The effect of central exit examinations on student achievement

Quasi-experimental evidence from TIMSS Germany

Hendrik Jürges, MEA, University of Mannheim*

Kerstin Schneider, University of Wuppertal and CESifo

Felix Büchel, Max Planck Institute for Human Development, Berlin

October 2004

forthcoming in *Journal of the European Economic Association*

Abstract: This paper makes use of the regional variation in schooling legislation within the German secondary education system to estimate the causal effect of central exit examinations on student performance. We propose a difference-in-differences framework that exploits the quasi-experimental nature of the German TIMSS middle school sample and discuss its identifying assumptions. The estimates show that students in federal states with central exit examinations clearly outperform students in federal states without such examinations. However, only part of this difference can be attributed to the existence of the central exit examinations themselves. Our results suggest that central examinations increase student achievement by the equivalent of about one-third of a school year.

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

JEL-Code: I28

---

* Corresponding author: Mannheim Research Institute for the Economics of Aging (MEA); University of Mannheim; L13, 17; 68131 Mannheim; Germany; Phone: +49-621-181-3519; Fax: +49-621-181-1863; Email: juerges@mea.uni-mannheim.de. We are grateful to Rainer Watermann, Max Planck Institute for Human Development, for managing administrative and technical access to the regional information used in this study, and to Jürgen Baumert for granting special permission to use the data. We would also like to thank two anonymous referees and the editor for their most helpful comments. The final manuscript was prepared in remembrance of our co-author, friend and colleague Felix Büchel who passed away in July 2004.

1. Introduction

The need to reform the German school system became clearly apparent after the publication of two large-scale international school studies: TIMSS (Third International Mathematics and Science Study) and PISA (OECD Programme for International Student Assessment). The poor performance of German students compared to their peers in other European countries prompted intense political discussion on the weaknesses of the German school system. From an economist's perspective, reforms can be targeted either at the allocation of financial resources to schools or at changing the institutional framework in which students, teachers and schools operate. The effects of school resources on student achievement are often small and sometimes even inconsistent (as exemplified by the class size discussion), and 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). Although experimental evidence shows positive effects on achievement in primary school, particularly for minority students (Krueger 1999), general reductions in class size are widely regarded as a questionable measure in terms of cost effectiveness.[1]

Simple economic reasoning suggests that the institutions of the school system affect the performance of students and teachers because they create incentives. Changing the institutional setup could thus be a more cost-efficient approach to reform. While the short-term costs of reforming the environment in which students, teachers and schools operate may be high, it is unlikely that new rules will remain costly once they are firmly established. School choice, competition between schools and school vouchers are among the issues that have received much attention in the United States in this context. While many economists are

[1] In a recent paper, Krueger (2003) challenges this view. He calculates the costs (more classrooms; higher teacher salary) and benefits (present value of student's lifetime earnings) of a reduction in the size of primary school classes from 22 to 15 students. Not surprisingly, whether the costs outweigh the benefits or vice versa is dependent on the assumed discount rate. Krueger computes an internal rate of return of 6.2 percent.

confident that competition between schools will have beneficial effects, the empirical findings thus far have been mixed (see e.g. Hoxby, 2003, Krueger and Zhu, 2002).

One 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 TIMSS and PISA. In fact, in response to TIMSS and PISA, Germany's state ministers of education agreed to work towards adopting national standards in the future. However, the question of whether states with CEEs outperformed non-CEE states in TIMSS and PISA purely because of the exit examinations or because of some other, omitted variable at the state level remains unanswered.

The theoretical literature almost unanimously shows that CEEs and hence 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 decentralised examinations (e.g. Bishop, 1997, 1999). Students, for example, benefit because the results of CEEs are more valuable as 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. Student test results can be used to monitor teacher and teaching quality on a regular basis. Whether incentives to improve teaching quality, arguably an important factor in the education production function, should come solely from reputation effects or in the form of higher pay for better teachers is open to discussion (Glewwe et al., 2003, Jürges et al., 2004, Lavy, 2002, 2003). Finally, the reputation of entire schools can be based on the achievement of its students, with good schools attracting good students (provided that aggregate CEE results are made available to the public).

Previous empirical findings of cross-country comparisons as well as single-country studies indicate that central exit examinations significantly improve student performance (Bishop, 1997, 1999). However, as Jürges and Schneider (2004) show, the positive effect of CEEs on achievement in cross-country analysis, identified on the basis of the international TIMSS database, is not robust. Besides the international evidence, Bishop (1997, 1999) also presents results from Canadian micro-data. In 1990-1991, Canada, like Germany, had a mixed system. Some provinces administered central exit exams in the final year of high school, whereas others did not. Bishop estimates the effect of CEEs on performance to be three-fifths of a US grade-level equivalent for science and four-fifths of a US grade-level equivalent for mathematics.

However, it is important to differentiate between simple correlations and causation. The possibility that countries or federal states with CEEs attach greater importance to education and academic achievement has to be taken into account. In this case, both high average student achievement and CEEs would simply reflect the higher value placed on education by the electorate of these areas. Earlier papers have tried to deal with this issue by asking whether CEE states also differ from non-CEE states along dimensions other than achievement, e.g. student discipline and absenteeism (Bishop, 1997). However, the data did not allow a convincing identification strategy and left important issues unresolved. Wößmann (2002) uses the international TIMSS micro-data to estimate the CEE effect as the difference-in-differences by grade, arguing that incentives should increase as the exit exams approach. His regressions suggest that students who take a central exam at the *end of upper secondary education* outperform students in states without central exams by about one grade-level equivalent in grade 7, and that the gap increases by another 50 percent when they move from grade 7 to 8. These estimates seem too large to be trusted (as a simple thought experiment, extrapolate this effect to the final school year).

We use the German federal education system as a source of exogenous variation to identify the causal effect of CEEs on student achievement. Earlier studies with German TIMSS data have looked at upper secondary students and found positive effects of CEEs for students in non-specialised mathematics courses only (Baumert and Watermann, 2000). However, as will be argued below, the German system of upper secondary education is not really suitable for analysing the CEE effect. Instead, 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 difference in the mathematics-science test score differential and interpret this difference-in-differences as the causal effect of CEEs on student achievement. We discuss the assumptions underlying this interpretation and try to substantiate them by means of additional analyses.

The remainder of the paper is organised as follows: In Section 2, we give a brief account of the German secondary education system. Section 3 describes the data, and Section 4 discusses issues of identification and estimation. The results are presented in Section 5 and Section 6 concludes.

2. German secondary education in perspective

In this section, we give a concise description of the German school system, trying to emphasise those aspects that are most relevant to an understanding of central exit examinations in the German context.[2] Figure 1 gives a stylised 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.

--- 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.[3]

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). *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.

---

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

[3] 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

--- Table 1 about here ---

Central exit examinations are most common at the end of upper secondary education. In 1995, 7 of the 16 German states had a central *Abitur* (*Gymnasium* leaving certificate) on the state level. These states are concentrated in the south (Baden-Württemberg, Bavaria, Saarland) and east (Mecklenburg-West Pomerania, Saxony, Saxony-Anhalt, Thuringia) of the country. 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 only four had them at the end of *Hauptschule*.

German exit examinations never cover all of the subjects taught at school. At the *Abitur* level, students can choose three or four subjects (within certain limits that vary 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 effect of CEEs on student achievement, we will thus concentrate on mathematics performance in *Hauptschule* and *Realschule* as the main outcome variable thought to be affected by CEEs. Other subjects assessed in central exit examinations are foreign languages (mostly English) or – less commonly – science.

3. Data description

The international data set of TIMSS Germany contains data on a total of 5763 7th and 8th grade students in 137 schools, collected in the 1994/95 school year. Data were collected in

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.

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 below, 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 2 contains the variable definitions and descriptive statistics by the type of exit examination. In contrast to publications focusing on international comparisons of student achievement, we do not use the internationally standardised mathematics (and in some analyses) science scores as dependent variables. For the sake of intra-German comparability, we chose the national Rasch scores, standardized to have a mean of 0 and a variance of 1. The size of our regression parameters can thus be directly interpreted in terms of standard deviations.

--- Table 2 about here ---

The most notable difference between students in states with and without CEEs is their achievement in mathematics and science. Students in states with CEEs score on average 0.4 standard deviations higher than those in states without CEEs. Student background, measured in terms of the number of books at home, differs only slightly in this respect – 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). Interestingly, the cumulative number of mathematics lessons – calculated from the official timetables of all federal states (Frenck 2001) – is considerably smaller for students in non-CEE states.

4. Identification and Estimation

The most basic approach to identifying the causal effect of CEEs on student achievement using German TIMSS data would seem to be estimating *simple differences* between average achievement in CEE states and non-CEE states, while controlling for student background and other variables of interest. However, simple differences have only limited value 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 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, last column).

The second potential drawback of simple differences is self-selection into treatment, which is one of the most frequent problems encountered by researchers trying to evaluate the causal effects of policy measures. First, parents might vote with their feet and move to another state in order to send their children to schools with a central exit examination. Second, parents in non-CEE states who live near a CEE-state may choose to send their children to school in the neighbouring 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. 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 preferred strategy is to isolate CEE effects from differential parental attitudes and other unobserved variables by exploiting further variation within states. For instance, a typical CEE state has central examinations for each type of leaving certificate (*Hauptschule, Realschule, Abitur*), whereas a typical non-CEE state has no central examinations at all. Some states, however, have mixed systems. In Mecklenburg-West Pomerania and Saxony-Anhalt, *Hauptschule* exams are not central, but *Realschule* exams and the *Abitur* are. In Saarland, only *Abitur* was centralised (since 2001/2002 all exams have been centralised). This variation in institutional settings can be exploited in the sense that students in CEE states who attend non-CEE schools can be used as a control group. Unfortunately, there are two problems with this estimation strategy. First, as mentioned above, the allocation of students to school types is not random, but based on prior academic achievement, which may in turn be correlated with unobservable but relevant variables, such as susceptibility to extrinsic motivation of all kinds. It is therefore advisable to take selection into school types into account. The second and more serious problem is lack of data: the "interesting" states such as Mecklenburg-West Pomerania and Saarland are small, and there are only few observations from these states in the sample.[4]

[4] Nevertheless, the fact that exams are not centralised in all types of schools in these states may affect our results. Students may self-select into schools without central exams if they expect to do poorly in CEEs, in which case the CEE effect might be overstated. As a robustness check, we thus ran all regressions reported below without Saxony-Anhalt and Mecklenburg-West Pomerania. Contrary to what the above argument suggests, the CEE effect increased somewhat, reinforcing our main hypothesis.

Another possibility for evaluating the effectiveness of central exit examinations would be *before-after comparisons*. Fuelled by the heated public debate after the publication of the TIMSS and, in particular, PISA results, some states have recently introduced CEEs or are planning to do so. Interestingly, no state actually plans to abolish CEEs. Since PISA is designed as a repeated cross-sectional study, the PISA data could be used to estimate the effect of these policy changes. However, the causal effect of CEEs will still be difficult to identify because there may also have been other policy changes in response to TIMSS/PISA, the publication of the results may have changed parental attitudes towards education, etc. Since only one cross-section of data is currently available, this possibility is mentioned mainly for sake of completeness.

Fortunately, the German secondary education system offers a unique source of exogenous variation that can be used to identify the causal effect of CEEs on student achievement. Table 1 shows which federal states have CEEs in which types of schools and in which subjects. Note that in *Haupt-* and *Realschule*, central exit examinations (if any) cover only German, mathematics and one foreign language (mostly English). Science is not tested in central examinations – with two exceptions. It is a compulsory subject for Saxony's *Realschule* certificate, and it is optional in Bavaria. Each year, roughly 40 percent of Bavarian students aiming at a *Hauptschule* certificate are tested in biology, chemistry and physics, i.e. all subjects covered by the TIMSS science test. Between 20 and 25 percent of those aspiring to a *Realschule* certificate take written exit examinations in physics only (roughly one-third of the TIMSS science items come from this domain). In states where mathematics is a CEE subject but science is not, the effect of CEEs on student achievement should be larger in mathematics than in science. Since TIMSS provides test results in both mathematics and science, we can estimate the difference-in-differences by subject. Formally, our estimator can be described as follows. Consider two regressions: one to explain a student's test result in mathematics y_i^m

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

and another to explain the test result in science y_i^s

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

where μ_i is some individual specific characteristic (e.g. general ability), X_i is a vector of covariates that might affect mathematics and science performance 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

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

where δ is the parameter of interest. The main advantage of this estimator is that each individual serves as his or her own control group. By taking differences, μ_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 or socio-economic background. Of course, in order to interpret this difference as the causal effect of CEEs on student performance, we still need identifying assumptions, most importantly $E[C_i(\varepsilon_i^m - \varepsilon_i^s)] = 0$. There are several ways in which this assumption might be violated. First, there could be systematic indirect effects in the form of spill-over from mathematics to science.[5] Second, CEE and non-CEE states could differ systematically in their relative preference for mathematics rather than science. Third, relative innate mathematics versus science skills could differ between federal states – at least within

[5] We do not consider science to mathematics spill-over, since we do not expect any effect of specific knowledge and skills (science) on more general skills (mathematics).

the sub-sample of students that we use. Finally, it is vital that mathematics and science test results be measured on the same scale, i.e., they must be comparable.

There will be positive spill-over from mathematics to science if good mathematics skills are a prerequisite for performing well in science, or – to be more precise – in the TIMSS science items. In this case, the difference-in-differences by subject framework will underestimate the effect of CEEs on achievement. However, we believe that spill-over from good mathematics skills to good performance in the TIMSS science test is likely to be very small. In order to assess the likelihood of such spill-over, we analysed the (released) set of TIMSS science items (IEA TIMSS, 1998). The released set contains 87 items, of which only four require mathematics skills, such as dividing by a fraction (see Appendix). Most science items are purely non-mathematical, e.g. "When a bird sings, it is most likely singing in order to (a) frighten away other types of birds, (b) mark the bird's territory against the same type of bird, (c) attract insects, or (d) wake up other animals".

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. This displacement effect is what a simple model of time allocation between learning mathematics, learning science, and leisure predicts, reflecting the incentives that are actually intended when introducing central exams. Given this displacement effect, the *difference-in-differences by subject* framework will overestimate the effect of a *general* introduction of CEEs. Strictly speaking, we are only able to measure the size effect of a *partial* introduction of CEEs unless we can keep constant all inputs invested in learning science. These inputs are only partly observable; e.g. as students' self-reports of the number of hours spent studying science outside school. Below, we make tentative use of these data, although we do believe that the results must be interpreted with caution. First, the quality of these self-reports is at best unclear; second, studying at home could itself be endogenous if weaker students need longer to do their homework than brighter ones.

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

While it is implausible to assume differences in relative innate mathematics versus science abilities between federal states when considering the entire student population, such differences do become more likely when examining selective sub-samples of students – as we do in the present study.[6] For example, 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 α, see above) may have better mathematics skills than their peers in low-α states, but comparable science skills. By controlling for α in our difference-in-differences framework, we are able to control directly for this type of composition effect.

Another important assumption for interpreting the difference-in-differences by subject as the effect of CEEs is that mathematics and science achievement in TIMSS are measured on

[6] We are particularly thankful to one referee for pointing out this possibility.

the same scale and that it is thus feasible to calculate the difference. Below, we examine the robustness of our estimates against violations of this assumption by converting the national Rasch scores to exact quantiles and using the differences therein as dependent variables.

Finally, one difficulty remains to be discussed. Since exit examinations in Saxony generally assess both mathematics and science, we excluded these cases from our analysis. (Alternatively, we could have used a dummy variable for Saxony, leading to only minor changes in our results.) Between 25 percent (in *Realschule*) and 40 percent (in *Hauptschule*) of all students in Bavaria take central exit examinations in science. Unfortunately, the data available do not indicate which of the students in the present sample will take the exam. One way of dealing with this problem would be to exclude all students in Bavaria from our regression. However, we are reluctant to do this for two reasons. First, Bavaria has the largest sample size of all CEE states, accounting for about 60 percent of all CEE observations. Second, if Bavaria is excluded from the data, all remaining CEE states are in East Germany. Since four of the five federal states in East Germany have CEEs, eliminating Bavaria from our sample would make it impossible to distinguish the CEE effect from a "former GDR" effect. This is important because schools in the former GDR appear to have a slightly different tradition in the way science is taught.[7]

Of course, estimates of the CEE effect in our difference-in-differences by subject framework will be biased downwards if part of our sample does in fact take a CEE in science. Thus, in addition to an estimate using all Bavarian students, which gives some lower bound for our parameter of interest, we provide two further estimates in which we eliminate those Bavarian students from the sample who are likely to choose science as a CEE subject. First, we discard *Hauptschule* students who strongly agree with the statement that they "usually do

[7] Note that the Eastern German students in the present sample entered primary school before the collapse of the Berlin Wall in 1989. Further, recent analyses from PISA show that students in East Germany perform significantly better on a specific national set of science items than on the international science items. In West Germany, no such difference was observed (Baumert et al. 2002).

well" in biology or physics (45 percent), and *Realschule* students who strongly agree with the statement that they "usually do well" in physics (13 percent). This selection is based on the assumption that those who *believe they do well* in science are most likely to take a CEE. An upper bound for the CEE effect can be obtained by discarding those Bavarian students who achieve the top 40 percent (*Hauptschule*) or 25 percent (*Realschule*) science scores in TIMSS. The assumption here is that those who *actually do well* in science are most likely to take a CEE. Roughly 15 percent of all cases are eliminated in both regressions. The two indicators for potential test taking thus overlap only partially.

5. Results

The results of our estimates are presented in Tables 3 and 4. Table 3, column (1) contains an estimate for the simple difference in mathematics achievement between students in states with and without a CEE in mathematics.[8] The difference amounts to .50 standard deviations, more than the equivalent of an entire school year (.42 standard deviations). Note that this difference is already estimated net of any student background and composition effects. Wößmann (2002) reports CEE effects of a similar magnitude in the international TIMSS micro-data, while Bishop's (1997) comparison of Canadian provinces with and without CEEs suggests the effect to be about four-fifth of a grade-level equivalent.

All of our background variables have the expected effects on the students' mathematics scores. *Realschule* students, selected on the basis of their primary school achievement, perform much better than their peers in *Hauptschule*. The number of books at home is used as a proxy for the parents' intellectual background because it usually has more explanatory power for the children's achievement than formal education does. In fact, it emerges that

[8] Saxony is already excluded from column (1). Moreover, we have limited the sample to all observations with non-missing values in the broadest set of variables used in our regressions (column (3)). Neither of these restrictions affect our results.

students from homes with over 200 books are almost one school year ahead of their peers from homes with less than ten books. Students from immigrant backgrounds perform slightly worse than others, as do students who already have repeated a grade. Male students tend to outperform female students, and West German students outperform their East German peers.[9]

--- Tables 3 and 4 about here ---

The correlations between central exit examinations and student achievement reported in column (1) could be driven by unobservables that are correlated with CEEs. In order to disentangle this correlation from causation, we now turn to our difference-in-differences by subject estimates.

As argued above, the main advantage of this estimator is that every student serves as his or her own control group, in that he or she takes a centralised examination in mathematics but not in science. As described above, we implemented this estimator simply by calculating the difference in the mathematics and science scores for each student and regressing this difference on a set of explanatory variables and a "CEE-in-mathematics" dummy. In contrast to the "levels" regression in column (1), we now use the ratio between the cumulative number of mathematics lessons and a proxy for the number of science lessons instead of the absolute number of mathematics lessons.[10] The estimates are listed in column (2) of Table 3. The first thing to note is that the coefficient for CEE remains positive and significant at the 10 percent level. However, its size drops from .50 standard deviations in the simple differences estimator to just .08 standard deviations, or about one-fifth of a grade-level equivalent. As discussed above, this estimate is likely to be biased downwards because some students in the sample do take a central exam in science. Hence, we consider this to be a lower-bound estimate of the true CEE effect.

---

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

[10] This proxy equals the total number of lessons minus mathematics lessons minus German lessons. Separate information on science lessons was not available.

In columns (4) and (6), we discarded Bavarian students who are likely to take a central exit examination in science. Elimination of those who say they "usually do well" results in a somewhat larger CEE effect of about one-third of a grade-level equivalent (0.13 standard deviations, see column (4)). An upper limit of the CEE effect is possibly provided in column (6), where Bavarian students who actually do well in science are excluded from the sample. Here, the effect is substantially larger than in the preceding columns, and well above one-half of a grade-level equivalent. This was to be expected, because the sample in question was limited to students with a poor absolute performance in science.

The amount of time that students spent studying science at home (columns (3), (5), and (7)) was controlled for in order to eliminate possible displacement effects from our estimates. If inputs into learning science are kept constant, the estimated effect of CEEs on the difference between mathematics and science achievement can be interpreted as the effect of a *general* introduction of CEEs. Otherwise, we are only able to identify the effect of a *partial* introduction, namely in mathematics and German. Contrary to our expectations, controlling for the time spent studying science at home does not change the estimated CEE effect substantially. The number of hours itself has no effect on the mathematics-science score differential (unless students claim to spend longer than an implausible five hours per day studying science). As we have already mentioned, given the fact that these are self-reported data and that the variable itself may be endogenous, these results should be treated with caution.

While the focus is clearly on the CEE variable, other parameter estimates in columns (2) to (7) are also worth noting. α is always positive and significant, implying that mathematics ability (or performance in primary school) is more important for sorting students into secondary school types than science ability. The *Realschule* dummy is positive and significant, indicating that *Realschule* students perform relatively better in mathematics than in science. The number of books dummies have coefficients that are negative and that increase slightly in absolute terms. In other words, students from better educated homes perform relatively better in science. Students with an immigration background are relatively better at mathematics. This may be due to an insufficient command of the German language, which is less relevant for mathematics than for science.

Table 4 replicates the regressions in Table 3, taking exact percentiles as the dependent variable. For example, the CEE coefficient of 0.145 in column (1) indicates that the average student from a CEE state represents a percentile on the common mathematics score distribution that is 14.5 percentage points above that of the average student from a non-CEE state. The coefficient of 0.024 in column (2) indicates a rank differential between mathematics and science that is 2.4 percentage points larger in CEE states than in non-CEE states. Note that this is only slightly more than one-fifth of a grade-level equivalent, which accounts for 10.6 percentage points on the mathematics distribution. Again, we observe increasing estimates for the CEE effect when the sample is confined to those who are likely not to take the science exam. The percentile regressions are again insensitive to the introduction of the "hours spent studying science at home" variable to control for a potential displacement effect. To summarize, converting the national Rasch scores into exact quantiles and using differences therein as dependent variables changes the quantitative nature of our results only slightly, thus lending further support to the claim that CEEs improve student achievement.

6. Conclusion

This paper discusses the benefits of central exit examinations (CEEs) for academic achievement in lower secondary education. The theoretical benefits of central examinations are convincing. However, it is not easy to identify the causal effect of CEEs empirically. Unlike earlier studies, we make use of regional institutional variation in Germany, allowing us to develop a unique identification strategy to estimate the causal effect of CEEs on academic performance. In the Germany school system, only some states have CEEs, mostly in the core subjects of 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. Various possible identification strategies, all difference-in-differences estimators, are discussed.

Comparison of simple test results indicates that students in German CEE states clearly outperform those in non-CEE states (by approx. 0.5 standard deviations or the equivalent of one and a quarter school years). However, this also applies to a somewhat lesser extent to subjects that are not tested in central examinations, such as science. We therefore propose a difference-in-differences estimator that interprets the difference in mathematics and science achievement in TIMSS in CEE states compared to the same difference in non-CEE states as the causal effect of central examinations on achievement. Depending on the sample definition and specification, the average causal effect of CEE on mathematics achievement is estimated to lie between a lower bound of one-fifth and an upper bound of two-thirds of a grade-level equivalent. Our preferred estimate of the CEE effect is about 0.13 standard deviations or one-third of a grade-level equivalent.

The gap between the raw difference between states with and without CEEs and what we identify as the causal effect of CEEs is fairly sizable. Thus, caution is warranted when interpreting observed differences between states with or without CEEs as the effect of CEEs on student achievement. Much (but not all) of the correlation between CEEs and student performance seems to be driven by the importance attached to a good school education in the various German states.

Still, our empirical findings suggest that introducing central exit examinations will raise average student achievement significantly. Central exit examinations would thus appear to be a viable and cost-effective alternative to other measures discussed for boosting student achievement, such as decreasing class sizes.

References:

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M.; Schiefele, U.; Schneider, W., Tillmann, K.-J., Weiß, M., 2002, *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske+Budrich.

Baumert, J., Watermann, R., 2000, Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. Ch. 8 in J. Baumert, W. Bos, and R. Lehmann (ed.), *TIMSS - Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske+Budrich.

Bishop, J.H., 1997, The effect of national standards and curriculum-based exams on achievement. *American Economic Review* 87, 260-264.

Bishop, J.H., 1999, Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review* 6, 349-401.

Costrell, R.M., 1994, A simple model of educational standards, *American Economic Review* 84, 956-971.

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

Effinger M.R., Polborn, M.K., 1999, A model of vertically differentiated education, *Journal of Economics* 69, 53-69.

Frenck, I., 2001, *Stundentafeln der Primar- und Sekundarstufe I im Ländervergleich - eine empirische Studie am Beispiel der Fächer Deutsch und Mathematik.* Unpublished master thesis, Universität Essen.

Glewwe, P., Ilias, N., Kremer, M., 2003, *Teacher Incentives*, NBER Working Paper 9671.

Hanushek, E.A., 1996, School Resources and Student Performance, pp.43-73 in Gary Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington, D.C.: Brookings Institution.

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

Hoxby, C.M., 2003, School Choice and School Productivity (Or, Could School Choice be a Rising Tide that Lifts All Boats, in C. Hoxby (ed) *The Economics of School Choic*e, Chicago: 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, G., Boele, K., 2001, *The Education System in the Federal Republic of Germany 2000.* German EURYDICE Unit, Bonn. http://www.kmk.org/dossier/dossier_2000_engl_e-book.pdf [2002, July 1].

Jürges, H., Schneider, K., 2004, International differences in student achievement: An economic perspective, *German Economic Review* 5, 357-380.

Jürges, H., Richter, W.F., Schneider, K., 2004, *Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality*, CESifo Working Paper 1296.

Krueger, A.B., 1999, Experimental Estimates of Education Production Functions, *Quarterly Journal of Economics* 114, 497-532.

Krueger, A.B., 2003, Economic Considerations and Class Size, *Economic Journal* 113, F34-F63.

Krueger, A.B., Zhu P., 2004, Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist* 47: 658-698.

Lavy, V., 2002, Evaluating the Effect of Teachers group performance incentives on pupil achievement, *Journal of Political Economy* 110, 1286-1317.

Lavy, V. 2003, *Paying for performance: The effect of individual financial incentives on teachers' productivity and students' scholastic outcome*. Mimeo, Hebrew University of Jerusalem.

Mullis, I.V.S, Martin, M. O., Fierros, E. G., Goldberg, A. L., and Stemler, S. E., 2000, *Gender Differences in Achievement.* International Study Center, Lynch School of Education, Boston College.

Wößmann, L., 2002, *Central Exams Improve Educational Performance*: International Evidence. Kiel Discussion Papers 397.

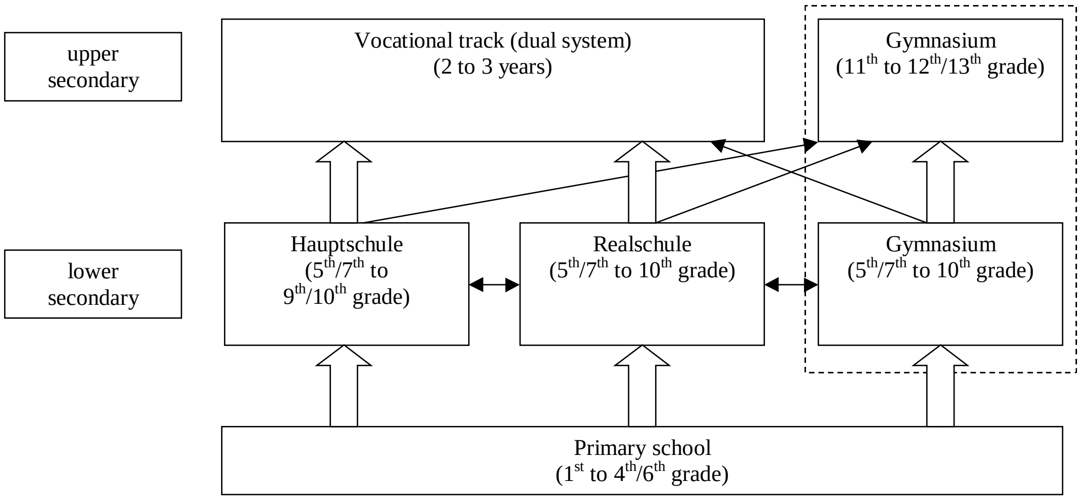


Figure 1: The German school system

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

	Hauptschule	*Realschule*	*Gymnasium* (*Abitur*)	8th grade students in *Hauptschule* tracks [c]	8th grade students in *Realschule* tracks [d]	8th grade students in *Abitur* tracks [e]
Baden-Württemberg (BW)	G/M/F/O	G/M/F	A	34.0	31.7	28.4
Bavaria (BY)	G/M/F/S[b]/O	G/M/F/S[b]/O	A	37.1	31.0	26.9
Berlin (BE)				21.5	31.1	42.8
Brandenburg (BB)				16.9	32.3	45.2
Bremen (HB)				26.8	30.9	36.1
Hamburg (HH)				22.6	26.2	43.3
Hesse (HE)				22.5	34.0	38.6
Lower Saxony (NI)				31.3	34.9	28.0
Mecklenburg-West Pomerania (MV)		G/M/F	A	15.7	48.9	29.3
North Rhine-Westphalia (NW)				29.3	30.6	35.3
Rhineland-Palatinate (RP)				37.2	28.4	29.5
Saarland (SL)			A	24.4	37.0	34.0
Saxony (SN)	G/M/F	G/M/S	A	12.8	51.0	30.1
Saxony-Anhalt (ST)		G/M	A	9.4	50.9	32.8
Schleswig-Holstein (SH)				31.4	34.4	28.5
Thuringia (TH)	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

[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: Descriptive Statistics (weighted)

	non-CEE		CEE		Mean Diff.	t-value [a]
Variable	Mean	StdDev	Mean	StdDev		
Mathematics score	-0.113	0.936	0.320	0.969	0.433	4.102**
Science score	-0.091	0.957	0.304	0.910	0.395	4.375**
Sex (1 = female)	0.489		0.447		-0.043	-1.303
Books at home: 0-10	0.178		0.186		0.009	0.499
Books at home: 11-25	0.295		0.330		0.035	1.775†
Books at home: 26-100	0.168		0.169		0.001	0.048
Books at home: 101-200	0.238		0.217		-0.021	-0.947
Books at home: 200+	0.121		0.098		-0.023	-1.301
Immigrant background (both parents born abroad)	0.189		0.093		-0.096	-3.181**
School type (1 = *Realschule*)	0.532		0.549		0.016	0.188
Grade (1 = 8th grade)	0.495		0.477		-0.018	-0.208
Repeated grade at least once	0.356		0.215		-0.141	-5.171**
Science at home=0 hours/day	0.171		0.189		0.018	0.785
Science at home<1 hours/day	0.644		0.633		-0.011	-0.456
Science at home=1-2 hours/day [b]	0.170		0.164		-0.006	-0.323
Science at home=3-5 hours/day	0.009		0.012		0.003	0.805
Science at home>5 hours/day	0.006		0.002		-0.004	-1.366
East Germany	0.040		0.306		0.266	4.045**
Cumulative maths lessons (in 1000s)	1.198	0.121	1.279	0.105	0.081	4.071**
N obs.	1767		1142		2909	

† p<10%; * p<5%; ** p<1%; [a] t-values allow for clustering on the class level; [b] response categories did not cover 2-3 hours.

Table 3: CEE effects on student achievement in mathematics

	simple diffs[a]	diff-in-diffs I[b]	diff-in-diffs I[b]	diff-in-diffs II[b,c]	diff-in-diffs II[b,c]	diff-in-diffs III[b,d]	diff-in-diffs III[b,d]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CEE	**0.499**	**0.082**	**0.083**	**0.132**	**0.133**	**0.256**	**0.257**
	(7.21)**	**(1.70)†**	**(1.73)†**	**(2.64)****	**(2.65)****	**(5.06)****	**(5.10)****
Grade	0.419	-0.009	-0.011	-0.006	-0.008	0.003	0.002
	(5.02)**	(0.25)	(0.29)	(0.17)	(0.21)	(0.09)	(0.06)
α	2.527	0.999	0.995	1.007	1.005	1.042	1.039
	(4.99)**	(3.17)**	(3.16)**	(3.19)**	(3.20)**	(3.25)**	(3.25)**
Realschule	0.743	0.159	0.161	0.134	0.137	0.117	0.120
	(13.25)**	(4.05)**	(4.13)**	(3.42)**	(3.51)**	(3.02)**	(3.10)**
11-25 books	0.082	-0.110	-0.104	-0.114	-0.108	-0.094	-0.089
	(1.53)	(2.22)*	(2.11)*	(2.22)*	(2.12)*	(1.85)†	(1.77)†
26-100 books	0.218	-0.146	-0.143	-0.126	-0.123	-0.122	-0.121
	(3.90)**	(2.66)**	(2.63)**	(2.22)*	(2.19)*	(2.18)*	(2.17)*
101-200 books	0.314	-0.156	-0.152	-0.156	-0.152	-0.146	-0.143
	(5.52)**	(2.71)**	(2.64)**	(2.61)*	(2.52)*	(2.52)*	(2.45)*
200+ books	0.349	-0.200	-0.200	-0.183	-0.184	-0.175	-0.177
	(6.06)**	(3.78)**	(3.79)**	(3.35)**	(3.38)**	(3.15)**	(3.20)**
Immigrant background	-0.075	0.336	0.335	0.350	0.349	0.323	0.322
	(1.37)	(7.12)**	(7.01)**	(7.23)**	(7.18)**	(6.71)**	(6.64)**
Repeated grade	-0.114	-0.078	-0.077	-0.075	-0.073	-0.071	-0.070
	(3.27)**	(2.16)*	(2.11)*	(1.99)*	(1.95)†	(1.95)†	(1.91)†
Female	-0.297	0.079	0.086	0.075	0.082	0.068	0.074
	(7.62)**	(2.42)*	(2.59)*	(2.19)*	(2.34)*	(2.05)*	(2.18)*
East	-0.489	-0.134	-0.137	-0.186	-0.187	-0.296	-0.298
	(4.65)**	(1.85)†	(1.89)†	(2.50)*	(2.52)*	(3.88)**	(3.91)**
Cum. maths lessons	-0.976						
	(2.51)*						
Relative cum. maths lessons		-0.670	-0.625	-0.200	-0.178	-0.035	-0.010
		(0.74)	(0.70)	(0.23)	(0.20)	(0.04)	(0.01)
Science at home:<1 hr/day			-0.037		-0.036		-0.018
			(0.73)		(0.65)		(0.32)
Science at home:1-2 hrs/day			0.001		-0.043		-0.056
			(0.01)		(0.44)		(0.54)
Science at home:3-5 hrs/day			-0.026		-0.030		-0.025
			(0.68)		(0.74)		(0.61)
Science at home:>5 hrs/day			0.722		0.703		0.680
			(3.01)**		(2.89)**		(2.67)**
Constant	-0.903	-0.408	-0.403	-0.526	-0.515	-0.580	-0.572
	(2.08)*	(1.40)	(1.39)	(1.86)†	(1.82)†	(2.05)*	(2.03)*
Observations	2909	2909	2909	2644	2644	2636	2636
R-squared	0.37	0.05	0.05	0.05	0.06	0.07	0.08

Robust t-statistics (accounting for clustering on the class level) in parentheses
† p<10%; * p<5%; ** p<1%
[a] dependent variable: maths score
[b] dependent variable: maths score minus science score
[c] excluding Bavarian students who claimed to be good at science (see text for explanation)
[d] excluding Bavarian students with best science scores (see text for explanation)

Table 4: CEE effects on student achievement in mathematics (percentile effects)

	simple diffs[a] (1)	diff-in-diffs I[b] (2)	diff-in-diffs I[b] (3)	diff-in-diffs II[b,c] (4)	diff-in-diffs II[b,c] (5)	diff-in-diffs III[b,d] (6)	diff-in-diffs III[b,d] (7)
CEE	**0.145**	**0.024**	**0.024**	**0.039**	**0.039**	**0.074**	**0.075**
	(7.66)**	**(1.66)†**	**(1.68)†**	**(2.59)***	**(2.59)***	**(4.71)****	**(4.72)****
Grade	0.106	-0.010	-0.010	-0.008	-0.008	-0.009	-0.009
	(4.43)**	(0.95)	(0.95)	(0.73)	(0.73)	(0.77)	(0.75)
α	0.696	0.277	0.275	0.279	0.279	0.284	0.282
	(4.62)**	(2.97)**	(2.93)**	(2.93)**	(2.93)**	(2.88)**	(2.84)**
Realschule	0.222	0.046	0.047	0.040	0.040	0.026	0.027
	(13.64)**	(3.94)**	(3.97)**	(3.43)**	(3.43)**	(2.15)*	(2.19)*
11-25 books	0.026	-0.027	-0.027	-0.029	-0.029	-0.023	-0.023
	(1.71)†	(2.02)*	(1.96)†	(2.03)*	(2.03)*	(1.58)	(1.57)
26-100 books	0.066	-0.038	-0.038	-0.035	-0.035	-0.034	-0.035
	(4.19)**	(2.42)*	(2.43)*	(2.14)*	(2.14)*	(2.04)*	(2.09)*
101-200 books	0.100	-0.032	-0.032	-0.033	-0.033	-0.035	-0.035
	(6.01)**	(1.93)†	(1.91)†	(1.90)†	(1.90)†	(1.99)*	(1.99)*
200+ books	0.107	-0.048	-0.049	-0.043	-0.043	-0.042	-0.044
	(6.17)**	(3.16)**	(3.22)**	(2.79)**	(2.79)**	(2.58)*	(2.67)**
Immigrant background	-0.017	0.092	0.092	0.096	0.096	0.094	0.094
	(1.04)	(6.76)**	(6.59)**	(6.77)**	(6.77)**	(6.76)**	(6.62)**
Repeated grade	-0.034	-0.020	-0.020	-0.017	-0.017	-0.020	-0.019
	(3.43)**	(1.88)†	(1.85)†	(1.47)	(1.47)	(1.69)†	(1.64)
Female	-0.084	0.033	0.034	0.033	0.033	0.033	0.034
	(7.46)**	(3.19)**	(3.24)**	(3.01)**	(3.01)**	(3.09)**	(3.11)**
East	-0.154	-0.046	-0.047	-0.061	-0.061	-0.095	-0.096
	(5.03)**	(1.97)†	(1.99)*	(2.50)*	(2.50)*	(3.80)**	(3.82)**
Cum. maths lessons	-0.216						
	(1.88)†						
Relative cum. maths lessons		0.027	0.036	0.142	0.142	0.315	0.324
		(0.10)	(0.13)	(0.52)	(0.52)	(1.11)	(1.15)
Science at home:<1 hr/day			-0.001	-0.002	-0.002		0.005
			(0.09)	(0.13)	(0.13)		(0.30)
Science at home:1-2 hrs/day			0.002	-0.009	-0.009		-0.007
			(0.07)	(0.28)	(0.28)		(0.20)
Science at home:3-5 hrs/day			-0.004	-0.004	-0.004		-0.000
			(0.38)	(0.36)	(0.36)		(0.04)
Science at home:>5 hrs/day			0.165	0.163	0.163		0.165
			(2.38)*	(2.26)*	(2.26)*		(2.09)*
Constant	0.161	-0.175	-0.174	-0.205	-0.205	-0.253	-0.255
	(1.23)	(2.05)*	(2.05)*	(2.44)*	(2.44)*	(2.94)**	(2.97)**
Observations	2909	2909	2909	2644	2644	2636	2636
R-squared	0.37	0.05	0.05	0.05	0.05	0.06	0.07

Robust t-statistics (accounting for clustering on the class level) in parentheses
† p<10%; * p<5%; ** p<1%
[a] dependent variable: maths score percentile
[b] dependent variable: maths score percentile minus science score percentile
[c] excluding Bavarian students who claimed to be good at science (see text for explanation)
[d] excluding Bavarian students with best science scores (see text for explanation)

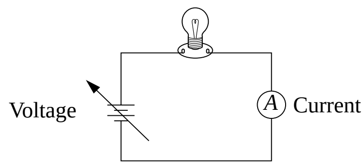
Appendix: TIMSS population 2 science items involving mathematics skills

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

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

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

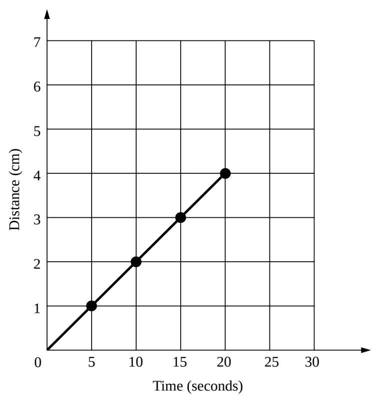
M12 – Some students used an ammeter *A* to measure the current in the circuit for different voltages.



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

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

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



If the ant keeps moving at the same speed, how far will it have traveled at the end of 30 seconds?
A. 5 cm
B. 6 cm
C. 20 cm
D. 30 cm

Z1 – It takes 10 painters 2 years to paint a steel bridge from one end to the other. The paint that is used lasts about 2 years, so when the painters have finished painting at one end of the bridge, they go back to the other end and start painting again.
a. Why MUST steel bridges be painted?
b. A new paint that lasts 4 years has been developed and costs the same as the old paint. Describe 2 consequences of using the new paint.
