SCHUMPETER DISCUSSION PAPERS

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

Hendrik Jürges
Kerstin Schneider

BERGISCHE
UNIVERSITÄT
WUPPERTAL

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

Hendrik Jürges & Kerstin Schneider

Hendrik Jürges
MEA
Universität Mannheim
L13,17
68131 Mannheim
Germany
Fax: +49-621-181-1863
Email: juerges@mea.uni-mannheim.de

Kerstin Schneider
Department of Economics
Universität Wuppertal
Gaußstr. 20
42097 Wuppertal
Germany
Fax: +49-202-439-2483
Email: schneider@wiwi.uni-wuppertal.de

Revised version: August 2008

Abstract: In response to PISA, all German federal states but one have adopted central exit examinations (CEEs) at the end of all secondary school tracks. Theoretically, the advantages of CEEs are fairly undisputed. CEEs make teaching and learning output observable and comparable across schools, and provide incentives for teachers and students to increase their effort. In line with earlier research, we confirm that CEEs have a positive causal effect on student performance. We also investigate what actually drives this effect. We find that the teachers' main reaction to CEEs is to increase the amount of homework, and to check and discuss homework more often. Students report increased learning pressure, which has sizeable negative effects on student attitudes towards learning. Students who take central exit exams in mathematics like mathematics less, think it is less easy and they are more likely to find it boring.

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 might 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 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 approach. But which institutions provide the right incentives in schools? In general, economists favor output-oriented governance of the public school system: define the goals of education, give incentives for attainment of these goals, and allow schools to choose the appropriate means to reach these goals. The main thrust is to introduce more competition into the school system and to develop indicators that allow comparing the performance of schools. Standardized tests are a requirement to for ranking schools *publicly* according to their average performance.

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 in the 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). A closely related issue that has received a great deal of attention is setting common

standards by establishing central exit examinations (CEEs) throughout the country. This discussion is of particular interest in Germany because federal states that already employed CEEs in the past generally outperformed non-CEE states in achievement tests. In response to PISA, all German federal states but one that have not had CEEs have introduced 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 tests in general) outperform non-CEE states 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 & Polborn 1999, Jürges, Richter & Schneider 2005). 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.

Much of the existing empirical literature has been devoted to estimating the effect of CEEs on student performance without trying to figure out how exactly CEEs work. In this paper, we aim at having a look into the black box and study the possible channels through which central exams raise performance. In particular, we look at changes in teacher and student effort as well as attitudes towards learning that might have beneficial effect on the learning process. For instance, teachers whose students have to face standardized examinations at the end of secondary school might adopt more efficient teaching styles but they might also simply increase the students' workload.

In this paper, we focus on the effect of exit exams at the end of lower secondary education, where a "natural experiment" provided by the German school system helps to infer the causal effect of CEES on performance and teaching practices. 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 scores, 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 described below, 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. Finally, we draw some conclusions in Section 7.

2. Institutional background

We now give a concise description of the German school system, trying to emphasize those aspects that are most relevant for understanding central exit examinations in the German context (a detailed description of the German school system can be found in Jonen and Eckardt (2004)). 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 (see Jürges & Schneider, 2007, for an empirical analysis of biases in the secondary school track allocation process).

--- Figure 1 about here ---

Hauptschule, *Realschule,* and *Gymnasium* are the three main types of secondary school; each leading 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 – at least theoretically – switch between school types.[1]

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 in which the TIMSS data were collected). Six states had central exit examinations at the end of *Realschule*, and only four had them at the end of *Hauptschule*.[2] 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.

Gymnasium students are not issued a leaving certificate after completing lower secondary level, but are admitted to the upper level of secondary education, which eventually leads to a university-entrance diploma (*Abitur*). Central exit examinations are most common at the end of upper secondary education. However, as Table 1 shows, decentralized systems of exit examinations at the end of upper secondary education exist as well. In the absence of central exist examinations teachers devise exams on their own, subject to the approval of the school supervisory authority.

--- Table 1 about here ---

German exit examinations never cover all of the subjects taught at school. For the *Abitur*, , students can choose four or five subjects (the choice is limited 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 the mathematics performance, teaching practices, student behavior and attitudes in *Hauptschule* and *Realschule* as the main outcome variables to be affected by CEEs.

---

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

[2] As mentioned in the introduction, CEEs have now been introduced in Saarland (2001), Hamburg (2005), Brandenburg (2005), Hesse (2006), Lower Saxony (2006), Berlin (2007), North Rhine-Westphalia (2007), Bremen (2007), Schleswig-Holstein (2008).

3. Conceptual framework

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. 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 costly for students and teachers, as more effort negatively affects utility. This could result in a more negative attitude 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 ---

While it is widely accepted that the main determinant of individual educational success is parental background, the influence of the parents, e.g. the electorate, on institutions is of importance as well. We indicate this by the (dashed) arrow from parental background to institutions. For estimating the causal effect of CEEs, this constitutes an important potential source of endogeneity of institutions, which we discuss in the next section.

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 critical ability levels α chosen to sort students. As a proxy for α, we will use $\Phi^{-1}(1-a)$, the *a* percent quantile of the standard normal distribution, where *a* is the proportion of 8th grade students aspiring to a high school diploma (see Table 1).

Besides the difficulties due to 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 purely random, this poses 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 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 even smaller (38). 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 (see the dashed arrow in Figure 2). If CEEs are correlated with such attitudes, simple differences between CEE and non-CEE states will be a biased measure of the causal 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 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)

$$y_i^m = \mu_i + X_i\beta + C_i\delta + \varepsilon_i^m, \tag{1}$$

and another to explain outcomes related to science y_i^s

$$y_i^s = \mu_i + X_i\gamma + \varepsilon_i^s, \tag{2}$$

where μ_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 ε_i^k, $k = m, s$ are i.i.d. error terms. Subtracting (2) from (1) yields

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

where δ is the parameter of interest. The main advantage of this estimator is that each student serves as her own control group. By taking differences, μ_i is swept out of the regression. Estimating (3) allows us to control for a lot of unobserved heterogeneity on the individual level, such as general ability, general attitudes towards learning and academic success, or socio-economic background. When looking at "subjective" outcome variables, the difference-

in-difference estimator has another important advantage: it will sweep out all differences between CEE- and non-CEE-states that are due to differences in survey response styles across both types of states.

In order for δ 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 variable. For instance, in the case of student test scores, there could be systematic indirect effects in the form of spillover from mathematics (more general skills) to science (more specific knowledge and skills). Negative spillovers from mathematics to science are 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, spillover 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 (e.g. innate mathematics and science skills) must not differ between federal states. Usually, one can plausibly assume that such characteristics are equally distributed across German states. 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 students are allocated 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 α, see above) may have better mathematics skills than their peers in low-α states, but comparable science skills. Finally, it is also important that mathematics and science outcomes are comparable.

The ideal comparison subject for our difference-in-difference-strategy is one (a) for which we have TIMSS test scores and a lot of ancillary information about the teaching process (b) that is "unrelated" to mathematics to avoid problems of knowledge spillover or of teachers teaching both subjects, and (c) that is not tested in central exams. Unfortunately, this subject does not exist. The subject that – in our opinion – comes closest to meeting all three requirements simultaneously is biology. However, biology test scores and separate information on hours spent learning biology at home are not available in the TIMSS data.

Thus we use the corresponding information for science in general for our comparisons of outcomes across exam types.

Jürges, Schneider & Büchel (2005) give a detailed discussion of the plausibility of our identifying assumptions with respect to student achievement as the outcome variable. They argue that spillover 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 (see IEA TIMSS 1998). 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 teaching time in primary schools and about one-seventh of teaching time in lower secondary schools and 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 α in our difference-in-differences framework.

5. Data description

The international data set of TIMSS Germany contains data on a total of 5763 7th and 8th 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. Moreover, we deleted data from Saxony (where both mathematics and biology are tested centrally) from our sample.[3] Our working sample consists of 1,976 students in non CEE-states and 1,219 students in CEE states. 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.

---

[3] As has been pointed out by one referee, Saxony would make good comparison state to corroborate our results if students in Saxony (passing central exams in mathematics and science) are compared to those with CEEs only in mathematics. The main problem with this strategy is that the number of independent observations (classes, not

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 eastern 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 question wordings 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

---

students) in Saxony is very low (12). However, there is evidence that in Saxony, biology homework is taken more seriously relative to mathematics homework than in other CEE states, which supports our hypothesis.

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 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, but 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. However, 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. that students use the same response categories in the same way, independent of the subject they refer to (mathematics or biology/science). Students may differ in the way they use these answer categories. For example, the construction of our dependent variables allows that "once in while" means the same frequency to one student as "pretty often" to another. Individual students' response styles may also differ across 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 the same questions related to mathematics classes and to biology classes.

6. Regression Results

Regression results are shown in Table 5. 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 set of explanatory variables. Besides CEE, we use the explantory variables described in Table 3 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*),

and 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 discussing our results, a note on the interpretation of the effects shown for multinomial logit models might be helpful. All models are 3-category models. The values shown in Table 6 are relative risks and their standard errors (computed with 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. With two four category outcome variables, there are 16 different combinations of answers. Take self-rated performance as an example. The original items read: "I usually do well in mathematics" and "I usually do well in biology", respectively. Students are asked whether they "strongly disagree", "disagree", "agree", or "strongly agree" to that statement. We reduce the 16 possible combinations to three.

(1) Agree less to do well in mathematics than to do well in biology (put differently: to state one does worse in mathematics than in biology)
(2) Agree equally to do well in mathematics and biology (put differently: to state one does about equally well in mathematics and in biology)
(3) Agree more to do well mathematics than to do well in biology (put differently: to state one does better in mathematics than in biology)

For example, a student who "disagreed" to both statements is assigned to the second category, a student who "disagreed" to the mathematics item but "agreed" to the corresponding biology item is assigned to the first category, and a student who "strongly agrees" to the math statement but "disagrees" with the biology statement is assigned to the third category, etc.

In the multinomial regressions, the middle category is always the baseline category. In Table5 we show two relative risks: the first mirrors 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: the second reflects the effect on the probability of thinking one does better in mathematics than in

biology relative to the probability of thinking that one does about equally well in both subjects. To illustrate, consider the the bivariate relationship between central exit exams and the relative self-evaluation in mathematics versus biology shown in Table 6.

--- about here Tables 5 and 6 ---

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 the differences in self-ratings 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 risk of thinking they do worse in mathematics than in biology relative to 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 simply state 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 a bit cumbersome 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, 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 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.

One potential objection against this result is that the regression include a small proportion of students (from Bavaria) who will pass CEEs in science. Earlier studies (Jürges et al. 2005) have addressed this issue by excluding students who are likely to take a central exit examination in science, e.g. those who (a) say they do well in science or (b) the 40% *Hauptschule* and 25% *Realschule* students who do best in science as measured by their TIMSS test score.[4] We have tried both sample restrictions. It turns out that imposing the restrictions results in an increased estimate of the effect of CEE on test scores. This is of course expected as we essentially censor the left-hand side variable. However, the results for the other dependent variables in this paper (discussed below) are very similar as those in the basic regressions including all Bavarian students.

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, studentes themselves appear to think the opposite. One explanation for this seemingly contraditory finding is that relative expectations are higher in CEE states, for example because teachers put more pressure on their students to perform well, knowing that the centralized exams are due in only one or two years.

6.2. Teaching practices

The results for teaching practices (as reported by the students) shows major differences between CEE- and non-CEE-states for all homework related items. Risk ratios smaller than one in the column labeled "RR(1)" indicate that it is less common in CEE-states to give less homework in mathematics than to give the same amount of homework in mathematics and biology. Risk ratios greater than one in the column labeled "(RR2)" indicate that it is more common in CEE-states to give more homework in mathematics than to give the same amount of homework in mathematics and biology. Both effects go in the same direction. It is thus much more common for teachers in CEE-states to give, check and discuss homework in class. Moreover, effect sizes (relative risks) for these items are substantial.

Overall, it seems as if the importance of homework is the main systematic differences between CEE- and non-CEE states. Of the other nine items, only two show significant

---

[4] In Bavaria between 25 percent (*Realschule*) and 40 percent (*Hauptschule*) of the students take the central exit examinations in science.

differences: how often teachers let students copy notes from the board, and how often they start a new topic by solving an example. The risk ratios smaller than one in the "RR(1)" column show that this is a less common practices in non CEE-states' mathematics lessons than in CEE-states' mathematics lessons.

6.3. Class behavior (student discipline)

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. Hence we use ordered logit models to compare levels of discipline in mathematics classes across states. The results do not suggest that there are any systematic differences between CEE and non-CEE states in student discipline in math classes. However, differential item funtioning across the two types of states could mask factual differences in discipline.

6.4 Student effort and motivation

TIMSS asked students how much time they spend outside school learning mathematics and science. 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 how much students agree to the statement that mathematics/biology is important in everyone's life and. whether students would like a job that involves mathematics/biology. Here we only find weak and/or inconstitent 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 that students in CEE-states 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.

6.5. 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, and this evidence clearly points into the direction that CEEs impose costs on students. 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 a worse attitude towards mathematics.

7. Summary and conclusion

This paper studies the costs and benefits of central exit examinations (CEEs) at the end of lower secondary school in Germany. The theoretical literature almost exclusively focuses on the benefits of central examinations, which arise in the form of higher student achievement. The costs, however, have been neglected so far by the economic literature. By costs we mean potentially negative effects on students' and teachers' morale and attitudes towards learning.

The identification of (positive or negative) causal effects of CEEs is by no means easy. Caution is warranted when interpreting observed differences between jurisdictions with and without CEEs as the effect of CEEs on student achievement, because CEEs are most likely the outcome of a political process (reflecting the preferences of the electorate) and thus potentially endogenous.

In this paper, we make use of some unique regional variation in Germany that 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 German school system, only some states have CEEs and these exams are restricted to core subjects such as German, mathematics and the first foreign language (mostly English). We use data from the Third International Mathematics and Science Study (TIMSS) 1995 to exploit this institutional variation and uncover the causal effect of CEEs on student achievement in mathematics, teaching practices and students' attitudes towards mathematics by comparing a range of outcome variables across subjects and types of exit examination. The fundamental idea is that a CEE affects only mathematics-related outcomes but not science-related outcomes. In most of our analyses, we use biology outcomes as our main comparison subject. Biology is almost never examined centrally and there is no mathematics involved in lower secondary biology topics.

There are three main insights from this study. First, CEEs have a small but statistically significant causal effect on student test scores. Second, teachers in CEE-states are more likely to give, check and discuss homework. Third, students in CEE-states do like mathematics less, find it less easy and find it more boring than those in non-CEE states. They are also somewhat

more diligent in 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 on students rather than employ more sophisticated or innovative teaching methods. However, giving, checking, and discussing homework certainly involves also increased teacher effort. But all in all, achievement gains in mathematics appear to result largely from increased student effort. One (certainly unintended) consequence is that students in CEE states less often think that mathematics is fun to do. This might actually offset some of the positive achievement effects of CEEs. Working harder but being less motivated could be less efficient than working hard but at the same enjoying it.

References:

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.

Costrell, Robert M. (1997). "Can educational standards raise welfare?" *Journal of Public Economics*, 65, 271-293.

Effinger Matthias R. & Mattias K. Polborn (1999). "A model of vertically differentiated education." *Journal of Economics*, 69, 53-69.

Frenck, Isabella (2001). "Stundentafeln der Primar- und Sekundarstufe I im Ländervergleich - eine empirische Studie am Beispiel der Fächer Deutsch und Mathematik." Unpublished master thesis, University of Essen.

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.

Holland, P.W., 1986, "Statistics and Causal Inference." *Journal of the American Statistical Association* 81; 945-960.

Hoxby, Caroline M. (2000). "The effects of class size on student achievement: New evidence from population variation." *Quarterly Journal of Economics*, 115, 1239-1285.

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 & Thomas Eckardt (2006). "The Education System in the Federal Republic of Germany 2004." Secretariat of the Standing Conference of the Ministers of Education (KMK). http://www.kmk.org/ dossier/dossier_en_ebook.pdf [2007, Jan 12].

Jürges, Hendrik & Kerstin Schneider (2004). "International differences in student achievement: An economic perspective." *German Economic Review*, 5, 357-380.

Jürges, Hendrik & Kerstin Schneider (2007). "What can go wrong will go wrong: Birthday effects and early tracking in the German school system" MEA Discussion Paper 138-07. University of Mannheim.

Jürges, Hendrik, Wolfram F. Richter & 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 & 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.

Kultusministerkonferenz (ed.) (2005): *Bildungsstandards der Kultusministerkonferenz Erläuterungen zur Konzeption und Entwicklung*. München/Neuwied: Luchterhand.

Wößmann, Ludger (2002). "Central Exams Improve Educational Performance: International Evidence." Kiel Discussion Papers 397, Kiel Institute for World Economics.
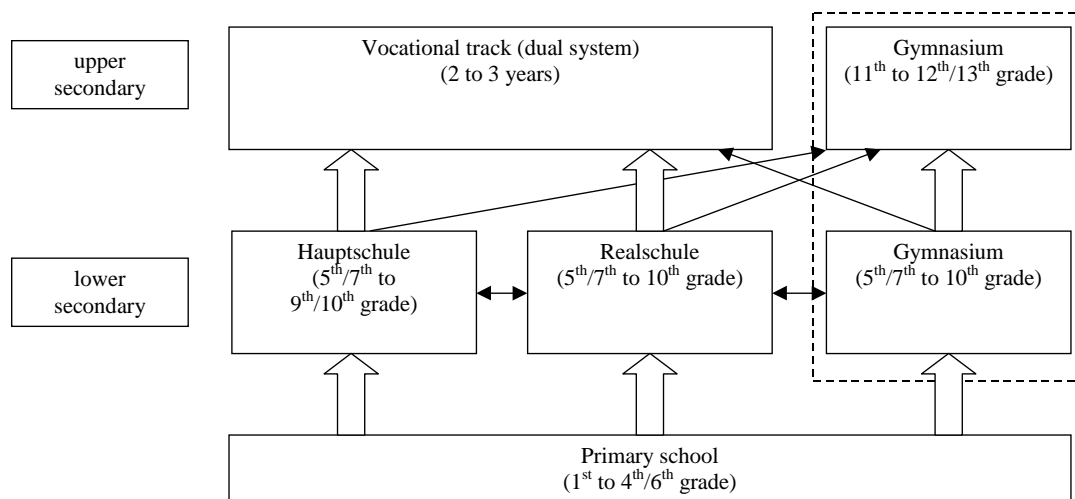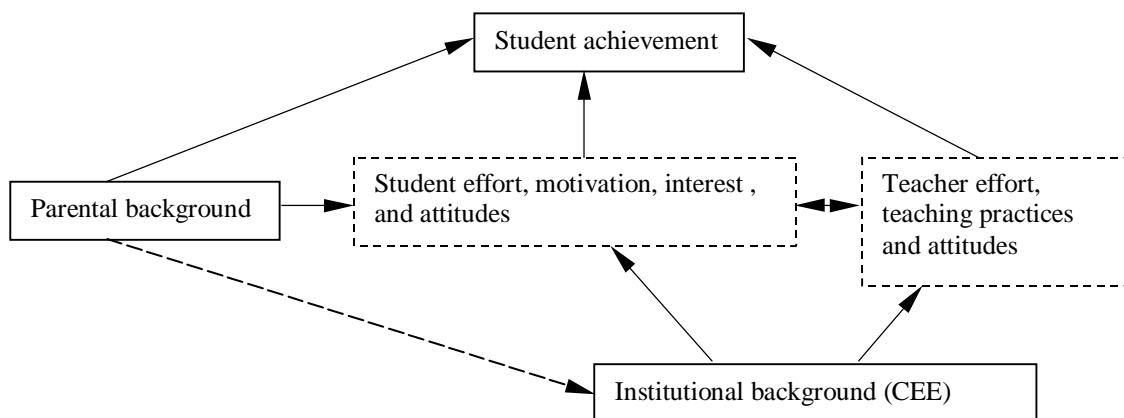
Figure 1: The German school system

Figure 2: Conceptual framework

Tables:

Table 1: CEE by federal state and type of certificate (as of 1995); proportion of students by school/type of certificate (in 1999)[a]

	Haupt-schule	Realschule	Gymna-sium (Abitur)	8th grade students in Hauptschule tracks [c]	8th grade students in Realschule tracks [d]	8th grade students in Abitur tracks [e]	8th grade students in other schools
CEE-States							
Baden-Württemberg	G/M/F/O	G/M/F	A	34.0	31.7	28.4	6.0
Bavaria	G/M/F/S[b]/ O	G/M/F/S[b]/ O	A	37.1	31.0	26.9	4.9
Mecklenburg-West Pomerania		G/M/F	A	15.7	48.9	29.3	6.2
Saarland			A	24.4	37.0	34.0	4.6
Saxony	G/M/F	G/M/S	A	12.8	51.0	30.1	6.1
Saxony-Anhalt		G/M	A	9.4	50.9	32.8	7.0
Thuringia	G/M	G/M/F	A	18.3	42.3	32.6	6.7
Average (unweighted)				21.7	41.8	30.6	5.9
Non-CEE states							
Berlin				21.5	31.1	42.8	4.5
Brandenburg				16.9	32.3	45.2	5.7
Bremen				26.8	30.9	36.1	6.3
Hamburg				22.6	26.2	43.3	7.9
Hesse				22.5	34.0	38.6	4.8
Lower Saxony				31.3	34.9	28.0	5.7
North Rhine-Westphalia				29.3	30.6	35.3	4.7
Rhineland-Palatinate				37.2	28.4	29.5	4.9
Schleswig-Holstein				31.4	34.4	28.5	4.8
Average (unweighted)				26.6	31.4	36.4	5.0

Notes:

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

[a] Percentages add up to less than 100. Students in special schools (e.g. for slow learners) are not listed.

[b] Optional subject.

[c] *Hauptschule* students and students in middle or comprehensive schools aspiring to the *Hauptschule* certificate.

[d] *Realschule* students and students in middle or comprehensive schools aspiring to the *Realschule* certificate.

[e] *Gynnasium* 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

Source: Bavarian Ministry of Education and Culture website.

Table 3: Descriptive Statistics, Student background variables

Variable	non-CEE	CEE	t-test of difference [a]
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 = *Realschule*)	0.459	0.569	1.28
Grade (1 = 8th 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%; [a] t-values allow for clustering on the class level

Table 4: Descriptive Statistics, Dependent variables

Variable	Mathematics			Biology (Science)		
	Non-CEE	CEE	t-test of difference [a]	Non-CEE	CEE	t-test of difference [a]
Achievement						
TIMSS test score	-0.224	0.364	5.78***	-0.184	0.298	5.51***
Self-rated performance [b]	0.732	0.716	-0.73	0.731	0.775	1.71*
Teaching practice						
Teacher shows how to do problems [c]	0.709	0.718	0.29	0.475	0.402	-2.57**
Copy notes from board [c]	0.610	0.729	3.60***	0.672	0.637	-0.83
Have quiz or test [c]	0.367	0.346	-0.61	0.553	0.549	-0.11
Work from textbooks on our own [c]	0.533	0.432	-2.70***	0.472	0.359	-3.51***
Use everyday things in solving problems [c]	0.280	0.272	-0.37	0.328	0.298	-1.48
Teacher gives homework [c]	0.832	0.754	-1.83*	0.500	0.262	-5.70***
Teacher checks homework [c]	0.700	0.754	1.75*	0.592	0.378	-4.90***
Teacher discusses homwork [c]	0.689	0.643	-1.48	0.465	0.272	-6.45***
Teacher explains rules and definitions [c]	0.734	0.714	-0.84	0.438	0.413	-0.85
Teacher discuss practical problems [c]	0.387	0.349	-1.60	0.425	0.421	-0.17
Teacher asks what students already know [c]	0.484	0.495	0.38	0.653	0.584	-2.72***
Try to solve examples related to problems [c]	0.698	0.737	2.05**	0.523	0.479	-2.02**
Class behaviour						
Students neglect school work [b]	0.554	0.545	-0.26			
Students are orderly and quiet [b]	0.418	0.433	0.42			
Students do as teacher says [b]	0.444	0.456	0.32			
Student effort and motivation						
Hours spend studying subject at home [d]	0.220	0.202	-0.78	0.185	0.183	-0.14
Subject important to everyone's life [b]	0.933	0.912	-1.33	0.570	0.540	-1.30
Likes to do job that involves subject [b]	0.341	0.321	-0.92	0.242	0.200	-2.17**
Student attitudes						
Like subject [e]	0.608	0.529	-2.52**	0.615	0.660	1.42
Enjoy learning subject [b]	0.496	0.350	-4.95***	0.543	0.534	-0.27
Subject is boring [b]	0.390	0.438	1.63	0.376	0.346	-1.01
Subject is easy [b]	0.414	0.344	-2.47**	0.626	0.697	2.38**

* $p<10\%$; ** $p<5\%$; *** $p<1\%$;

[a] t-values allow for clustering on the class level (student variables) or the teacher level (teacher variables)

[b] 1 = agree, strongly agree; 0 = disagree, strongly disagree

[c] 1 = pretty often, always; 0 = never, once in a while

[d] 0 = no time, less than 1 hour, 1-2 hours; 1 = 3-5 hours, more than 5 hours

[e] 1 = like, like a lot; 0 = dislike, dislike a lot

[f] 1 = yes; 0 = no

[g] 1 = quite a lot, a great deal; 0 = not at all, a little

[h] in percent; [j] in minutes

Table 5: Estimated effects of central exit examinations on achievement, student behavior and student attitudes

Dependent Variable	Estimation method	RR (1) (math<bio)	Standard error	RR (2) (math>bio)	Standard error
Achievement					
TIMSS test score	OLS	0.110*	0.064		
Self-rated performance [b]	Mult. Logit	1.631***	0.242	0.839	0.131
Teaching practice					
Teacher shows how to do problems [c]	Mult. Logit	0.903	0.172	1.235	0.163
Copy notes from board [c]	Mult. Logit	0.635**	0.114	1.134	0.220
Have quiz or test [c]	Mult. Logit	0.828	0.150	0.850	0.169
Work from textbooks on our own [c]	Mult. Logit	0.859	0.123	1.039	0.162
Use everyday things in solving problems [c]	Mult. Logit	0.862	0.114	0.837	0.111
Teacher gives homework [c]	Mult. Logit	0.708	0.150	2.512***	0.604
Teacher checks homework [c]	Mult. Logit	0.558***	0.099	2.425***	0.446
Teacher discusses homwork [c]	Mult. Logit	0.640**	0.117	1.626***	0.247
Teacher explains rules and definitions [c]	Mult. Logit	0.864	0.162	1.018	0.122
Teacher discuss practical problems [c]	Mult. Logit	1.121	0.155	0.920	0.128
Teacher asks what students already know [c]	Mult. Logit	0.902	0.119	1.231	0.188
Try to solve examples related to problems [c]	Mult. Logit	0.724**	0.118	1.071	0.122
Class behaviour					
Students neglect school work [b]	Ord. Logit	0.891	0.140		
Students are orderly and quiet [b]	Ord. Logit	1.143	0.160		
Students do as teacher says [b]	Ord. Logit	1.156	0.186		
Student effort and motivation					
Hours spend studying subject at home [d]	Mult. Logit	0.964	0.154	1.329*	0.217
Subject important to everyone's life [b]	Mult. Logit	1.096	0.276	0.983	0.128
Likes to do job that involves subject [b]	Mult. Logit	1.143	0.147	1.219*	0.129
Student attitudes					
Like subject [e]	Mult. Logit	1.868***	0.301	0.875	0.135
Enjoy learning subject [b]	Mult. Logit	1.813***	0.285	0.772*	0.104
Subject is boring [b]	Mult. Logit	0.966	0.133	1.337**	0.175
Subject is easy [b]	Mult. Logit	1.655***	0.260	0.769*	0.118

Notes: Standard errors for odds ratios are computed by the delta method.

Significance levels based on original t-statistics: * p<10%; ** p<5%; *** p<1%

Control variables: number of books at home, immigrant, sex, school type, repeated grade, grade, East German state, state proportion of students attending Gymnasium.

RR denotes the relative risk that the rating pertaining to mathematics is larger of smaller then the same rating pertaining to biology.

Table 6: 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

Appendix:

Table A1: Description of dependent variables

Student achievement	
5 Plausible values in mathematics **5 Plausible values** in science	mean = 0 stddev = 1
How well do you usually do in mathematics and science at school? I usually do well in <mathematics/biology>	(1) strongly disagree (2) disagree (3) agree (4) strongly agree
Teaching practice (Student questionnaire)	
How often does this happen in your <mathematics/biology> lessons 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
When we begin a new topic in <mathematics>, we begin by… ... 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
Class behavior	
In my mathematics class... ... 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
Student effort and motivation	
On a normal school day, how much time do you spend before or after school doing each of these things? 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
What do you think about <mathematics>? <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
Student attitudes	
How much do you like... <mathematics>	(1) dislike a lot (2) dislike (3) like (4) like a lot
What do you think about <mathematics>? 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