# The effect of central exit examinations on student achievement

## Quasi-experimental evidence from TIMSS Germany

[preliminary draft, do not cite without permission]

Hendrik Jürges, MEA, Universität Mannheim

Kerstin Schneider, Universität Dortmund *

Felix Büchel, Max Planck Institute for Human Development, Berlin

November 2002

**Abstract [98 words]:** 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 alternative difference-in-differences estimators that exploit the quasi-experimental nature of the national 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 this difference is due to central exit examinations themselves. Our most preferred estimator suggests that such examinations increase student achievement by at least one third school year equivalent.

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

**JEL-Code:** I28

---

* Corresponding author: Department of Economics; Universität Dortmund; 44221 Dortmund; Germany. Tel. +49-231-755-3440. Fax: +49-231-755-5404. Email: kerstin.schneider@wiso.uni-dortmund.de. We are grateful to Rainer Watermann, Max Planck Institute for Human Development, Berlin, for editing the regional information used in this study, and Jürgen Baumert for granting special permission to use this data. Regional TIMSS data is currently strictly protected and will be made available to interested researchers after being factually anonymised.

## 1. Introduction

The poor performance of German students in TIMSS and PISA results gave rise to 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, Wößmann and West, 2002, Jürges and Schneider, 2002). Even if small positive effects of financial resources can be identified, they are very costly. (example)

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 in 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

[1] See Jürges and Schneider (2002) for a discussion in the context of the international TIMSS database.

teaching quality on a regular basis. Whether incentives to improve teaching quality, an 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 claim 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 (2002) 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, both 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 outline a theoretical framework which is targeted at the German system. A brief account of German secondary education is given in Section 3. Section 4 describes the data, and Section 5 discusses issues of identification and estimation. The results are presented in Section 6 and Section 7 concludes.

## 2. Theoretical framework

The theoretical literature on CEEs or central standards deals with issues of whether centrally set standards are higher than decentral standards, whether this translates in welfare improvements and the optimal level of centralisation (Costrell, 1994, 1997; Effinger and Polborn, 1999). While central standards are not unanimously preferred by this literature, the general tendency is more in favour of central standards than against them.

Here we present a simple explanation for why CEEs improve the average performance of students that reflects the German practice of setting standards. Note that we do not compare standards across school types but only within school types. In a decentralised system, it is on the whole the teachers who decide on the standard in the exams. Hence exams are written and graded such that the grade distribution is acceptable. While teachers have to obey the binding guidelines of the curriculum, the standard is heavily influenced by the ability of both students within the class and teacher. If, however, standards are set centrally for every school in a federal state, the ministry of education collects proposals for exit exams from the teachers. A commission then decides on the central exam. Thus the central exam reflects the average ability of the students and the average ability of the teachers in the federal state. In the following we assume that the curriculum is the same for all federal states. Moreover, assume for the time being that sorting effects can be ignored.

A crucial question is how standards are determined. In a system without central standards, academic standards in a class depend on factors like the ability of students, the socio-economic background of students, the motivation of the teachers etc., which we summarise in an index $a^j$. Since the standards depend on the ability of the class, standards might be different for each class but they are determined according to the same set of rules, $s = s(a^j)$ with $s_{a^j} > 0$. Hence, a class with higher $a^j$ does not have a lower standard than a class with lower ability students. Given a class specific standard, $s^j$, the academic achievement of class *j* under a decentralised system can be written as $l^j = f(a^j, s(a^j))$. We assume that $f$ is strictly concave with $f_a, f_s > 0$ and $f_{aa}, f_{ss} < 0$. Thus standards raise academic achievement at a decreasing rate.

Assuming *n* different types of classes with different $a^j$ and different standards, the average academic performance in a decentralised system is

$$l_{decentral} = \sum_{j=1}^{n} \gamma^j l^j = \sum_{j=1}^{n} \gamma^j f(a^j, s(a^j)), \qquad (1)$$

with weights $\gamma^j$ – reflecting relative class size – and $\sum_{j=1}^{n} \gamma^j = 1$.

If standards are determined according to the same rules as in a decentralised system but are based on the characteristics of the population instead of the class, average academic performance in a system with central standards is

$$l_{central} = \sum_{j=1}^{n} \gamma^j f\left( a^j, s\left( \sum_{j=1}^{n} \gamma^j a^j \right)\right) = \sum_{j=1}^{n} \gamma^j f(a^j, s(\overline{a})) . \tag{2}$$

Comparing academic achievement in a decentralised and a centralised system, we get the following

> **Proposition:** In the absence of sorting, academic achievement is higher with central standards if classes are heterogeneous and the education production function *f* is strictly concave.

To see this define

$$\begin{aligned} d &= l_{central} - l_{decentral} \\ &= \sum_{j \neq i} \gamma^j f(a^j, s(\overline{a})) + \gamma^i f(a^i, s(\overline{a})) - \sum_{j \neq i} \gamma^j f(a^j, s(a^j)) - \gamma^i f(a^i, s(a^i)), \end{aligned} \tag{3}$$

with $\overline{a} = \sum_{j \neq i} \gamma^j a^j + \gamma^i a^i$. Clearly, in the absence of heterogeneity $\overline{a} = a^i = a^j$ , and $d = 0$, thus academic achievement is the same in both regimes. Now assume that $a^i > \overline{a}$ . It follows that $\sum_{j \neq i} \gamma^j s(a^j) < s(\overline{a}) < s(a^i)$, hence

$$\sum_{j \neq i} \gamma^j f(a^j, s(a^j)) < \sum_{j \neq i} \gamma^j f(a^j, s(\overline{a})) < f(a^i, s(\overline{a})) < f(a^i, s(a^i))$$

Since *f* is strictly concave, it follows that any convex combination of $\sum_{j \neq i} \gamma^j f(a^j, \mathrm{s}(\overline{\mathrm{a}}))$ and $f(a^i, s(\overline{a}))$ is greater than a convex combination of $\sum_{j \neq i} \gamma^j f(a^j, \mathrm{s}(\mathrm{a}^j))$ and $f(a^i, s(a^i))$

Figure 1 illustrates the proposition. The argument for $a^i < \overline{a}$ follows the same line of reasoning.

--- about here Figure 1 ---

## 3. 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 2 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.

--- about here Figure 2 ---

*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). This type of school prepares for

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

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

--- about here Table 1 ---

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* (upper secondary leaving degree giving access to higher education) 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 had central exit examinations at the end of *Realschule* and only four had them at the end of *Hauptschule*.

Exit examinations do not always cover all subjects taught at school. At the *Abitur* level, students can choose three or four subjects (within certain limits). At *Realschule* and *Hauptschule*, German and mathematics are always part of the exit examinations. In order to

[3] A forth 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

assess the effect of CEEs on student achievement, we will concentrate on mathematics performance in *Hauptschule* and *Realschule* as our main outcome variable: performance measures in mathematics are available in our data (TIMSS) and mathematics is compulsory for all students passing exit examinations in these two school types. Other subjects in central exit examinations are languages (mostly English) or – less common – science.

### 4. Data description

TIMSS Germany contains data on a total of 5763 students from 7$^{th}$ and 8$^{th}$ 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 mainly make use of the data from *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, although ignoring weights does not lead to significant changes in our results. 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

---

upper secondary school. Comprehensive schools also offer upper-secondary education. It only plays a minor role have mean zero and variance one. The size of our regression parameters are thus directly interpretable in terms of standard deviations.

--- about here Table 2 ---

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. However, immigrant children are far 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).

Two more differences between students with and without central exit examinations are worth being mentioned: first, the cumulative number of math lessons – calculated from official time tables of all federal states (Frenck 2001) – is considerably smaller for students within non-CEE states. Second, students without central exit examinations strongly agree to the statement "I usually do well in mathematics" more often. Measured by their achievement in TIMSS, they do worse than their counterparts with central exams. This form of biased self-assessment may be explained by higher standards in CEE states against which students have to match themselves.

in most federal states. Less than 10 percent of all students in grade 8 are in a comprehensive school.

## 5. Identification and Estimation

With German TIMSS data plus information on federal state and school type, there are several possible ways to identify the causal effect of CEE on student achievement with different underlying assumptions and different strengths and weaknesses.

The most basic approach is 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 descriptive value because they ignore two potential confounding effects. One is a composition effect: as was already described above, 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. Below, 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 $\alpha$ chosen to separate students. As a proxy for $\alpha$, we will use $\Phi^{-1}(1-a)$, the *a* percent quantile of the standard normal distribution, where *a* is the proportion of 8$^{th}$ grade students heading for a high school diploma.

A more important confounding effect than differential composition arises when CEE states differ in unobserved aspects from non-CEE states. Parental attitudes towards education and achievement in school can also differ systematically between states. A CEE might thus be simply an indicator of the electorate's priority on education. Simple differences will thus always be vulnerable to the objection that they merely reflect correlation, not causation.

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 central exit examination, this seems rather unlikely. We can therefore assume that the treatment status is in fact exogenous.

Our sample of middle school students consists of 7$^{th}$ and 8$^{th}$ graders. In *Haupt*-and *Realschule*, CEEs take place at the end of lower secondary education, i.e. in 9$^{th}$ or 10$^{th}$ grade, respectively. In *Gymnasium*, CEEs take place at the end of upper secondary education, i.e. in 12$^{th}$ or 13$^{th}$ grade. The strength of the incentive effect of CEEs should become stronger the closer a student is to his or her expected final grade. The effect of CEEs on student achievement should thus be stronger for students in *Hauptschule* and *Realschule* than in *Gymnasium*. However, in this framework (*difference-in-differences by school type*) it is very important to control for composition effects, since students are allotted to control and treatment groups on the basis of their ability or prior academic achievement. For this reason, those with the intense treatment (*Haupt-* and *Realschule*) and those with a milder treatment (at *Gymnasium*) are not very similar, which might become a problem. Students who are sensitive to all kinds of incentives might have been more hard-working in primary school, so that *Gymnasium* students are systematically responding stronger to incentives – for instance in the form of CEE. For this reason, we will also compare 7$^{th}$ and 8$^{th}$ graders in *Haupt-* and *Realschule* only. If the effect of CEEs increases when students and their teachers come closer to the final school year, students in 8$^{th}$ grade should perform relatively better than those in 7$^{th}$ grade. We call the corresponding estimator *difference-in-differences by grade*.

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. However, some states have mixed systems. For instance, 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. 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 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 is a data problem: the "interesting" states such as Mecklenburg-Western Pomerania and Saarland are small and have only very few cases in the study.

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.

Let us now discuss our most preferred identification strategy. 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. Since students in Gymnasium have a lot more discretion concerning the choice of subjects for written exams, we will concentrate on the other two school types.

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 the *difference-in-differences by subject.* Note that with the available data (math and science results for one and the same student), student composition effects are no longer an issue, but correlation between both measures on the individual level must and can be accounted for. The main advantage of this estimator is that each individual serves as his/her own control group, i.e. we are able to control for heterogeneity on the individual level, for instance general ability.[4] In order to interpret this difference as the causal effect of CEEs on student performance, one needs two identifying assumptions: the absence of any indirect effects in the form of spill-over from mathematics to science, and that students attending schools in CEE and non-CEE states do not differ systematically in their relative preference of maths over science.[5]

[4] One remaining source of heterogeneity are of course relative innate math versus science skills, but these should be randomly distributed across states.

[5] 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).

Spill-over can be either positive or negative. It will be positive if good math skills are a prerequisite for performing good in science, or – to be more precise – in TIMSS science items. In this case the difference-in-differences by subject framework will *under*estimate 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).

Spill-over from math to science will be negative if students divert resources away from learning science to learning maths because the latter is tested against an external standard (displacement effect). In that case the *difference-in-differences by subject* framework will *over*estimate the effect of CEEs on achievement.

To assess the plausibility of a displacement effect, we have examined self-reported student behaviour: Table 3 shows some bivariate probit estimates, where we use the probability of studying math and science at home and the probability of taking extra lessons in math or science outside school as dependent variables. These variables are regressed on a CEE dummy and some explanatory variables that mirror student ability and background. If there was a displacement effect at work, students in CEE states should have a higher relative probability of studying math than science or taking private lessons in math than in science.

--- about here Table 3 ---

However, as the results in Table 3 show this is not the case. On the contrary, students in CEE states have even smaller probabilities of taking extra lessons in math than others, both absolutely, which is surprising, and relative to science, which is evidence against a displacement effect. Regarding studying at home, students in CEE states are relatively more likely to so in maths than in science, but this difference is not significant.

The second identifying assumptions 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 teaching time in primary schools and about one seventh of 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 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.[6]

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.

## 6. Results

Our estimation results are listed in Table 4. 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.

[6] 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 difference can be observed (Baumert et al. 2002).

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.[7]

--- about here Table 4 ---

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 disentangle this correlation from causation, we now turn to our difference-in-differences estimates. Difference-in-differences by grade are measured by the interaction effect of CEE and grade in column (2) of Table 4. This interaction effect would be significantly larger than zero if the difference in student achievement between students with and without central exit examinations increased from 7$^{th}$ to 8$^{th}$ grade, i.e. when students move closer to their exit examinations. However, quite surprisingly we observe a small negative and insignificant coefficient indicating that if there are operative incentives of CEEs, they do not increase when the actual examinations comes closer. A comparison by school types that makes use of the fact that students heading for *Abitur* have four to six years left until their CEE (compared to one to three years for our main study population), also reveals no effects of closeness to exam on the CEE/non-CEE difference (column (3)). What remains in both cases is a strong and highly significant main effect of central exit examinations.

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

We now discuss the results of our most preferred estimator, the difference-in-differences by subject. As already mentioned, 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 the usual set of explanatory variables. One exception is that we use the ratio between cumulative math lessons and a proxy for science lessons instead of the number of math lessons.[8] The estimates are listed in column (4) of Table 4. 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 CEE-tested subjects (mathematics) than in others (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 (5)). 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.

While the focus is clearly on the CEE variable, other parameter estimates in columns (4) and (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, less needed in maths than in science.

---

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

Finally, a possible objection against calculating differences between mathematics and science scores is that they are not measured on the same scale and thus not directly comparable measures. We have therefore examined the robustness of our estimates by converting the national Rasch scores into exact quantiles and using differences therein as dependent variables. The qualitative nature of the corresponding regressions is in fact very similar to the results reported above. The average student's quantile difference between math and science is significantly higher in CEE states (by about 3.3 percentiles or again about one third grade year equivalent; not reported in Tables).

## 7. 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 between federal states 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 approx. 0.5 standard deviations or one and a quarter grade years, respectively). However, this also applies to a somewhat lesser extent to subjects, such as science, that are not tested in central examinations. Our most preferred estimator 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 or displacement effects 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. *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, *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., 2002. *What explains international differences in student achievement?* 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., West, M. R., 2002. *Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS*, IZA Discussion Paper No. 485.

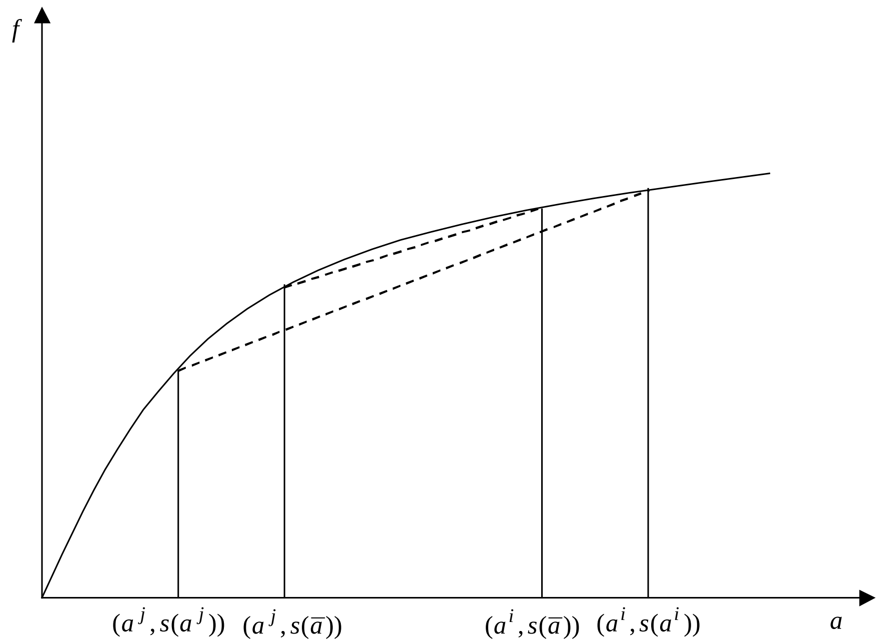



**Figure 1**: Education production with central and decentral standards

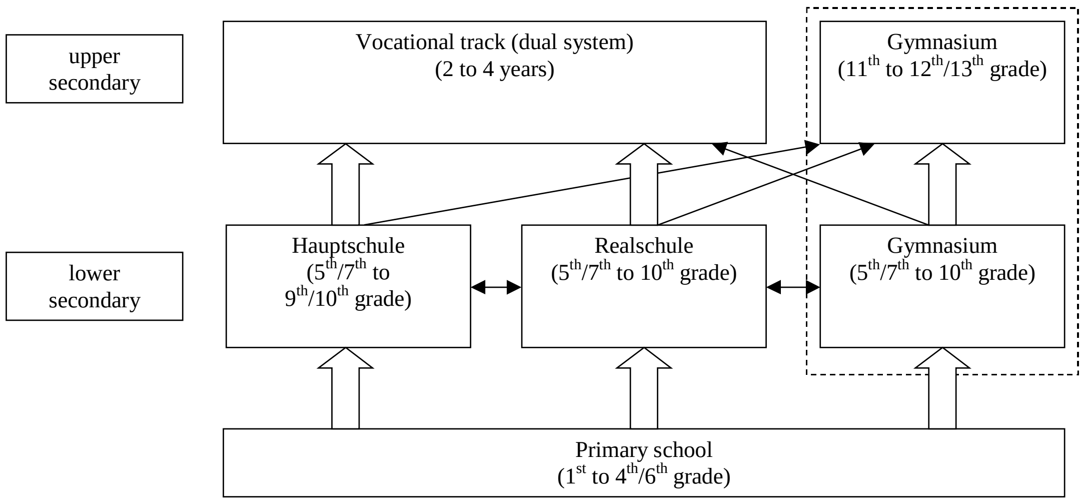



**Figure 2:** A stylised model of the German school system

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

| | *Hauptschule* degree | *Realschule* degree | *upper secondary leaving* degree (*Abitur*) | Students in 8th grade *Haupt-/ Realschule* | 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 |
| Sachsen (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.

**Table 2:** Descriptive Statistics (weighted)

| | non-CEE | | CEE | |
|---|---|---|---|---|
| Variable | 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 = *Realschule*) | 0.522 | | 0.541 | |
| Grade (1 = 8th grade) | 0.498 | | 0.478 | |
| Repeated class at least once | 0.359 | | 0.246 | |
| Does homework in maths | 0.877 | | 0.838 | |
| Does homework in science | 0.829 | | 0.784 | |
| Extra lessons maths | 0.149 | | 0.096 | |
| Extra lessons science | 0.077 | | 0.058 | |
| Usually good in maths: strongly agree | 0.400 | | 0.335 | |
| Usually good in biology: strongly agree | 0.315 | | 0.358 | |
| East Germany | 0.042 | | 0.455 | |
| Cumulative math lessons (in 1000s) | 1.198 | 0.121 | 1.292 | 0.105 |
| N obs. | 1834 | | 1363 | |

**Table 3:** Bivariate probit models of the likelihood to take extra lessons or study at home

| | extra lessons math | extra lessons science | studies math at home | studies science at home |
|---|---|---|---|---|
| CEE | -0.290 | -0.190 | -0.007 | -0.102 |
| | (3.69)** | (2.08)* | (0.08) | (1.21) |
| Usually good: strongly agree | -0.521 | -0.259 | 0.371 | 0.578 |
| | (4.68)** | (1.76) | (2.78)** | (4.86)** |
| Usually good: agree | -0.271 | -0.204 | 0.345 | 0.464 |
| | (2.58)** | (1.33) | (2.70)** | (3.71)** |
| Usually good: disagree | 0.052 | -0.143 | 0.213 | 0.417 |
| | (0.49) | (0.97) | (1.56) | (3.17)** |
| Repeated class | 0.092 | 0.163 | -0.249 | -0.191 |
| | (1.35) | (1.82) | (3.33)** | (2.48)* |
| 11-25 books | 0.056 | -0.166 | 0.506 | 0.458 |
| | (0.48) | (1.22) | (3.98)** | (4.13)** |
| 26-100 books | 0.177 | 0.020 | 0.608 | 0.493 |
| | (1.44) | (0.13) | (5.98)** | (5.39)** |
| 101-200 books | 0.212 | 0.243 | 0.603 | 0.539 |
| | (1.57) | (1.64) | (4.48)** | (5.07)** |
| 200+ books | 0.099 | -0.077 | 0.520 | 0.464 |
| | (0.80) | (0.53) | (4.19)** | (4.44)** |
| Immigrant child | 0.054 | 0.420 | 0.202 | 0.087 |
| | (0.55) | (4.38)** | (2.17)* | (0.96) |
| Girl | 0.040 | -0.046 | 0.322 | 0.238 |
| | (0.55) | (0.49) | (4.24)** | (3.45)** |
| Constant | -1.013 | -1.380 | 0.305 | 0.017 |
| | (6.85)** | (8.48)** | (1.81) | (0.12) |
| Rho | 0.817** | | 0.864** | |
| Observations | 2759 | | 2839 | |
| Diff-in-diffs | -0.100 | | 0.095 | |
| Wald-Chi2 (df=1) | 1.63 | | 2.18 | |

Robust z-statistics (accounting for clustering on the class level) in parentheses;* significant at 5% level; ** significant at 1% level

**Table 4:** Central exit examinations effects on student achievement in mathematics

| | simple diffs account. for composition effect[a] | diff-in-diffs by grade[a] | diff-in-diffs by school type[a] | diff-in diffs by subject I[b] | diff-in diffs by subject II[b,c] |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| CEE | 0.472 | 0.502 | 0.331 | 0.123 | 0.174 |
| | (6.78)** | (6.21)** | (3.87)** | (2.93)** | (3.87)** |
| Grade | 0.372 | 0.397 | 0.230 | -0.005 | -0.002 |
| | (4.27)** | (4.09)** | (3.67)** | (0.14) | (0.06) |
| Composition correction (α) | 2.131 | 2.124 | 1.327 | | |
| | (3.47)** | (3.44)** | (3.62)** | | |
| CEE*Grade | | -0.062 | | | |
| | | (0.55) | | | |
| CEE*School type | | | -0.020 | | |
| | | | (0.22) | | |
| *Realschule* | 0.715 | 0.716 | 0.585 | 0.138 | 0.116 |
| | (12.88)** | (12.84)** | (12.15)** | (3.27)** | (2.70)** |
| *Gymnasium* | | | 1.365 | | |
| | | | (20.91)** | | |
| 11-25 books | 0.129 | 0.130 | 0.116 | -0.116 | -0.122 |
| | (2.25)* | (2.26)* | (2.41)* | (2.10)* | (2.11)* |
| 26-100 books | 0.293 | 0.294 | 0.256 | -0.132 | -0.112 |
| | (4.82)** | (4.81)** | (5.46)** | (2.25)* | (1.85) |
| 101-200 books | 0.384 | 0.384 | 0.356 | -0.132 | -0.129 |
| | (5.97)** | (5.97)** | (7.01)** | (2.00)* | (1.87) |
| 200+ books | 0.446 | 0.446 | 0.422 | -0.160 | -0.137 |
| | (7.69)** | (7.69)** | (8.90)** | (2.72)** | (2.31)* |
| Immigrant child | -0.142 | -0.142 | -0.125 | 0.275 | 0.294 |
| | (2.47)* | (2.49)* | (2.86)** | (4.92)** | (5.17)** |
| Repeated class | -0.130 | -0.130 | -0.135 | -0.066 | -0.065 |
| | (3.32)** | (3.32)** | (4.60)** | (1.68) | (1.58) |
| Girl | -0.264 | -0.265 | -0.215 | 0.090 | 0.091 |
| | (7.00)** | (7.03)** | (7.54)** | (2.58)* | (2.57)* |
| East | -0.489 | -0.490 | -0.295 | -0.294 | -0.342 |
| | (4.63)** | (4.62)** | (4.35)** | (3.68)** | (4.16)** |
| Cum. math lessons | -0.410 | -0.403 | 0.153 | | |
| | (0.96) | (0.94) | (0.53) | | |
| Relative cum. math lessons | | | | 0.590 | 0.926 |
| | | | | (0.60) | (0.95) |
| Constant | -1.404 | -1.421 | -1.898 | -0.176 | -0.261 |
| | (3.10)** | (3.10)** | (5.38)** | (0.71) | (1.08) |
| Observations | 3197 | 3197 | 5020 | 2995 | 2727 |
| R-squared | 0.35 | 0.35 | 0.47 | 0.04 | 0.04 |

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-science score; [c] excluding Bavarian students who claim being good in science

## Appendix: TIMSS population 2 Science items involving math skills

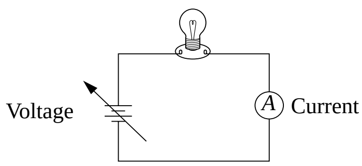
**L4** – Machine A and Machine B are each used to clear a field. The table shows how large an area each cleared in 1 hour and how much gasoline each used.

| | 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.

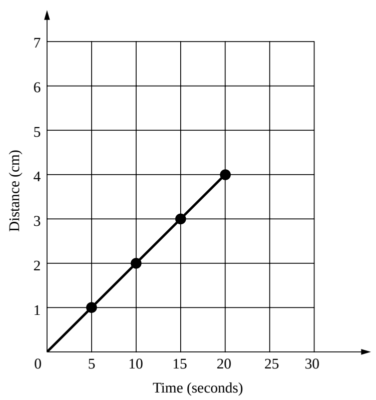



The table shows some of the results. Complete the table.

| 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.

---