Assessment drives learning

The effect of central exit exams on curricular knowledge and mathematical literacy

Hendrik Jürges, University of Wuppertal, Germany

Kerstin Schneider, University of Wuppertal and CESifo, Germany[1]

Martin Senkbeil, IPN, Kiel, Germany

Claus H. Carstensen, University of Bamberg, Germany

Final version - July 2011

Abstract: In this paper, we use data from the German PISA 2003 sample to study the effects of central exit examinations on student performance and student attitudes. Unlike earlier studies we use (i) a value-added measure to pin down the effect of central exit exams on learning in the last year before the exam and (ii) separate test scores for mathematical literacy and curriculum-based knowledge. The findings indicate that central exit exams improve curriculum-based knowledge but do not affect mathematical literacy. Students, although showing a better performance, are less intrinsically motivated in school.

Keywords: Central exit exams, value added, achievement, student motivation

JEL-Codes: I28, H42, D02

[*] Corresponding author.
Kerstin Schneider, Schumpeter School of Business and Economics, University of Wuppertal
Gaußstr. 20, 42119 Wuppertal, Germany, Phone: +49-202-439-2483, Fax: +49-202-439-3649
E-mail addresses: juerges@wiwi.uni-wuppertal.de (H. Jürges), schneider@wiwi.uni-wuppertal.de (K. Schneider), senkbeil@ipn.uni-kiel.de (M. Senkbeil), claus.carstensen@uni-bamberg.de (C. Carstensen)

1. Introduction

In this paper, we present new evidence on the effects of state-mandated graduation exams or central exit examinations (CEEs) on educational outcomes in German secondary schools. The existing literature on CEE effects reaches fairly unanimous conclusions. Theoretically, centralized examinations yield positive effects on student achievement because of incentive effects on teachers and students (Costrell, 1997; Effinger and Polborn, 1999; Jürges, Richter, and Schneider, 2005). This prediction is matched by the empirical literature, although estimated effect sizes vary substantially depending on the data and identification strategy (Alexander, 2000; Bishop, 1997, 1999; Jürges, Schneider, and Büchel, 2005; Wößmann, 2005). Given the beneficial effect on student achievement and provided that introducing central exit exams requires little additional monetary resources, one might be tempted to close the case. But low monetary costs do not mean that central exit exams are costless. Achievement gains might be costly for students and teachers if they put forth more effort. It is certainly of interest to find out if such costs of higher achievement exist, what they are and who eventually bears them. For instance, one would like to know whether students actually work harder or whether learning is becoming less fun (Jürges and Schneider, 2010).

The present study makes several contributions to the empirical literature on central exit exams. First, we use longitudinal data from the PISA-I Plus study, an extension of PISA 2003 – with performance measured in ninth and tenth grade – to estimate the effect of central exit exams on student performance. Thus, in contrast to earlier studies, we can measure student achievement as value-added in the last year before the exit examination in nonacademic tracks. Second, we control for general cognitive skills of the students, as PISA-I 2003 also includes a cognitive ability test. Our third contribution is to differentiate achievement on test items referring to basic skills (mathematical literacy) and test items that reflect the German mathematics curriculum. This difference is important, because "PISA considers student knowledge […] not in isolation but in relation to students' ability to reflect on their knowledge and experience and to apply them to real world issues." (OECD, 2003, p. 24). This type of knowledge is different from the German curriculum, which is the binding standard for schools. In contrast to the regular PISA questions, the German curriculum requires substantial technical and conceptual modeling abilities in mathematics (Blum et al., 2004). With our data, we are able to test whether mathematical literacy and curricular knowledge are similarly affected by external standards. In that respect we can add new evidence on the "teaching to the test" question, or put it differently, on the importance of the curriculum as the binding standard.

Fourth, we exploit the fact that the German secondary school system is characterized by a fairly rigid system of early academic tracking. Our data contain information on students in nonacademic tracks (who graduate after grade 10) and students in academic tracks (who continue schooling until grade 12 or 13). While for nonacademic tracks we measure value-added immediately before exit examinations, exit exams in academic tracks are still two or three years ahead. For students in nonacademic tracks who are in 9th and 10th grade, the incentive effects of central exit exams are therefore expected to be stronger than for their peers in the academic track. Moreover, nonacademic track students in non-CEE states graduate without taking an additional *exit* exam. A leaving certificate is awarded upon successfully completing the final grade. At the end of the academic track, however, there are exit exams in all German states, but exit exams can be administered centrally or drafted and administered locally by teachers and schools. Finally, we aim at disentangling the effect of central exit exams on student achievement by studying student attitudes and behaviour as potentially important mediators in the relationship between central exit exams and student performance.

Our findings are in line with earlier results on CEE-effects on student achievement in mathematics (Jürges, Schneider, and Büchel, 2005; Jürges and Schneider, 2010) and yield new insights into the mechanics of external standards. We show that the effects are significant only for curricular knowledge, whereas mathematical literacy appears not to be affected. Further, central exit exams appear to matter significantly only for students in nonacademic tracks. Students in 9th and 10th grade of the academic track, where students graduate after grade 12/13, are hardly affected by central exit exams. Moreover, earlier findings that students bear costs in form of being less intrinsically motivated and feeling higher pressure are confirmed.

The paper proceeds as follows: in Section 2, we describe relevant features of the German school system. In Section 3 we give a brief description of the PISA-I-Plus data. Sections 4 and 5 show our empirical strategy and estimation results. Finally, we draw some conclusions in Section 6.

2. A primer on the German school system and recent reforms

In the context of our analysis, the German school system has two important characteristics (a comprehensive description of the German school system can be found in Jonen and Eckardt (2004)). First, it is characterized by a distinct federal structure. The 16 federal states have far-reaching autonomy in all education matters. The federal influence on education policy is rather weak and has been further weakened by the latest reform of federalism in Germany in 2006. As a result, the German school system is characterized by strong regional differences that tend to inhibit the mobility of families and teachers across German states. The diversity across states is often criticized, but the coexistence of various school systems within one country can be exploited for evaluation studies.

One important cross-state difference is the existence of central exit examinations. In some of the states, central exit examinations exist since the end of World War II. Although central exit examinations have been an ideological battlefield in Germany ever since, there was no reform until very recently. While it was always argued that students from states with central exit examinations performed much better, reliable comparisons of student achievement across states with and without central exit exams (not to speak of causal analyses) have been virtually non-existent. This was mainly due to the lack of nationwide standardized tests. In response to this lack of data, the Conference of State Education Ministries (KMK) commissioned an extension to the OECD PISA 2000 study (PISA-E), boosting sample sizes and including questions that were more specific to the German schooling system. The results of PISA-E in 2000 did indeed reveal large differences in test scores between states and in particular between states with and without exit exams. In response, all except one federal state have introduced central exit exams, however, without any reference to empirical evidence that observed differences in achievement are causally linked to central exit examinations. Further, a group of seven (since 2008: twelve) German federal states have introduced regular standardized tests of student skills at different grades in primary and secondary schools.

One reason for this recent zeal in education policy was the tremendous effect of the publication of the international PISA 2000 results on the German public. In contrast to German self-perception, student achievement proved to be at most average in international comparison. Since then, concerns were growing about a declining quality of education in German schools and a decreasing international competitiveness of the German labor force.

A second salient feature of the German school system is the rigid tracking in secondary schools. After primary school (usually at the age of ten), students are allocated to one of three types of secondary schools: two nonacademic tracks (basic and intermediate), and an academic track. Requirements for graduation differ across these tracks. Graduating from the academic track (in grade 12 or 13), which is equivalent to a general university-entrance certificate, requires passing an exit exam in every state. In some states, these exams are centrally administered, and in others they are designed at each individual school (subject to state-wide guidelines). A typical exit examination in academic tracks consists of four to five subjects. The choice of subjects is limited and varies from state to state, but mathematics was not mandatory in 2003 (the year that is relevant for our analysis).

--- about here Table 1 ---

Leaving certificates for basic tracks are awarded upon completion of grade 9 or 10, depending on the state. Leaving certificates for intermediate tracks are awarded after grade 10. In some states, the leaving certificate requires passing a central exit examination. Table 1 describes the situation in 2003/04, the years in which the PISA-I-Plus data were collected. Hesse and Brandenburg were excluded from the analysis because they introduced CEEs in 2003/04, the year of the PISA-I study. Hence students were in grade 9 immediately before CEEs were introduced and in grade 10, when CEEs just had been introduced. It is therefore not clear how to treat the two states. Seven states had central exit examinations at the end of the intermediate track, and six had CEEs at the end of the basic track. In contrast to academic tracks, there is only limited choice of subjects in the exit exams. Written exams in German and mathematics are compulsory subjects in all central exit examinations.

3. Data

We use data from PISA-I-Plus, an extension to PISA 2003 for Germany. Compared to the original PISA study design, this extension has several distinct characteristics. First, in contrast to the original PISA sample design that includes 15 year olds, the PISA-I-Plus baseline sample consisted exclusively of 9th-graders. Second, students had to sit two test days. On the first day, they were given the common set of international mathematics, science, and reading items. On the second day, students were given national test items. While the international test items focused on "mathematical literacy" as defined by the OECD, i.e. the capacity to use mathematical skills in everyday contexts, the national items were designed to test curricular knowledge. Third, students were tested again one year later in 2004, when they were in 10^{th} grade. This allows us to use achievement gains as outcomes.

Although our data have many advantages compared to the regular PISA data, the longitudinal design also has some limitations. Students who repeat 9^{th} grade, who change schools, or who leave school after grade 9 are not followed. Due to the latter restriction, we excluded all students in basic tracks from our analyses. The sample thus becomes selective in terms of socioeconomic background and achievement. Compared to the full sample of 9^{th} graders, there are more girls than boys, less immigrant students and higher average test scores. To account for selectivity based on observable covariates, we use survey weights, so that the samples are representative for the secondary school types in Germany (Prenzel et al., 2006).

The sample used in the analysis consists of 4,928 students, of which 2,599 students are enrolled in nonacademic track schools (with graduation after grade 10) and 2,329 students are enrolled in the academic track school (with graduation after grade 12 or 13 years). The data are described in more detail below.

4. Effect of exit exams on student achievement

Empirical strategy

In this section we report empirical estimates for the effect of exit examinations on student achievement. The basic specification is the simple education production function:

(1) $$y_{iGc} = a_{iGc} + \beta d_{iGc} + \gamma X_{iGc} + e_{iGc}$$

This equation models mathematics achievement *y* of student *i* in grade *G* (grade 10) in classroom *c* as a function of unobserved skill *a*, exam regime *d*, family and other background variables *X*, and a random (measurement) error term *e*. For sake of simplicity, we leave out teacher, school, peer, and federal state effects. Note that unobserved skill, *a,* is modeled as a time-varying variable, i.e., skill is explicitly allowed to evolve over time. We can think of unobserved skill as being a function of the history of all external influences on ability (family, peers, teachers, schools) and of innate learning capacity. A value-added specification of equation (1) that includes achievement in grade $G-1$ (i.e. grade 9) along with contemporaneous family characteristics and a contemporaneous exam regime dummy is specified in (2).

(2) $$y_{iGc} = \alpha_{iG} + \theta y_{i(G-1)c} + \beta d_{iGc} + \gamma X_{iGc} + e_{iGc}$$

where α_{iG} denotes innate learning capacity of student *i* in grade *G* and θ is a parameter that measures how past experiences and acquired knowledge persist into the future. Note that $y_{i(G-1)c}$ also contains past effects of the exam regime. Hence by estimating equation (2), we control for long-term effects of central exit examinations on student attainment. More importantly, including achievement in grade 9 helps to identify the causal effect of CEEs. The potential policy endogeneity of CEEs that has been discussed in detail in the literature (e.g. Jürges, Schneider, and Büchel, 2005) is taken care of in equation (2). For instance, if the decision to introduce CEEs depends on the valuation of education in a state and the attitude towards education also affects academic achievement, simple differences between CEE and non-CEE states do not estimate the causal effect of CEEs. Including achievement in grade 9, however, controls for the influence of unobserved parental preferences for education up to grade 9. But clearly, even achievement gains from grade 9 to 10 might depend on differences between states that are unrelated to CEEs. Therefore, we include additional variables to control for state differences. In particular we include an indicator variable for former East Germany, per capita spending on schooling in lower secondary school, and the average class size in lower secondary school. The East Germany dummy is included because most teachers and parents in the Eastern states where raised and educated in the former East Germany. This might have affected their attitude towards education. In addition, educational spending in 2003 could be viewed as an indicator of how much education is valued. Spending per student was between 4,800 € in the Saarland and 6,500 € in Hamburg. Similarly, average class sizes in lower secondary schooling vary as well. The range was between 17 in the Eastern state of Saxony-Anhalt and 24 in Hamburg. Hence, utilizing the value added nature of our data and including additional state controls, we can be fairly confident that our coefficient of interest, β, shows the effect of central exit examinations on achievement in grade 10.

In the following, we estimate four different versions of equation (2) – with standard errors that account for stratification and clustering of the sample. First, we differentiate between students in nonacademic and in academic tracks. For each track type, we then estimate a production function for mathematical literacy and for curricular knowledge. For each individual, we have two test scores for each concept, one in grade 9 and one in grade 10. In equation (2), y_{iGc} represents the test score in grade 10, and $y_{i(G-1)c}$ represents the test score in grade 9. Note that test items in grades 9 and 10 were not the same. Each test was an independent assessment of mathematical literacy and curricular knowledge – standardized to mean 50 and standard deviation 10. Hence we do not measure absolute but relative achievement gains. Additional information on the test and the test items can be found in Blum et al. (2004). Descriptive statistics for test scores, by exam and track type, are shown in Table 2. Two results stand out. First, academic track students have on average much higher PISA test scores in all tests, independent of the exam type and the grade in which the test is taken. The difference amounts to one standard deviation in the overall test score distribution. Second, students in CEE states have higher average PISA test scores, independent of the track type. The difference is between 0.1 and 0.2 standard deviations.

For innate learning capacity α_{iG}, we use two proxy variables: an index for cognitive skills (measured in grade 9 that is based on test items on figure analogies, as part of a non-verbal IQ-test (German revised adaptation of the Thorndike and Hagan-cognitive abilities test (Heller and Perleth, 2000)), and a dummy variable that reflects whether a student already had to repeat a grade. For general cognitive skills, Table 2 shows a large difference between nonacademic and academic track students – independent of the exam type. Repeating classes is more common in nonacademic than in academic tracks. However, the general cognitive skills score difference between the two types of tracks is smaller than the test score difference. An explanation for this finding is that the tracked school system in Germany reinforces innate ability differences between students. Another important finding is that cognitive skills do not differ significantly between students in CEE and non CEE states.

As indicators of the social background we use the international socioeconomic index of the highest parental occupational status (HISEI) (Ganzeboom *et al.,* 1992, OECD, 2005) and an immigrant dummy variable that has the value of one if at least one parent was not born in Germany. Table 2 shows that students in academic track schools have considerably higher values of the socioeconomic index. Further, 12 percent of the academic-track students in CEE states are immigrants, which compares to 14 percent in non-CEE-states. This difference is more pronounced for students in the nonacademic track. While in CEE states only 10 percent are immigrants, in non-CEE states the percentage amounts to 21 percent.

--- about here Table 2 ---

4.2 Results

Table 3 shows our regression results. The first two models show regression results with mathematical literacy in grade 10 as dependent variable. In model (1) we restrict the sample to students in nonacademic tracks, whereas model (2) includes only students in academic track schools. The main finding is that there are only small and insignificant coefficients of central exit examinations on mathematical literacy in grade 10. Hence, there is no evidence that central exit examinations in mathematics improve mathematical literacy.

As expected, the strongest predictor for mathematical literacy in grade 10 is mathematical literacy in grade 9. With respect to the other control variables we briefly note that general cognitive ability is a highly significant predictor of mathematical literacy. Also, students who did not repeat a class until grade 9 have significantly higher achievement gains. Students with parents of higher socioeconomic status have slightly higher test scores in mathematical literacy but the differences are not significant. Immigrant students in the nonacademic track have significantly lower test scores on the literacy test and the effect is quite sizable. The coefficient is insignificant for academic track students. Hence, once immigrants manage to attend the academic track educational disparities are no longer significant. The inequality is most visible at the nonacademic basic track, where the immigrant students are also overrepresented. Most likely, the effects of the social background (HISEI) are largely absorbed by general cognitive ability and by test scores measured in grade 9. The gap in mathematical literacy is not further widened in grade 10 (e.g., Ai, 2002; Scott et al., 1995). Finally, coefficients of state level variables are statistically insignificant, except for class size in nonacademic track and regressions pertaining to mathematical literacy. States with on average smaller classes in nonacademic schools do not produce higher value added than states with larger classes. This finding is in accordance with large parts of the class size literature (cf. Hanushek 2002).

--- about here Table 3 ---

In models (3) and (4), the dependent variable is the score on the curricular knowledge test. In contrast to our results for mathematical literacy, we now find a highly significant positive effect of central exit exams on achievement gains in the curriculum-based knowledge test. However, this finding is constrained to students who graduate from the non-academic track, i.e. after grade 10. For academic track students, who pass their final exams after grades 12 or 13, the CEE effect remains insignificant and fairly weak. There are some possible explanations for this lack of effect. First, academic track students have more time until graduation which weakens any incentive effects if students discount the future. Second, even if exams at the end of the academic track school are not administered centrally, there is an exit exam. These exams are subject to the approval by the supervisory authority, so that teachers are not entirely free in setting up the test problems. Hence there might be some achievement effect of the non-central exit exams as well. And third, students in academics track schools might be simply intrinsically motivated to learn, even in the absence of external standards. Hence they cannot further benefit from the extrinsic motivation by CEEs.

In terms of control variables, we find again that general cognitive ability and test scores in grade 9 are the strongest predictors for test scores in grade 10. However, compared to the results obtained for mathematical literacy in (1) and (2), the relative importance of the grade 9 test score grows and the influence of general cognitive skills decreases. This is a plausible result, because we expect general cognitive ability to be a better predictor of general skills (mathematical literacy) than of specific skills (curricular knowledge). Also, nonacademic track students who did not repeat a class until grade 9 have significantly higher achievement gains.

For academic track students, family background is statistically significant in explaining achievement gains in curricular knowledge. For nonacademic track students family background remains insignificant. This is noteworthy because students in academic track schools tend to come from families with comparably high socioeconomic status and only a small proportion of immigrant students attend the academic track. Hence one might expect the differences in socioeconomic background to be more relevant for nonacademic track students. This is not the case. The socioeconomic gap in achievement on the curricular test continues to widen for academic track schools, possibly because parents with a high socioeconomic status are more familiar with the requirements of academic track schools, they can either help their children themselves or organize and finance coaching to prepare for the exams. By the same token, being immigrant is detrimental for progress in curriculum-based knowledge for academic track students. Hence while immigrant students show the same progress in literacy, they still perform worse on the curricular tests, which will be reflected in school grades.

In summary, our analysis in this section suggests that achievement gains in mathematical literacy between grades 9 and 10 are not larger in states with central exit exams. Gains in curricular knowledge, however, are significantly larger in states with central exit exams, but only for students who finish school after grade 10. This insight is important for our understanding of the mechanics of central exams. If students are tested centrally and students and teachers know that the command of curricular knowledge is tested, curricular knowledge increases. Mathematical literacy (as a broader concept) is not in the focus of the German curriculum and hence not tested in the exit exam. Thus there are no strong incentives for students and teachers to improve on the students' broader skills. This is consistent with the education literature: Teachers teach the curriculum if the curriculum is tested (Au, 2007; Jacob, 2005).

We interpret our result as an argument in favor of CEEs, as the results show that students have a better knowledge of the material they are tested on if it is defined in the curriculum. It should be noted that in Germany the curriculum is binding on the state level and teachers cannot adjust the number of mathematics lessons. Thus they cannot divert resources from other subjects to mathematics. Moreover, mathematics teachers rarely teach a second subject in the same class, which also reduces opportunities to teach more mathematics instead of other subjects. Hence with our data, we get an estimate of the net CEE effect.

The analysis also shows that one has to be precise about the contents of the curriculum or the abilities that students are expected to have at the end of schooling. Mathematical literacy as defined by the OECD is not automatically improved by CEEs that test curricular knowledge if the curriculum requires fairly abstract technical and conceptual modeling abilities. However, if CEEs ought to increase mathematical literacy, one might think about means on how to include literacy in the exit exams and hence also in the curriculum.

5. Effect of exit exams on student attitudes

5.1 Empirical Strategy

Having identified the positive effects of CEEs on achievement on curricular knowledge, we now provide some suggestive evidence on the effects of external standards on student attitudes. If students are better in doing math and are at the same time more motivated and have a better self esteem, we would interpret this as additional evidence in favor of external standards. However, while the estimated effect on CEEs on achievement is reliable, the evidence we are providing here is more of descriptive nature. Causal effects cannot be estimated given the PISA-I data.

We use data from the student questionnaires to construct four indices of student attitudes in (grade 9) with respect to the learning climate in mathematics: *Emotions* (anger, anxiety, boredom, despair, achievement pressure), *work habits* (effort, completion of homework, attention in lessons), *motivation* (instrumental motivation, enjoyment, performance orientation), and *self esteem* (self efficacy, self concept). Table 4 contains variable definitions, sample items and descriptive statistics by type of exit examinations and school type. Note that values for each item were standardized to have mean zero and unit standard deviation. Positive values correspond to higher than average agreement to the respective items, negative values correspond to less than average agreement.

Our basic regression model is:

(3) $$r_{iGc} = a_{iGc} + \beta d_{iGc} + \gamma X_{iGc} + e_{iGc},$$

where attitudes r reported by student i in grade G in class c are modeled as a function of general cognitive ability a, the exam regime d and background variables X (socioeconomic and migration background). Again, cognitive ability is approximated by the cognitive skills score and a dummy variable that indicates whether the student has repeated a class. Ability and background are included as control variables, because they might not only affect achievement but also motivation and the self perception with respect to mathematics. As before, standard errors account for stratification and clustering of the sample.

--- about here Table 4 ---

5.2 Results

Table 5 summarizes the results. It turns out that the differences in student traits and attitudes between exam regimes are fairly small in magnitude. Nevertheless, students in states with central exit exams show significantly more often negative emotions like anger, anxiety, and despair. Differences in achievement pressure, however, are not significant.[2] The difference in

[2] One might note, however, that the results in Table 5 are partly sensitive to the inclusion of state-level covariates (here, the East Germany dummy and average class size are significant). For instance, estimates for achievement pressure are substantially larger, positive and – in the case of non-academic students – significant at the 5% level

(intrinsic) motivational orientation as measured by enjoyment is also significant for all school tracks. That does not necessarily imply that students in CEE states are less motivated to learn, but they seem to have less desire to learn for the sake of learning. Moreover, instrumental motivation is also significantly lower in CEE states. The self-concept in mathematical competence is relatively weak for academic students in CEE states but not for nonacademic students, who show higher levels of self efficacy in CEE states.

While the results cannot necessarily be interpreted as causal, it is still interesting to see that negative student attitudes are more prevalent in CEE than in non-CEE states. Moreover, unlike the results for achievement tests, differences between CEE and non-CEE states prevail for nonacademic as well as for academic track students.

Our results are in line with recent studies in which standardised tests (or high-stake tests) are understood as a highly controlling extrinsic motivation strategy. Although this strategy was often found to have the desired effect (better performance), it often also had a number of undesirable side effects, e.g. loss of intrinsic subject motivation, increased test anxiety, increased pressure to perform, lower self-efficacy (Abrams et al., 2003; Ryan, Ryan, Arbuthnot and Samuels, 2007; Ryan and Sapp, 2005). The comparatively small effect size may indicate that central exit examinations do not have a homogenous effect on all students: while they provide an (external) incentive to achieve their maximum performance in a test situation for some students, other students (in particular, those who do not have much self-confidence in their own abilities) may see them as a threat or a hurdle which they cannot manage and this then leads to the undesirable side effects on intrinsic motivation mentioned above (cf., e.g. Abrams et al., 2003; Kellaghan et al., 1996; Ryan et al., 2007).

--- about here Table 5 ---

6. Summary and Conclusion

---

when state-level covariates are excluded. In contrast, the coefficients for instrumental motivation become much larger, and for non-academic students statistically significant. This could be explained by generally larger achievement pressure and higher instrumental motivation in the East.

Using data from the German PISA-I-Plus study of 2003, we have analyzed the effect of central exit exams in mathematics on students' mathematics performance and students' attitudes towards mathematics. The PISA-I-Plus data are superior to data used in earlier studies for two reasons. First, they contain repeated measures of performance in grades 9 and 10 (when exams take place), so that we could measure actual gains in achievement. Second, the data also allow differentiating between achievement on a mathematical literacy test and a test of curricular knowledge. Our analysis shows that central exit exams are significantly related to better curricular knowledge. However, we found no significant effect on mathematical literacy – the type of mathematical knowledge that is regarded to be relevant in everyday situations. This paper thus qualifies earlier findings on the central exam effect in Germany in an important way: measured (incentive) effects of central exit examinations appear to be larger when outcome measures are more in line with what is actually tested in central exit examinations, i.e. the curriculum.

Only few previous studies have attempted to shed light on the question what is driving the positive achievement gains of central exit exams. Theoretically, teachers as well as students should react to central exit exams by increasing effort. Using student self-ratings on behavior, motivation, and attitudes towards learning mathematics, Jürges and Schneider (2010) find differences in general student motivation. There appears to be a downside to CEEs because students in CEE-states do like mathematics less. They find it less easy and more boring than those in non-CEE states. The present study confirms findings that central exit exams are associated with negative student attitudes: students in CEE-states are generally more anxious, are more often in despair, and are actually less motivated to learn.

One important result of our analyses is that differences in achievement due to central exit exams are typically found only in nonacademic tracks. Students and teachers in academic track schools appear to be less affected. There are various possible explanations for this result. One explanation is the difference in years until graduation that is reducing the CEE-effect. Another explanation is that students in academics track schools might be intrinsically more motivated to learn, even in the absence of external standards. Hence they would not further benefit from the extrinsic motivation by CEEs. Rather extrinsic motivation might even crowd out intrinsic motivation. Further, (centrally approved) exit exams existed for students at academic track schools in all German states, even though they were not centrally drafted and administered. Thus, the curriculum was more binding at academic tracks schools even without central exit exams. The most degrees of freedom with respect to the academic level required for graduation exist in nonacademic secondary schools. In non-CEE states no formal exit exam is administered for nonacademic track students. Students graduate as long as they have passing grades on the (teacher drafted and graded) exams during the final year. Hence, not finding CEE effect for academic track students is not a persuasive argument against positive effects of CEEs on student achievement. The significant effect for nonacademic students, however, can be regarded as evidence in favor of achievement raising CEEs. Hence overall, the results of this study support the monitoring of student achievement by standardized tests at all levels of schooling and all levels of tracks, as that enhances performance.

The results also suggest that if achievement is monitored e.g. by central exit exams, teachers do in fact teach the curriculum, which is what they are expected to do. Mathematical literacy (as a broader concept and not part of the curriculum) does not automatically improve. Notably, these differences between curricular and literacy tests are in accordance with recent findings from the US (e.g. Amrein and Berliner, 2002; Klein et al., 2000; Neil and Gaylor, 2001), where state-mandated exams were found to increase performance measured in terms of the curriculum but not necessarily in terms of a better understanding of mathematical concepts that enables students to actually use mathematics in everyday life, as defined in PISA. To further improve literacy by central exit exams thus may require including strategies of knowledge transfer into the curriculum.

References

Abrams, L. M., Pedulla, J. J., Madaus, G. F., 2003. Views from the classroom: Teachers` opinions of statewide testing programs. Theory into Practice 42(1), 18-29.

Ai, X., 2002. Gender differences in growth in mathematics achievement: Three-level longitudinal and multilevel analysis of individual, home, and school influences. Mathematical Thinking and Learning 4, 1-22.

Alexander, N., 2000. The missing link: An econometric analysis on the impact of curriculum standards on student achievement. Economics of Education Review 19(4), 351-361.

Amrein, A., Berliner, D. C., 2002. High-stakes testing, uncertainty, and student learning. Education Policy Analysis Archives, 10 (18). Retrieved April 23, 2009 from http://epaa.asu.edu/epaa/v10n18/.

Au, W., 2007. High-stakes testing and curricular control: A qualitative metasynthesis. Educational Researcher 36 (5), 258-267.

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.

Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F., Carstensen, C. H., 2004. Mathematische Kompetenz [Mathematical competence], in: PISA-Konsortium Deutschland (Eds.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Waxmann, Münster, pp. 47-92.

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.

Ganzeboom, H. B. G., de Graaf, P., Treiman, D. J., 1992. A standard international socio-economic index of occupational status. Social Science Research 21, 1-56.

Hanushek, E., 2002. Evidence, politics, and the class size debate, pp. 37-65 in Mishel, L., Rothstein, R. (eds.), The class size debate. Washington, DC: Economic Policy Institute.

Hanushek, E., 2011. The economic value of higher teacher quality. Economics of Education Review 30, 466-479.

Heller, K. A., Perleth, C., 2000. Kognitiver Fähigkeitstest (KFT 4-12+R). (German revised adaptation of the Cognitive Abilities Test; Thorndike and Hagan, 1954). Hogrefe, Göttingen.

Jacob, B. A., 2005. Accountability, Incentives And Behavior: The Impact Of High-Stakes Testing In The Chicago Public Schools. Journal of Public Economics 89, 761-796.

Jonen, G., Eckardt, T., 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, H., Richter, W.F., Schneider, K., 2005. Teacher quality and incentives, Theoretical and empirical effects of standards on teacher quality. FinanzArchiv 61 (3), 298-326.

Jürges, H., Schneider, K., Büchel, F., 2005. The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. Journal of the European Economic Association 3 (5), 1134-1155.

Jürges, H., Schneider, K., 2010. Central exit examinations increase performance...but take the fun out of mathematics. Journal of Population Economics 23, 497-517.

Kellaghan, T., Madaus, G. F., Raczek, A., 1996. The use of external examinations to improve student motivation. Washington, DC: American Educational Research Association.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., 2000. What do test scores in Texas tell us? Education Policy Analysis Archives, 8 (49). Retrieved April 23, 2009 from http://epaa.asu.edu/epaa/v10n18/.

Kunter, M., Baumert, J., 2006. Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. Learning Environmental Research 9, 231-251.

Neil, M., Gaylor, L., 2001. Do high-stakes graduation tests improve learning outcomes? In G. Orfield and M. L. Kornhaber (Eds.), Raising standards or raising barriers? Inequality and high-stakes testing in public education. The Century Foundation Press, New York, pp. 107-126.

OECD, 2003. The PISA 2003 assessment framework. Mathematics, reading, science and problem solving knowledge and skills. OECD, Paris.

OECD, 2005. PISA 2003. Technical report. OECD, Paris.

Prenzel, M., Carstensen, C. H., Schöps, K., Maurischat, C., 2006. Die Anlage des Längsschnitts bei PISA 2003 [The design of the longitudinal study of PISA 2003]. In PISA-Konsortium Deutschland (Hrsg.), PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres. Waxmann, Münster, pp. 29-62.

Ryan, K. E., Ryan, A. M., Arbuthnot, K., Samuels, M., 2007. Students` motivation for standardized math exams. Educational Researcher 36(1), 5-13.

Ryan, R. M., Sapp, A., 2005. Considering the impact of test-based reforms: A self-determination theory perspective on hight stakes testing and student motivation and performance. Unterrichtswissenschaft 33, 143-159.

Senkbeil, M., 2006. Die Bedeutung schulischer Faktoren für die Kompetenzentwicklung in Mathematik und in den Naturwissenschaften [The importance of school factors for skill acquisition in mathematics and the natural sciences]. In PISA-Konsortium Deutschland (Hrsg.), PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres. Waxmann, Münster, pp. 277-308.

Scott, L. A. et al., 1995. Two years later: Cognitive gains and school transitions of NELS:88 eighth graders. National Center for Education Statistics, Washington.

Wößmann, L., 2005. The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA. Education Economics 13 (2), 143-169.

Table 1. Overview of Central Exit Exams in Germany in 2003

	Basic track	Intermediate track	Academic track
Baden-Württemberg	X	X	X
Bavaria	X	X	X
Mecklenburg-West Pomerania		X	X
Saarland	X	X	X
Saxony	X	X	X
Saxony-Anhalt	X	X	X
Thuringia	X	X	X

No Central Exit Exams in Berlin, Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, Schleswig-Holstein. Brandenburg and Hesse introduced central exit exams in 2003/2004 and are excluded from the analysis.

Table 2. Description of student sample

	Nonacademic students				Academic students			
	CEE (N = 1453)		No CEE (N = 1146)		CEE (N = 1079)		No CEE (N = 1250)	
	Mean	(S.E.)	mean	(S.E.)	mean	(S.E.)	mean	(S.E.)
Mathematical literacy test grade 9	46.24	(0.44)	44.55	(0.44)	56.39	(0.47)	54.96	(0.39)
Mathematical literacy test grade 10	46.40	(0.43)	45.47	(0.47)	55.80	(0.48)	54.34	(0.44)
Mathematical curricular test grade 9	46.11	(0.46)	44.48	(0.51)	56.88	(0.50)	54.75	(0.47)
Mathematical curricular test grade 10	46.80	(0.45)	43.74	(0.49)	56.53	(0.50)	54.83	(0.44)
Cognitive skills	46.87	(0.40)	47.22	(0.47)	54.20	(0.55)	53.96	(0.38)
HISEI	47.83	(0.58)	49.63	(0.62)	59.36	(0.69)	59.61	(0.69)
Student is immigrant	0.10	(0.01)	0.21	(0.02)	0.12	(0.02)	0.14	(0.01)
Repeat class	0.21	(0.02)	0.22	(0.02)	0.07	(0.01)	0.10	(0.01)

CEE: central exit exams

S.E.: standard error of the mean

Table 3. Value-added regression of student achievement

	Mathematical literacy test				Curricular tests			
	Nonacademic track (Model 1)		Academic track (Model 2)		Nonacademic track (Model 3)		Academic track (Model 4)	
	b	(S.E.)	b	(S.E.)	b	(S.E.)	b	(S.E.)
State-level variables								
Central exit exam	0.71	(0.56)	0.45	(0.55)	2.56***	(0.47)	0.29	(0.46)
Eastern Germany	0.58	(0.71)	0.29	(0.69)	0.82	(0.78)	-1.05	(0.83)
Educational expenditure (in 1,000€)	0.32	(0.79)	0.25	(0.68)	0.21	(0.58)	0.04	(0.48)
Class size	0.33**	(0.16)	0.07	(0.15)	0.15	(0.16)	0.21	(0.14)
Individual level variables								
Mathematical literacy - grade 9	0.60***	(0.05)	0.62***	(0.06)				
Curricular skills - grade 9					0.67***	(0.06)	0.68***	(0.05)
Cognitive skills	0.19***	(0.03)	0.17***	(0.03)	0.11***	(0.03)	0.08***	(0.03)
No repeated class	1.08***	(0.33)	1.36**	(0.57)	0.86**	(0.37)	0.90	(0.58)
HISEI	0.01	(0.01)	0.02	(0.01)	0.02	(0.01)	0.03***	(0.01)
Student is immigrant	-0.80**	(0.39)	-0.68	(0.56)	-0.58	(0.42)	-1.15***	(0.41)
R^2	0.58		0.50		0.61		0.57	
N	1453		1146		1079		1250	

S.E.: standard error

[a] Reference group: no central exit exam

* p < 0.10

** p < 0.05

*** p < 0.01

Table 4. Sample description of students` attitudes

				Nonacademic students				Academic students			
				CEE		Non CEE		CEE		Non CEE	
	# of items	Alpha	Example item	mean	(S.E.)	mean	(S.E.)	mean	(S.E.)	mean	(S.E.)
Emotions											
Anger	5	0.82	I am annoyed that mathematics is so hard.	0.14	(0.03)	-0.12	(0.05)	0.06	(0.05)	-0.13	(0.05)
Anxiety	9	0.89	Right before mathematics exams I am very nervous.	0.15	(0.04)	-0.08	(0.04)	0.02	(0.04)	-0.16	(0.04)
Boredom	5	0.86	I am very bored during mathematics classes.	0.02	(0.04)	-0.08	(0.04)	0.08	(0.04)	-0.03	(0.06)
Despair	5	0.88	During mathematics exams the idea to give up is very appealing.	0.14	(0.03)	-0.10	(0.04)	0.01	(0.04)	-0.11	(0.04)
Achievement pressure	6	0.79	It is very important to my parents for me to be good at mathematics.	0.18	(0.04)	0.05	(0.06)	-0.11	(0.06)	-0.19	(0.05)
Work habits											
Effort	5	0.80	I put a lot of effort into mathematics to understand everything.	0.01	(0.03)	0.05	(0.03)	-0.05	(0.04)	-0.03	(0.04)
Completion of homework	10	0.72	I always try to solve all the problems in my mathematics homework.	-0.01	(0.04)	0.14	(0.03)	-0.17	(0.03)	0.02	(0.04)
Attention in lessons	3	0.76	Even if the subject in mathematics is very difficult I put all my effort into understanding it.	-0.05	(0.04)	0.13	(0.03)	-0.13	(0.04)	0.06	(0.05)
Motivation											
Instrumental motivation	4	0.82	I see that mathematics will be very expedient in my future job, so I apply myself to it.	0.00	(0.03)	0.08	(0.04)	-0.11	(0.05)	0.01	(0.04)
Enjoyment	6	0.92	Mathematics is fun.	-0.04	(0.05)	0.17	(0.04)	-0.16	(0.05)	0.03	(0.06)
Performance orientation	3	0.62	I want my grades to be good so I make an effort to achieve that.	0.05	(0.04)	0.03	(0.03)	-0.05	(0.03)	-0.05	(0.04)
Self Esteem											
Self efficacy	8	0.81	I can calculate how much cheaper a TV set is at a 30% discount.	-0.24	(0.03)	-0.20	(0.04)	0.25	(0.05)	0.31	(0.04)
Self concept	5	0.92	I keep up easily in mathematics.	-0.09	(0.04)	0.13	(0.04)	-0.11	(0.04)	0.10	(0.04)

CEE: central exit exams, S.E.: standard error of the mean. (Cronbach's) Alpha is the standard psychometric measure of the reliability of a scale measuring some latent construct (such as anger, self esteem or motivation). It is based on the inter-item correlation. A value of alpha bigger than 0.7 indicates good reliability.

Table 5. Differences in student attitudes between CEE and non-CEE states.

Outcomes	Nonacademic students		Academic students	
	coef.	s.e.	coef.	s.e.
Emotions				
Anger	0.23***	0.09	0.25***	0.08
Anxiety	0.13*	0.07	0.17**	0.07
Boredom	0.00	0.03	0.04	0.10
Despair	0.18**	0.08	0.19***	0.06
Achievement pressure	0.03	0.09	-0.01	0.08
Work habits				
Effort	-0.02	0.07	0.01	0.07
Completion of homework	-0.06	0.07	-0.12*	0.06
Attention in lessons	-0.12	0.08	-0.26***	0.08
Motivation				
Instrumental motivation	-0.15**	0.06	-0.18**	0.08
Enjoyment	-0.17*	0.10	-0.17*	0.10
Performance orientation	-0.01	0.05	-0.01	0.11
Self esteem				
Self efficacy	0.15**	0.06	0.04	0.08
Self concept	-0.12	0.08	-0.21***	0.08

Note: Each coefficient pertains to a separate regression of student attitudes on a CEE dummy, controlling for socioeconomic background, whether the students is an immigrant, cognitive skills of the student, whether the student has repeated a class, whether the state is in former East Germany, average class size, and educational expenditure per student

* $p < 0.10$

** $p < 0.05$

** $p < 0.01$