

Long-term effects of the implementation of state-wide exit exams: a multilevel regression analysis of mediation effects of teaching practices on students' motivational orientations

Katharina Maag Merki¹ · Britta Oerke²

Received: 3 July 2015 / Accepted: 26 July 2016 / Published online: 17 August 2016
© Springer Science+Business Media New York 2016

Abstract This study extends previous research investigating the effects of state-wide exit exams by studying the change from a class-based to a state-wide exit exam system over 5 years, using multilevel analyses and examining mediating effects of teachers' practices on students' motivational orientations. In this multi-cohort study, we analyzed in particular the effects on students' interest, scholastic self-efficacy, and persistence in advanced level English courses ($N = 1835$) and mathematics courses ($N = 1336$) in two states in Germany (28 schools). Descriptive analyses, multivariate hierarchical regression analyses, and differences-in-differences analyses were carried out. The results revealed long-term effects of the implementation of state-wide exit exams particularly in the advanced level English courses. Here, a close relationship between the change in all analyzed motivational orientations and teacher support perceived by the students can be identified. These results show the ambivalent effects of state-wide exit exams: Due to the increased teacher competence support, students' interest is enhanced in the long term. However, scholastic self-efficacy and persistence might have been negatively affected by state-wide exams, if teacher competence support had not increased over time. In the advanced level mathematics courses, the results are mixed. Implications for further research are discussed.

Keywords State-wide exit exams · Multilevel regression analyses · Interest · Scholastic self-efficacy · Persistence

✉ Katharina Maag Merki
kmaag@ife.uzh.ch

¹ Institute of Education, University of Zurich, Freiestrasse 36, CH-8032 Zürich, Switzerland

² Institute of Education, University of Bielefeld, Universitätsstraße 25, D-33615 Bielefeld, Germany

1 Introduction

In recent years, standardized testing of educational outcomes has been used more and more for education governance (EACEA/Eurydice 2009). This trend increased also in German-speaking countries, once studies revealed that student performance was not meeting expectations and that assessments by teachers were not comparable across schools and classes (e.g., Baumert and Watermann 2000). To smooth out differences between classes and schools and to increase students' performance, there was therefore a demand for procedures that would assure a higher level of standardization and support better learning (Klieme 2004). To reach these aims, several states in Germany instituted state-wide exit exams at the end of academic upper secondary school (*Gymnasium*). In contrast to the previous class-based exit exams, the exams are now developed and administered by the states for each domain and for all schools and courses in the entire state and not by the individual teacher for his or her own classes (for more information on the differences between class-based and state-wide exit exams, see in Section 2). From a political perspective, this strategy should allow increased comparability between the exit exams in the classes in terms of subject-specific requirements, topics, scope, and achievement level. Furthermore, due to an increased value of the state-wide exit exams as compared to the class-based exit exams, it is assumed that teachers will better support student learning, which should then lead to increased student performance.

This paper looks in particular at students' motivational orientations in the context of the implementation of state-wide exit exams in two German states, Bremen and Hesse. Some previous studies analyzed the impact of state-wide exit exams on students' motivational orientations, since these are crucial and influence significantly students' subject-specific competence (Chiu and Xihua 2008; Winne and Hadwin 2008). However, the results were mixed. Some studies found that state-wide exit exams were associated with negative motivational dimensions of learning (e.g., Pedulla et al. 2003). In contrast, other studies revealed no effects (e.g., Baumert and Watermann 2000). Further, there is a lack of studies on the change in the exam system from class-based to state-wide, especially on the effect of the change on central motivational dimensions of student learning, and in the few studies analyzing this, only short-term effects were considered (Maag Merki 2011). Finally, the results of previous studies are undifferentiated, as the studies did not sufficiently consider differences between school subjects or between basic level and advanced level courses.

Taking into account these research deficits, our study over 5 years in Germany analyzes the long-term effects of the implementation of state-wide exit exams on students' domain-specific interest, scholastic self-efficacy, and persistence while they were preparing for the exit exams. These dimensions are part of the motivational self-regulation of student learning (Baumert 2000; Winne and Hadwin 2008) and have a significant impact on student achievement. Although the selection of these three concepts is limited in relation to the complex framework of self-regulated learning, and other important concepts like self-concept of ability, test anxiety, or goal orientation would have been important to analyze as well, the investigations realized will deepen our understanding of the effects of state-wide exit exams on students' motivational orientations. Whereas domain-specific interest and scholastic self-efficacy belong to the most effective dimensions of motivational self-regulation on achievement (Chiu and

Xihua 2008), the analysis of whether or not students are able to maintain their learning process for the exit exam is crucial, since some of the theoretical frameworks (Bishop 1999) (see below) argue that the implementation of state-wide exit exams particularly has a positive effect on student effort.

Our analyses are conducted in the domains mathematics and English in courses with advanced requirements. The courses selected for this study are the courses most frequently chosen by the students and are highly important for successful completion of academic upper secondary schools.

In the following, we clarify the state-wide exit exam in Germany. We then discuss the relationship between the implementation of state-wide exit exams and student motivation from a theoretical and empirical perspective. These analyses bring us to an overview of significant research gaps. In the next step, we present our research questions and hypotheses, provide information on our methods, present our results, and discuss them in the last section.

2 State-wide exit exams in Germany

In Germany, state-wide exit exams at the end of academic upper secondary schools (ISCED Level 3A) are end-of-course exams that focus on curriculum content. The exit exams are mandatory for graduation and are very important for students, as they are the basis for regulation of access to university places. Nevertheless, compared to the exam systems in the USA or other OECD countries, the system in Germany shows a rather low level of standardization (Klein and Van Ackeren 2011). For example, in the German system, the final exam grade includes not only the results of the single tests at the end of upper secondary schools but also the students' grades during the last 2 years of upper secondary schools. Further, in contrast to other exit exam systems, students can choose their exam subjects to a certain extent, which provides the students with at least partial autonomy.

These aspects did not change when some states in Germany implemented the new state-wide exit exam system.¹ However, in other aspects, the old and the new systems differ widely from each other. The differences are basically fourfold:

- 1) Class-based exams encompassed the topics covered in the school subjects in the last 2 years in academic upper secondary schools. Within this framework, each teacher chose the contents of the exit exam questions/tasks for his/her single course at the latest one half-year before the exit exams took place. The students were not informed systematically about the exam topics selected (some teachers informed the students on the selected topics or content areas, some did not). In state-wide exit exams, the general topics that will be on the exams are announced publicly 2 years prior to the date of the exit exams. Basically, the information includes a short description of the topic, an outline of the competences that will be required,

¹ Whereas in some states in Germany, state-wide exit exams have already been implemented for a long time, and one state has not implemented the new exams at all, there are eight states that have implemented the new system.

- and sometimes a list of mandatory literature that has to be prepared. Importantly, the degree of differentiation of this information varies subject specifically.
- 2) In class-based exams, teachers of each school subject and course designed their own exams for their classes, with only little control by the education ministry that had to approve the teachers' proposed exams. Today, an external expert group develops state-wide exams for each school subject for all schools and courses in the entire state. They are administered at the same time in all schools and classes within the same state.
 - 3) As with the former class-based exam systems, the exams are still corrected and graded by the individual teachers. In contrast to the old system, however, teachers now correct and grade the exams following uniform correction and grading guidelines for the entire state.
 - 4) The new system changes the relationship between teachers and their students, as the teachers are no longer "those in the know." Teachers are just as unaware of the concrete exam questions/tasks as their students are.

In the two states where we conducted our study, state-wide exit exams were implemented in Bremen in 2007 for basic level courses and in Hesse in 2008 for all courses. In Bremen, state-wide exams were implemented for advanced level courses in 2008.² Due to substantial overlaps in the two systems, the state-specific type of implementation (year and extent of the implementation) allows quasi-experimental comparisons over the years. Of particular interest are the advanced courses in English and mathematics, since for these courses the implementation type differs between Bremen and Hesse. Whereas in Bremen we can analyze the change from a class-based (2007) to a state-wide exit exam (2008), the change in Hesse happened already from 2006 to 2007. Accordingly, the system in Hesse may be taken as a control group, since from 2007 to 2011 we are able to compare the stable system in Hesse (without change of the exit exams) with a system where the change was implemented from 2007 to 2008 (Bremen) (see also Section 6.1 for more details).

3 Theoretical perspective on the relationship between the implementation of state-wide exit exams and students' motivational orientations

From a theoretical perspective, why should the implementation of state-wide exit exams have an impact on students' motivational orientations? Several theoretical models exist to explain the relationship between the implementation of state-wide exit exams and the development of students' motivational dimensions of learning. Bishop (1999) argues that state-wide exit exams have a positive effect on student learning effort and as a consequence on student achievement, because the value and the reward of the grades in the state-wide exit exams are higher than in class-based exit exams, due to the higher comparability of the grades across classes, schools, and regions (Piopiunik et al. 2014). Additionally, due to the higher reward of the grades, not only individual students put more effort into learning

² Advanced level courses have more hours of instruction per week, and the outcomes have a higher impact on the average exam grade than the outcomes of the basic level courses do.

but also the class and teachers put more weight on learning and set higher standards. Further, it is assumed that also parents will put greater effort into trying to induce their children to study regularly. Accordingly, the higher level of student effort on state-wide exit exams is not only explained by the (formal) implementation of state-wide exit exams and the perceived higher rewards of the grading but also by a complex interplay of parent and peer pressure as well as by teachers' standards and stronger engagement in promoting students' learning.

In line with Bishop (1999), Jürges and Schneider (2010) assume that student effort but also motivation, interest, and attitudes are influenced by the implementation of central exit exams, mediated by the increased teacher effort, teacher practices, and attitudes and by parental background. In total, they assume mixed effects on student motivation. On the one hand, increasing effort may also be costly for students and teachers. This may lead to more negative student attitudes towards school. In contrast, better knowledge may increase students' interest and result in a more positive attitude, and teachers might find it more enjoyable to teach more highly motivated students (Jürges and Schneider 2010, p. 503).

Ryan and Sapp (2005) criticize the positive potential of high-stakes state-wide exit exams for the development of intrinsic motivation. They argue that the increased extrinsic incentives of these exams and the higher level of pressure will probably lead to increased extrinsic motivation, which is not related to deep understanding and which may suppress the development of intrinsic motivation. Furthermore, since the pressure towards a specific outcome is stronger in high-stakes (where the test outcomes have severe consequences for teachers and students) than in low-stakes systems, high-stakes testing systems can undermine best teaching practices and enhance the use of controlling strategies in classes, which again lead to degraded forms of motivation.

Although there are significant differences between the conceptual models presented, there is one relevant overlap between the models: The interrelationship between the exit exam system and teaching practices can be identified as relevant mechanism to explain students' motivational orientations in the context of exit exams. Accordingly, students' motivational orientations are not only influenced by the type of exit exam (Ryan and Sapp 2005) but are also the result of the interdependency between the type of exit exam and student perceptions of teaching practices. Based on previous results on the effects of state-wide exit exams on teacher practices (e.g., Pedulla et al. 2003; Au 2007; Faxon-Mills et al. 2013), this means that the effect of the implementation of state-wide exit exams on students' motivational dimensions should be at least partially mediated by teacher practices. As especially relevant forms of teaching practices that have an effect on students' motivational orientations, Deci and Ryan's (1985) well-established and empirically proven self-determination theory identifies teacher autonomy support and competence support as well as students' relatedness in classes. Additional studies on teacher professionalism (e.g., Kunter et al. 2008, 2013; Rakoczy 2008) found that teacher practices have a significant impact on motivational orientations support as well. For this reason, in this study, we focus on ability to motivate, competence support, and autonomy support as forms of teaching practices.

Taking these general descriptions of the relationship between the type of exit exam, teaching practices and student motivation as a starting point of our argumentation, we now specify this framework for the three concepts analyzed in our study: "interest," "scholastic self-efficacy," and "persistence."

3.1 Interest

Interest is understood as individual interest in a learning topic, a domain of knowledge, or subject matter (Krapp 2005). It has an influence on how intensively a person will engage in a certain domain of knowledge (2005). Additionally, interest is closely interrelated with intrinsic motivation (Krapp 2002). Many studies have found that high interest in a subject has a positive effect on student learning performance (e.g., Chiu and Xihua 2008).

Given that information on the exit exams' topics is announced 2 years prior to the exams, which allows students to work on the topics over a long time on a deep level, e.g., by forming learning groups, looking for additional material, or discussing the topic with parents and peers at other schools, there is support for positive effects on student interest. In contrast, due to the higher standardization and the pressure connected with the state-wide exit exams and the lower level of leeway for the students to participate in defining the learning content and to experience autonomy and competence (Deci and Ryan 1985; Krapp 2002), student scholastic interest might be reduced and extrinsic motivation might be increased.

This assumption might be additionally supported by the fact that an increased emphasis on the exit exam topics without considering other relevant aspects of the domain could decrease student interest. However, if teaching practices increase, particularly teacher autonomy and competence support, meeting students' basic human needs (Deci and Ryan 1985; Krapp 2002), the negative effects on student interest might be weakened or the positive effects might be strengthened. In contrast, if teachers' enthusiasm and motivation decrease (due to their loss of autonomy to conduct the exit exams and the higher pressure involved in being controlled, e.g., Pedulla et al. 2003; Au 2007; Faxon-Mills et al. 2013), the quality of instruction might decrease as well (e.g., Kunter et al. 2013; Roth et al. 2007), which would have an additional negative effect on student interest.

3.2 Scholastic self-efficacy

Scholastic self-efficacy (Bandura 1997; Schwarzer and Jerusalem 2002) is an important factor affecting student learning. An important prerequisite for successful learning processes is students' belief that they can meet scholastic demands successfully through their own abilities and effort (e.g., Bandura et al. 2001; Chiu and Xihua 2008; Sitzmann and Yeo 2013). According to Bandura (1997) or Schwarzer and Jerusalem (2002), there are several factors that influence self-efficacy, including mastery experience, social persuasion, vicarious experience, physiological and affective states, and reachable goals that are relevant and can be achieved through a student's own effort.

In the context of state-wide exit exams, it might be argued that knowing what topics will be on the exit exam for 2 years prior to the exams could lead to deeper learning, mastery experience, and accordingly higher perceived self-efficacy. This effect might be even strengthened, if teachers provide competence support and continuous information on students' current competence levels and specific needs for improvement (e.g., Bandura 1997; Chan and Lam 2010; Schwarzer and Jerusalem 2002). Additionally, increased instructional quality and supportive teacher behavior due to teachers' higher engagement under state-wide exit exam

systems could again support students' belief that they can meet the requirements of the exit exams successfully.

However, particularly the increased loss of control and higher perceived pressure but also the fact that students have to prepare for 2 years for the exit exams could be reasons why students in state-wide exit exams may fail to set realistic goals, use effective learning strategies, and attribute success to internal conditions and effort (Schwarzer and Jerusalem 2002). Subsequently, they will experience a lower level of self-efficacy regarding performing successfully on the exit exams. But supportive teaching and the increased quality of teaching due to teachers' higher engagement might interact with student self-efficacy in the sense that it could compensate for the higher level of insecurity regarding meeting the requirements of the state-wide exit exams. This could lead again to increased student self-efficacy or at least hinder a decrease in self-efficacy.

3.3 Persistence

Persistence—or the ability to maintain the learning process, willingness to put forward effort, perseverance, and suppression of emotions that hinder learning—is an essential component of action control and has been found to be significantly associated with learning performance (e.g., Chiu and Xihua 2008). The phase of intention realization is centrally important in whether a desired objective, such as preparation for the exit exams, can in fact be realized (Heckhausen 1989; Kuhl 1987; Rheinberg 2004).

Due to the higher level of rewards of grades on the state-wide exit exams (Piopiunik et al. 2014) and given the fact that the general topics that will be on the exam are announced significantly earlier than in the class-based testing system, increased persistence to prepare for the state-wide exit exams could be expected. In contrast, rather negative effects on student persistence to prepare for the exams could be assumed due to the fact that their teachers no longer create their own exam questions/tasks. Whereas with class-based exams, students were used to their teacher's specific style of exam questions/tasks, this is no longer the case with state-wide exams. This means that students may experience a higher loss of control of their learning effectiveness. Additionally, students are more likely to depend on the quality of preparation for the exit exam on the part of their teachers (Oerke et al. 2011), which again could lead to an increased loss of control and therefore a lower level of persistence when preparing for the state-wide exit exams. Moreover, the teachers' higher level of perceived pressure in state-wide exit exams (Oerke 2012) could reduce their capacity to support students. Accordingly, students' ability to put effort into their learning might be reduced. In contrast, if teaching is well structured and perceived as supportive and enthusiastic, or if teachers improve their teaching quality in reaction to the implementation of state-wide exit exams, negative effects may be compensated or it might help students to maintain their persistence despite possible higher standards and pressure (Kunter et al. 2013).

4 Effects of state-wide exit exams on students' motivational orientations

Although the effects of state-wide exit exams on student achievement are quite well analyzed, the literature is limited when it comes to investigating effects on students'

motivational orientations, and this is even more evident for effects on student interest, scholastic self-efficacy, and persistence. Furthermore, studies that analyzed the relationship between state-wide exit exams and motivational orientations most often emphasize emotional dimensions of learning (e.g., test anxiety, pressure), and in most studies the operationalization is based only on single items. Accordingly, the following literature review presents an overview on existing results on the effects on motivational orientations in general and not differentiated for every single concept that is relevant in our current analysis, starting with results in the USA and followed by results of studies in Europe and in German-speaking countries. This differentiation is important because the exit exam systems differ widely from each other, particularly in terms of level of standardization (Klein and Van Ackeren 2011).

Empirical results in the context of high-stakes state-wide exit exam systems particularly in the USA revealed basically negative or inconsistent effects on motivational and emotional aspects of students' learning. Ryan et al. (2007) conducted semi-structured interviews with 33 moderate and high achieving mathematics students in the USA on their experiences with mathematics high-stakes test taking. The analyses revealed that students respond differently to standardized tests. Further, the tests have a negative impact on motivational and emotional dimensions of learning also for moderate and high achieving students (e.g., test anxiety).

A study of students in the USA who had participated in high-stakes tests (Wheelock et al. 2000) confirmed these results and showed that a substantial group of students portrayed themselves as anxious, angry, bored, pessimistic, or withdrawn from testing. The results of a comparative study by Abrams et al. (2003), although considering the view of teachers, also support the assumption that a higher level of student anxiety and pressure to perform well can be expected in high-stakes test systems than in low-stakes testing systems.

Another study by Richman et al. (1987) examined possible changes in students' self-esteem and some personality dimensions as a function of minimum competency test taking (MCT) in dependency upon performance status and test result. Richman et al. did not find any effects for low-risk students, but they found negative effects especially for high-risk-fail students. These are students who were rated by school personnel prior to the MCT as being at high risk to fail the MCT and who in reality failed the MCT. The results revealed for these students a "marked increase in neuroticism and apprehension with a corresponding decrease in general self-esteem" (Richman et al. p. 14). However, the study did not analyze whether these negative effects lead to more drop-outs or subsequent success in graduating or not. That was analyzed, however, by Reardon et al. (2010), using data from the California High School Exit Exam (CAHSEE). They investigated whether failing the state-mandated public high school high-stakes exit exams had an effect on subsequent high school course-taking, achievement, persistence,³ or graduation rates for students near the exit exam passing score; the results revealed no effects. Whether these students experienced negative emotional effects as analyzed by Richman et al. (1987) was not investigated in this study, however.

³ Reardon et al. operationalized "persistence" as whether students are present in the district in the spring semester 2 years after they first took the CAHSEE in tenth grade. Although we use the same term in our study, the meaning of "persistence" is different than the meaning used by Reardon et al.

The results of the studies not only in the USA but also in Germany seem to be ambiguous in terms of effects of state-wide exit exams. Using data from the Third International Mathematics and Science Study (TIMSS) at the end of the lower secondary level in Germany, a comparison of German states with and without state-wide exit exams showed that students in states with state-wide exit exams are “consistently less likely to like or enjoy mathematics, or to find it an easy subject, but they are more likely to find it boring” (Jürges and Schneider 2010, p. 514). The findings of the Programme for International Student Assessment (PISA) study are comparable (Jürges et al. 2012). Jürges et al. (2012) saw these negative effects as associated with the increased pressure on students exerted by teachers.

In contrast to these results, a cross-sectional German study based on the TIMSS data as well, but at the end of the upper secondary level, compared students in states with and without state-wide exit exams (Baumert and Watermann 2000). The study did not identify any effects on test anxiety in physics courses but reported by trend lower test anxiety in mathematics in states with state-wide exit exams than without state-wide exit exams. The reason for these results is seen in the new teacher-student relationship in state-wide exit exams, since both teachers and their students do not know what will be on the tests; this seems to lead to a stronger coalition between teachers and students and to a reduction of anxiety.

In line with these results, a survey study administered in academic upper secondary schools in two German states, analyzing the short-term effects of the implementation of state-wide exit exams, revealed no general effects on students’ motivational dimensions but differential effects in terms of school subject; particularly for English advanced courses, positive effects could be identified. No negative effects on student motivation were found (Maag Merki 2011).

Also cross-sectional analyses by Bishop (1999), taking into account several international comparative data (e.g., TIMSS) and science and mathematics scores in nine Canadian provinces, found little support that state-wide exit exams are systematically associated with problematic motivational burdens in students. Inconsistent results were found on the relationship between state-wide exit exams and spending more time on homework, an indicator used to assess student effort or persistence, however. A relationship seems to be true for Canadian students, but in the TIMSS analyses, a negative association between these two aspects was found. Further, the TIMSS data revealed that whether students liked mathematics and science was unrelated to the existence of state-wide exit exams.

The studies presented above revealed information on the direct effects of state-wide exit exams on students’ motivational orientations. However, only a very limited number of studies have examined mediating or moderation effects in dependence on teaching practices. Putwain et al. (2012) found that teacher practices, e.g., fear appeals, lead to a higher level of student’s test-related anxiety. Other studies revealed that student characteristics might moderate the relationship between the importance of a task and text anxiety, such as their academic self-efficacy (Lie et al. 2011), goal-setting (Flanagan et al. 2015), or academic buoyancy (Putwain et al. 2015).

The studies therefore support the theoretical assumption that it is important to analyze the effects on motivational orientations by taking into account teacher practices in classes. At the same time, it is obvious that research is needed to better understand the interrelationship between implementation of state-wide exit exams, teacher

practices, and the development of students' motivational orientations. This research deficit is all the more serious, as previous research neglected the multilevel structure of the education system. Accordingly, the effects were investigated independently of teaching characteristics, which are important influencing factors in motivational and emotional dimensions of student learning (e.g., Putwain et al. 2012; Deci and Ryan 1985; Kunter et al. 2007; Seidel 2006). Furthermore, analyses at the upper secondary level are rare and the results, based on previous research, are mixed. Particularly for high-stakes state-wide exit exams, negative effects on student motivation and emotions occur. Further, international comparative studies like TIMSS or PISA and inner state comparative studies revealed negative effects for students on motivational dimensions at the end of the lower secondary level. However, previous results for students at the upper secondary level in Germany and international comparative cross-country analyses do not support the negative effects. Furthermore, the effects seem to vary between subjects.

One reason for the discrepancies might be methodological: the fact that international comparative analyses do not take into sufficient account that the characteristics of the exit exams differ widely across countries (Klein and Van Ackeren 2011). Additionally, the German study at the end of upper secondary level analyzed only one dimension of student motivation. Further, one of the most important gaps in the previous studies can be seen in the fact that only cross-sectional data are available for analyzing the effect of state-wide exit exams on student motivation. There is therefore a need for multivariate analyses of the change from a class-based to state-wide exam system at the end of academic upper secondary schools, considering changes over several years and analyzing the interaction effects between the implementation of state-wide exit exams and teacher practices on students' motivational dimensions. The analyses presented in this paper are in line with these research deficits and aim to investigate the long-term effects of the implementation of state-wide exit exams over 5 years.

5 Research questions and hypotheses

In this study, our two research questions were

- 1) How does the implementation of state-wide exit exams in academic upper secondary schools influence students' motivational orientations—in particular interest, scholastic self-efficacy, and persistence—in advanced level mathematics and English courses in Bremen in the long term?
- 2) Are the effects mediated by teacher practices—measured as ability to motivate, competence support, and autonomy support—as perceived by students?

Considering the differences between the old class-based exam system and the new state-wide exam system (see Section 1) and taking into account previous theoretical models and empirical results (see Sections 2 and 3), we expect that the implementation of state-wide exams will not have general but rather differential long-term effects on student interest, scholastic self-efficacy, and persistence and assume that there will be more school subject-specific effects for English than for mathematics courses (H1). Additionally, we expect that the impact of the implementation of state-wide exit exams

on motivational orientations is at least partially mediated by teacher practices (H2). However, due to the mixed results of previous studies, no clear hypotheses can be drawn in terms of the direction of the effects.

6 Methods

6.1 Design and data collection

The data used were collected within a multi-year cohort study on the effects of state-wide exit exams in the German states of Bremen and Hesse; the study was supported by the German Research Foundation (DFG) and by the two states. The study was conducted in 2007, 2008, 2009, and 2011—in each year 1 month before the written exit exams. In 2010, no data were collected in order to reduce the workload for the schools. In the current analyses, only the data from 2007, 2008, and 2011 were used, to examine long-term changes (2007–2011) and to check whether they occurred directly after the implementation of state-wide exams in Bremen or later. The questionnaires were administered in the schools by trained test administrators. For more details, see Maag Merki (2012).

In both Bremen and Hesse, written exit exams are administered for two advanced level courses and one basic level course. State-wide exams were implemented in 2007 in basic level courses in Bremen and in all courses in Hesse. In advanced level courses in Bremen such as mathematics and English, state-wide exams were implemented in 2008. Thus, in these courses, class-based exams (in 2007) can be compared with state-wide exams in the long term (2007–2011). The main focus of our analyses was therefore on the change of the exam system in Bremen. Due to the fact that in Hesse the exit exams also changed, but 1 year prior to the start of data collection in this study, Hesse can be considered a “control group.” It is assumed that if there are effects of the implementation of state-wide exit exams on students’ motivational orientations, the change of the system in Bremen (2007–2008; 2007–2011) will be associated with a greater change in these motivational orientations than will be found in Hesse where the analyzed system in this period is stable.

This research strategy is supported by the fact that other than the above-described different type of implementation, the exit exam systems in the two states do not differ substantially before and after the change. Furthermore, if system changes are investigated in educational science, a randomized sample strategy is not feasible, so that a quasi-experimental design is the best possible design (Slavin 2010; Borman 2009). Following Jürges et al. (2010), we used difference-in-differences analyses by comparing the development over 5 years in two states to identify possible effects of the implementation of state-wide exit exams. However, due to the non-random strategy of sampling, we have to be cautious about interpreting possible effects as causal. Nevertheless, it is a valid strategy to contribute towards minimizing the existing research deficit and to answer our research questions.

6.2 Sample

In every school and every year, students in their last year (Grade 13) of academic upper secondary schools (*Gymnasium*) participated in the study. The data were collected by

written questionnaires in all 19 academic upper secondary schools in Bremen. In Hesse, nine academic upper secondary schools participated, they are representative for all academic upper secondary schools in Hessen in terms of school type and region. In total, 28 schools were included in our analyses.

In total, between 751 and 1157 students (19 schools) per year participated in the study in Bremen and between 510 and 680 (9 schools) participated in Hesse. The response rate was satisfactory to good, with 51 % (2007) to 74 % (2011) in Bremen and 73 % (2007) to 81 % (2011) in Hesse.

Students had to respond to the items separately for each of their three written exam subjects, including an advanced level mathematics or English course.

In this study, advanced level English and mathematics courses were analyzed. These courses belong to the most often selected exam subjects in Bremen and Hesse and are highly relevant for successful graduation from academic upper secondary school. Within the 3 years analyzed, 617 students in Hesse (2007: $N=204$; 2008: $N=181$; 2011: $N=232$) and 1218 students in Bremen (2007: $N=330$; 2008: $N=401$; 2011: $N=487$) took the advanced level English course; 489 students in Hesse (2007: $N=158$; 2008: $N=150$; 2011: $N=181$) and 847 students in Bremen (2007: $N=220$; 2008: $N=289$; 2011: $N=338$) took the advanced level mathematics course. In the English courses in Hesse, 61 to 65 % of the participants were young women; in the mathematics courses, 40 to 46 % were young women. In the English courses in Bremen, 60 to 63 % were young women; in the mathematics courses, 30 to 39 % were young women. The percentage of female students in the data set corresponds to the percentage of female students in the population. Because cognitive ability was controlled for in the multilevel models, students with missing values in cognitive ability were excluded (Hesse: mathematics 27 in 2007, 28 in 2008, Bremen: 78 in 2007, 21 in 2008). By this exclusion, in Bremen, the percentage of young women did not change more than 1 %; in Hesse, it changed 2.6 % at maximum (2008 in mathematics, 2007 in English), with more young women participating in the smaller sample.

6.3 Instruments

To examine the research questions, the motivational dimensions interest, scholastic self-efficacy, and persistence were measured using standardized scales one month before the exit exams. Additionally, students were asked to assess their teachers' supportive teaching practices. The first dimension, perceived "teachers' ability to motivate," is related to Kunter et al.'s (2013) concept of teacher enthusiasm and assesses the extent to which students believe that the teaching stretches/challenges and inspires, and whether their teachers enjoy teaching. The other two dimensions, "teacher autonomy support" and "teacher competency support," were selected based on Deci and Ryan's (1985) self-determination theory. According to this theory, teacher autonomy support and competence support are relevant prerequisites for the development of students' intrinsic motivation. Previous studies found positive effects of perceived teachers' autonomy support and competence support on students' intrinsic motivation, self-efficacy, and behavioral and cognitive engagement in school (e.g., Assor et al. 2002; Kunter et al. 2007, 2013; Roth et al. 2007; Schwarzer and Jerusalem 2002). Further, the results of a study of Maag Merki et al. (2010) revealed that the implementation of state-wide exit exams was correlated with a higher level of teacher

support. They are therefore relevant constructs to assess our main research questions. The response scale for all indicators ranged from 1 (“not true at all”) to 4 (“very true”). Reliability was merely satisfactory only for the indicator “perceived teacher autonomy support.” For all of the other dimensions, the reliabilities were acceptable to good.

6.3.1 Scales for students' motivational orientations

- *Interest* (3 items): Example item: “When I work on this subject, I sometimes forget everything else around me” (source: PISA-Konsortium Deutschland 2000). Year- and state-specific Cronbach's alpha values for mathematics were 0.79 to 0.86; year- and state-specific Cronbach's alpha values for English were 0.75 to 0.80.
- *Scholastic self-efficacy* (4 items): Example item: “I can answer even the difficult questions in class, if I try hard” (source: Jerusalem and Satow 1999). State- and year-specific Cronbach's alpha values for mathematics were 0.85 to 0.87; Cronbach's alpha values for English were 0.76 to 0.84.
- *Persistence* (3 items): Example item: “Even when I run into difficulty when preparing for the test, I remain determined and keep at it” (source: Grob and Maag Merki 2001). Year- and state-specific Cronbach's alpha values for mathematics were 0.73 to 0.80; year- and state-specific Cronbach's alpha values for English were 0.70 to 0.77.

6.3.2 Scales for perceived motivating teaching practices

All items were adapted from Leutwyler and Maag Merki (2005). The original sources are named for the individual scales.

- *Perceived teachers' ability to motivate* (5 items): Example item: “In the classroom my teacher's enthusiasm often motivates me” (source: Items were adapted from Baumert et al. 1997; Prenzel et al. 1996). Year- and state-specific Cronbach's alpha values for mathematics were 0.77 to 0.81; year- and state-specific Cronbach's alpha values for English were 0.80 to 0.87.
- *Perceived autonomy support by teacher* (4 items): Example item: “In class, I can often decide myself how to work” (source: Prenzel et al. 1996). Year- and state-specific Cronbach's alpha values for mathematics were 0.64 to 0.68; Cronbach's alpha values for English were 0.63 to 0.72.
- *Perceived competency support by teacher* (3 items): Example item: “In mathematics class my mathematics teacher keeps me regularly informed on my progress” (source: Prenzel et al. 1996). Year- and state-specific Cronbach's alpha values for mathematics were 0.73 to 0.79; year- and state-specific Cronbach's alpha values for English were 0.77 to 0.80.

To test for possible cohort effects in the change in student motivation, we controlled for students' sex and cognitive ability. For cognitive ability, we used a 25-item figure analogies subtest from a German cognitive ability test (KFT; Heller and Perleth 2000).

Based on the scale values, missing values were imputed using multiple imputations in SPSS 18 (Graham 2009; Lüdtke et al. 2002). The descriptive statistics reported in the

following combine the values of the individual data sets using the formula developed by Rubin (1987).

6.4 Analysis strategies

Descriptive analyses were carried out in Bremen and Hesse for each of the two courses (advanced level English and mathematics courses). We calculated difference-in-differences analyses by “year” and “state,” taking into account the multilevel structure of the data. Here, the possible effect of state-wide exit exams was examined by comparing the change in the two states over the years considered. The assumption is that there are differences between the two states in the students’ development of the motivational orientations, if the type of exams changed in one state (Bremen) but not in the other (Hesse). In the following, different effects of “year” appear that can be identified empirically via the cross-level interaction effects between the variables “year” and “state.”

Long-term (2007–2011) changes in student interest, scholastic self-efficacy, and persistence were tested using multilevel analysis in the statistics program HLM (version 6.06). The evaluations were based on two two-level models, one for each course (mathematics, English). Since students from the same school with the same subject and course level attended several courses in the surveyed schools, they were part of the same classes only per chance; therefore, we defined the schools and not the courses as level 2 (level 1: students). In both models the domain-specific motivation variable was set as the dependent variable. Their intraclass correlations (ICCs), and those for further level 1 variables, varied from 0.9 % (“persistence” in advanced level mathematics courses) to 4.8 % (“interest” in advanced level mathematics courses). Even for such rather low ICCs, multilevel analysis has its advantages in that it provides correct standard errors and allows for a possible variation of effects on level two.

6.4.1 Question 1

The first model (model 1) tested whether students’ motivational variables changed in the long-term (main effect) and whether these changes were different for the two German states (Bremen and Hesse). At level 1, the independent variables were the two dummy variables “Y0708” (0 = 2007, 1 = 2008) and “Y0711” (0 = 2007, 1 = 2011) to determine the changes from year 2007 to 2008 and from year 2007 to 2011, and at level 2, the variable “state” (1 = Bremen, 0 = Hesse). To make sure that potential changes between the years did not occur due to a cohort effect, students’ sex (0 = female, 1 = male) and cognitive ability (ranging from 0 to 25) were entered as control variables on level 1, sex also uncentered, and cognitive ability grand mean centered. Both the fixed and the random effects were included in the regression equation. By means of the interaction effect between state and year dummy variables, we checked if the changes in the motivational variables were specific to Bremen or occurred in both states. This is important because only in Bremen can we observe the change from class-based (2007) to state-wide exams (2008 and later) and therefore analyze the longer term effect of the implementation of state-wide exams from 2007 to 2011. We expect to find a larger change in students’ motivational dimensions in Bremen than in Hesse.

6.4.2 Question 2

To test for a mediator effect (Baron and Kenny 1986; Zhang et al. 2009), first it is necessary to test for an effect of state-wide exit exams on the dependent variables, as was done in model 1. Second, we need to check for an effect of state-wide exit exams on teaching practices. This was done in model 2, following the analysis strategies described above in model 1. Third, we need to test whether the effect of state-wide exit exams disappears or becomes smaller in a model that also considers the teaching practices as mediator variables. This last step was done in model 3. Here, student perceptions of teachers' ability to motivate and to provide autonomy support and competence support were added grand mean centered on the students' level. The teaching variables were added on the students' level, because empirical studies have shown that the individual level—meaning the students' individual perceptions of whether teachers are motivating, enthusiastic, and supportive with regard to their learning—is the most important level for the development of motivational dimensions of learning (Kunter et al. 2007; Rakoczy 2008). The slopes of the teaching variables were allowed to vary between schools.

The aim of model 3 was to find out if a Bremen-specific change of motivation orientations was mediated by teachers' motivating teaching, which should be reflected in the disappearance of the interaction effect. In the case of a suppressor effect, however, an interaction effect may not disappear but occur when the teaching variables are considered. This can happen, if there is a positive relationship between the implementation of state-wide exit exams and teaching but a negative relationship between state-wide exit exams and motivation, for example, because motivational variables in one state increase less or decrease more than in the other state in association with to teachers' motivational teaching.

This points to the effectiveness of teaching practices in relation to students' development of interest, self-efficacy, and persistence and to the possibility that there are influences on motivational variables that are independent of teaching practices, since the interrelationship between teacher practices and student motivation is rather weak (Rakoczy 2008; Seidel 2006).

7 Results

In this section, we present the descriptive statistics for all variables and then the outcomes of the multilevel analyses, focusing first on research question 1, concerning the state-specific change in students' motivational orientations, and then on the mediation of this change by teacher practices (research question 2, models 2 and 3).

7.1 Descriptive statistics for motivational dimensions

Table 1 shows the descriptive statistics of the motivational orientations for the two German states in advanced level English and mathematics courses. Short- and long-term changes in Bremen refer to comparisons between the time before (2007) and after (2008 and 2011) the change to state-wide exams. Significance was tested by means of variance analyses. The outcomes for English in Bremen showed a

Table 1 Descriptive statistics of interest, scholastic self-efficacy and persistence for advanced level English and math courses

		Advanced English courses					
		Interest		Scholastic self-efficacy		Persistence	
		Hesse	Bremen	Hesse	Bremen	Hesse	Bremen
2007	<i>M</i>	2.86	2.67	3.11	3.08	2.95	3.01
	<i>SD</i>	0.72	0.78	0.55	0.70	0.59	0.67
	<i>SE</i>	0.05	0.04	0.04	0.04	0.04	0.04
	<i>N</i>	204	330	204	330	204	330
2008	<i>M</i>	2.92	2.86	3.04	3.07	2.94	2.98
	<i>SD</i>	0.77	0.71	0.59	0.62	0.72	0.61
	<i>SE</i>	0.06	0.04	0.05	0.03	0.06	0.03
	<i>N</i>	181	401	181	401	181	401
2011	<i>M</i>	2.78	2.88	3.26	3.19	2.97	2.97
	<i>SD</i>	0.72	0.74	0.55	0.60	0.66	0.68
	<i>SE</i>	0.05	0.03	0.04	0.03	0.04	0.03
	<i>N</i>	232	487	232	487	232	487
Effect size (Cohen's <i>d</i>)							
	Y0708	0.08	0.26**	-0.12	-0.02	-0.02	-0.05
	Y0711	-0.11	0.28***	0.27**	0.17*	0.03	-0.06
		Advanced math courses					
		Interest		Scholastic self-efficacy		Persistence	
		Hesse	Bremen	Hesse	Bremen	Hesse	Bremen
2007	<i>M</i>	3.07	2.80	2.87	2.77	3.08	3.03
	<i>SD</i>	0.74	0.88	0.71	0.75	0.63	0.71
	<i>SE</i>	0.06	0.06	0.06	0.05	0.05	0.05
	<i>N</i>	158	220	158	220	158	220
2008	<i>M</i>	3.06	2.87	2.89	2.72	3.09	2.90
	<i>SD</i>	0.72	0.82	0.67	0.77	0.63	0.66
	<i>SE</i>	0.06	0.05	0.05	0.05	0.05	0.04
	<i>N</i>	150	289	150	289	150	289
2011	<i>M</i>	3.16	2.95	3.00	2.85	3.11	2.97
	<i>SD</i>	0.80	0.79	0.70	0.74	0.67	0.68
	<i>SE</i>	0.06	0.04	0.05	0.04	0.05	0.04
	<i>N</i>	181	338	181	338	181	338
Effect size (Cohen's <i>d</i>)							
	Y0708	-0.01	0.08	0.03	-0.07	0.02	-0.19*
	Y0711	0.12	0.18*	0.18	0.11	0.05	-0.09

Y0708 = change from year 2007 to 2008; Y0711 = change from year 2007 to 2011

M combined mean, *SD* combined standard deviation (naive pooling), *SE* standard error of the mean

* $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

small (following Cohen 1988) increase in interest that started directly after the change to state-wide exams and remained in the long term and a small- to medium-sized increase

in scholastic self-efficacy in the long term (2007 to 2011). Changes in Hesse refer to comparisons between the first year and the second and fifth years of state-wide exams. Here we also found a long-term increase in scholastic self-efficacy.

For mathematics, we likewise found a statistically long-term increase in interest and a short-term decrease in persistence in Bremen and no changes in Hesse. Thus, a long-term positive change in interest occurred in both school subjects with only the change to state-wide exams in Bremen. No long-term negative changes in motivational variables were found.

Table 2 shows descriptive statistics for the three student-reported measures of teaching practices. In advanced level English courses, all three measures—teachers' ability to motivate, autonomy support, and competence support—increased from the year before (2007) to the year after (2008) the change to state-wide exams in Bremen, and these positive changes remained in the long term (2011). Compared to that, in Hesse, without a change to state-wide exams, there was no significant change in student perceptions of teaching practices. In the mathematics courses, the results were the reverse. Here, no changes in supportive and motivating teaching were observed with the change to state-wide exams in Bremen, but long-term changes of teacher autonomy support and competence support were found in Hesse.

Table 3 shows that the motivational orientations were highly correlated with each other; the same was true for the teaching dimensions. Motivational and teaching dimensions were also correlated with each other; here, the effect was medium to large, with smaller correlations with persistence and larger correlations with interest.

7.2 Multilevel analyses: advanced level English courses

In the multilevel analyses, an interaction effect between state and year dummies indicated whether a change of motivational variables was specific to Bremen or occurred in both states, controlling for sex and cognitive ability. For student interest, model 1a in Table 4 shows a significant interaction effect ($p < 0.05$) between state and year dummy (2007 to 2011) that indicates a long-term increase in Bremen, where the system changed to state-wide exams, but no change in Hesse. Table 5 shows the outcomes of model 2 (the second step of the test for mediation). The Bremen-specific increase in interest was found in parallel with a Bremen-specific increase in teacher competence support in advanced level English courses from 2007 to 2011. Similar Bremen-specific increases in teachers' ability to motivate and autonomy support were not significant. No main effect for change in teaching practices over time was found, but the random slopes indicated a high variance between schools.

Table 4 shows the last step, model 1b. It provides two important pieces of information. First, all teaching scales positively predicted student interest. Second, the interaction effect for interest disappeared if the teaching behavior variables were included in the model. These outcomes support the hypothesis that the change in interest in Bremen was mediated by teaching, especially teacher competence support, the only instructional behavior that changed significantly over time.

Scholastic self-efficacy also changed positively in the long term; however, it changed in both states (main effect), so this increase was not specific to Bremen, where

Table 2 Descriptive statistics of ability to motivate, autonomy support, and competence support for advanced level English and math courses

		Ability to motivate		Autonomy support		Competence support	
		Hesse	Bremen	Hesse	Bremen	Hesse	Bremen
Advanced English courses							
2007	<i>M</i>	2.53	2.24	2.44	2.27	2.58	2.34
	<i>SD</i>	0.75	0.71	0.58	0.66	0.65	0.71
	<i>SE</i>	0.05	0.04	0.04	0.04	0.05	0.04
	<i>N</i>	204	330	204	330	204	330
2008	<i>M</i>	2.55	2.53	2.50	2.43	2.54	2.47
	<i>SD</i>	0.73	0.70	0.63	0.63	0.67	0.69
	<i>SE</i>	0.06	0.04	0.05	0.03	0.05	0.03
	<i>N</i>	181	401	181	401	181	401
2011	<i>M</i>	2.41	2.53	2.38	2.46	2.53	2.59
	<i>SD</i>	0.80	0.72	0.63	0.60	0.71	0.68
	<i>SE</i>	0.05	0.03	0.04	0.03	0.05	0.03
	<i>N</i>	232	487	232	487	232	487
Effect size (Cohen's <i>d</i>)							
Y0708	0.03	0.41***	0.10	0.25***	-0.06	0.19*	
Y0711	-0.15 ⁺	0.41***	-0.10	0.30***	-0.07	0.36***	
Advanced math courses							
		Ability to motivate		Autonomy support		Competence support	
		Hesse	Bremen	Hesse	Bremen	Hesse	Bremen
2007	<i>M</i>	2.74	2.57	2.45	2.52	2.59	2.48
	<i>SD</i>	0.58	0.69	0.59	0.63	0.63	0.66
	<i>SE</i>	0.05	0.05	0.05	0.05	0.05	0.05
	<i>N</i>	158	220	158	220	158	220
2008	<i>M</i>	2.62	2.50	2.48	2.50	2.58	2.48
	<i>SD</i>	0.66	0.66	0.64	0.64	0.59	0.67
	<i>SE</i>	0.05	0.04	0.05	0.04	0.05	0.04
	<i>N</i>	150	289	150	289	150	289
2011	<i>M</i>	2.87	2.50	2.76	2.52	2.76	2.51
	<i>SD</i>	0.68	0.74	0.61	0.65	0.61	0.65
	<i>SE</i>	0.05	0.04	0.05	0.04	0.05	0.04
	<i>N</i>	181	338	181	338	181	338
Effect size (Cohen's <i>d</i>)							
Y0708	-0.19	-0.10	0.05	-0.03	-0.02	0.00	
Y0711	0.20 ⁺	-0.10	0.52***	0.00	0.27**	0.05	

Y0708 = change from year 2007 to 2008; Y0711 = change from year 2007 to 2011

M combined mean, *SD* combined standard deviation (naive pooling), *SE* standard error of the mean

⁺ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

class-based exams changed to state-wide exams. However, when teaching practices were added to model 1b, a negative interaction effect emerged that pointed to a higher

Table 3 Inter-correlation across 2007, 2008, and 2011 between motivational and instructional dimensions (math and English)

Measure	1	2	3	4	5	6
1. Interest	–	0.59***	0.56***	0.44***	0.42***	0.47***
2. Scholastic self-efficacy	0.44***	–	0.54***	0.28***	0.33***	0.53***
3. Persistence	0.47***	0.37***	–	0.31***	0.30***	0.37***
4. Ability to motivate	0.48***	0.41***	0.28***	–	0.57***	0.56***
5. Autonomy support	0.44***	0.23***	0.28***	0.65***	–	0.57***
6. Competence support	0.48***	0.44***	0.32***	0.58***	0.59***	–

Inter-correlations for advanced level math courses are presented above the diagonal, and inter-correlations for advanced level English courses are presented below the diagonal

*** $p < 0.001$

long-term increase in scholastic self-efficacy in Hesse than in Bremen, when teaching practices were considered. Indeed, the increasing teacher competence support found mainly in Bremen had a positive effect on student self-efficacy, so that these outcomes suggest a suppressor effect. In contrast to competence support, teachers' higher ability to motivate was related to lower student self-efficacy.

Student persistence did not change over time in advanced level English courses, and there was also no interaction effect. However, as with scholastic self-efficacy, if teacher practices were considered in model 1b, there was a negative interaction effect that pointed to a decrease in persistence in Bremen and no change in Hesse. All three teaching practices had a statistically significant positive impact on student persistence, but perceived competence support, increasing especially in Bremen, was the teaching behavior with the largest effect. As for self-efficacy, these outcomes suggest a suppressor effect of the growing competence support in Bremen, compensating for probably decreasing persistence in the long term.

7.3 Multilevel analyses: advanced mathematics courses

Table 6 presents the outcomes of the multilevel analyses for advanced level mathematics courses. Whereas in the descriptive statistics a long-term increase in interest was found in Bremen, there was no significant main effect of the year dummy variables in model 1a of the multilevel analysis and no interaction effect between state and year. Table 5 shows the outcomes of the second step of test for mediation, model 2 for mathematics courses. There was an increase in teacher autonomy support in the long term that was much larger for Hesse, however, as indicated by a negative interaction effect. For teacher competence support, a general positive development of marginal significance was found. As shown in model 1b (Table 6), in mathematics, as in English courses, all teachers' motivating teaching behaviors had a positive impact on interest. If these practices were included into the model, a positive interaction effect between state and change over time occurred, indicating a long-term increase in student interest in Bremen but not in Hesse.

Students in both states perceived a long-term increase in scholastic self-efficacy ($p < 0.01$). Likewise, no interaction effect between year and state was found. Model 1b shows a large impact of teacher competence support on self-

Table 4 Effects on interest, scholastic self-efficacy, and persistence in advanced level English courses

German AC	Interest		Scholastic self-efficacy		Persistence	
	Model 1a	Model 1b	Model 1a	Model 1b	Model 1a	Model 1b
Fixed effects						
Intercept	2.94** (0.09)	2.88** (0.03)	3.13** (0.05)	3.10** (0.05)	3.05** (0.07)	3.00** (0.03)
Level 1						
Y0708	0.07 (0.10)	0.04 (0.06)	-0.08 (0.05)	-0.06 (0.04)	-0.02 (0.10)	-0.01 (0.07)
Y0711	-0.10 (0.07)	-0.03 (0.05)	0.15* (0.05)	0.19** (0.05)	-0.03 (0.08)	0.04 (0.05)
Sex (1 = m)	-0.15** (0.03)	-0.13** (0.03)	-0.04 (0.03)	-0.03 (0.03)	-0.21** (0.04)	-0.19** (0.03)
CAT	-0.01 (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.01* (0.00)	-0.01* (0.00)
Motiv	-	0.24** (0.04)	-	-0.12* (0.03)	-	0.09* (0.03)
Aut sup	-	0.13* (0.04)	-	0.04 (0.04)	-	0.08* (0.03)
Comp sup	-	0.31** (0.03)	-	0.45** (0.02)	-	0.21** (0.03)
Level 2 (schools)						
State (St)	-0.17 (0.11)	-0.02 (0.05)	-0.04 (0.08)	0.04 (0.06)	0.05 (0.08)	0.15* (0.05)
Interactions						
Y0708 × St	0.08 (0.12)	0.02 (0.07)	0.07 (0.09)	0.02 (0.07)	-0.03 (0.11)	-0.10 (0.08)
Y0711 × St	0.28* (0.12)	0.05 (0.06)	-0.04 (0.09)	-0.18* (0.07)	-0.01 (0.10)	-0.17* (0.06)
Random parameters						
Intercept U0	0.05** (0.23)	0.00 (0.07)	0.04** (0.20)	0.02* (0.14)	0.02* (0.14)	0.00 (0.05)
Y0708 slope	0.03 (0.17)	0.02 (0.13)		0.02 ⁺ (0.14)	0.02* (0.15)	0.01 (0.09)
Y0711 slope	0.08** (0.28)	0.01 (0.08)	0.04* (0.19)	0.01 (0.12)	0.03* (0.17)	0.01 (0.08)
Sex slope	0.00 (0.04)	0.00 (0.06)	0.03* (0.17)	0.00 (0.05)	0.01 ⁺ (0.11)	0.01 ⁺ (0.08)
CAT slope	0.00 (0.00)	0.00 (0.01)	0.00 (0.07)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)
Motiv slope	-	0.02* (0.16)	0.00 (0.00)	0.01 (0.10)	-	0.01 (0.10)
Aut sup slope	-	0.02 (0.13)		0.02* (0.15)	-	0.01 (0.09)
Com sup slope	-	0.01 (0.10)		0.00 (0.05)	-	0.00 (0.07)
Level 1 R	0.52 (0.72)	0.36 (0.60)	0.35 (0.59)	0.27 (0.52)	0.40 (0.63)	0.36 (0.60)
ICC1	1.9 %		2.8 %		1.2 %	

St state (1 = Bremen, 0 = Hesse), *Motiv* ability to motivate, *Aut sup* autonomy support, *Comp sup* competence support, *CAT* cognitive ability test (Heller and Perleth 2000); unstandardized coefficients (standard error)

⁺ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, sex (1 = male, 0 = female)

efficacy that, if included in the model, reduced the long-term change such that it was no longer significant, thus pointing to a mediation effect of teaching practices. However, the general increase in competence support that was found was only marginally significant. Additionally, the growth of student self-

Table 5 Effects on competence support, autonomy support and ability to motivate

Parameter	Advanced level math courses			Advanced level English courses		
	Competence support	Autonomy support	Ability to motivate	Competence support	Autonomy support	Ability to motivate
	Model 2	Model 2	Model 2	Model 2	Model 2	Model 2
Fixed effects						
Intercept	2.51*** (0.06)	2.46*** (0.08)	2.81*** (0.08)	2.63*** (0.11)	2.50*** (0.10)	2.56*** (0.16)
Level 1						
Y0708	-0.02 (0.12)	0.03 (0.12)	-0.18 (0.16)	-0.09 (0.10)	0.05 (0.10)	0.01 (0.17)
Y0711	0.17 ⁺ (0.09)	0.29** (0.08)	0.10 (0.13)	-0.11 (0.09)	-0.09 (0.09)	-0.17 (0.17)
Sex (1 = m)	0.12* (0.04)	-0.02 (0.03)	-0.04 (0.04)	-0.02 (0.03)	-0.10*** (0.02)	-0.03 (0.03)
CAT	0.01* (0.00)	0.01* (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
Level 2 (Schools)						
State (St)	-0.08 (0.08)	0.09 (0.09)	-0.18 (0.12)	-0.26* (0.12)	-0.14 (0.12)	-0.16 (0.18)
Interactions						
Y0708 x St	-0.00 (0.15)	-0.07 (0.15)	0.08 (0.21)	0.18 (0.12)	0.04 (0.12)	0.13 (0.21)
Y0711 x St	-0.14 (0.12)	-0.25* (0.11)	-0.14 (0.19)	0.32* (0.12)	0.22 (0.13)	0.32 (0.23)
Random parameters						
Intercept U0	0.02* (0.14)	0.05** (0.22)	0.07*** (0.26)	0.08*** (0.28)	0.08*** (0.28)	0.18*** (0.42)
Y0708 slope	0.09*** (0.29)	0.14*** (0.37)	0.31*** (0.56)	0.07*** (0.27)	0.06*** (0.25)	0.24*** (0.49)
Y0711 slope	0.08*** (0.27)	0.08*** (0.28)	0.27*** (0.52)	0.08*** (0.28)	0.10*** (0.32)	0.37*** (0.61)
Sex slope	0.02** (0.14)	0.01 (0.10)	0.01 (0.10)	0.01 (0.08)	0.00 (0.05)	0.00 (0.06)
CAT slope	0.00 (0.01)	0.00 (0.01)	0.00* (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.01)
Level 1 R	0.37 (0.61)	0.34 (0.58)	0.35 (0.59)	0.43 (0.65)	0.35 (0.59)	0.42 (0.65)
ICC1	0.018	0.065	0.089	0.064	0.062	0.089

St state (1 = Bremen, 0 = Hesse), CAT cognitive ability test (Heller and Perleth 2000), unstandardized coefficients (standard error)

⁺ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, sex (1 = male, 0 = female)

efficacy was not specific to Bremen, so it is not clear if it was triggered by the change to state-wide exams.

Finally, no main effect of the year and no interaction effect were found for student persistence in the short or long term, in spite of the statistically significant short-term decrease in persistence of $d = -0.19$ in Bremen in Table 1. The inclusion of motivating teaching practices in model 2 resulted in an effect of competence support on persistence but did not change the impact of the year. Accordingly, no mediation effects occurred.

Table 6 Effects on interest, scholastic self-efficacy, and persistence in advanced level math courses

Math AC	Interest		Scholastic self-efficacy		Persistence	
	Model 1a	Model 1b	Model 1a	Model 1b	Model 1a	Model 1b
Fixed effects						
Intercept	3.13** (0.08)	3.13** (0.05)	2.75** (0.06)	2.78** (0.06)	2.82* (1.04)	2.82* (1.04)
Level 1						
Y0708	-0.01 (0.08)	-0.01 (0.05)	0.02 (0.06)	0.02 (0.06)	0.01 (0.07)	0.02 (0.06)
Y0711	0.09 (0.06)	-0.08 (0.07)	0.16* (0.05)	0.07 (0.04)	0.05 (0.09)	-0.04 (0.09)
Sex (1 = m)	-0.09 ⁺ (0.05)	-0.12* (0.04)	0.16* (0.05)	0.09* (0.03)	-0.10 ⁺ (0.05)	-0.13* (0.06)
CAT	0.01* (0.01)	0.01 (0.00)	0.02** (0.00)	0.01* (0.00)	0.01 (0.01)	0.00 (0.00)
Motiv	-	0.24** (0.03)	-	-0.04 (0.03)	-	0.08 ⁺ (0.04)
Aut sup	-	0.21** (0.05)	-	0.06 (0.04)	-	0.10 ⁺ (0.05)
Comp sup	-	0.31** (0.05)	-	0.58** (0.04)	-	0.25* (0.10)
Level 2 (Schools)						
State (St)	-0.25* (0.10)	-0.20* (0.06)	-0.06 (0.08)	-0.02 (0.07)	-0.02 (0.06)	0.01 (0.06)
Interactions						
Y0708 × St	0.06 (0.13)	0.09 (0.09)	-0.10 (0.11)	-0.07 (0.09)	-0.14 (0.11)	-0.13 (0.10)
Y0711 × St	0.06 (0.11)	0.23* (0.08)	-0.10 (0.11)	-0.01 (0.08)	-0.09 (0.11)	-0.02 (0.10)
Random parameters						
Intercept U0	0.04* (0.21)	0.02 (0.13)	0.02 (0.15)	0.02 (0.14)	0.02 (0.14)	0.02 (0.13)
Y0708 slope	0.06* (0.25)	0.04 ⁺ (0.19)	0.04 (0.20)	0.01 (0.10)	0.04 (0.20)	0.04 (0.21)
Y0711 slope	0.03 (0.17)	0.04 (0.19)	0.04 ⁺ (0.20)	0.01 (0.09)	0.04 ⁺ (0.20)	0.04* (0.21)
Sex slope	0.01 (0.08)	0.01 (0.09)	0.01 (0.11)	0.01 (0.08)	0.01 (0.10)	0.01 (0.09)
CAT slope	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.04)	0.00 (0.05)
Motiv slope	-	0.00 (0.06)	-	0.00 (0.06)	-	0.01 (0.09)
Aut sup slope	-	0.03 ⁺ (0.17)	-	0.01 (0.11)	-	0.01 (0.11)
Com sup slope	-	0.02 ⁺ (0.15)	-	0.01 (0.11)	-	0.01 (0.09)
Level 1 R	0.60 (0.77)	0.44 (0.66)	0.50 (0.71)	0.36 (0.60)	0.38 (0.62)	0.32 (0.56)
ICC1	4.8 %		1.7 %		0.9 %	

St state (1 = Bremen, 0 = Hesse), Motiv ability to motivate, Aut sup autonomy support, Comp sup competence support, CAT cognitive ability test (Heller and Perleth 2000), unstandardized coefficients (standard error)

⁺ $p < 0.10$; * $p < 0.05$ or $p < 0.01$; ** $p < 0.001$, sex (1 = male, 0 = female)

8 Discussion

In this 5-year study (2007–2011), we analyzed the long-term effects of the implementation of state-wide exit exams at the end of academic upper secondary schools in Bremen, Germany, on students' domain-specific interest, scholastic self-efficacy, and

persistence while they were preparing for the exit exams. In Bremen, state-wide exit exams were implemented in 2008, which allowed us to investigate the change in the exam system from a class-based to a state-wide testing system. By calculating difference-in-differences analyses by “year” and “state,” it was possible to estimate the effect of the change to state-wide exams in Bremen taking Hesse as a control group. We hypothesized differing long-term effects in dependence on school subject as well as mediating effect by teaching practices.

To sum up, the results corroborate our assumption partially. As expected in our first hypothesis, we do not find general but rather subject-specific effects. Long-term effects are identified particularly in advanced level English courses (H1). Furthermore, the effects in advanced level English courses are mediated by the teaching practices as perceived by the students, particularly by perceived teacher competence support. This is in line with our second hypothesis (H2). The results for the English courses reveal that without considering teaching practices, the implementation of state-wide exit exams has a positive effect on only students’ subject-specific interest. Considering teaching practices, this effect is mediated by the increased teacher competence support. In contrast, student scholastic self-efficacy would not have been affected by the implementation of state-wide exit exams, if perceived teaching practices had not increased over time in Bremen. Additionally, student persistence would have been negatively affected, if increased teacher support had not compensated for this negative effect.

For the mathematic courses, the results are not that clear. Positive long-term effects can be identified on student self-efficacy in both states, but there are no effects on student interest and persistence. The effects on self-efficacy are influenced by teaching practices—but also in both states. A positive effect on interest occurs only if the lower teaching quality in Bremen is considered. In the following section, we discuss these results in detail for each dimension.

8.1 Effects of the change to a state-wide exit exam system

8.1.1 Interest

First, the outcomes for student interest in the advanced level English courses support the hypothesis that with the change of the exit exam system in Bremen, student interest increases in the long term. Second, the results show that this increase is mediated by supportive teaching practices. This can be seen in the Bremen-specific increase in teacher competence support and in the fact that consideration of these teaching practices reduces the change in interest, so that it is no longer significant. These results are in line with previous research that identified supportive teaching as a significant influencing factor on the development of interest (Deci and Ryan 1985).

In advanced level mathematics courses, no change in student interest occurs. Only when we consider the rather not increasing teaching support in Bremen, does it become clear that compared to Hesse, there is a positive Bremen-specific increase in mathematics interest from 2007 to 2011. The outcomes show an increase in teacher autonomy support only in Hesse. Accordingly, in contrast to advanced level English courses, the positive change in interest in mathematics courses in Bremen is not related to a positive change in teaching practices. Other factors might have had an impact on student interest

in Bremen. For instance, the announcement of the exit exams topics 2 years prior to the exams as well as the public availability of old exam questions/tasks gives students the opportunity to study for the exit exams autonomously. Further results revealed that these preparation strategies increased significantly after the implementation of state-wide exit exams. This is consistent with Bishop's (1999) finding that under central exit exam systems, students put more effort into learning outside school, supported by their parents and classmates. It would be interesting to examine whether these learning strategies are implemented also by the students who experienced strong support from their teachers or if instead, these school-external strategies compensate for lack of support at school.

In both courses, our findings do not corroborate Ryan and Sapp's (2005) assumption that state-wide exams could have negative effects on students' intrinsic motivation as a consequence of increased extrinsic motivation. One reason could be that the motivation to succeed on exit exams corresponds to a more autonomously driven form of extrinsic motivation that is regulated, for example, through identification, with studying for the exams being accepted as personally important (Deci and Ryan 1985). Further, the pressure—and thus extrinsic motivation—on students may be not too high, as the exit exam system is only standardized on a low level (Klein and Van Ackeren 2011) and the final decision on failure or success is still based to a significant degree on previous achievement and grades in academic upper secondary schools. Additionally, transparency of the exit exam system, the exit exam topics, and publicly available material to prepare for the exit exam could lead students to feel that there is not much difference between the old and the new system. Further, this feeling could be reinforced by reports in public media that the ratio of failure and success in the exit exam has been quite stable over the years.

8.1.2 Scholastic self-efficacy

Scholastic self-efficacy increased in both states in advanced level English courses in the long term. However, the increase in student self-efficacy in the English courses turns out to be rather specific to Hesse and not to Bremen, when perceived teaching practices are considered as a mediating variable. The mediating effect occurs due to the substantial increase in perceived teacher competence support in Bremen. Teacher competence support, such as through giving students self-referenced feedback, has a high impact on student self-efficacy (Chan and Lam 2010; Schwarzer and Jerusalem 2002). This might be an indication that a positive effect of state-wide exams occurs in Bremen mediated by more supportive teaching practices.

In Hesse, there is also a long-term increase but without a change of systems. Since in Hesse teachers' motivating teaching practices did not increase over the years, other factors might have influenced this result. Again, students might have benefited from the growing amount of old exam questions/tasks and preparation materials that are publicly available for all students. Preparation materials normally also provide the solutions to the questions/tasks; therefore, they could be used as a formative assessment (Chan and Lam 2010). The information provided on the respective performance level helps students deal with challenging exit exam goals and has a positive influence on their development of self-efficacy. Further, social persuasion, particularly from peers and parents (Bandura 1997), could compensate for the lack of support from teachers and

could lead to a higher level of student self-efficacy, positively influenced by the possibility to prepare for the exit exams over the period of 2 years.

The assumption that the implementation of state-wide exit exams could negatively affect students' belief that they are capable of successfully meeting the requirements in class is not supported by the empirical results, however, since we did not find negative effects in Bremen on student self-efficacy.

As in English courses, self-efficacy in mathematics courses increased in both states, an effect that is influenced by perceived teacher competence support, which shows marginally significant growth in the long term. However, as this change was not specific to Bremen, we cannot conclude that it is caused by the change in the type of exit exams.

8.1.3 Persistence

When teaching practices are not included in the analyses, no long-term effects on student persistence in advanced level English courses can be identified—neither in Bremen, where there was a change in the exit exam system, nor in Hessen, where there was no change in the same period. However, if teaching practices are taken into account, there is a long-term decrease in persistence in Bremen; whereas in Hesse, there is still no change over the years. These results point to a suppressor effect of teacher competence support on persistence. In contrast to English courses, in advanced level mathematics courses we found no long-term main effects or interaction effects. In both English and mathematics courses, teacher competence support is the most important aspect in relation to student persistence, but only in the English courses is there a significant increase in teacher competence support in Bremen.

These results do not support the assumptions that the higher level of rewards of the grades on state-wide exit exams (Piopiunik et al. 2014) as supposed by Bishop (1999) and the earlier announcement of the general exam topics enhance student persistence in mathematics courses. Instead, they suggest that there is even lower student engagement in preparing for the exit exam in the English courses, if teachers do not simultaneously provide increased support in their classes. Student attributions reported by Oerke et al. (2011) also indicate that students preparing for state-wide exit exams tend to rely to a higher degree on their teachers' explanations, which might be a consequence of teachers' tendency to focus on the exam topics, so that students gain the impression that they are well prepared.

Particularly for the advanced level English courses, it might be, rather, that the increased loss of control for the students—due to a higher level of insecurity with regard to the type and the content of the exam questions/tasks (Maag Merki 2011) or the higher level of pressure perceived by the teachers in state-wide exit exam systems (Oerke 2012)—negatively affects students' ability to sustain the learning process. In further research, these assumptions have to be analyzed in detail.

8.2 Conclusion

In the advanced level English courses, the study results identified a close relationship between the change to state-wide exams in all students' motivational variables analyzed and the change in teaching practices, particularly the Bremen-specific increase in

teacher competence support from 2007 to 2011. This means that increased teacher competence support in Bremen enhances student interest in the long term. This result may indicate a positive effect of the implementation of state-wide exams in Bremen and contradicts especially the fear that because of increased extrinsic incentives and increased control, intrinsic motivation might decrease after implementing state-wide exams, as predicted by the self-determination theory (Deci et al. 2001). However, scholastic self-efficacy and persistence in Bremen might have been negatively affected or not affected by state-wide exams, if teacher competence support had not improved and perhaps compensated for these effects. These trans-intentional results show the ambivalent effects of state-wide exit exams. On the one hand, negative effects occur, as many studies have shown (e.g., Jürges et al. 2012), but on the other hand, teacher support increases, which can be identified as a positive effect of the implementation of state-wide exit exams. Further analyses could examine this ambivalent dynamic of student learning and teaching in the context of state-wide exit exams in greater depth. Following this research aim, it would be important to take into account additional mediating variables not only on the student level (e.g., uncertainty about passing the exit exam; Maag Merki 2011) but also on the teacher (e.g., professional competence of teachers) and school level (e.g., collective self-efficacy or cooperation practices of teachers) that are relevant factors in dealing effectively with state-wide exit exams (Jäger et al. 2012).

For the advanced level mathematics courses, the outcomes are not so clear. A positive development of supportive teaching practices does not occur in Bremen but in Hesse only, and thus there is no Bremen-specific increase in scholastic self-efficacy and persistence. For interest, however, there is some indication of an increase that was not related to teacher support. With regard to that, there is a need for investigation of what factors besides teaching lead to the positive effect on student interest in Bremen mathematics courses. To examine this question, it would be interesting to include further variables in the calculations, such as student learning outside the classroom and at home (Bishop 1999). It would also be important to include the amount of “teaching to the test,” which has already been found to be a consequence of implementing state-wide exams and to be negatively related to interest (Jäger et al. 2012; Oerke et al. 2013). This would allow investigation of how positive and negative teacher practices interact with each other and influence students’ interest.

Teacher competence support (Deci and Ryan 1985) especially can be interpreted as a relevant mediating factor that influences the development of student interest—and further self-efficacy and persistence—in English courses when they are preparing for exit exams. Since this factor is positively affected by the implementation of state-wide exit exams, it has the potential to compensate for negative effects of state-wide exams. Therefore, for teaching practice, it would be important to implement teacher training programs that provide teachers with opportunities to enhance their own professional competency.

Additionally, to understand this mediator effect better, it would be important to conduct longitudinal studies analyzing the interaction effect between teaching and learning while students are preparing for exit exams, e.g., cross-lagged panel analyses and structural equation modeling, taking into account several motivational variables simultaneously. This would also lead to a better understanding of the domain-specific effects, since the effects in advanced level English courses are stronger than in

advanced level mathematics courses. In advanced level mathematics courses, students and teachers appear to be more independent of the change in the exam system. This is in line with a study by Baumert and Watermann (2000) at the end of upper secondary school that found no significant differences between state-wide and class-based exams systems on test anxiety in advanced level mathematics courses. One reason might be that subject-specific specifications in the old class-based exam system differed for mathematics and English, so that the change from the old to the new exam system was more substantial for English than for mathematics courses with regard to content and standards. Accordingly, the change to state-wide exams and the risk of losing control in the learning process to prepare for the exit exams could be more substantial for English than for mathematics students.

From a theoretical perspective, our analyses support the assumption of a complex interplay between student motivation, teacher practices, and the implementation of state-wide exit exams.

However, previous theoretical models are too vague and did not clearly fit the outcomes of this study. We also did not find any support for a decrease in student interest because of increasing extrinsic motivation, or for a decrease in teacher support as predicted by self-determination theory (see Ryan and Sapp 2005). Further, Bishop's educational economic perspective (Bishop 1999) could not be corroborated generally, as for mathematics students no effect on persistence was found and in English there was even a negative effect. Both theories thus do not seem to be satisfactory for description of what happens to students when state-wide exit exams are implemented. A stronger fit with existing theoretical models (Bishop 1999; Jürges and Schneider 2010) can be seen in the (partially) positive change in teacher support. However, as previous results revealed as well, there are not only positive changes but also increased teaching to the test as a consequence of implementing state-wide exams, which negatively influences student interest (Jäger et al. 2012; Oerke et al. 2013). Therefore, for more consistent theories on the interrelationship between exam system, teacher practices, and individual and class factors, more in-depth analyses of complex teacher practices will be needed.

For this reason, further theoretical models should be specified by including multivariate and non-linear pathways, considering intentional and trans-intentional direct and indirect effects, and taking into account not only mediation effects but also moderation effects, e.g., course level and domains. Additionally, the school and state context has to be considered more carefully, because dealing effectively with state-wide exit exams is not only influenced by individuals (e.g., teachers, students) but also by schools and the type of exit exams (Klein 2013).

8.3 Methodological limitations

There are various methodological limitations that should be kept in mind when interpreting the results. First of all, although we conducted a long-term study, we were not able to test a causal link between the implementation of state-wide exit exams, teacher practices, and students' motivational orientations. As we only compared different student cohorts and not a longitudinal sample in the analyses, we cannot be totally sure that we did not overlook important control variables. Control is given for student sex and cognitive ability, however, and the comparison between Bremen and Hesse allows us to a special extent to attribute

especially the changes in the English courses to the implementation of state-wide exams. However, it might be that the relationship between the implementation of state-wide exit exams and student motivation is mediated not only by perceived teacher practices but also by other dimensions.

Additionally, it would be important to conduct a multi-cohort longitudinal study to be able to examine student learning in the last years of academic upper secondary schools, comparing class-based and state-wide exam systems. This strategy could also address another limitation of our study: For the advanced courses, we were able to compare only 1 year with class-based exit exams to the 4 years of state-wide exit exams.

Further, when comparing the year 2007 with 2008 and 2011 in Bremen, it has to be considered that in 2007, state-wide exit exams were already implemented in the basic level courses. In the worst case, this could have increased student motivation in the advanced level courses already in 2007 without state-wide exit exams (e.g., transfer effect), with the consequence that a possible long-term effect could not have been detected (type II error). On the other hand, if the implementation of state-wide exit exams in the basic courses in 2007 had reduced student motivation also in the advanced courses, a possible long-term effect would have been overestimated (type I error). In both cases, however, this probability of error should be rather small, as the advanced courses count more for the total grade in the exams than basic level courses. Furthermore, our results show that there are no short-term effects on motivational orientations but instead positive long-term effects. Hence, it is not really plausible that in 2007, students in advanced courses without state-wide exit exams differed significantly from students in the old system.

Additionally, when comparing outcomes in Bremen with those in Hesse, it has to be considered that state-wide exam systems in Hesse were implemented in 2007 in all courses. This makes it still possible, however, to uncover short-term effects, as they should occur only in Bremen and not in Hesse, as well as long-term effects developing continuously, as they should be larger in Bremen than in Hesse. Nevertheless, we should keep in mind that we do not know in detail which school development strategies have been implemented by the teachers and principals that would have had an impact on students' motivational dimensions and teaching practices. However, to the best of our knowledge, having discussed the results in both states with teachers, principals and state administrators, there are no hints that other state related reforms had been implemented in Bremen and Hesse in the time between 2008 and 2011 that could explain changes in teaching practices in the states. Therefore, the cautious interpretation of these effects in relation to the implementation of state-wide exit examinations seems to be valid.

Finally, another problem of our study is the rather small subject-specific sample and the fact that we conducted this study in only two states. Before generalizing the findings, it would be important to corroborate these results in other states as well.

Acknowledgments We would like to express our great appreciation to Prof. Dr. Eckhard Klieme, Dr. Monika Holmeier, Dr. Daniela J. Jäger, and Elisabeth Maué for their substantial contributions to the whole project.

Compliance with ethical standards

Disclosure of potential conflicts of interest This work was supported by the German Research Foundation [MA 4184/3-1 and KL 1057/12-1] and by the two states Bremen and Hesse, Germany. Research independency was contractually agreed. Accordingly, we have no conflict of interest.

References

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the class-room: teachers' opinions of statewide testing programs. *Theory Into Practice*, *42*(1), 18–29.
- Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, *72*, 261–278.
- Au, W. (2007). High-stakes testing and curricular control: a qualitative metasynthesis. *Educational Researcher*, *36*(5), 258–267.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman and Company.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, *72*, 187–206.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Baumert, J. (2000). Lebenslanges Lernen und internationale Dauerbeobachtung der Ergebnisse von institutionalisierten Bildungsprozessen [Lifelong learning and international monitoring of the results of institutionalized educational processes]. In F. Achtenhagen & W. Lempert (Eds.), *Lebenslanges Lernen im Beruf—seine Grundlegung im Kindes- und Jugendalter. Band 5: Erziehungstheorie und Bildungsforschung* (pp. 121–127). Leske+Budrich: Opladen.
- Baumert, J., & Watermann, R. (2000). Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe [Institutional and regional variability and assuring common standards in the gymnasium upper level]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (pp. 317–372). Opladen: Leske+Budrich.
- Baumert, J., Gruehn, S., Heyn, S., Köller, O., & Schnabel, K. U. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation, Band 1. Skalen Längsschnitt I, Welle 1-4 [Educational trajectories and psycho-social development in adolescence (BIJU). Documentation, Vol. 1. Scales, longitudinal study I, waves 1–4]*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Bishop, J. H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review*, *6*, 349–398.
- Borman, G. D. (2009). The use of randomized trials to inform education policy. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of policy research* (pp. 129–138). New York: Routledge.
- Chan, J. C. Y., & Lam, S.-f. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, *38*, 37–58.
- Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: analyses of students in 41 countries. *Learning and Instruction*, *18*, 321–336.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.
- Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: reconsidered once again. *Review of Educational Research*, *71*(1), 1–27.
- EACEA/Eurydice. (2009). *National testing of pupils in Europe: objectives, organisation and use of results*. Brussels: Eurydice.
- Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. M. (2013). *New assessments, better instruction? Designing assessment systems to promote instructional improvement*. Santa Monica: RAND.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

- Grob, U., & Maag Merki, K. (2001). *Überfachliche Kompetenzen. Theoretische Grundlegung und empirische Erprobung eines Indikatorensystems [Cross-curricular competencies: Theoretical foundation and empirical analysis of an indicator system]*. Bern: Peter Lang.
- Heckhausen, H. (1989). *Motivation und Handeln [Motivation and action]*. Berlin: Springer.
- Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest KFT 4-12+ R (für 4. bis 12. Klassen, Revision) [Cognitive ability test KFT 4-12+ R (for grades 4 to 12, revised)]*. Göttingen: Beltz-Test GmbH.
- Jäger, D. J., Maag Merki, K., Oerke, B., & Holmeier, M. (2012). Statewide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German states. *Assessment in Education: Principles, Policy & Practice*, 19(4), 451–467.
- Jerusalem, M., & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung [Scholastic self-efficacy belief]. In R. Schwarzer & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs „Selbstwirksame Schulen“ [03.July 2015]* (pp. 15–16). Berlin: Freie Universität Berlin. Verfügbar unter: http://web.fu-berlin.de/gesund/skalen/Kollektive_Selbstwirksamkeit/kollektive_selbstwirksamkeit.htm.
- Jürges, H., & Schneider, K. (2010). Central exit examinations increase performance, but take the fun out of mathematics. *Journal of Population Economics*, 23, 497–517.
- Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: the effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, 31, 56–65.
- Klein, E. D. (2013). *Statewide exit exams, governance, and school development: an international comparison*. Münster: Waxmann.
- Klein, E. D., & Van Ackeren, I. (2011). Challenges and problems for research in the field of statewide exams. A stock taking of differing procedures and standardization levels. *Studies in Educational Evaluation*, 37, 180–188.
- Klieme, E. (2004). Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde [Rational, implementation and effects of performance standards: current discussion and empirical results]. *Zeitschrift für Pädagogik*, 50(5), 625–634.
- Krapp, A. (2002). An educational-psychological theory of interest and its relation to SDT. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 405–427). Rochester: University of Rochester Press.
- Krapp, A. (2005). Basic need and the development of interest and intrinsic motivational orientations. *Learning and Instruction*, 15, 381–395.
- Kuhl, J. (1987). Action control: the maintenance of motivational states. In F. Halisch & J. Kuhl (Eds.), *Motivation, intention, and volition* (pp. 279–291). Berlin: Springer.
- Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17, 494–509.
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, 18, 468–482.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.
- Leutwyler, B., & Maag Merki, K. (2005). Mittelschulerhebung 2004. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Schülerinnen und Schüler: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation [Assessment in high schools 2004. Indicators for context variables. Students' perspectives: Experiences in school and classes. Documentation of scales and items]. Zurich: Forschungsbereich Schulqualität & Schulentwicklung, Institute of Education, University of Zurich.
- Lüdtke, O., Robitzsch, A., & Köller, O. (2002). Statistische Artefakte bei Kontexteffekten in der pädagogisch-psychologischen Forschung [Handling missing values in psychological research: Problems and solutions]. *Zeitschrift für Pädagogische Psychologie*, 16(3/4), 217–232.
- Maag Merki, K. (2011). Effects of the implementation of state-wide exit exams on students' self-regulated learning. *Studies in Educational Evaluation*, 37, 196–205.
- Maag Merki, K., Holmeier, M., Jäger, D. J., & Oerke, B. (2010). Die Effekte der Einführung zentraler Abiturprüfungen auf die Unterrichtsgestaltung in Leistungskursen in der gymnasialen Oberstufe [The effects of the implementation of state-wide exit examinations on teaching in advanced courses in academic upper secondary schools]. *Unterrichtswissenschaft*, 34(2), 173–192.

- Oerke, B. (2012). Emotionaler Umgang von Lehrkräften und Schüler/-innen mit dem Zentralabitur: Unsicherheit, Leistungsdruck und Leistungsattribution [Teachers' and students' emotional dealings with state-wide exit exams: Uncertainty, pressure to achieve and attribution]. In K. Maag Merki (Ed.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland* (pp. 119–153). Wiesbaden: Springer VS.
- Oerke, B., Maag Merki, K., Holmeier, M., & Jäger, D. J. (2011). Changes in student attributions due to the implementation of central exit exams. *Educational Assessment, Evaluation and Accountability*, 23(3), 223–241.
- Oerke, B., Maag Merki, K., Maué, E., & Jäger, D. J. (2013). Zentralabitur und Themenvarianz: Lohnt sich Teaching-to-the-Test? [State-wide exit examination and content variance: Does teaching-to-the test pay off?]. In D. Bosse, F. Eberle, & S.-T. Barbara (Eds.), *Standardisierung in der gymnasialen Oberstufe* (pp. 27–50). Wiesbaden: Springer VS.
- Pedulla, J., Abrams, L. M., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: findings from a national survey of teachers*. Chestnut Hill: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.
- Piopiunik, M., Schwerdt, G., & Wössmann, L. (2014). Zentrale Abschlussprüfungen, Signalwirkung von Abiturnoten und Arbeitsmarkterfolg in Deutschland [Central school-leaving exams, signaling effects and labor-market outcomes in Germany]. *Zeitschrift für Erziehungswissenschaft*, 17, 35–60.
- PISA-Konsortium Deutschland. (2000). *Dokumentation der Erhebungsinstrumente [Documentation of the survey instruments] (Vol. 72)*. Berlin: Max-Planck Institut für Bildungsforschung.
- Prenzel, M., Kristen, A., Dengler, P., Ettl, R., & Beer, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung [Self-determinedly motivated and interested learning in vocational business school]. *Zeitschrift für Berufs- und Wirtschaftspädagogik. Beiheft*, 13, 108–127.
- Putwain, D. W., Connors, L., Woods, K., & Nicholson, L. J. (2012). Stress and anxiety surrounding forthcoming standard assessment tests in English schoolchildren. *Pastoral Care in Education*, 30(4), 289–302.
- Putwain, D. W., Daly, A. L., Chamberlain, S., & Sadreddini, S. (2015). Academically buoyant students are less anxious about and perform better in high-stakes examinations. *British Journal of Educational Psychology*, 85, 247–263.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht [Motivational support in mathematics classes]*. Münster: Waxmann.
- Reardon, S. F., Arshan, N., & Atteberry, A. (2010). Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. *Educational Evaluation and Policy Analysis*, 32(4), 498–520.
- Rheinberg, F. (2004). *Motivationsdiagnostik [Motivation diagnostic]*. Göttingen: Hogrefe.
- Richman, C. L., Brown, K., & Clark, M. (1987). Personality changes as a function of minimum competency test success or failure. *Contemporary Educational Psychology*, 12, 7–16.
- Roth, G., Assor, A., Kanat-Mayom, Y., & Kaplan, H. (2007). Autonomous motivation for teacher: how self-determined teaching may lead to self-determined learning. *Journal of Educational Psychology*, 99(4), 761–774.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Ryan, R. M., & Sapp, A. (2005). Considering the impact of test-based reforms: a self-determination theory perspective on high stakes testing and student motivation and performance. *Unterrichtswissenschaft*, 33(2), 143–159.
- Ryan, K. E., Ryan, A. M., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, 36(1), 5–13.
- Schwarzer, R., & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit [The concept of self-efficacy]. *Zeitschrift für Pädagogik*, 44. Beiheft, 28–53.
- Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research*, 9(3), 253–271.
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology*, 66, 531–568.
- Slavin, R. E. (2010). Experimental studies in education. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 102–114). London: Routledge.
- Wheelock, A., Bebell, D. J., & Haney, W. (2000). What can student drawings tell us about high-stakes testing in Massachusetts? Teachers College Record, ID Number: 10634.

- Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Mahwah: Erlbaum.
- Zhang, Z., Zyphur, M., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. problems and solutions. *Organizational Research Methods, 12*(4), 695–719.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.