

# Central exit examinations increase performance but take the fun out of mathematics

Hendrik Jürges, University of Mannheim\*  
Kerstin Schneider, University of Wuppertal\*\*

Preliminary Version, February 2007

**Abstract:** In response to PISA, all except one German federal state have decided to adopt central exit examinations at all secondary school levels. Theoretically, the advantages are fairly undisputed. CEEs make teaching and learning output observable and comparable across schools and thus raise teachers' and students' effort, so that average student performance increases. However, finding convincing empirical evidence for positive causal effects of CEEs is difficult. Using TIMSS 1995 data, we employ a difference-in-difference framework to analyze cross-state variations in mathematics performance and student and teacher attitudes and effort. We find that, while CEEs increases student performance, they have negative effects on student attitudes towards mathematics. Students who pass central exit examinations like mathematics less, find it less easy and are more like to find it boring than others. This is largely due to teachers increasing the learning pressure on students rather than employing more sophisticated or innovative teaching methods.

**Keywords:** German school system, high-stakes testing, student achievement, teacher quality  
**JEL-Codes:** I28, H42, D02

---

\* juerges@mea.uni-mannheim.de

\*\* schneider@wiwi.uni-wuppertal.de

## 1. Introduction

Since the publication of TIMSS and PISA test results, school reform has gained renewed interest in the German public. In particular the results of PISA have sparked intense political discussions about the need to reform the German school system. Part of the discussion has focused on insufficient financial resources flowing into the school system, as exemplified by repeated complaints about too large class sizes. Although increasing financial inputs into the education system will most likely raise outputs (measured e.g. as average student performance) to some degree, it has to be kept in mind that the education system operates under decreasing marginal returns. In a developed country like Germany it is at most unclear if marginal the returns are sufficiently high to warrant a general increase in the education budget. In fact, estimated effects of school resources on student achievement are often small and sometimes even inconsistent (as exemplified by the class size discussion). Increasing resources *alone* does not appear to be a very promising approach, especially when dealing with a broad target population (e.g. Hanushek, 1996, Hoxby, 2000).

An alternative to an input-oriented approach is to change the institutional setup of the school system (or the education system in general). From an economist's point of view, creating the right incentives for students and schools can increase the average performance with given financial inputs. Thus, changing the institutional setup appears to be a more cost-efficient. Although the short-term costs of changing the environment in which students, teachers, and schools act can be high, it is unlikely that new rules cost much once they are firmly established. But what do economists mean when they talk about creating the right incentives in schools? In general, they favor output-oriented governance of the public school system: provide clear definitions of the education goals, incentives for attainment of these goals, but give schools freedom of choice regarding appropriate means to reach these goals. The main thrust is to introduce more competition into the school system. Competition is assumed to be beneficial in the sense that it raises performance in good and bad schools alike. An important ingredient are standardized tests that are used to rank schools *publicly* according to their average performance. Public schools are exposed to additional competition from private schools by granting everyone access to the latter via school vouchers. If parents are informed about the performance of individual schools, and if parents' school choice is not subject to financial or institutional constraints, parents will send their children to the best available schools, thereby exerting competitive pressure on all schools.

In fact, German education policy has reacted to the "PISA shock" with what is sometimes termed a "paradigm shift" (see Kultusministerkonferenz 2005): the move from the old input-oriented to a more output-oriented governance. One key element of this new paradigm are national performance standards, which have become mandatory with school year 2005/2006, and which define expected competencies and performance levels for students at different ages and in different secondary school tracks (see Section 2 for a description of the German school system). The national performance standards replace the former content standards, also known as "education plans". A closely related issue that has received a great deal of attention in Germany is setting common standards by establishing central exit examinations (CEEs) throughout the country. This discussion is of particular interest in Germany because the federal states that already employ CEEs have generally outperformed non-CEE states in achievement tests. In fact, in response to TIMSS and PISA, all German federal states that have not had CEEs (with a single exception) have introduced or plan to introduce central exit examinations. Further, a group of seven German federal states have recently introduced regular standardized tests of student skills at different grades in primary and secondary schools (VERA).

The implicit assumption behind these policy changes is that states with CEEs (or standardized test in general) outperform non-CEE states purely because of the beneficial effects of central exit examinations and not because of some other, omitted variable at the federal state level. The theoretical literature almost unanimously shows that CEEs and central standards improve student performance and might even raise welfare (Costrell, 1997, Effinger and Polborn, 1999). Central exit examinations are purported to function better as incentives for students, teachers and schools than decentralized examinations (e.g. Bishop, 1997, 1999). Students, for example, benefit because the results of CEEs are more valuable signals on the job market than the results of non-central examinations, simply because the former are comparable. Furthermore, students who have to meet an external standard at the end of their school career have no incentive to establish a low-achievement cartel in class, possibly with the tacit consent of the teachers.

Estimating the effects of central exams is not straightforward, because it is typically decided on the country level whether to have or not to have central exams. Thus within countries there is hardly any variation in exam types which makes it difficult to estimate the effects of central

standards using national data.<sup>1</sup> Germany is an exception because, due to its federal structure, there has been a long-standing tradition of testing against external standards at the end of secondary schooling in some federal states and of having no standardized tests in others. Following our earlier studies (Jürges et al. 2005a, Jürges et al 2005b), we use the German federal education system as a source of exogenous variation to identify the causal effect of CEEs on teaching practices and teaching outcomes as measured in TIMSS 1995. One advantage of the TIMSS data is the availability of student and teacher information on teaching practices. Teaching practices are arguably very important inputs in the education production function. Hence, there is a broad consensus that academic achievement of students can be raised if the higher quality teaching practices are employed.

We focus on the effect of exit exams at the end of lower secondary education, when a "natural experiment" situation helps us to infer the relevant causal effect. In CEEs, students are generally examined in only one of the two subjects tested in TIMSS, namely mathematics. We calculate the between-state differences in the mathematics-science differences in test score, teaching practices (as perceived by students and teachers), student behavior and attitudes, as well as teaching careers and teacher attitudes. Under fairly weak identifying assumptions, these differences-in-differences can be interpreted as the causal effect of CEEs on outcomes.

The paper proceeds as follows: in Section 2 we describe the relevant features of the German school system. Section 3 gives a schematic overview of our conceptual framework of the learning process and Section 4 explains our identification strategy in detail. In Section 5 we give a brief description of the German TIMSS 1995 data and Section 6 contains the estimation results for differences-in-differences in a large number of education process outcomes. We draw some conclusions in Section 7.

## **2. Institutional background**

---

<sup>1</sup>Using international data, the effects could theoretically be estimated (Bishop 1997, 1999; Wößmann 2002) but the drawbacks are manifold (Jürges and Schneider, 2004; Jürges, Schneider and Büchel, 2005). Another example of a country with a federal structure and central exams in some but not all provinces is Canada (Bishop 1997).

We now give a concise description of the German school system, trying to emphasize those aspects that are most relevant to an understanding of central exit examinations in the German context.<sup>2</sup> Figure 1 gives a stylized overview of primary and secondary education in Germany.

All children in Germany attend primary school, which covers grades 1 to 4, or in some states grades 1 to 6. There is no formal exit examination at the end of primary schooling. Rather, students are generally allocated to one of the three secondary school types on the basis of their ability and performance in primary school. If the primary school considers a student suitable for a certain type of school, he or she will be admitted without any special admission procedure. If the primary school's recommendations conflict with the parents' wishes, however, the final decision about the future course of education lies either with the parents, the secondary school, or the school supervisory authority, depending on the laws of the state in question (see Jürges & Schneider, 2006, for an empirical analysis of biases in the secondary school track recommendation and allocation process).

--- Figure 1 about here ---

The *Hauptschule*, *Realschule* and *Gymnasium* are the three main types of secondary school; each leads to a specific leaving certificate. The *Hauptschule* provides its students with basic general education, and usually comprises grades 5 to 9 (or 10 in some states). The *Realschule* provides a more extensive general education, usually comprising grades 5 to 10. The *Gymnasium* provides an in-depth general education covering both lower and upper secondary level, and usually comprises grades 5 to 13 (or 12 in some former GDR states). Depending on their academic performance, students can switch between school types.<sup>3</sup>

At the end of lower secondary level, *Hauptschule* and *Realschule* students who complete grade 9 or 10 successfully are awarded a leaving certificate. They are only required to take central exit examinations in some states (Table 1 describes the situation in 1995, the year the TIMSS data were collected). Central exit examinations are most common at the end of upper secondary education. The other states had decentralised systems, in which teachers devise their own questions for the exit examinations, subject to the approval of the school supervisory authority. Six states had central exit examinations at the end of *Realschule*, and

---

<sup>2</sup> A detailed description of the German school system can be found in Jonen and Eckardt (2004).

<sup>3</sup> A fourth type of school, the *Gesamtschule* (comprehensive school), does not appear in our figures. This type of secondary school offers all lower secondary level leaving certificates, as well as providing upper secondary

only four had them at the end of *Hauptschule*.<sup>4</sup> *Gymnasium* students are not issued a leaving certificate after completing lower secondary level, but are admitted to the upper level of the *Gymnasium*. Students leaving *Hauptschule* and *Realschule* usually embark on vocational training in the "dual" system, so called because it combines part-time education in a vocational school with on-the-job training with a private or public sector employer.

--- Table 1 about here ---

German exit examinations never cover all of the subjects taught at school. At the *Abitur* level, students can choose three or four subjects (the choice is limited that and varies from state to state). This leads to self-selection problems, which are unlikely to be solved convincingly with the available TIMSS data. At *Hauptschule* and *Realschule*, German and mathematics are always tested in the exit examinations, i.e., mathematics is compulsory for all students in these two school types taking exit examinations. In order to assess the causal effect of CEEs, we will thus concentrate on mathematics performance, teaching practices, student behavior and attitudes in *Hauptschule* and *Realschule* as the main outcome variables to be affected by CEEs.

### 3. Conceptual background

A stylized conceptual model of the education process underlying our study is shown in Figure 2. Student achievement is typically viewed as the main outcome of the education process and education policy is often evaluated based on this outcome only. What is often missing is an analysis of how education policy and institutions are affecting the process of teaching and learning. In the case of CEEs, the earlier empirical literature has mainly analyzed the effect of CEEs on student achievement and has tried to identify the causal effect of external exams. The theoretical models, however, also consider the channels through which CEEs work in more detail. For instance, CEEs are thought to affect students and teachers effort and thereby improve student achievement. But clearly, raising effort is being costly for students and teachers, as more effort negatively affects utility. This could result in a more negative attitude

---

education. It only plays a minor role in most federal states, however, with less than 10 percent of all students in grade 8 attending a comprehensive school.

<sup>4</sup> As mentioned in the introduction, CEEs have been introduced or will be introduced in Saarland (2001), Hamburg (2005), Brandenburg (2005), Hesse (2006), Lower Saxony (2006), Berlin (2007), North Rhine-Westphalia (2007), Bremen (2007), Schleswig-Holstein (2008).

of students towards school. However, a better knowledge of for instance mathematics might well increase the student's interest in mathematics and result in a more positive attitude. And even teachers might find it more enjoyable to teach more motivated students. If this is the case, CEEs promise to grant a "free lunch". In this paper, we try to find out how CEEs affect effort, motivation and attitudes of students and teachers, and hope to shed more light on the costs and benefits of CEEs.

--- Figure 2 about here ---

#### 4. Identification

The most basic approach to identify the effect of CEEs on any outcome would be to estimate *simple differences* between average outcomes in CEE states and non-CEE states, controlling for student background and other variables of interest. Simple differences in outcomes across CEE and non-CEE states are of limited value, however, because they ignore two potentially confounding effects: a composition effect and endogeneity of CEEs. The first problem, the composition effect, stems from the fact that in CEE states more students attend *Haupt-* and *Realschule* and fewer students attend *Gymnasium* than in non-CEE states. Since students are selected into secondary schools mainly on the basis of their achievement in primary school, student achievement in CEE states (conditional on school type) will be higher simply because there are, 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 compositions of the student body in German secondary schools across states are interpreted as the result of different ability cutpoints  $\alpha$  chosen to sort 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<sup>th</sup> grade students aspiring to a high school diploma (see Table 1, last column).

Besides difficulties stemming from a composition effect, the attempt to estimate the effect of CEE is subject to the fundamental problem of causal inference, namely that it is impossible to observe the individual treatment effect (Holland, 1986). One cannot observe the same teacher or student at the same time as being teacher or student in a state with and without CEE. Only if selection into treatment is random, this is no problem. However, self-selection into treatment is one of the most frequent problems encountered by researchers trying to evaluate the causal effects of policy measures. In our context, this can happen if parents vote with their

feet and move to another state in order to send their children to schools with a central exit examination (or to avoid CEEs). Parents in non-CEE states who live near a CEE-state may choose to send their children to school in the neighboring state. However, this will not apply to many parents. In the short run, the treatment status might be considered exogenous, given the institutional arrangement in each state. In the long run, however, institutions can change and that would affect all parents. But clearly, not only parents can vote with their feet; teachers might well be more mobile than parents when deciding where to work. However, the between state-mobility of teachers, who are mostly state civil servants, is rather limited. As an example, consider the mobility between Bavaria (one of the large southern CEE-states) and the rest of Germany (see Table 2). In 2001 a total of 102 teachers applied to be transferred from a non-Bavarian school to a Bavarian school. Only 22 teachers were granted the transfer. The number of teachers from Bavaria who applied to be transferred to another German state was with 38 even smaller. Moreover, the observed mobility has been mainly between Bavaria and neighboring Baden-Württemberg, which is another large CEE state.

--- Table 2 about here ---

Even if mobility of parents and teachers is low, the existence of CEEs might reflect unobserved variables such as the importance attached to education by the electorate of a particular state, i.e., parental attitudes towards education and achievement in school. If CEEs are correlated with such attitudes, simple differences between CEE and non-CEE states will be a biased measure of the CEE effect.

Our strategy to isolate CEE effects from differential parental attitudes and other unobserved variables draws on variation within states. As explained above, the fact that CEEs only apply to a narrow range of subjects offers a unique source of exogenous variation that can be used to identify the causal effect of CEEs. When mathematics is a CEE subject but science is not and if CEEs have a causal effect, the observed outcome differences should be larger in mathematics than in science.

Formally, our estimator can be described as follows (for simplicity, let us assume for a moment that all outcomes are measured continuously). Consider two regressions: one to explain outcomes related to mathematics  $y_i^m$  and measured at the student level (the index  $i$  denotes the student)

$$y_i^m = \mu_i + X_i\beta + C_i\delta + \varepsilon_i^m, \quad (1)$$

and another to explain outcomes related to science  $y_i^s$

$$y_i^s = \mu_i + X_i\gamma + \varepsilon_i^s, \quad (2)$$

where  $\mu_i$  is some student-specific characteristic (e.g. general ability),  $X_i$  is a vector of covariates that might affect mathematics and science outcomes differently,  $C_i$  is a dummy variable for central exams in mathematics, and  $\varepsilon_i^k$ ,  $k = m, s$  are i.i.d. error terms. Subtracting (2) from (1) yields

$$d_i = y_i^m - y_i^s = X_i(\beta - \gamma) + C_i\delta + (\varepsilon_i^m - \varepsilon_i^s), \quad (3)$$

where  $\delta$  is the parameter of interest. The main advantage of this estimator is that each student serves as his or her own control group. By taking differences,  $\mu_i$  is swept out of the regression, so we are able to control for a lot of heterogeneity on the individual level, such as, for instance, general ability, general attitudes towards learning and academic success, socio-economic background. In order for  $\delta$  to identify the causal effect of CEEs on outcomes, we need identifying assumptions, specifically  $E[C_i(\varepsilon_i^m - \varepsilon_i^s)] = 0$ . There are several ways in which this assumption might be violated, depending on the outcome under study. For instance, in the case of student test scores, there could be systematic indirect effects in the form of spill-over from mathematics (more general skills) to science (more specific knowledge and skills). Negative spill-over from mathematics to science is also conceivable if students divert resources away from learning science to learning mathematics because the latter is tested against an external, and possibly higher, standard. If mathematics teachers also teach science, spill-over can be thought of as teachers transferring more successful teaching strategies from one subject to another. In the analysis of test scores and student attitudes, the above assumption can be violated if CEE and non-CEE states differ systematically in their relative preference for mathematics rather than science. Also, unobserved student background (eg. innate mathematics and science skills) must not differ between federal states. Usually, one can plausibly assume that such characteristics are equally distributed across German regions. But as was mentioned in the discussion of the *composition effect*, we use selective sub-samples of the student population. Mathematics skills may be more important than science skills when it comes to allocating students to secondary school types. If the *Gymnasium* skims off the students with the best mathematics skills (and mathematics ability is not perfectly correlated with science ability), students in states with a high proportion of students in *Hauptschule* and *Realschule* (high  $\alpha$ , see above) may have better mathematics

skills than their peers in low- $\alpha$  states, but comparable science skills. Finally, it is also important that mathematics and science outcomes be comparable. This could indeed become a problem when the outcome variables are measured on an ordinal scale, such as (1) strongly disagree (2) disagree (3) agree (4) strongly agree, and when there are systematic differences in the way students from CEE- and non-CEE-states use such scales. This problem (differential item functioning or DIF) and the way we deal with it in this paper is discussed in detail below.

Elsewhere, we give a detailed discussion of the plausibility of these assumptions with respect to student achievement as outcome variable (Jürges, Schneider & Büchel, 2005). In brief, we believe that spill-over from good mathematics skills to good performance in the TIMSS science test is likely to be very small, because of the 87 (released) science items, only four require mathematics skills, such as dividing by a fraction. Negative spillover is likely, so that strictly speaking, we are only able to measure the size effect of a *partial* introduction of CEEs (that includes the effect of students to divert time away from non-tested to tested subjects). Sizeable spillover on the teacher level is probably less of a problem. Less than 15 percent of the teachers teach both mathematics and biology. Relative preferences for mathematics versus science are most likely to be very similar in CEE and non-CEE states. Mathematics are 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, there are no significant differences in *relative* teaching time between CEE and non-CEE states (Frenck 2001). Finally, we can account for the possibility of relative composition or selection effects by controlling for  $\alpha$  in our difference-in-differences framework.

Due to the limited number of observations of teachers in the sample who teach mathematics and science, the proposed identification strategy described in (3) cannot be applied directly to the information from the teacher questionnaire. The number of teachers in the sample with answers on the questions on teaching practices etc. is due to high non-response of teachers only between 202 and 212. Only about 15% of those teachers are sampled as mathematics and science teachers. Thus we cannot continue with a difference-in-difference analysis on the individual teacher level. However, we can differentiate between the groups of mathematics and science teachers. Our analysis of the teacher information (the index  $j$  denotes the teacher) will be based on

$$y_j = \mu + C_j \delta_1 + M_j \delta_2 + M_j \times C_j \delta_3 + X_{ij} \beta + \varepsilon_j \quad (4)$$

As before,  $X_{ij}$  controls for the covariates of the student  $i$  in teacher  $j$ 's class.  $C_j$  is a dummy variable for central exams in mathematics, and  $\varepsilon_j$  are i.i.d. error terms.  $M_j$  is a dummy variable for a mathematics teacher, and the interaction term  $C_j \times M_j$  is the product of the two dummy variables. In (4),  $\delta_3$  is the parameter of interest.

## 5. Data description

The international data set of TIMSS Germany contains data on a total of 5763 7<sup>th</sup> and 8<sup>th</sup> grade students and 566 teachers in 137 schools, collected in the 1994/95 school year. Data were collected in 14 of the 16 German states (Baden-Württemberg and Bremen did not participate), and from all major types of secondary schools. However, for reasons explained above, we consider only the *Haupt-* and *Realschule* data. In addition to the actual test results in mathematics and science, the TIMSS data contain a wide range of context variables on student backgrounds and attitudes, as well as on teachers and schools.

Despite the wealth of data available, we take 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 3 contains variable definitions and descriptive statistics, by the type of exit examination, for these variables. Student background, measured in terms of the number of books at home, differs only slightly by exit examination type – the proportion of students within each range is very similar in CEE and non-CEE states. There are far more students with an immigrant background in the non-CEE group than in the CEE group. This is largely attributable to the relatively low rates of immigration to East Germany, where most states have central exit examinations (a legacy of the former GDR education system). Another major difference between students in CEE and non-CEE states is that in the latter, a larger proportion of students have repeated class at least once.

--- Table 3 about here ---

Table 4 contains variable descriptive statistics for our dependent variables. Exact definitions and operationalizations are shown in Table A1 in the Appendix. The most notable difference between students in states with and without CEEs is their achievement in mathematics and science (scores were standardized to have a mean of 0 and a variance of 1, differences can thus be interpreted in terms of standard deviations). In mathematics, students in states with CEEs score on average nearly 0.6 standard deviations higher than those in states without

CEEs. In science the difference is somewhat less than 0.5 standard deviations. In both types of states, roughly three quarters of the students agree or agree strongly to the statement that they are usually good in mathematics or biology.

--- Table 4 about here ---

There are a number of statistically significant differences between CEE and non-CEE states in teaching practices – as reported by the students. For instance, in mathematics it appears that CEE students more often copy notes from the board but less often work from textbooks or worksheets on their own. Teachers also appear to give homework less often but homework is more often checked. Overall, however, the percentage differences are relatively small. In biology, the differences are much larger, in particular with respect to giving, checking, and discussing homework. While about half of the students in non-CEE states say that teachers give, check, and discuss homework pretty often or always, 25 to 37 percent of the students in CEE states do so. Such large differences shed some doubt on the cross-state comparability of the ordinal response scales such as the one used for these question. It rather seems as if there is some sort of differential item functioning at work, i.e. "pretty often" may mean different things in absolute terms depending on whether a student lives in a CEE or a non-CEE state.

Another noteworthy difference between students in CEE and non-CEE states is their attitude towards mathematics. CEE students are consistently less likely to like or enjoy mathematics, or to find it an easy subject. On the other hand, they are more likely to find it boring. Differences with respect to biology are smaller and less often statistically significant. Again, differential item functioning might be an issue here, but we believe that for our difference-in-difference analysis of *causal effects* of CEEs, this is less of a problem as it might seem at first. Since we use intra-student variation in ordinal judgements, the only measurement assumption we make is that of response consistency, i.e. students use the same response categories in the same way, independent of the subject they refer to (mathematics or biology/science). Students may have differential item functioning in the way they use answer categories. For example, our procedure allows that "once in while" means the same frequency to one student as "pretty often" to another. Individual students may even have DIF across different questions. For instance, "pretty often" may mean a different frequency when used with the statement "We have a quiz or test" rather than with "The teacher gives us homework". We only require that "pretty often" means the same when used for questions related to mathematics classes and questions related to biology classes.

The teacher variables show less significant differences between CEE and non-CEE-states than suggested by the student data. While teachers in CEE-states report less often that they would change their career if they had a chance to do so, even though they don't feel as appreciated by society as their colleagues in non-CEE states, the differences are not statistically significant. A significant difference can be found in the assessment of teachers regarding the importance of the natural talent one has for mathematics and science. Teachers in CEE-states believe more often that this is the case. However, unlike the students in CEE-states who do not consider themselves good math students, teachers think that their mathematics students perform better than other students in Germany. For science this appears not to be true. Science teachers in CEE-states think that only 11% of their students are in top third, whereas 15% of the teachers consider their students to be good in science. This is an interesting result, as TIMSS in 1995 has been the first student assessment in Germany that allows comparing achievement of students across states. As expected, teachers in CEE-states are more familiar with the curriculum in math and tend to rely more heavily on external exams when planning their lessons. The difference is not significant for science. Finally note that students in CEE-states spend significantly more time in math lectures. The difference in weekly minutes in science is small and insignificant.

## **6. Regression Results**

Regression results are shown in Table 5 below. We only report the coefficients for the CEE-dummy variable, which measure the effect of CEEs on various dimensions of student achievement, teacher and student behavior and student attitudes. In other words, Table 5 shows the results of 24 regressions with different dependent variables but the same explanatory variables. Besides CEE, we use the explanatory variables described in Table 2 above: the number of books in the student's home, student's sex, grade, and immigration background, whether a student repeated class, region (East/West Germany), type of school (*Realschule/Hauptschule*), Alpha, the variable that reflects the selectivity of the student body in *Real/Hauptschule* in the respective federal state. With the exception of class behavior, each dependent variable measures achievement, behavior and attitudes in mathematics relative to biology (or in some instances, science in general). Thus the CEE coefficient identifies differences-in-differences, as explained above.

Before actually going through our results, a note on the interpretation of the effects shown for multinomial logit models might be necessary. All models are 3-category models. The values shown in Table 5 are relative risks and their standard errors (computed via the delta method). Significance levels are based on t-tests using the original logit coefficients, however.

The three categories of the dependent variable are defined in the same generic way:

- (1) mathematics value "smaller" than biology value
- (2) mathematics value "equal to" biology value
- (3) mathematics value "larger" than biology value

Take the variable GOOD as an example. The original items read: "I usually do well in mathematics" and "I usually do well in biology", respectively. Students were asked to state whether they "strongly disagree", "disagree", "agree", or "strongly agree". A student who "disagreed" to both statements is assigned to the second category, a student who "disagreed" to the first item but "agreed" to the second is assigned to the first category, and a student who "strongly agrees" to the first statement but "disagrees" with the second is assigned to the third category, and so on.

Thus for the variable GOOD the three categories mean:

- (1) the student agrees to a lesser extent to the statement that she does well in mathematics than to the statement that she does well in biology. In brief: she thinks she does worse in mathematics than in biology.
- (2) the student thinks is as good in mathematics as in biology.
- (3) the student thinks she does better in mathematics as in biology.

In the multinomial regressions, the middle category is always the baseline category. The relative risks shown in Table 5 thus mirror the effect of central exit examinations on the probability of thinking one does worse in mathematics than in biology relative to the probability of thinking that one does about equally well in mathematics than in biology. To illustrate, consider the the bivariate relationship between central exit exams and the relative self-evaluation in mathematics versus biology shown in Table 5.

--- about here Table 5 ---

The relative risk of those in CEE-states to judge themselves worse in math than in biology can be computed as:

$$RR_1 = \frac{P(Y = 1 | CEE = 1)}{P(Y = 2 | CEE = 1)} \bigg/ \frac{P(Y = 1 | CEE = 0)}{P(Y = 2 | CEE = 0)} = \frac{32.45}{41.16} \bigg/ \frac{25.62}{41.95} \approx 1.291$$

A multinomial regression of GOOD on CEE *without covariates* would yield exactly the same result. Since  $RR_1$  is larger than one this means that students in CEE-states have a higher relative risk of thinking they do worse in mathematics than in biology than non-CEE students.  $RR_2$  equals 0.829, i.e. students in CEE-states have a lower relative risk of thinking they do better in mathematics than in biology than non-CEE students. As a shorthand, we will often say that students in CEE states are less likely to think they do well in mathematics than students in non-CEE states, bearing in mind that this need not be true in absolute terms but relative to biology or science in general. Although the difference-in-differences risk ratios are admittedly hard to interpret, they have the advantage that we can assume that all ordinal variables are measured on the same scale, i.e. we think that under fairly weak assumptions we have no differential item functioning problem.

### 6.1. Objective and self-perceived student achievement

We now discuss our results one by one, starting with student achievement. TIMSS test scores were rescaled to have mean zero and a standard deviation of one. Thus the difference in mathematics scores between CEE and non-CEE states is 0.11 standard deviations larger than the same difference in science scores. CEE state students thus do relatively better than non-CEE state students, which clearly indicates that there is some causal effect of CEEs on achievement. Comparing this to the one grade year differences in mathematics scores of 0.28 shows that the effect amounts to a little more than one third of a school year.

Returning to the students self-assessment shows that controlling for covariates does only slightly change the results presented above. Thus despite the fact that the relative mathematics performance of students in CEE-states is superior to that of their peers in non-CEE states, students themselves appear to think the opposite. One explanation for this seemingly contradictory finding could be that relative expectations are higher in CEE states, for example because teachers put more pressure on their students to perform well, knowing that the centralised exams are due in only one or two years.

--- about here Table 6 ---

## *6.2. Teaching practice*

We restrict the discussion on the items in which we find statistically significant differences-in-differences. The variable NOTE indicates how often students copy notes from the board. The odds ratio smaller than one in column 4 shows that this is a less common practice in non CEE-states' mathematics lessons than in CEE-states' mathematics lessons. Giving homework, checking homework, and discussing homework in mathematics classes are practices that are more common in CEE states. Finally, CEE teachers less often introduce a new topic by trying to solve an example related to the new topic.

## *6.2. Class behavior (student discipline)*

The results of the ordered logit regressions do not suggest that there are any systematic differences between CEE and non-CEE states in student discipline in class. Note however that the items concerning the behavior of the students in class can only be analysed in terms of simple differences between CEE and non-CEE states. This is because there are no corresponding items for science or biology classes in the data. Differential item functioning across the two types of states could perhaps mask any factual differences in discipline.

## *6.3 Student effort and motivation*

The variable DAY indicates how much time students spend outside school learning mathematics and science (subjects are not asked for separately in this question). Our results indicate that students in CEE-states spend relatively more time learning mathematics at home than their peers in non-CEE states. Two variables that aim at capturing the general motivation for learning mathematics and science are LIFE and JOB. LIFE indicates how much students agree to the statement that mathematics / biology is important in everyone's life. JOB indicates whether students would like a job that involves mathematics / biology. Here we only find weak and/or inconstituent relationships with the presence of central exit examinations. Students in CEE-states have a slightly lower chance to think that mathematics is importance in everyone's life but the relationship is not significant. There is also a higher probability of

students in CEE-states to want to get a job that involves mathematics rather than biology, but they are also more likely to want to have it the other way round.

#### *6.4. Student attitudes*

The final set of items measures the difference in individual attitudes towards mathematics and science. Here we find strong and consistent evidence for causal effects of central exit examinations. However, the direction of these effects is unexpected. Students in CEE-states are consistently less likely to like mathematics, to enjoy doing mathematics, and to find that mathematics is an easy subject. They are also more likely to find mathematics boring. Thus despite the better performance, CEE state students have much worse attitudes towards mathematics.

#### *6.5 Teaching career and teacher attitudes*

Table 7 summarizes the results on the teacher variables. Teachers in CEE-states generally feel more often to be appreciated by their students. Mathematics teachers in CEE- and non-CEE-states feel more appreciated by society than science teachers but less so in CEE states. Thus CEEs might not result in an upgrading of the value of teaching. This is an interesting result, because in the past the teaching profession in Germany has been suffering from diminishing appraisal, which is a problem for recruiting new teachers. In CEE-states, a natural talent of students is generally viewed to be more important for good performance but talent is viewed to be less important for mathematics. This finding can be interpreted as the teacher's attempt to justify poor student performance due to increased pressure to perform in CEE-states. On the other hand, teachers in CEE-states generally report to have fewer students who are in the bottom third of the distribution. In mathematics but not in science CEE-students are, according to the assessment of their teachers, more likely to be in the top third of the achievement distribution.

--- about here Table 7 ---

#### *6.6 Teaching Practice*

Judging from the regression results, CEEs appear to have only a limited effect on teaching practice. The only differences seem to be that teachers in CEE-states less often tend to correct students in front of the class and more often use assessment information in CEE-subjects to diagnose student's learning problems. Interestingly, teachers do not confirm the student's impression that teachers assign more homework, even though the (insignificant) coefficient on the interaction term indicates the positive effect of CEEs on the amount of homework assigned. The same is true for the variable HOMER. Teachers in CEE states appear to be more familiar with the national curriculum and rely on textbooks to a larger extent. However, we get this effect for mathematics and science as well. Moreover, external exams play a significant role for planning mathematics lessons in CEE-states, which was to be expected.

In summary, the teacher results are not strong enough to allow for strong conclusions. Maybe due to the small number of observations, we find only few significant results. However, the findings from the student responses indicate that mathematics is less fun and more work in CEE states while the achievement levels are higher. This finding is to some extent supported by the analysis of the teacher variables.

## **7. Conclusion**

This paper discusses the benefits of central exit examinations (CEEs) for academic achievement in lower secondary education. The theoretical literature almost exclusively sees the benefits of central examination. However, identifying the causal effect of CEEs empirically is by no means easy. Caution is warranted when interpreting observed differences between states with or without CEEs as the effect of CEEs on student achievement, because CEEs are the outcome of a political process (reflecting the preferences of the electorate) and thus likely to be endogenous.

Unlike earlier studies, we make use of regional institutional variation in Germany, which allows us to develop a difference-in-differences identification strategy to estimate the causal effect of CEEs on academic performance, teaching practices, and student attitudes: in the Germany school system, only some states have CEEs and these exams are restricted to core subjects German and mathematics. We use data from the Third International Mathematics and Science Study (TIMSS) to exploit this institutional variation and uncover the causal effect of CEEs on student achievement in mathematics by comparing outcomes across subjects and types of exit examination. The fundamental idea is that a CEE affects only mathematics-related outcomes but not science (e.g. biology) related outcomes.

Our study has the following findings: CEEs have a small but statistically significant effect on student test scores. Teachers in CEE-states are more likely to give, check and discuss homework. Students in CEE-states do like mathematics less, find it less easy and find it boring than those in non-CEE states. They are also somewhat more diligent learning mathematics at home.

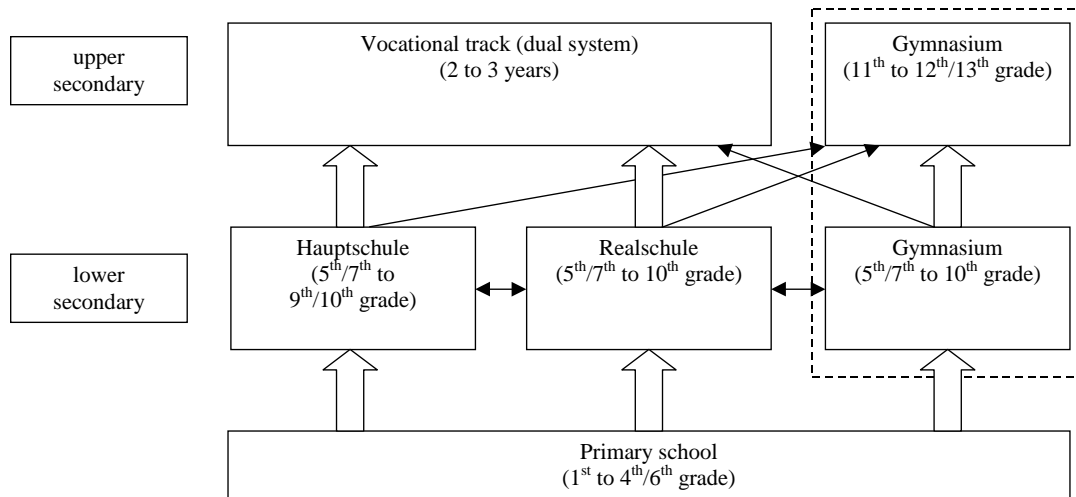
We find only little difference in (student-reported) teaching practices other than those that are homework-related, little difference in student behavior in class, and little difference in general student motivation to learn mathematics.

Broadly speaking, this evidence is consistent with the view that that the main effect of CEEs is that teachers increase the pressure exerted on students rather than employing more sophisticated or innovative teaching methods. Surely giving, checking, and discussing homework involves some effort on the part of the teacher, but on the whole, achievement gains in mathematics are simply realized by increased effort on the part of the students. One perhaps unintended consequence is that students in CEE states less often find that mathematics is fun to do and teachers more often think that learning mathematics and science depends on a natural talent.

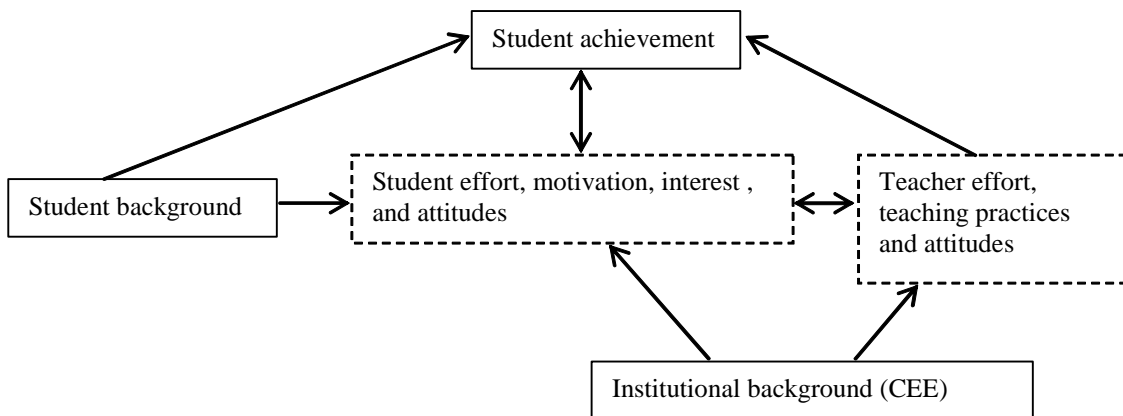
## References:

- Allen, N.L. Donoghue, J.R., Schoeps, T.L., 2001, The NAEP 1998 Technical Report. Washington, DC: National Center for Education Statistics.
- Armstrong, M., Cowan, S., Vickers, J., 1994, Regulatory Reform, Economic Analysis and British Experience, MIT Press.
- Adams, R. /Wu, M., 2002, PISA 2000 Technical Report. OECD: Paris.
- Angrist J., J. Guryan, 2003, "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements", NBER Working Paper 9545.
- Baumert, Jürgen., Cordula Artelt, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Klaus-Jürgen Tillmann and Manfred. Weiß (2002). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Leske+Budrich.
- Baumert, Jürgen und Rainer Watermann (2000). "Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe." In *TIMSS - Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*, edited by Jürgen Baumert, Wilfried Bos and Rainer Lehmann. Leske+Budrich.
- Bishop, John H. (1997). "The effect of national standards and curriculum-based exams on achievement." *American Economic Review*, 87, 260-264.
- Bishop, John H. (1999). "Are national exit examinations important for educational efficiency?" *Swedish Economic Policy Review*, 6, 349-401.
- Blundell R, Costa Dias M., 2000, "Evaluation Methods for Non-Experimental Data." *Fiscal Studies* 21, 427-268.
- Costrell, Robert M. (1997). "Can educational standards raise welfare?" *Journal of Public Economics*, 65, 271-293.
- Effinger Matthias R. and Mattias K. Polborn (1999). "A model of vertically differentiated education." *Journal of Economics*, 69, 53-69.
- Frenck, Isabella (2001). "Studentafeln der Primar- und Sekundarstufe I im Ländervergleich - eine empirische Studie am Beispiel der Fächer Deutsch und Mathematik." Unpublished master thesis, University of Essen.
- Glewwe, Paul, Nauman Elias and Michael Kremer (2003). "Teacher Incentives", NBER Working Paper 9671.
- Hanushek, Eric A. (1996). "School Resources and Student Performance." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Gary Burtless. Brookings Institution.
- Hanushek, E.A., Kain, J.F., Rivkin, S.G., 2004, "Why Public Schools Lose Teachers", *Journal of Human Resources* 39, 326-354.
- Hanushek, E.A., Rivkin, S.G., 2004, "How to Improve the Supply of High Quality Teachers", *Brookings Papers on Education Policy*, 7-25.
- Holland, P.W., 1986, "Statistics and Causal Inference." *J Am Stat Assoc* 81; 945-960.
- Heckman J.J., Ichimura H., Todd P., 1998, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Rev Econ Stud* 64: 605-654.
- Hoxby, Caroline M. (2000). "The effects of class size on student achievement: New evidence from population variation." *Quarterly Journal of Economics*, 115, 1239-1285.
- Hoxby, Caroline M. (2003). "School Choice and School Productivity (Or, Could School Choice be a Rising Tide that Lifts All Boats)." In *The Economics of School Choice*, edited by Caroline M. Hoxby. University of Chicago Press.
- 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, Gerdi and Thomas Eckardt (2006). "The Education System in the Federal Republic of Germany 2004." Secretariat of the Standing Conference of the Ministers of Education (KMK).
- Cultural Affairs of the Länder in the Federal Republic of Germany. [http://www.kmk.org/dossier/dossier\\_en\\_ebook.pdf](http://www.kmk.org/dossier/dossier_en_ebook.pdf) [2007, Jan 12].

- Jürges, Hendrik and Kerstin Schneider (2004). "International differences in student achievement: An economic perspective." *German Economic Review*, 5, 357-380.
- Jürges, Hendrik and Kerstin Schneider (2006). "Age at school entry and teacher's recommendations for secondary school track choice in Germany" Paper presented at the 20<sup>th</sup> ESPE conference 2006 in Verona.
- Jürges, Hendrik and Kerstin Schneider (2007). "Fair Rankings of Teachers." *Empirical Economics*, forthcoming.
- Jürges, Hendrik, Wolfram F. Richter and Kerstin Schneider (2005). "Teacher quality and incentives, Theoretical and empirical effects of standards on teacher quality", *FinanzArchiv* 61 (3) 2005, 298-326
- Jürges, Hendrik, Kerstin Schneider and Felix Büchel (2005). "The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany." *Journal of the European Economic Association* 3 (5) 2005, 1134-1155.
- Krueger, Alan B. (1999). "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114, 497-532.
- Krueger, Alan B. (2003). "Economic Considerations and Class Size." *Economic Journal*, 113, F34-F63.
- Krueger, Alan B. and Pei Zhu (2004). "Another Look at the New York City School Voucher Experiment." *American Behavioral Scientist*, 47: 658-698.
- Kultusministerkonferenz (ed.) (2005): *Bildungsstandards der Kultusministerkonferenz Erläuterungen zur Konzeption und Entwicklung*. München/Neuwied: Luchterhand.
- Lavy, Victor (2002). "Evaluating the Effect of Teachers' group performance incentives on pupil achievement." *Journal of Political Economy*, 110, 1286-1317.
- Lavy, Victor (2003). "Paying for performance: The effect of individual financial incentives on teachers' productivity and students' scholastic outcome." Working paper, Hebrew University of Jerusalem.
- Lavy, V., 2004, "Performance Pay and Teachers' Effort, Productivity and Grading Ethics", NBER Working Paper 10622.
- Lazear, E. P., 2004, "Speeding, tax fraud, and teaching to the test." NBER Working Paper 10932.
- Mislevy, R. J. 1991, "Randomization-based inference about latent variables from complex samples." *Psychometrika*. 56, 177-196.
- Mullis, Ina V.S, Michael O. Martin, Edward G. Fierros, Amie L. Goldberg, , and Steven E. Stemler, (2000). "Gender Differences in Achievement." Working paper, Lynch School of Education, Boston College.
- Rosenbaum, P.R. and D.B. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41-55.
- Wößmann, Ludger (2002). "Central Exams Improve Educational Performance: International Evidence." Kiel Discussion Papers 397, Kiel Institute for World Economics.



**Figure 1: The German school system**



**Figure 2: Conceptual background**

## Tables:

**Table 1:** CEE by federal state and type of certificate (as of 1995); proportion of students by school/type of certificate (in 1999)<sup>a</sup>

	<i>Hauptschul e</i>	<i>Realschule</i>	<i>Gymnasiu m (Abitur)</i>	8 <sup>th</sup> grade students in <i>Hauptschul e</i> tracks <sup>c</sup>	8 <sup>th</sup> grade students in <i>Realschule</i> tracks <sup>d</sup>	8 <sup>th</sup> grade students in <i>Abitur</i> tracks <sup>e</sup>
Baden-Württemberg	G/M/F/O	G/M/F	A	34.0	31.7	28.4
Bavaria	G/M/F/S <sup>b</sup> / O	G/M/F/S <sup>b</sup> / O	A	37.1	31.0	26.9
Berlin				21.5	31.1	42.8
Brandenburg				16.9	32.3	45.2
Bremen				26.8	30.9	36.1
Hamburg				22.6	26.2	43.3
Hesse				22.5	34.0	38.6
Lower Saxony				31.3	34.9	28.0
Mecklenburg-West Pomerania		G/M/F	A	15.7	48.9	29.3
North Rhine-Westphalia				29.3	30.6	35.3
Rhineland-Palatinate				37.2	28.4	29.5
Saarland			A	24.4	37.0	34.0
Saxony	G/M/F	G/M/S	A	12.8	51.0	30.1
Saxony-Anhalt		G/M	A	9.4	50.9	32.8
Schleswig-Holstein				31.4	34.4	28.5
Thuringia	G/M	G/M/F	A	18.3	42.3	32.6

Notes:

G = German; M = Mathematics; F = Foreign Language (mostly English); S = Science; O = Other; A = Any subject chosen for the written exams

<sup>a</sup> Percentages add up to less than 100. Students in special schools (e.g. for slow learners) are not listed.

<sup>b</sup> Optional subject.

<sup>c</sup> *Hauptschule* students and students in middle or comprehensive schools aspiring to the *Hauptschule* certificate.

<sup>d</sup> *Realschule* students and students in middle or comprehensive schools aspiring to the *Realschule* certificate.

<sup>e</sup> *Gymnasium* students and students in comprehensive schools aspiring to the *Abitur*.

**Table 2: Cross-state mobility of teachers (2001)**

	Applications		Transfers	
	To Bavaria	From Bavaria	To Bavaria	From Bavaria
Baden-Württemberg	21	16	11	11
Berlin	5	2	0	1
Brandenburg	1	1	0	0
Bremen	1	0	0	0
Hamburg	2	1	1	1
Hesse	15	4	3	2
Lower Saxony	5	3	1	1
Mecklenburg-West Pomerania	3	0	0	0
North Rhine-Westphalia	20	5	3	2
Rhineland-Palatinate	6	2	1	0
Saarland	0	0	0	0
Saxony	8	1	2	1
Saxony-Anhalt	4	0	0	0
Schleswig-Holstein	3	3	0	0
Thuringia	8	0	0	0

**Table 3: Descriptive Statistics, Student background variables**

Variable	non-CEE	CEE	t-test of difference <sup>a</sup>
Sex (1 = female)	0.476	0.435	-1.27
Books at home: 0-10	0.139	0.099	-2.36**
Books at home: 11-25	0.196	0.179	-1.08
Books at home: 26-100	0.292	0.330	2.16**
Books at home: 101-200	0.161	0.176	0.83
Books at home: 200+	0.211	0.216	0.22
Immigrant background	0.212	0.095	-3.96***
School type (1 = <i>Realschule</i> )	0.459	0.569	1.28
Grade (1 = 8 <sup>th</sup> grade)	0.498	0.479	-0.22
Repeated grade	0.375	0.207	-6.60***
East Germany	0.041	0.262	3.67***
N obs.	1976	1219	

\* p<10%; \*\* p<5%; \*\*\* p<1%; <sup>a</sup> t-values allow for clustering on the class level

**Table 4:** Descriptive Statistics, Dependent variables

Variable	Mathematics		t-test of difference <sup>a</sup>	Biology (Science)		t-test of difference <sup>a</sup>
	Non-CEE	CEE		Non-CEE	CEE	
<i>Student variables</i>						
SCORE	-0.224	0.364	5.78***	-0.184	0.298	5.51***
GOOD <sup>b</sup>	0.732	0.716	-0.73	0.731	0.775	1.71*
PROB <sup>c</sup>	0.709	0.718	0.29	0.475	0.402	-2.57**
NOTE <sup>c</sup>	0.610	0.729	3.60***	0.672	0.637	-0.83
TEST <sup>c</sup>	0.367	0.346	-0.61	0.553	0.549	-0.11
WSHT <sup>c</sup>	0.533	0.432	-2.70***	0.472	0.359	-3.51***
EVLF <sup>c</sup>	0.280	0.272	-0.37	0.328	0.298	-1.48
HWGV <sup>c</sup>	0.832	0.754	-1.83*	0.500	0.262	-5.70***
HWTC <sup>c</sup>	0.700	0.754	1.75*	0.592	0.378	-4.90***
HWDS <sup>c</sup>	0.689	0.643	-1.48	0.465	0.272	-6.45***
RULE <sup>c</sup>	0.734	0.714	-0.84	0.438	0.413	-0.85
PRAC <sup>c</sup>	0.387	0.349	-1.60	0.425	0.421	-0.17
ASK <sup>c</sup>	0.484	0.495	0.38	0.653	0.584	-2.72***
EG <sup>c</sup>	0.698	0.737	2.05**	0.523	0.479	-2.02**
NEGL <sup>b</sup>	0.554	0.545	-0.26			
QUIET <sup>b</sup>	0.418	0.433	0.42			
SAYS <sup>b</sup>	0.444	0.456	0.32			
DAY <sup>d</sup>	0.220	0.202	-0.78	0.185	0.183	-0.14
LIFE <sup>b</sup>	0.933	0.912	-1.33	0.570	0.540	-1.30
WORK <sup>b</sup>	0.341	0.321	-0.92	0.242	0.200	-2.17**
LIKE <sup>e</sup>	0.608	0.529	-2.52**	0.615	0.660	1.42
ENJY <sup>b</sup>	0.496	0.350	-4.95***	0.543	0.534	-0.27
BORE <sup>b</sup>	0.390	0.438	1.63	0.376	0.346	-1.01
EASY <sup>b</sup>	0.414	0.344	-2.47**	0.626	0.697	2.38**
<i>Teacher Variables</i>						
CHNGE	0.225	0.157	-0.80	0.169	0.151	-0.28
SOAP	0.566	0.454	-1.06	0.427	0.487	0.63
STAP	0.837	0.836	-0.02	0.745	0.845	1.42
LIKET	3.149	3.187	0.23	3.291	3.096	-1.68*
TALNT	1.852	2.000	2.95***	2.645	2.880	2.24**
DISR	2.205	2.342	0.81	2.568	2.415	-0.94
LOW3	34.439	27.694	-1.87*	33.040	27.992	-1.43
UP3	12.368	15.693	1.45	15.992	10.616	-3.12***
PLAN	4.385	4.511	0.79	4.316	4.483	1.05
READ	2.915	2.943	0.14	3.071	3.046	-0.12
DIAG	3.081	3.272	1.69*	2.943	2.878	-0.64
CORCT	2.020	1.914	-0.55	2.128	1.962	-0.99
HELP	2.914	3.047	0.97	2.890	2.877	-0.12
HOMET	2.930	2.979	0.54	2.536	2.431	-0.80
HOMER	2.474	2.804	1.90*	2.581	2.585	0.02
FAM	3.391	3.713	3.17***	3.477	3.639	1.53
TEXT	2.894	2.784	-0.55	2.387	2.738	1.72*
RELY	1.686	2.267	3.10***	1.552	1.739	1.30
TIME	175.677	197.726	2.28**	78.361	80.958	0.30

\* p&lt;10%; \*\* p&lt;5%; \*\*\* p&lt;1%;

<sup>a</sup> t-values allow for clustering on the class level (student variables) or the teacher level (teacher variables)<sup>b</sup> 1 = agree, strongly agree; 0 = disagree, strongly disagree<sup>c</sup> 1 = pretty often, always; 0 = never, once in a while<sup>d</sup> 1 = no time, less than 1 hour, 1-2 hours; 0 = 3-5 hours, more than 5 hours<sup>e</sup> 1 = like, like a lot; 0 = dislike, dislike a lot

**Table 5:** Crosstabulation of central exit examination indicator and relative performance self-ratings (column percentages)

Subjective performance	Non-CEE state	CEE state
1 (does worse in mathematics)	25.62	32.45
2 (does about equally well)	41.95	41.16
3 (does better in mathematics)	32.44	26.39
Total	100.00	100.00

**Table 6:** Estimated effects of central exit examinations on achievement, student behavior and student attitudes

Var	cmd	comparison	b1	se1	b3	se3
Achievement						
SCORE	regress	science	0.110*	0.064		
GOOD	mlogit	biology	1.631***	0.242	0.839	0.131
Teaching practice						
PROB	mlogit	biology	0.903	0.172	1.235	0.163
NOTE	mlogit	biology	0.635**	0.114	1.134	0.220
TEST	mlogit	biology	0.828	0.150	0.850	0.169
WSHT	mlogit	biology	0.859	0.123	1.039	0.162
EVLF	mlogit	biology	0.862	0.114	0.837	0.111
HWGV	mlogit	biology	0.708	0.150	2.512***	0.604
HWTC	mlogit	biology	0.558***	0.099	2.425***	0.446
HWDS	mlogit	biology	0.640**	0.117	1.626***	0.247
RULE	mlogit	biology	0.864	0.162	1.018	0.122
PRAC	mlogit	biology	1.121	0.155	0.920	0.128
ASK	mlogit	biology	0.902	0.119	1.231	0.188
EG	mlogit	biology	0.724**	0.118	1.071	0.122
Class behaviour						
NEGL	ologit		0.891	0.140		
QUIET	ologit		1.143	0.160		
SAYS	ologit		1.156	0.186		
Student effort and motivation						
DAY	mlogit	science	0.964	0.154	1.329*	0.217
LIFE	mlogit	biology	1.096	0.276	0.983	0.128
JOB	mlogit	biology	1.143	0.147	1.219*	0.129
Student attitudes						
LIKE	mlogit	biology	1.868***	0.301	0.875	0.135
ENJY	mlogit	biology	1.813***	0.285	0.772*	0.104
BORE	mlogit	biology	0.966	0.133	1.337**	0.175
EASY	mlogit	biology	1.655***	0.260	0.769*	0.118

**Table 7:** Estimated effects of central exit examinations on teaching career, teacher attitudes and teaching practice

Var	cmd	CEE	se(CEE)	MATH	se(MATH)	CEE*MATH	se(CEE*MATH)
<i>Teaching career and teacher attitudes</i>							
CHNGE	ologit	0.455	0.263	1.320	0.637	0.815	0.576
SOAP	ologit	1.975	0.902	2.047**	0.702	0.419*	0.207
STAP	ologit	3.883**	2.212	1.836	0.792	0.447	0.303
LIKET	ologit	0.504*	0.181	0.727	0.281	1.910	1.095
TALNT	ologit	2.921**	1.447	0.011***	0.008	0.991	0.534
DISR	ologit	0.528	0.247	0.395***	0.120	2.322	1.193
LOW3	regress	-9.898***	3.652	1.298	2.829	-1.725	3.909
UP3	regress	-4.846**	1.970	-4.050**	1.774	8.576***	2.657
<i>Teaching Practice</i>							
PLAN	ologit	1.095	0.478	1.068	0.342	1.143	0.560
READ	ologit	0.875	0.386	0.626	0.190	1.325	0.622
DIAG	ologit	0.811	0.405	1.777	0.756	2.912*	1.772
CORCT	ologit	0.496*	0.211	0.812	0.279	1.005	0.535
HELP	ologit	1.028	0.430	1.053	0.394	1.663	0.892
HOMET	ologit	0.604	0.265	5.175***	2.221	1.690	0.956
HOMER	ologit	0.969	0.460	0.693	0.217	1.832	0.871
FAM	ologit	2.434*	1.128	0.729	0.269	1.807	1.107
TEXT	ologit	2.304*	1.043	2.367***	0.770	0.414	0.228
RELY	ologit	0.845	0.330	1.523	0.466	2.476*	1.225
TIME	regress	7.665	7.952	97.192***	9.585	18.397	12.990

**Appendix:**

**Table A1: Definition of dependent variables**

<i>Student achievement</i>		
SCORE	<b>5 Plausible values</b> in mathematics <b>5 Plausible values</b> in science	mean = 0 stddev = 1
GOOD	<b>How well do you usually do in mathematics and science at school?</b> I usually do well in <mathematics/biology>	(1) strongly disagree (2) disagree (3) agree (4) strongly agree
<i>Teaching practice (Student questionnaire)</i>		
PROB NOTE TEST WSHT. EVLF HWGV HWTC HWDS	<b>How often does this happen in your &lt;mathematics/biology&gt; lessons</b> The teacher shows us how to do <mathematics/biology> problems. We copy notes from the board. We have a quiz or test. We work from worksheets or textbooks on our own. We use things from everyday life in solving <mathematics/biology> problems The teacher gives us homework. The teacher checks homework. We discuss our completed homework.	(1) Never (2) once in a while (3) pretty often (4) always
RULE PRAC ASK EXMP	<b>When we begin a new topic in &lt;mathematics&gt;, we begin by...</b> ... having the teacher explain the rules and definitions ... discussing a practical or story problem related to everyday life ... having the teacher ask us what we know related to the new topic ... trying to solve an example related to the new topic	(1) Never (2) once in a while (3) pretty often (4) always
<i>Class behavior</i>		
NEGL QUIET SAYS	<b>In my mathematics class...</b> ... students often neglect their school work. ... students are orderly and quiet during lessons ... students do exactly as the teacher says.	(1) strongly disagree (2) disagree (3) agree (4) strongly agree
<i>Student effort and motivation</i>		
DAY	<b>On a normal school day, how much time do you spend before or after school doing each of these things?</b> studying <mathematics/science> or doing <mathematics/science> homework after school	(1) no time (2) less than 1 hour, (3) 1-2 hours (4) 3-5 hours (5) more than 5 hours
LIFE JOB	<b>What do you think about &lt;mathematics&gt;?</b> <mathematics/biology> is important to everyone's life. I would like a job that involves using <mathematics/biology>	(1) strongly disagree (2) disagree (3) agree (4) strongly agree
<i>Student attitudes</i>		
LIKE	<b>How much do you like... &lt;mathematics&gt;</b>	(1) dislike a lot (2) dislike (3) like (4) like a lot
ENJOY BORE EASY	<b>What do you think about &lt;mathematics&gt;?</b> I enjoy learning <mathematics/biology> <mathematics/biology> is boring <mathematics/biology> is an easy subject.	(1) strongly disagree (2) disagree (3) agree (4) strongly agree

<i>Teaching career and attitudes</i>		
CHNGE	<b>Would you change to another career if you had the opportunity?</b>	(1) Yes (0) No
SOAP	<b>Do you think that society appreciates your work?</b>	
STAP	<b>Do you think that students appreciate your work?</b>	
LIKET	<b>To what extent do you agree or disagree with the following statements?</b> A liking for and understanding of students are essential for teaching <mathematics /science>	(1) strongly disagree (2) disagree (3) agree (4) strongly agree
TALNT	Some students have a natural talent for <mathematics /science> and others do not.	
DISR	<b>In your view to what extent do the following limit how you teach your &lt;mathematics /science&gt; class?</b>	(1) not at all (2) a little (3) quite a lot (4) a great deal
LOW3	<b>Compared with other students in Germany at this grade level, estimate</b> ... what percent of students in your class have low achievement levels (bottom third nationally)	percent
UP3	... what percent of students in your class have high achievement levels (top third nationally)	
<i>Teaching Practice</i>		
PLAN	<b>Approximately how many hours per week do you normally spend on each of the following activities outside the formal school day?</b> ... planning lessons by yourself	None Less than 1 hour 1-2 hours 3-4 hours More than 4 hours
READ	... professional reading and development activity	
DIAG	<b>How often do you use assessment information to diagnose students' learning problems?</b>	1 none 2 little 3 quite a lot 4 a great deal
CORCT	<b>After a wrong answer,...</b> ... how often do you correct the student in front of the class?	(1) Never or almost never (2) some lessons (3) most lessons (4) every lesson
HELP	... how often do you ask another student to help?	
HOMET	<b>How many minutes of homework do you usually assign?</b>	(1) i do not assign homework (2) less than 15 minutes (3) 15-30 minutes (4) 31-60 minutes (5) 61-90 minutes
HOMER	<b>How often do you collect, correct and return homework assignments?</b>	(1) I do not assign homework (2) never (3) rarely (4) sometimes (5) always
FAM	<b>How familiar are you with the national curriculum guide for &lt;mathematics /science&gt;?</b>	(1) no such document (2) not familiar (3) fairly familiar (4) very familiar
TEXT	<b>What percentage of your teaching time is based on textbooks?</b>	0-25% 26-50% 51-75% 76-100%
RELY	<b>In planning &lt;mathematics /science&gt; lessons, how much do you rely on external examinations or standardized tests?</b>	(1) never (2) rarely (3) sometimes (4) always
TIME	<b>How many minutes per week do you teach &lt;mathematics /science&gt; to your class?</b>	minutes