, both in math and science. Those with central exams score on average one half standard deviation higher than those without. Student background, measured by the number of books at home differs only slightly. The proportion of students within each range are quite the same in CEE and non-CEE states. Immigrant children are more frequent in the non-CEE group than in the CEE group. This is to be explained mainly by the fact that immigrants are relatively rare in East Germany, where most federal states have central exit examinations (a heritage of the former GDR education system). Another interesting fact is that the cumulative number of math lessons – calculated from official time tables of all federal states (Frenck 2001) – is considerably smaller for students in non-CEE states.

4. Identification and Estimation

Using German TIMSS data, the most basic approach to identify the causal effect of CEE on student achievement would seem to estimate *simple differences* between average achievement in CEE states and non-CEE states, controlling for student background and other variables of interest, e.g. the total time devoted to math and science education. By total time, we do not only mean hours in the current school year but hours accumulated over all school years up to the examination period. Simple differences have only limited value because they ignore two potentially confounding effects: endogeneity of CEEs and a composition effect. The composition effect stems from the fact that CEE states have on average more students in *Haupt-* and *Realschule* and less students in *Gymnasium* than non-CEE states. Since students are selected into secondary schools mainly on the basis of achievement in primary school, this fact gives rise to the interesting effect that student achievement in CEE states (conditional on school type) will be higher simply by having on average relatively more able students in each type of school. We will use information on the proportion of students in each school type to account for this kind of composition effect. Different student compositions in German secondary schools across federal states are interpreted as the result of different ability cutpoints α chosen to separate students. As a proxy for α , we will use $\Phi^{-1}(1-a)$, the a percent quantile of the standard normal distribution, where a is the proportion of 8th grade students heading for a high school diploma.

Self-selection into treatment is one of the most frequent problems when researchers try to evaluate causal effects of certain policy measures. Although it cannot be ruled out completely that parents vote with their feet and move between federal states in order to send their children to schools with a central exit examination, this seems rather unlikely. We therefore assume that the treatment status is exogenous given the institutional arrangement in

each federal state. However, in the long run institutions can change. The existence of CEEs might reflect unobserved variables such as of the electorate's priority on education, that is parental attitudes towards education and achievement in school. When CEEs are correlated with such attitudes, simple differences between CEE and non-CEE states are a biased measure of the CEE effect.

To isolate CEE effects from differential parental attitudes or other unobserved variables, our strategy is to exploit further variation within states. For instance, a typical CEE state has central examinations for each type of leaving certificate (*Haupt-, Realschulabschluss, Abitur*), a typical non-CEE state has no central examinations at all. Some states, however, have mixed systems. In Mecklenburg-Western Pomerania, *Hauptschule* exams are not central, but *Realschule* exams and *Abitur* are. In Saarland, only *Abitur* was a CEE (since 2001/2002 all exams are central). This variation in institutional settings can be used in the sense that those in non-CEE schools living in CEE states can be used as a control group. Unfortunately, there are two problems with this estimation strategy. First, as mentioned before, the allocation of students to school types is not random but rather on the basis of prior academic achievement, which in turn might be correlated with unobservable but relevant variables, such as susceptibility to extrinsic motivation of all kind. It is therefore desirable to take selection into school types into account. The second and more important problem is lack of data: the "interesting" states such as Mecklenburg-Western Pomerania and Saarland are small and there are not many observations from both states in the sample.

A further possibility to evaluate the effectiveness of central exit examinations are *before-after comparisons*. Fuelled by a heated public discussion after the publication of TIMSS and in particular the PISA study, some states have newly introduced CEEs or are planning to introduce them. Quite interestingly, no federal state actually has plans to abolish CEEs. Since PISA is designed as a repeated cross-section, this data could be used to estimate

the effect of these policy changes. In this framework, the causal effect of CEEs will still be difficult to identify because there might also have been other policy changes as a response to TIMSS/PISA. The publication of the test results might have changed parental attitudes towards education, etc. Since we have only one cross-section of data, we mention this possibility mainly for sake of completeness.⁴

Fortunately, the German secondary education system offers a unique source of exogenous variation that can be used to identify the causal effect of CEEs on student achievement. Table 1 shows not only which federal states have CEEs in which types of schools, but also which subjects are covered. Note that in *Haupt-* and *Realschule*, central exit examinations (if any) cover only German, mathematics, and one foreign language (mostly English). Science is not tested in central examinations – with two exceptions. For Saxony's *Realschule* degree, science is a compulsory subject. In Bavaria, science is optional. Each year, roughly 40 percent of all students aiming at a *Hauptschule* degree are tested in biology, chemistry and physics, i.e. all subjects covered by the TIMSS science test. Between 20 and 25 percent of those aiming at the *Realschule* degree actually have written exit examinations in physics only, which accounts for roughly one third of the TIMSS science items. If mathematics is a subject in central examinations and science is not, the effect of CEEs on student achievement should be larger in math than in science classes. Since TIMSS provides test results both in math and in science, we can estimate *difference-in-differences by subject*. The main advantage of this estimator is that with the available data (math and science results for one and the same student) each individual serves as his/her own control group. By accounting for the correlation between both measures on the individual level, we are able to

⁴ Wößmann (2002) uses the international TIMSS data set to estimate difference-in-differences by grade, arguing that incentives should increase as the exit exams approach. His regressions suggest that the *end of upper secondary* CEE effect increases by about 50% when students move from grade 7 to 8. We have two objections against this approach. First, we simply believe that this estimates is too large to be believed in (as a simple thought experiment, extrapolate this effect to the final school year). Second, how incentives increase as one approaches the final exam is an interesting question in its own right and would have to be studied in a dynamic

control for individual heterogeneity, for instance general ability or parental background.⁵ Of course, in order to interpret this difference as the causal effect of CEEs on student performance, one still needs identifying assumptions. First, there must be no systematic indirect effects in the form of spill-over from mathematics to science.⁶ Second, CEE and non-CEE states must not differ systematically in their relative preference of maths over science. Third, mathematics and science test results must be measured on the same scale, i.e., they must be comparable.

There will be positive spill-over from math to science if good math skills are a prerequisite for performing well in science, or – to be more precise – in TIMSS science items. In this case the difference-in-differences by subject framework will *underestimate* the effect of CEEs on achievement. However, we believe that spill-over from good maths education to good performance in the TIMSS science test is likely to be very small. In order to assess the likelihood of such spill-over, we analysed the (released) set of TIMSS science items (IEA TIMSS 1998). The released set contains 87 items of which only four require quite basic mathematics skills, such as dividing by a fraction (see Appendix).

One can also think of some kind of negative spill-over from math to science if students divert resources away from learning science to learning maths because the latter is tested against an external and thus higher standard (displacement effect). This is what a simple model of time allocation between learning mathematics, learning science, and leisure would predict, reflecting the incentives that are actually intended when introducing central exams. Given this displacement effect the *difference-in-differences by subject* framework will *overestimate* the effect of a general introduction of CEEs. Strictly speaking, we are only able to measure the size effect of a partial introduction of CEEs unless we can keep constant all

optimisation framework. Without such a model it remains unclear at which rate relative performance will increase from year to year.

⁵ One remaining source of heterogeneity are of course relative innate math versus science skills.

inputs invested in learning science. These inputs are only partly observable, e.g. as the students' self-reported hours of studying science outside school. Below, we will make tentative use of this data although we believe that the results must be interpreted with caution. First, the quality of these self-reports is at best unclear and second, studying at home itself could be endogenous provided less than average performing students need more hours to do their homework than brighter ones.

The second identifying assumption is that relative preferences are the same in both types of states. The fact that most CEE-states test mathematics but not science in exit examinations indicates that mathematics skills are generally more valued than science skills. It does not allow to conclude that the *relative* preference is stronger in CEE states than in others. Mathematics appears to be a core subject in every state, accounting for roughly one fifth of official teaching time in primary schools and about one seventh of official teaching time in lower secondary schools. However, between CEE and non-CEE states, there are no significant differences in relative teaching time. In CEE states, the proportion of math lessons among all lessons is 14.3 and 13.7 percent in *Hauptschule* and *Realschule*, respectively. In non-CEE states, the corresponding figures are 14.6 and 13.7 percent, i.e. the average percentage of math lessons is even slightly higher (Frenck 2001).

According to the complementary evidence presented above, the identifying assumptions of our difference-in-differences by subject framework appear to be plausible. Still, one potential problem remains to be discussed: Since exit examinations in Saxony generally comprise maths and science, we exclude these cases from this part of our analysis. (Alternatively, Saxony can be treated as a non-CEE state, which leads only to minor changes in our results). Between 25 and 40 percent of all students in Bavaria have central exit examinations in science. Although it seems reasonable to assume that those who are good in

⁶ We do not consider science to math spill-overs, since we do not expect any effect of specific knowledge and skills (science) to more general skills (mathematics).

science will choose science as a topic for their exit examination, we are not able to tell for sure who will eventually do which kind of exam. One possibility to deal with this problem is to discard all students from Bavaria from our regression. However, we are reluctant to do so for two reasons: First, Bavaria has the largest sample size among all CEE states, accounting for about 60 percent of all CEE observations. Second, if Bavaria is discarded from the data, all remaining CEE states are in East Germany. Since four out of five federal states in East Germany have CEEs, eliminating Bavaria from our sample would make it impossible to distinguish the CEE effect from a "former GDR" effect. This is important because schools in the former GDR appear to have a slightly different tradition in the way science is taught than the rest of Germany.⁷

Estimates of the CEE effect in our difference-in-differences by subject framework will of course be affected when some of the students in our sample will eventually have to pass CEEs in science, but this will bias our estimates downwards. We are therefore able to give some kind of lower bound for the causal effect of CEEs. However, we will also provide a simple robustness check by discarding those Bavarian students from the sample that are most likely to choose science as a subject of their exit examination. These are students in *Hauptschule* who strongly agree to the statement that they "usually do well" in biology and/or physical science (chemistry/physics), and students in *Realschule* who strongly agree to the statement that they "usually do well" in physical science. Since using the full sample of Bavarian students gives us a lower bound of the CEE effect, we expect larger effects when we exclude those doing well in science.

The third assumption needed to interpret the difference-in-differences by subject as the size effect of CEEs is that mathematics and science achievement in TIMSS are measured on the same scale and that calculating the difference is thus feasible. Below, we have examined

⁷ Recent analyses from PISA show that students in East Germany perform significantly better when given a specific national science item set than when given the international science items. In West Germany, no such

the robustness of our estimates against violations of this assumption by converting the national Rasch scores into exact quantiles and using differences therein as dependent variables.

5. Results

Our estimation results are listed in Tables 3 and 4. Table 3, column (1) contains an estimate for the simple difference in maths achievement between students with and without a central exit examination in mathematics. The difference is as large as .472 standard deviations, more than the equivalent of an entire school year (.372 standard deviations). Note that this difference is already estimated net of any student background and composition effects. Bishop (1997) reports the CEE effect to be only about one-half of a US grade-level equivalent when comparing Canadian provinces with and without CEEs.

All our background variables have the expected effects on the students' math scores. Since they have been selected on the basis of primary school achievement, *Realschule* students perform much better than those in *Hauptschule*. The number of books at home is used as a proxy for the intellectual background of the parents, which usually has much better explanatory power for and stronger impact on children's achievement than formal education. In fact, the difference between those with less than ten books at home and those with more than 200 is larger than one school year. Immigrant children perform slightly worse than others, those who already have repeated a class are also doing less well, girls perform on average worse than boys, and East German students perform worse than West Germans.⁸

The correlation between central exit examination and student achievement reported in column (1) could well be driven by unobservables that are correlated with CEEs. In order to

difference can be observed (Baumert et al. 2002).

⁸ For a detailed analysis of gender differences in student abilities by type of TIMSS task see Mullis et al. (2000).

Table 3: Central exit examinations effects on student achievement in mathematics

	simple diffs account. for composition effect ^a	diff-in-diffs I ^b	diff-in-diffs I ^b	diff-in-diffs II ^{b,c}	diff-in-diffs II ^{b,c}
	(1)	(2)	(3)	(4)	(5)
CEE	0.472 (6.78)**	0.123 (2.93)**	0.139 (3.38)**	0.174 (3.87)**	0.189 (4.31)**
Grade	0.372 (4.27)**	-0.005 (0.14)	-0.021 (0.55)	-0.002 (0.06)	-0.019 (0.47)
Alpha	2.131 (3.47)**				
Realschule	0.715 (12.88)**	0.138 (3.27)**	0.150 (3.51)**	0.116 (2.70)**	0.128 (2.94)**
11-25 books	0.129 (2.25)*	-0.116 (2.10)*	-0.129 (2.39)*	-0.122 (2.11)*	-0.134 (2.40)*
26-100 books	0.293 (4.82)**	-0.132 (2.25)*	-0.139 (2.41)*	-0.112 (1.85)	-0.115 (1.95)
101-200 books	0.384 (5.97)**	-0.132 (2.00)*	-0.127 (1.98)*	-0.129 (1.87)	-0.120 (1.80)
200+ books	0.446 (7.69)**	-0.160 (2.72)**	-0.181 (3.10)**	-0.137 (2.31)*	-0.157 (2.68)**
Immigrant child	-0.142 (2.47)*	0.275 (4.92)**	0.291 (5.12)**	0.294 (5.17)**	0.313 (5.45)**
Repeated class	-0.130 (3.32)**	-0.066 (1.68)	-0.065 (1.63)	-0.065 (1.58)	-0.063 (1.50)
Girl	-0.264 (7.00)**	0.090 (2.58)*	0.102 (2.82)**	0.091 (2.57)*	0.104 (2.77)**
East	-0.489 (4.63)**	-0.294 (3.68)**	-0.305 (3.69)**	-0.342 (4.16)**	-0.350 (4.14)**
Cum. math lessons	-0.410 (0.96)				
Relative cum. math lessons		0.590 (0.60)	0.471 (0.47)	0.926 (0.95)	0.787 (0.79)
Science at home<1 hour/day			-0.048 (0.86)		-0.048 (0.81)
Science at home=1-2 hours/day			0.035 (0.34)		-0.015 (0.14)
Science at home=3-5 hours/day			-0.026 (0.64)		-0.027 (0.63)
Science at home>5 hours/day			0.707 (2.89)**		0.691 (2.80)**
Constant	-1.404 (3.10)**	-0.176 (0.71)	-0.129 (0.51)	-0.261 (1.08)	-0.211 (0.85)
Observations	3197	2995	2909	2727	2644
R-squared	0.35	0.04	0.05	0.04	0.05

Robust t-statistics (accounting for clustering on the class level) in parentheses;

* significant at 5% level; ** significant at 1% level; ^a dependent variable: math score; ^b dependent variable: math score -science score; ^c excluding Bavarian students who claim being good in science

disentangle this correlation from causation, we now turn to our difference-in-differences by subject estimates.

As argued before, the main advantage of this estimator is that every student serves as his or her own control group, being examined centrally in mathematics but not in science. We have implemented this estimator by simply calculating each individual's difference between his/her math and his/her science score and regressing this difference on a set of explanatory variables. In contrast to "levels"-regression in column (1) we use the ratio between cumulative math lessons and a proxy for science lessons instead of the number of math lessons.⁹ The estimates are listed in column (2) of Table 3. The first thing to note is that the coefficient for CEE remains positive and significant. However, it drops from .472 standard deviations in the simple differences estimator to .123 standard deviations, or one third school year equivalent. Students in CEE states show better relative performance in mathematics than in science. This is consistent with the claim that CEEs improve student performance. When Bavarian students with great interest in science (and who are thus likely to choose a central exit examination in science) are eliminated from the sample, the estimate for the CEE effect increases (column (4)). This is in line with our expectations because we take out students whose relative performance is less likely to be affected by CEEs than the average.

The rationale to control for the students' amount of time spent on learning science at home (columns (3) and (5)) is to eliminate possible displacement effects from our estimates. If we keep inputs into learning science constant, the estimated effect of CEEs on the difference between math and science achievement can be interpreted as the effect of a general introduction of CEEs. Otherwise we are only able to identify the effect of a partial introduction, namely in mathematics and German. Contrary to our expectations, controlling

for the time spent on learning science at home slightly increases our estimate of the CEE effect and the number of hours itself has no effect on the math-science score differential (unless if learning science longer than incredible 5 hours per day). As we already mentioned these results have to be treated with caution, given the fact that this is self-reported data and the potential endogeneity of the variable itself.

While the focus is clearly on the CEE variable, other parameter estimates in columns (2) to (5) are worth noting. The *Realschule* dummy is positive and significant, hence students at *Realschule* perform relatively better in math than in science. The book variables are negative and slightly increasing in magnitude with the number of books at home. Thus students from better educated households perform relatively better in science. Children with an immigration background are relatively better math students, which might be due to an insufficient command of German, which is less important for maths than for science.

Table 4 replicates the regressions in Table 3, taking exact percentiles as the dependent variable. For example, the CEE coefficient of 0.135 in column (1) indicates that the average student from a CEE state represents a percentile of the common math score distribution that is 13.5 percentage points above that of the average non-CEE student. The coefficient of 0.035 in column (2) shows that the rank differential between math and science is 3.5 percentage points larger in CEE states than the same rank differential in non-CEE states. Note that this is again in the range of about one third grade year which accounts for 9.9 percentage points in the math distribution. Note also that the percentile regressions are less sensitive to the introduction of "science hours at home" to control for a potential displacement effect. To summarize, the conversion of the national Rasch scores into exact quantiles and using differences therein as dependent variables does change the quantitative nature of our results only slightly, supporting our claim that CEEs improve student achievement.

⁹ This proxy is total lessons minus math lessons minus German lessons. Separate information on science lessons was not available.

Table 4: Central exit examinations effects on student achievement in mathematics (percentile effects)

	simple diffs account. for composition effect ^a	diff-in-diffs I ^b	diff-in-diffs I ^b	diff-in-diffs II ^{b,c}	diff-in-diffs II ^{b,c}
	(1)	(2)	(3)	(4)	(5)
CEE	0.135 (6.80)**	0.035 (2.83)**	0.035 (2.89)**	0.050 (3.80)**	0.050 (3.85)**
Grade	0.099 (3.70)**	-0.010 (0.90)	-0.011 (0.96)	-0.009 (0.73)	-0.010 (0.80)
Alpha	0.602 (3.26)**				
Realschule	0.217 (12.65)**	0.044 (3.31)**	0.044 (3.32)**	0.038 (2.81)**	0.038 (2.82)**
11-25 books	0.034 (1.94)	-0.036 (2.27)*	-0.035 (2.20)*	-0.037 (2.28)*	-0.037 (2.21)*
26-100 books	0.082 (4.58)**	-0.037 (2.20)*	-0.037 (2.19)*	-0.033 (1.88)	-0.033 (1.88)
101-200 books	0.121 (6.19)**	-0.025 (1.29)	-0.025 (1.28)	-0.023 (1.16)	-0.023 (1.15)
200+ books	0.129 (7.18)**	-0.046 (2.64)**	-0.047 (2.73)**	-0.036 (2.15)*	-0.038 (2.24)*
Immigrant child	-0.030 (1.62)	0.081 (4.86)**	0.078 (4.57)**	0.087 (5.17)**	0.084 (4.92)**
Repeated class	-0.040 (3.42)**	-0.024 (1.91)	-0.023 (1.84)	-0.021 (1.63)	-0.020 (1.54)
Girl	-0.080 (6.72)**	0.031 (2.77)**	0.032 (2.83)**	0.031 (2.73)**	0.033 (2.81)**
East	-0.153 (4.81)**	-0.097 (3.47)**	-0.099 (3.57)**	-0.111 (3.83)**	-0.113 (3.93)**
Cum. math lessons	-0.083 (0.65)				
Relative cum. math lessons		0.301 (0.91)	0.329 (1.00)	0.382 (1.14)	0.411 (1.23)
Science at home<1 hour/day			-0.000 (0.01)		-0.002 (0.09)
Science at home=1-2 hours/day			0.034 (1.06)		0.023 (0.72)
Science at home=3-5 hours/day			-0.002 (0.17)		-0.001 (0.11)
Science at home>5 hours/day			0.197 (3.16)**		0.194 (2.97)**
Constant	0.048 (0.34)	-0.084 (1.03)	-0.091 (1.10)	-0.109 (1.33)	-0.116 (1.40)
Observations	2982	2791	2791	2539	2539
R-squared	0.34	0.04	0.05	0.05	0.05

Robust t-statistics (accounting for clustering on the class level) in parentheses;

* significant at 5% level; ** significant at 1% level; ^a dependent variable: math score percentile; ^b dependent variable: math score percentile-science score percentile; ^c excluding Bavarian students who claim being good in science

6. Conclusion

This paper discusses the benefits of central exit examinations (CEEs) for academic achievement at lower secondary education. Theoretically, the benefits from central examinations are at hand. However, it is not straightforward to identify the causal effect of CEEs empirically. Unlike earlier studies, we make use of institutional variation in Germany that allows us to develop an identification strategy to estimate the causal effect of CEEs on academic performance. In Germany's school system, only some states have CEEs, mostly in the core subjects German and Mathematics. We use data from the Third International Mathematics and Science Study (TIMSS) and exploit this institutional variation to uncover the causal effect of CEEs on student achievement in mathematics. Several possible identification strategies, all difference-in-differences estimators, are discussed.

Comparing simple test results, students in German CEE states clearly outperform those in non-CEE states (by approximately 0.5 standard deviations or one and a quarter grade years). However, this also applies to a somewhat lesser extent to subjects that are not tested in central examinations, such as science. We propose a difference-in-differences estimator that interprets the difference in math and science achievement in TIMSS in CEE states compared to the same difference in non-CEE states as the causal effect of central examinations on achievement. Under the assumption of no spill-over effects and identical *relative* preferences between math and science, the average causal effect of CEE on math achievement is estimated to be about 0.12 standard deviations or one third grade year.

The difference between the raw difference between states with and without CEEs and what we identify as the causal effect of CEEs is fairly sizable. Thus caution is warranted when interpreting observed differences between states with or without CEEs as the effect of CEEs on student achievement. Much (but not all) of the correlation between CEEs and student

performance seems to be driven by general preferences for education in the German federal states.

Still, our empirical findings suggest that the introduction of central exit examinations will raise average student achievement significantly. Although the estimated increase will not completely level out raw differences between states with and without CEEs, policy makers in German federal states should seriously consider CEEs in order to provide students and teachers in their states with incentives to adhere to higher standards. Compared to other measures discussed to raise student achievement such as decreasing class sizes, central exit examinations seem to be a lot more cost effective.

References:

- Baumert, J., Artelt, C., Klieme, E., et al., 2002, PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Leske + Budrich, Opladen.
- Baumert, J., Watermann, R., 2000, Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. Ch. 8 in J. Baumert, W. Bos, and R. Lehmann (ed.), TIMSS - Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske+Budrich.
- Bishop, J.H., 1997, The effect of national standards and curriculum-based exams on achievement. *The American Economic Review* 87, 260-264.
- Bishop, J.H., 1999, Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review* 6, 349-401.
- Costrell, R.M., 1994, A simple model of educational standards, *The American Economic Review* 84, 956-971.
- Costrell, R.M., 1997, Can educational standards raise welfare?, *Journal of Public Economics* 65, 271-293.
- Effinger M.R., Polborn, M.K., 1999, A model of vertically differentiated education, *Journal of Economics* 69, 53-69.
- Hanushek, E. A., 1999, The Evidence on Class Size, pp. 131-168 in S. E. Mayer and P. Peterson (ed.), *Earning and Learning: How Schools Matter*. Washington, D.C.: Brookings Institution.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G., 1998, Teachers, schools, and academic achievement, NBER Working Paper 6691.
- Hoxby, C. M., 2000a, The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics* 115, 1239-1285.

- Hoxby, C. M., 2000b, Does competition among public schools benefit students and taxpayers, *The American Economic Review* 90, 1209-1239.
- Frenck, I., 2001, Stundentafeln der Primar- und Sekundarstufe I im Ländervergleich - eine empirische Studie am Beispiel der Fächer Deutsch und Mathematik. Unpublished master thesis, Universität Essen.
- IEA TIMSS, 1998, TIMSS science items: Released set for population 2 (seventh and eighth Grades). <http://timss.bc.edu/timss1/timsspdf/bsitems.pdf> [2002, March 1].
- Jonen, G., Boele, K., 2001, The Education System in the Federal Republic of Germany 2000. German EURYDICE Unit, Bonn. http://www.kmk.org/dossier/dossier_2000_engl_e-book.pdf [2002, July 1].
- Jürges, H., Schneider, K., 2003, International differences in student achievement: An economic perspective, Working paper. Department of Economics, University of Dortmund.
- Mullis, I.V.S, Martin, M. O., Fierros, E. G., Goldberg, A. L., and Stemler, S. E., 2000, Gender Differences in Achievement. International Study Center, Lynch School of Education, Boston College.
- Wößmann, L., 2002, Central Exams Improve Educational Performance: International Evidence. Kiel Discussion Papers 397.

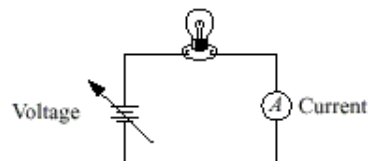
Appendix: TIMSS population 2 Science items involving math skills

L4 – Machine A and Machine B are each used to clear a field. The table shows how large an area each cleared in 1 hour and how much gasoline each used.

	Area of field cleared in 1 hour	Gasoline used in 1 hour
Machine A	2 hectares	3/4 liter
Machine B	1 hectare	1/2 liter

Which machine is more efficient in converting the energy in gasoline to work? Explain your answer.

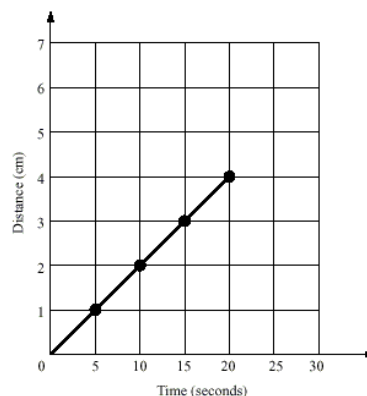
M12 – Some students used an ammeter A to measure the current in the circuit for different voltages.



The table shows some of the results. Complete the table.

Voltage (volts)	Current (milliamperes)
1.5	10
3.0	20
6.0	

P1 – The graph shows the progress made by an ant moving along a straight line



If the ant keeps moving at the same speed, how far will it have traveled at the end of 30 seconds?

- A. 5 cm
 - B. 6 cm
 - C. 20 cm
 - D. 30 cm
-

Z1 – It takes 10 painters 2 years to paint a steel bridge from one end to the other. The paint that is used lasts about 2 years, so when the painters have finished painting at one end of the bridge, they go back to the other end and start painting again.

- a. Why **MUST** steel bridges be painted?
- b. A new paint that lasts 4 years has been developed and costs the same as the old paint. Describe 2 consequences of using the new paint.

CESifo Working Paper Series

(for full list see www.cesifo.de)

- 873 Eytan Sheshinski, Optimum and Risk-Class Pricing of Annuities, February 2003
- 874 Willi Leibfritz, Paul O'Brien and Jean-Christophe Dumont, Effects of Immigration on Labour Markets and Government Budgets – An Overview, February 2003
- 875 M. Hashem Pesaran and Allan Timmermann, How Costly is it to Ignore Breaks when Forecasting the Direction of a Time Series?, February 2003
- 876 Thorvaldur Gylfason and Gylfi Zoega, Education, Social Equality and Economic Growth: A View of the Landscape, February 2003
- 877 Robin Boadway and Jean-François Tremblay, Public Economics and Startup Entrepreneurs, February 2003
- 878 Erkki Koskela and Roope Uusitalo, The Un-Intended Convergence: How the Finnish Unemployment Reached the European Level, February 2003
- 879 Robert Fenge and Volker Meier, Pensions and Fertility Incentives, February 2003
- 880 Eytan Sheshinski, Note on Income Taxation and Occupational Choice, February 2003
- 881 A B Atkinson, Income Inequality in OECD Countries: Data and Explanations, February 2003
- 882 Thomas Gehrig and Rune Stenbacka, Venture Cycles: Theory and Evidence, February 2003
- 883 Ralf Becker and Thomas Hellmann, The Genesis of Venture Capital - Lessons from the German Experience, March 2003
- 884 Eytan Sheshinski, Note on the Optimum Pricing of Annuities, March 2003
- 885 Paul De Grauwe and Magdalena Polan, Globalisation and Social Spending, March 2003
- 886 F. van der Ploeg, Do Social Policies Harm Employment and Growth?, March 2003
- 887 Mirjam van Praag, Initial Capital Constraints Hinder Entrepreneurial Venture Performance: An empirical analysis, March 2003
- 888 Bernard Steunenberg, Coordinating Sectoral Policymaking: Searching for Countervailing Mechanisms in the EU Legislative Process, March 2003
- 889 Eytan Sheshinski, Optimum Delayed Retirement Credit, March 2003

- 890 Frederick van der Ploeg, Rolling Back the Public Sector – Differential effects on employment, investment and growth, March 2003
- 891 Paul De Grauwe and Marc-Alexandre S negas, Monetary Policy in EMU when the Transmission is Asymmetric and Uncertain, March 2003
- 892 Steffen Huck and Kai A. Konrad, Strategic Trade Policy and the Home Bias in Firm Ownership Structure, March 2003
- 893 Harry Flam, Turkey and the EU: Politics and Economics of Accession, March 2003
- 894 Mathias Hoffmann and Ronald MacDonald, A Re-examination of the Link between Real Exchange Rates and Real Interest Rate Differentials, March 2003
- 895 Badi H. Baltagi, Espen Bratberg, and Tor Helge Holm s, A Panel Data Study of Physicians' Labor Supply: The Case of Norway, March 2003
- 896 Dennis C. Mueller, Rights and Citizenship in the European Union, March 2003
- 897 Jeremy Edwards, Gains from Trade in Tax Revenue and the Efficiency Case for Trade Taxes, March 2003
- 898 Rainer Fehn and Thomas Fuchs, Capital Market Institutions and Venture Capital: Do They Affect Unemployment and Labour Demand?, March 2003
- 899 Ronald MacDonald and Cezary W jcik, Catching Up: The Role of Demand, Supply and Regulated Price Effects on the Real Exchange Rates of Four Accession Countries, March 2003
- 900 R. Selten, M. Schreckenberg, T. Pitz, T. Chmura, and S. Kube, Experiments and Simulations on Day-to-Day Route Choice-Behaviour, April 2003
- 901 Stergios Skaperdas, Restraining the Genuine *Homo Economicus*: Why the Economy Cannot be Divorced from its Governance, April 2003
- 902 Yin-Wong Cheung, Menzie D. Chinn, and Antonio Garcia Pascual, What Do We Know about Recent Exchange Rate Models? In-Sample Fit and Out-of-Sample Performance Evaluated, April 2003
- 903 Mika Widgr n, Enlargements and the Principles of Designing EU – Decision-Making Procedures, April 2003
- 904 Phornchanok Cumperayot, Dusting off the Perception of Risk and Returns in FOREX Markets, April 2003
- 905 Kai A Konrad, Inverse Campaigning, April 2003
- 906 Lars P. Feld and Stefan Voigt, Economic Growth and Judicial Independence: Cross Country Evidence Using a New Set of Indicators, April 2003

- 907 Giuseppe Bertola and Pietro Garibaldi, The Structure and History of Italian Unemployment, April 2003
- 908 Robert A.J. Dur and Otto H. Swank, Producing and Manipulating Information, April 2003
- 909 Christian Gollier, Collective Risk-Taking Decisions with Heterogeneous Beliefs, April 2003
- 910 Alexander F Wagner, Mathias Dufour, and Friedrich Schneider, Satisfaction not Guaranteed – Institutions and Satisfaction with Democracy in Western Europe, April 2003
- 911 Ngo Van Long, Raymond Riezman, and Antoine Soubeyran, Trade, Wage Gaps, and Specific Human Capital Accumulation, April 2003
- 912 Andrea Goldstein, Privatization in Italy 1993-2002: Goals, Institutions, Outcomes, and Outstanding Issues, April 2003
- 913 Rajshri Jayaraman and Mandar Oak, The Signaling Role of Municipal Currencies in Local Development, April 2003
- 914 Volker Grossmann, Managerial Job Assignment and Imperfect Competition in Asymmetric Equilibrium, April 2003
- 915 Christian Gollier and Richard Zeckhauser, Collective Investment Decision Making with Heterogeneous Time Preferences, April 2003
- 916 Thomas Moutos and William Scarth, Some Macroeconomic Consequences of Basic Income and Employment Subsidies, April 2003
- 917 Jan C. van Ours, Has the Dutch Miracle Come to an End?, April 2003
- 918 Bertil Holmlund, The Rise and Fall of Swedish Unemployment, April 2003
- 919 Bernd Huber and Marco Runkel, Optimal Design of Intergovernmental Grants under Asymmetric Information, April 2003
- 920 Klaus Wälde, Endogenous Business Cycles and Growth, April 2003
- 921 Ramon Castillo and Stergios Skaperdas, All in the Family or Public? Law and Appropriative Costs as Determinants of Ownership Structure, April 2003
- 922 Peter Fredriksson and Bertil Holmlund, Improving Incentives in Unemployment Insurance: A Review of Recent Research, April 2003
- 923 Bernard M.S. van Praag and Adam S. Booij, Risk Aversion and the Subjective Time Discount Rate: A Joint Approach, April 2003
- 924 Yin-Wong Cheung, Kon S. Lai, and Michael Bergman, Dissecting the PPP Puzzle: The Unconventional Roles of Nominal Exchange Rate and Price Adjustment, April 2003

- 925 Ugo Trivellato and Anna Giraldo, Assessing the ‘Choosiness’ of Job Seekers. An Exploratory Approach and Evidence for Italy, April 2003
- 926 Rudi Dornbusch and Stanley Fischer, International Financial Crises, April 2003
- 927 David-Jan Jansen and Jakob de Haan, Statements of ECB Officials and their Effect on the Level and Volatility of the Euro-Dollar Exchange Rate, April 2003
- 928 Mario Jametti and Thomas von Ungern-Sternberg, Assessing the Efficiency of an Insurance Provider – A Measurement Error Approach, April 2003
- 929 Paolo M. Panteghini and Guttorm Schjelderup, Competing for Foreign Direct Investments: A Real Options Approach, April 2003
- 930 Ansgar Belke, Rainer Fehn, and Neil Foster, Does Venture Capital Investment Spur Employment Growth?, April 2003
- 931 Assar Lindbeck, Sten Nyberg, and Jörgen W. Weibull, Social Norms and Welfare State Dynamics, April 2003
- 932 Myrna Wooders and Ben Zissimos, Hotelling Tax Competition, April 2003
- 933 Torben M. Andersen, From Excess to Shortage – Recent Developments in the Danish Labour Market, April 2003
- 934 Paolo M. Panteghini and Carlo Scarpa, Irreversible Investments and Regulatory Risk, April 2003
- 935 Henrik Jacobsen Kleven and Claus Thustrup Kreiner, The Marginal Cost of Public Funds in OECD Countries. Hours of Work Versus Labor Force Participation, April 2003
- 936 Klaus Adam, George W. Evans, and Seppo Honkapohja, Are Stationary Hyperinflation Paths Learnable?, April 2003
- 937 Ulrich Hange, Education Policy and Mobility: Some Basic Results, May 2003
- 938 Sören Blomquist and Vidar Christiansen, Is there a Case for Public Provision of Private Goods if Preferences are Heterogeneous? An Example with Day Care, May 2003
- 939 Hendrik Jürges, Kerstin Schneider, and Felix Büchel, The Effect of Central Exit Examinations on Student Achievement: Quasi-experimental Evidence from TIMSS Germany, May 2003