

RESEARCH

Open Access

The role of test-taking motivation for students' performance in low-stakes assessments: an investigation of school-track-specific differences

Christiane Penk*, Claudia Pöhlmann and Alexander Roppelt

* Correspondence:
christiane.penk@iqb.hu-berlin.de
Institute for Educational Quality
Improvement (IQB), Humboldt
University Berlin, Unter den Linden
6, D-10099 Berlin, Germany

Abstract

Background: Low-stakes assessments do not have consequences for the test-takers. Currently, motivational research indicates that a lack of test-taking motivation can decrease students' performance in low-stakes assessments. However, little research has explored the domain-specific and situation-specific aspects of motivation simultaneously. Research examining differences in test-taking motivation among students in different types of schools is also limited. Our study therefore addressed the motivational determinants of test performance in low-stakes assessments, in general, as well as school-track-specific differences in particular.

Method: Drawing on national data from students who participated in a cross-national study of educational achievement, we conducted multiple regression analyses to predict the students' test performance and the effort they invested in that test. We conducted the analyses for the entire sample as well as for the students in that sample separated according to the school track they were attending.

Results: The results showed that, after we had controlled for self-concept in mathematics, test-taking motivation was significantly, but relatively weakly, associated with test performance: Students achieved higher test performance the more effort they invested and the less worry they experienced during the test. We also found school-track-specific differences for invested effort. Test attractiveness seems to be a more important inducement to invest effort for students in nonacademic-track schools than for students in academic-track schools.

Conclusions: The weak relationship between test-taking motivation and performance supports the validity of the applied low-stakes test. However, it seems that invested effort and worry are indispensable constructs for performance in low-stakes tests. For students of nonacademic tracks especially, an attractive and enjoyable test seems a crucial aspect of motivating them to expend their best effort. Implications for constructing low-stakes tests are discussed.

Keywords: Test-taking motivation; Low-stakes assessments; Effort; School-track-specific differences

Background

Over the last two decades, an increasing number of education systems (hereafter countries) have found their participation in large-scale cross-national educational assessments a more and more relevant part of quality evaluation of their school systems. Examples of these studies are the Programme for International Student Assessment (PISA), conducted by the Organisation for Economic Co-operation and Development (OECD), and the Trends in International Mathematics and Science Study (TIMSS), conducted by the International Association for the Evaluation of Educational Achievement (IEA). Germany is one of the countries that regularly takes part in these comparative studies.

The results of these studies allow countries not only to assess how well their own students are performing, on average, but also to assess that performance against the average performance of students in the other participating countries. These rankings play an important role in government-led educational decision-making, which forms the basis for reforms. In order to draw valid conclusions about students' abilities during this process, test-takers need to be motivated to expend full effort throughout the entire testing session. However, such tests have no positive or negative consequences for the test-takers, no matter how successfully or unsuccessfully they perform.

These tests are often referred to as *low-stakes* tests. Accordingly, it is uncertain whether the students do actually expend full effort; it could be that the students' results do not depict their true level of ability due to low motivation. Therefore, the results of low-stakes assessments may not constitute a valid measure of students' abilities. In this case, a valid interpretation of the test results is threatened. Our aim in this study is to provide a closer look at the role of test-taking motivation on student performance in low-stakes assessments. Before describing our research questions in detail, we define test-taking motivation and provide an overview of previous research.

Literature

Test-taking motivation

Test-taking motivation is a specific type of achievement motivation that can be understood as an active process by which goal-oriented activity is initiated and maintained (Schunk et al. 2008). It is assumed that students have domain-specific achievement motivation (e.g., motivation to engage in mathematics) and situation-specific achievement motivation (e.g., motivation to work hard in a specific school-based assessment). Domain-specific motivational constructs such as self-concept in mathematics cover a relatively stable personal trait, while situation-specific motivational constructs cover a state that can differ (e.g., depend on how the student feels "on the day"). Test-taking motivation is assigned to the latter motivational constructs, because taking a test is a specific situation for students. Baumert and Demmrich (2001) define this type of motivation as "the willingness to engage in working on test items and to invest effort and persistence in this undertaking" (p. 441).

In high-stakes tests, test-takers typically show high motivation to perform well because of the positive or negative consequences of their performance on that test (Barry and Finney 2009). Research exploring test-taking motivation relative to low-stakes assessments presents a less clear picture. Most of these studies show a connection between test-taking motivation and performance on the one hand, and between test-taking motivation and test stakes on the other (Cole et al. 2008; Eklöf 2007, 2008; Thelk et al. 2009; Wise and DeMars 2005; Wolf and Smith 1995). However, some studies have found no such relationships (Baumert

and Demmrich 2001; O'Neil et al. 1995, 2005). In the following subsections, we describe studies that have detected associations between test-taking motivation, performance, and test stakes, and those that have not. These studies include some of those just listed.

Studies showing associations

The investigation by Eklöf (2007, 2008) focused on the test-taking motivation of Swedish Grade 8 students in TIMSS 2003, deemed a low-stakes assessment, and examined both domain-specific and situation-specific aspects of motivation. In this study, the following motivational scales explained 31% of the variance in the students' average mathematics achievement scores: *mathematics self-concept* and *value of mathematics* as domain-specific factors of motivation as well as *test-taking motivation* as a situation-specific aspect of motivation. Of these variables, mathematics self-concept was the most important predictor. However, after controlling for the domain-specific factors of motivation, Eklöf no longer found a significant relationship for the situation-specific aspect of motivation. Eklöf assumed that test-taking motivation had no effect because most of these Swedish Grade 8 students, having not previously experienced receiving grades or taken external tests, did not perceive the test as a low-stakes one.

Eklöf and Nyroos's (2013) analyses of data pertaining to performance of Grade 9 students on the Swedish national test of science achievement in 2009 supported the findings from the 2003 TIMSS data: a significant relationship between performance in science and (a) reported effort ($r = 0.25$), (b) perceived importance of the test ($r = 0.20$), and (c) test anxiety ($r = -0.10$). However, the authors could not consider the domain-specific aspects of motivation in their analyses because data on this matter were not collected during the assessment.

Cole et al. (2008) investigated the relationship of the following situation-specific aspects of motivation to the mathematics test performance of undergraduate students: interest, effort, and perceived usefulness and importance of the test. The results of the path analyses revealed that usefulness and importance of the test were strong predictors of effort (e.g., $R^2 = 0.26$ for mathematics), which in turn was an important predictor of test performance.

Lau et al. (2009) tried to vary test-taking effort in a low-stakes assessment by changing the behavior of the test proctors (invigilators). The proctors were trained to point out the importance and usefulness of the test to the students and to encourage them to work hard. The proctors were also asked to create a productive working environment. The research team investigated the students' effort in testing sessions before (traditional sessions) and after implementation of the proctor-strategies (strategic sessions). Student effort was higher and less variable in the strategic sessions than in the traditional sessions (effect sizes between $d = 0.35$ and $d = 0.57$). The effect of increased effort on performance could not be analyzed because the tests before and after the implementation were slightly different in content, making performance on them noncomparable.

Other studies that have found a strong relationship between test-taking motivation and performance include those by Thelk et al. (2009) and Wise and DeMars (2005). The latter two authors showed from their synthesis of 12 empirical studies that motivated students outperformed their unmotivated classmates by more than one-half of a standard deviation. However, Wise and DeMars cautioned that the relationship between test performance and test-taking motivation could have been distorted by academic ability as a mediator variable.

Studies showing no associations

One of the studies that found no relationship between test-taking motivation, performance, and test-stakes is that by O'Neil et al. (2005). They analyzed the effect of financial incentives on test-taking motivation and performance in mathematics, and divided their sample of test-takers into two groups. The Group 1 students were told they would receive a financial incentive of \$10 per item correct. Also, in order to increase the credibility of the study, test-takers immediately received \$20 if they got two simple items at the beginning of the test correct. Group 2 received no incentives for their participation. Group 1 reported significantly higher levels of test-taking effort and self-efficacy than Group 2 did. However, despite the high reward and the higher level of reported effort for the incentive group, there was no significant difference in performance between the treatment and the control group. The authors assumed that this outcome was due to the lack of correlation between effort and performance for the whole sample.

Similar results were found in a PISA 2000 pilot study in Germany (Baumert and Demmrich 2001). The study examined whether increasing the test's stakes led to a higher level of test-taking motivation and a higher level of performance. Using an experimental design, the researchers manipulated the test conditions across four different groups of test-takers. The incentive for Group 1 was informational feedback, for Group 2 it was grades, and for Group 3 a financial reward. The fourth group was positioned as a reference group. Its members received the usual instructions accompanying PISA assessments and also had emphasized to them the social importance of tests in international comparative studies. In all groups, the invested effort was high, and the personal value of a successful test and the perceived usefulness of the test were the same. Furthermore, the authors found no treatment effects on test performance. While they considered many situation-specific aspects of motivation in their analyses, no domain-specific aspects of motivation were included.

The importance of investigating school-track-specific differences for tracked school systems

Before describing research on specific differences in motivation across types of school, which is one focus of our study, we consider it useful to explain Germany's tracked school system. After completing elementary school (grade 4 or grade 6, depending on the federal state), German students are assigned to different school tracks, primarily according to their scholastic performance. The academic track is the *Gymnasium*. The intermediate track has several school types, such as the *Realschule*, and the lower track is the *Hauptschule*. Of these school types, the *Gymnasium* (academic track) is the only one that exists in *all* German federal states.

One of the rationales for tracking in Germany is that school lessons can be better optimized according to student requirements if students are in homogeneous learning groups. For instance, because students in homogenous learning groups assumedly require similar learning time, groups with high achievers can cover more learning topics as well as topics with higher cognitive demands (Köller and Baumert 2001, 2012). In short, the supposition is that students attain higher learning outcomes in homogeneous learning groups than in heterogeneous ones.

Significant differences in mean achievement occur across the schools in the three different tracks, while mean achievement in schools of the same track is generally similar (Trautwein et al. 2006). One investigation, for example, showed students in grade ten

in academic-track schools outperforming students in intermediate-track schools and in lower-track schools in a mathematic test even after the researchers had controlled for math achievement in grade 7 at individual and school levels (Köller and Baumert 2001). The differences between the schools in each track were only minor. Köller and Baumert suggested that one reason for the superior performance of the academic-track schools is because of their instruction culture, seen partly as a consequence of the teacher training (Köller and Baumert 2001, 2012). Differentiation in student performance in any one school track or school will still occur, of course, commensurate with socioeconomic, psychosocial, motivational and cognitive variables. However, because achievement covaries with socioeconomic status to a very strong extent, social segregation is an undesirable ancillary effect of tracking. In essence, the different tracks “act” as developmental environments differentially influencing student performance (Baumert et al. 2003).

Reference to one of the studies already discussed in this paper—that by Baumert and Demmrich (2001)—is useful at this point. In addition to looking at the influence of incentives on test-taking motivation, the authors also compared the effort students in the lower-track *Hauptschule* and the academic-track *Gymnasium* put into their work on the particular test. The intended effort turned out to be the same for both school types, but the invested effort was lower for the *Hauptschule* students than for the *Gymnasium* students. The students in the academic-tracked schools reported a more positive emotional state and less task-irrelevant cognitions than the students in the lower-track schools. For the entire sample, self-reported effort and worry were the most powerful predictors of test performance. However, there was no investigation of the interplay between the motivational variables and their effects on performance for the different school tracks conducted.

Research on differences in domain-specific and trait-like motivational constructs across the school types has shown mixed results. Two investigations provide useful examples. Artelt et al. (2001) found no differences across school tracks in students’ mathematics self-concept or interest in the subject. The absence of self-concept differences suggests the “big fish little pond” effect may have been at play here (Marsh 1987; Trautwein et al. 2006). According to this effect, students construct their self-concept by comparing themselves with their schoolmates; not by comparing themselves with all students of their age. Thus, students with a similar level of performance will report lower self-concepts if they are in a high-achieving environment (such as the academic track) than in a low-achieving environment. Consequently, despite students in academic-track schools knowing that their performance is higher than the performance of students in lower-track schools, they do not show a corresponding higher self-concept (Artelt et al. 2001).

In contrast, Baumert et al. (2006) found specific differences in students’ self-efficacy beliefs across the school tracks. The authors used national data from the German extension sample of PISA 2000 to explore the influence of school structure on the emergence of differentiated learning environments. They also found evidence of the big fish little pond effect in that the self-efficacy beliefs of students with a similar level of achievement decreased as the track level of the school increased. Baumert and colleagues also conjectured that the larger proportion of class repeaters in the lower than higher tracks might lead to lower self-efficacy beliefs among the students in the lower-track schools. Although this effect did not reach significance, the results nevertheless suggest that the concentration of underachievers in lower tracks can affect students’ effort.

Study objectives

The current state of research indicates that there is a relationship between test performance and test-taking motivation in low-stakes assessments. However, consideration of situation-specific and domain-specific aspects of motivation is lacking in most of the aforementioned studies. Moreover, the lack of research on school-track-specific differences in test-taking motivation and the mixed results of the cited studies on these differences points to the need for more investigation of test-taking motivation across school tracks. We therefore examined the relationship between different motivational aspects and students' performance in general and across school tracks in particular. Our initial research questions were the following:

- 1a) To what extent do domain-specific and situation-specific aspects of motivation predict students' performance in a low-stakes mathematics test?
- 1b) Are there school-track-specific differences in the relationship between performance in mathematics and domain-specific and situation-specific aspects of motivation?

In many studies, test-taking motivation is mainly operationalized through questions about students' invested effort, which covers the main element of the test-taking motivation definition. Accordingly, in a second step, we examined whether invested effort was influenced by other motivational aspects and again considered different school tracks in our research questions:

- 2a) To what extent do situation-specific aspects of motivation predict the invested effort of test-takers in a low-stakes test?
- 2b) Are there school-track-specific differences in the proportion of invested effort?

In summary, our research questions addressed two separate matters. The first focused on the relationship between performance and domain-specific as well as situation-specific aspects of motivation. The second focused on invested effort and its relationship with situation-specific aspects of motivation. Both sets of research questions also required us to consider differences in student performance across Germany's school tracks.

Method

Participants

We used the German extension sample of the PISA 2000 study (Deutsches PISA-Konsortium 2003; Kunter et al. 2002; OECD and UNESCO Institute for Statistics 2003) to investigate test-taking motivation and its relationship with students' performance in mathematics. The sample, nationally representative of German ninth-graders, consisted of 31,740 students. Half of the sample (50%) were female, and the average age of the students was 15.7 years ($SD = 0.56$). Thirty percent of the students were attending academic-track schools and 70% nonacademic-track schools. Eighty-eight percent of the students reported that German was their first language. The random sampling of schools was conducted by the IEA DPC (IEA Data Processing and Research Center), which is responsible for collecting PISA data in Germany.

Procedure

The PISA test took place in the spring of 2000. On the first day of testing, German students took the international standard assessment. On the following day, they took the national PISA extension assessment. The motivational questions used in our study were administered on this second day of testing. Students spent approximately three hours in total on the international and national administrations (two hours of performance tests and 30 minutes of student questionnaires accompanied by questions on cross-curricular competencies).

Measurement instruments

Motivation subscale

Although in PISA 2000, questions on test-taking motivation (situation-specific aspects of motivation) were administered before and after the test, we did not analyze test-taking motivation until *after* the test because the motivation scales varied slightly between the two measurements. For example, task-irrelevant cognition, one of the important predictor variables in the study by Baumert and Demmrich (2001), was not measured until the end of the test.

The post-test subscales assessed various aspects of test-taking motivation: emotional state, invested effort, test attractiveness, and usefulness of the test. The subscales were based on the items in the Online Motivation Questionnaire (Boekaerts and Otten 1993). Items assessing task-irrelevant cognitions, namely worry and distraction, were also administered. These questions were derived from the Test Anxiety Inventory (Hodapp et al. 1982). All self-reported items were measured on a four-point Likert scale, with ratings ranging from 1 = strongly agree to 4 = strongly disagree. Negatively worded items within a positive scale were recoded. Table 1 provides examples of the items and also the subscales' internal consistencies.

Although the subscales contained only a few items, the internal consistencies of the situation-specific subscales of motivation were all acceptable. The invested effort subscale assessed students' test-taking motivation defined according to Baumert and Demmrich's (2001) definition—willingness to engage on test items. We assigned the other

Table 1 Means, standard deviations, and internal consistencies of the test-taking motivation subscales

Aspect of motivation	Subscale	Item no.	Example item	M	SD	Internal consistency
Domain-specific	Self-concept in mathematics	3	I've always been good at math.	2.48	0.94	0.89
Situation-specific	Invested effort	3	How much effort you've given during the test?	2.89	0.58	0.83
	Test attractiveness	3	How much fun did you have during the test?	2.56	0.62	0.82
	Usefulness of the test	1	How useful do you find these kinds of tests?	2.84	0.83	-
	Emotional state	4	I'm in a good mood.	2.87	0.67	0.81
	Worry	3	I have doubted my abilities.	2.06	0.69	0.73
	Distraction	2	My thoughts wandered from the tasks.	2.06	0.83	0.75

subscales to the test-related facets (test attractiveness, usefulness of the test) and to the person-related facets (emotional state, worry, distraction) of test-taking motivation.

Student background questionnaire

Students completed this instrument with its self-report scales after they had taken the test. Among other constructs, this questionnaire, Marsh's (1990) Self Description Questionnaire, assessed students' self-concept in mathematics as a domain-specific aspect of motivation. Responses were measured on a four-point Likert scale, with ratings ranging from 1 = strongly disagree to 4 = strongly agree. Internal consistency was good (see Table 1). Studies by Brunner et al. (2009) and Chen et al. (2013) show that it is possible to distinguish both general and domain-specific dimensions of students' academic self-concept. We were mainly interested in our study in identifying any relationships among motivational constructs and test performance in mathematics, which is why we used only mathematical self-concept as the domain-specific component of academic self-concept.

Achievement test

The achievement test assessed reading, mathematical, and scientific literacy. In the present study, we drew on data from the national PISA test in mathematical literacy. The results were reported on an international scale with a mean of 500 and a standard deviation of 100. The PISA test is considered a low-stakes test because test-takers do not receive information about their performance and their results do not count towards their grades.

Analyses

In order to answer our research questions, we used Mplus 6 software (Muthén and Muthén 1998-2010) to conduct multiple regression analyses with five plausible values (PVs). PVs are ability estimates, which we derived from an item response theory analysis (Yen and Fitzpatrick 2006) conducted via ConQuest software (Wu et al. 2007). Due to the structure of the student sample and the fact that students belonged to different classes, we used a clustering method to correct the standard errors. We also weighted the students for the population size. In order to gain a better interpretation of the results, we reported the unstandardized regression coefficient b , which reflects points on the international PISA achievement scale. In addition, because of the large sample size, we focused only on highly significant effects with a p -value below 0.001.

Results

Before presenting the findings pertaining to our research questions, we provide information about the students' test performance and their scores on the subscales of domain-specific and situation-specific motivational aspects. The weighted mean of the mathematics scores was 500.75 ($SD = 79.50$). The standard deviation differed from the international metric because we computed the performance of the ninth-graders in the PISA German sample instead of all 15-year-olds in it.

As we anticipated, the academic-track students ($M_{at} = 573.83$; $SD_{at} = 59.00$) outperformed the nonacademic-track students ($M_{nt} = 470.30$; $SD_{nt} = 65.95$) on the mathematics test. As evident in Table 1, the students invested effort and concentrated on the items, enjoyed taking the test, and found the test useful. Accordingly, the students

reported a positive emotional state, little worry, and little distraction. The self-concept scores fell within a medium range, as did test attractiveness. The correlations between the several subscales ranged from $r = 0.00$ between worry and usefulness of the test to $r = 0.59$ between emotional state and test attractiveness. In summary, the pattern of students' ratings indicated that they were motivated to do well on the PISA 2000 test.

Prediction of test performance with domain-specific and situation-specific aspects of motivation

To answer our first research question (1a), the extent to which domain-specific and situation-specific aspects of motivation explained performance in the low-stakes PISA test, we examined the relationship between domain-specific and situation-specific aspects of motivation and test performance in mathematics. To accomplish this, we conducted a multiple regression analysis with mathematics performance as the criterion. The predictors were self-concept as the *domain-specific* aspect of motivation and the diverse test-taking motivation subscales as the *situation-specific* aspects of motivation. The procedure we used here followed the approach proposed by Eklöf (2008).

We added the predictors to the regression model in the following manner: first, self-concept as the domain-specific aspect of motivation; second, effort as the main element of test-taking motivation. We then added the test-related facets and the person-related facets, respectively. In a second step, we were interested in school-track-specific differences. Here, we conducted the regression analysis separately for students of two tracks: the academic-tracked schools (*Gymnasium*) and the nonacademic-tracked schools. Our decision to compare just two school tracks was because, as mentioned earlier, the *Gymnasium* is the only type of school that exists across all federal states.

Table 2 shows the results of the multiple regressions. In Model 1, self-concept in mathematics explained approximately 8% of the variance in mathematics scores. The regression coefficient ($b = 24.37$) was significant and indicated that an increase of 1 on the self-concept scale entailed an increase of approximately 24 score points on the PISA mathematics achievement scale. We then added invested effort as the first situation-specific aspect of motivation (Model 2). The variance explained increased slightly to 11%, and the effect of invested effort was significant. The third model included the test-related facets of test attractiveness and test usefulness. Usefulness had a significant but small coefficient, and the variance remained stable. The last model

Table 2 Multiple regression of mathematics performance on domain-specific and situation-specific aspects of motivation

	Model 1		Model 2		Model 3		Model 4	
	<i>b</i>	(<i>SE</i>)						
Self-concept	24.37*	(0.79)	22.99*	(0.78)	22.85*	(0.79)	19.45*	(0.80)
Invested effort			19.64*	(1.33)	15.76*	(1.51)	11.53*	(1.56)
Test attractiveness					1.83	(1.47)	-5.39	(1.55)
Usefulness of the test					4.78*	(0.97)	5.93*	(0.96)
Emotional state							1.04	(1.26)
Worry							-19.65*	(1.12)
Distraction							-9.98*	(0.89)
R^2 (<i>SE</i>)	0.08*	(0.01)	0.11*	(0.01)	0.11*	(0.01)	0.15*	(0.01)

Note: * $p < .001$.

contained all aspects of motivation. The overall explained variance was 15%, of which the domain-specific aspect of motivation, represented by students' self-concept in mathematics, contributed 8%: thus, the higher the students' self-concept in mathematics, the higher their performance in mathematics.

Overall, all subscales other than test attractiveness and emotional state significantly predicted test performance. The most important situation-specific aspects of motivation were (in order of size) worry, invested effort, distraction, and perceived usefulness of the test. These findings indicate that as the students' performance in mathematics improved, the (a) less worried they were, (b) more effort they invested, (c) less distracted they were, and (d) more useful they perceived the test to be. Thus, the situation-specific aspects of motivation—worry and distraction as well as invested effort—showed a relationship with performance, as did the domain-specific aspect of motivation, despite the small amount of variance that it explained.

Our second research question (1b) within this focus referred to the differences in the relationship between explained performance in mathematics and domain-specific and situation-specific aspects of motivation across school tracks. For a clearer presentation of the results, we chose only two models (shown in Table 3): the model with self-concept in mathematics as the domain-specific aspects of motivation (Model 1), and the complete model with the domain- and situation-specific aspects of motivation (Model 4).

With the first model, we established differences between students in academic-tracked schools and students in nonacademic-tracked schools. Self-concept in mathematics explained 24% of the variance in the mathematics scores of the students in the first group of schools, but only 10% of the variance in the mathematics scores of the students in the second group.

In the complete model, the model to which we added the situation-specific aspects of motivation, the explained variance increased marginally, by 5%, for the students attending academic-track schools. Not only self-concept in mathematics but also worry and invested effort became relevant at this juncture, meaning that (a) the higher the self-concept of these students, (b) the less worried they were, and (c) the more they invested effort in the test, the better their performance on it.

The explained variance also increased marginally, again by 5%, in the complete model for the students in the nonacademic track. Here again, in addition to self-

Table 3 Multiple regression of mathematics performance on domain-specific and situation-specific aspects of motivation, separated by school track

	Model 1				Model 4			
	Nonacademic track		Academic track		Nonacademic track		Academic track	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Self-concept	22.39*	(0.86)	29.64*	(0.79)	19.06*	(0.88)	25.83*	(0.85)
Invested effort					10.11*	(1.65)	9.69*	(2.01)
Test attractiveness					-2.50	(1.65)	3.96	(1.95)
Usefulness of the test					1.10	(1.06)	-2.15	(1.19)
Emotional state					-2.32	(1.31)	3.50	(1.68)
Worry					-16.64*	(1.16)	-12.78*	(1.31)
Distraction					-4.59*	(0.96)	-0.06	(1.20)
<i>R</i> ² (SE)	0.10*	(0.01)	0.24*	(0.01)	0.15*	(0.01)	0.29*	(0.01)

Note: * *p* < .001.

concept, the situation-specific aspects of motivation (i.e., worry and invested effort) were significantly associated with performance. In contrast to the findings for the students in the academic-track, distraction also significantly predicted performance. Thus, for the students in the nonacademic-track schools, the higher their (a) self-concept and (b) invested effort, and the less their (c) worry and (d) distraction, the better they performed.

Our next step was to run a new regression for the complete model to determine if any of the interactions between motivational variables and school track were significant. Four of the seven interactions became statistically significant (ordered by size of the coefficients): test attractiveness ($b = 7.35$), self-concept ($b = 5.95$), emotional state ($b = 5.76$), and distraction ($b = 4.17$). We were surprised to find the interaction of emotional state and test attractiveness reaching significance given that the main effects of these two variables on test performance were not significant. We accordingly decided not to overemphasize these interactions given that these subscales did not seem to predict mathematics scores in either school track.

In summary, the interactions supported our results: self-concept and distraction demonstrated school-track-specific differences and were also significant predictors of test performance. The test-taking motivation scales explained the same amount of variance in performance in both school tracks.

Prediction of invested effort with situation-specific aspects of motivation

In order to answer the research question focusing on test-takers' invested effort in a low-stakes test by situation-specific aspects of motivation (2a), we used invested effort as criterion and the test-related (Model 1) and person-related facets of test-taking motivation (Model 2) as predictors. Because we were interested in the effects of situational aspects on effort, we did not include the domain-specific aspect of motivation. In a second step, we conducted a regression analysis, using the same approach as for the first set of research questions—that is, separately for the academic-track students and the nonacademic-track students.

Table 4 illustrates the results. In the first model, both of the test-related facets had a significant effect on invested effort whereby the coefficient of test attractiveness was bigger than the coefficient of usefulness of the test. Both explained 35% of the variance. In the complete model (i.e., containing the person-related facets), all subscales significantly predicted the invested effort and explained 40% of the variance in that effort. The test-related facets of test-taking motivation (test attractiveness, test usefulness) and

Table 4 Multiple regression of invested effort on situation-specific aspects of motivation

	Model 1		Model 2	
	<i>b</i>	(SE)	<i>b</i>	(SE)
Test attractiveness	0.45*	(0.01)	0.39*	(0.01)
Usefulness of the test	0.13*	(0.01)	0.10*	(0.01)
Emotional state			0.04*	(0.01)
Worry			0.06*	(0.01)
Distraction			-0.17*	(0.01)
R^2 (SE)	0.35*	(0.01)	0.40*	(0.01)

* $p < .001$.

Note: * $p < .001$.

distraction had the strongest numerical values of the regression coefficients. The meaning that can be taken from this pattern is that the more (a) attractive and (b) useful the students perceived the test to be and (c) the less distracted they were, the higher their level of invested effort.

The second question (2b) of this research focus referred to school-track-specific differences in the relationship between invested effort and situation-specific aspects of motivation. Table 5 contains the results. In Model 1, test attractiveness and test usefulness significantly predicted the invested effort for both school tracks. However, the coefficient of test attractiveness for nonacademic-track students was higher than for academic-track students. Correspondingly, the two test-related facets of test-taking motivation explained approximately 23% of the variance in the effort invested by the academic-track students, and 39% of the variance in effort invested by the nonacademic-track students.

Model 2, the model to which we added the person-related facets to the test-related facets, explained 31% of the variance in the effort the academic-track students invested in the mathematics assessment: test attractiveness, distraction, and usefulness of the test all showed significant coefficients. The pattern, then, was that (a) the more attractive and (b) useful the academic-track students perceived the test to be, and (c) the less distracted they were, the more effort they put into it. For nonacademic-track students, the complete model explained 43% of the variance in invested effort; all subscales significantly predicted that effort.

When we looked at the coefficients (those exceeding ± 10.0), we found that the pattern for the nonacademic-track students in Model 2 was similar to the pattern for the academic-track students. In order to assess whether the differences between the school tracks were statistically significant, we conducted a regression with interaction effects for the complete model. Again, as anticipated, the interaction between test attractiveness and school track showed a significant coefficient ($b = -0.15$) as did the interaction between worry and school track ($b = -0.05$). However, we do acknowledge that the latter coefficient is relatively small. In summary, the results relating to our second set of research questions suggests that the attractiveness of the test differed according to whether the students were from the academic-track schools or from the nonacademic-track schools.

Table 5 Multiple regression of invested effort on situation-specific aspects of motivation, separated by school track

	Model 1				Model 2			
	Nonacademic track		Academic track		Nonacademic track		Academic track	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Test attractiveness	0.49*	(0.01)	0.35*	(0.01)	0.43*	(0.01)	0.28*	(0.02)
Usefulness of the test	0.13*	(0.01)	0.10*	(0.01)	0.10*	(0.01)	0.08*	(0.01)
Emotional state					0.04*	(0.01)	0.03	(0.01)
Worry					0.06*	(0.01)	0.02	(0.01)
Distraction					-0.16*	(0.01)	-0.19*	(0.01)
<i>R</i> ² (SE)	0.39*	(0.01)	0.23*	(0.01)	0.43*	(0.01)	0.31*	(0.01)

* $p < .001$.

Note: * $p < .001$.

Discussion

This study examined two sets of research questions focused on the relationship of various aspects of test-taking motivation, performance, and effort as well as school-track-specific differences within this relationship.

Prediction of test performance with domain-specific and situation-specific aspects of motivation

The first set of research questions examined the relationship between domain-specific and situation-specific aspects of motivation and test performance in mathematics. The results showed that nearly all situation-specific aspects of motivation predicted mathematics scores even after we had controlled for the domain-specific aspect of motivation. Self-concept as the domain-specific aspect of motivation explained slightly more variance than the situation-specific aspects of motivation. Along with self-concept, invested effort as well as worry and distraction as person-related facets of test-taking motivation had the greatest impact on the mathematics test scores (Research Question 1a).

These results do not support Eklöf (2008) findings. In her study, test-taking motivation showed no significant effect on test performance when considered with domain-specific aspects of motivation. In order to explain these differences, we note that Eklöf (2008) examined a relatively small sample ($N = 343$) of Swedish eighth-graders, whereas we used a nationally representative sample of German ninth-graders. As mentioned in the theoretical section of this paper, Eklöf assumed that students probably did not perceive the test as low-stakes because they had not yet experienced receiving grades or taking external tests. Hence, it is likely that test-taking motivation varies across countries due to cultural differences or different response behaviors. Thus, cross-country comparisons of test-taking motivation on low-stakes tests constitute an important area of further research.

Our results support the findings of Baumert and Demmrich (2001). In their study, as in ours, effort and worry were the most powerful predictors of test performance. However, these authors were able to explain nearly twice as much variance in performance on the basis of the two situation-specific aspects of motivation than we could with *all* of our domain- and situation-specific aspects of motivation. Unfortunately, they did not explicitly describe their analyses, which is why we were not able to compare these differences more concretely. Here, further research is necessary.

According to the school-track-specific differences, the results indicated that these differences are primarily due to the domain-specific aspect of motivation—students' self-concept of their mathematics ability. For academic-track students, self-concept had a stronger relationship with mathematics performance than it did for the students from the nonacademic track (Research Question 1b). These results correspond with the big fish little pond effect (Marsh 1987; Trautwein et al. 2006). Trautwein and colleagues concluded on the basis of their study that students construct their self-concept by comparing themselves with their schoolmates and not by comparing themselves with all students of their age. With respect to our study, this effect implies that even though students in the high-achieving environment (the *Gymnasium*) knew their achievement was higher on average than that of students in the lower-achieving environments, their self-concept was, on average, not higher than that of their lower-tracked peers.

Looked at another way, this pattern could mean that for the academic-track students, high self-concept actually corresponds with good performance (therefore the higher R^2), whereas for the students from the nonacademic track high self-concept does not necessarily lead to good performance (therefore the lower R^2). This hypothesis is supported by the correlation between self-concept and mathematics performance, which was higher for the academic-track students ($r = 0.48$) than for their counterparts from the nonacademic track ($r = 0.32$). When we used the Fisher's z test, we found this difference was highly significant ($z = -14.75, p < .001$).

The other school-track-specific difference we examined concerned the person-related facets of test-taking motivation. For nonacademic-track students, our findings suggest that it is more important that they do not doubt their abilities when taking tests and that they are focused on the tasks. With respect to the task-irrelevant cognitions, our findings correspond with the results of Baumert and Demmrich (2001), who found that academic-track students had a more positive emotional state and less task-irrelevant cognitions than students attending lower-track schools. Our results furthermore show that worry and distraction had a greater negative effect on performance for nonacademic-track students than for academic-track students. Thus, it is especially important that nonacademic-track students undergo testing in a distraction-free environment, with steps having been made to mitigate anxieties so that they are motivated to do their best. It may also be beneficial for further investigations to include questions assessing anxiety in their test-taking motivation scale (see, in this regard, Nie et al. 2011; Putwain and Daniels 2010).

Prediction of invested effort with situation-specific aspects of motivation

In regard to the second set of research questions, we found that the test- and person-related facets of test-taking motivation predicted invested effort, with test attractiveness emerging as the most powerful predictor. Distraction and usefulness of the test showed a smaller relationship with invested effort (Research Question 2a), a finding that aligns with work by Cole et al. (2008). They found that perceived usefulness and importance of the test were strong predictors of effort. For the nonacademic-track students, test attractiveness was more relevant than for the academic-track students (Research Question 2b); a positive image of low-stakes assessments and a calm working atmosphere appear to have been essential aspects of a favorable test environment for this first group of students. This finding can be regarded as "good news" because it suggests that test-related facets of test-taking motivation can be positively influenced by making low-stakes tests interesting and appealing. Even if the test has no consequences for the test-takers, it is nonetheless important that they find it an enjoyable experience.

Our study furthermore found that performance in low-stakes tests was slightly influenced by different motivational aspects of test-taking motivation. These results imply that students are likely to achieve higher test performance the more effort they invest and the less worry they experience during the testing session. These small effects support the general validity of this low-stakes assessment in Germany. Thus, educational policy decision-making processes based on the results of low-stakes assessments can be supported for this sample. However, we do not know whether the small effects depend on the country in which it is administered, or on the particular

test. It is thus important to take into account motivational measures, such as students' invested effort, when endeavoring to draw valid conclusions about students' performance. The school-track-specific differences in self-concept in mathematics and in invested effort that we found imply that for students attending nonacademic-tracked schools especially, an attractive and enjoyable test is crucial to motivate them to do their best. This consideration should be kept in mind by researchers when constructing low-stakes tests items.

Limitations and conclusion

A limitation of the present study is the number of items per subscale. For example, the test usefulness subscale had just one item, while the distraction subscale contained only two. The internal consistencies of these two subscales could be improved by adding further items to them. Due to the restricted testing time and the large number of questions in the student questionnaire, more items could not be implemented. However, good and substantial reliabilities confirmed the homogeneity of these scales.

Another limitation concerns the fact that the students completed the full motivational questionnaire after they took the test. Thus, it is possible that their responses to the self-report questionnaire were confounded by their perceived test performance. According to attribution theory, it seems likely that students reported lower invested effort to justify their lower perceived test performance (Weiner 1986). Whether or not the reported level of test-taking motivation corresponded with the actual test-taking motivation during the test is therefore uncertain. We intend to undertake further research to explore reported test-taking motivation before a test and its relationship with test performance. We also intend to compare reported test-taking motivation before a test with the test-taking motivation after it using the same motivational subscales.

In general, further investigations similar to the experimental study of Baumert and Demmrich (2001) are necessary. They found no effect of raising the stakes on effort and performance. However, their study was conducted before the first PISA survey, which was administered in 2000. Over the intervening years, the frequency of international and national tests in German schools has greatly increased; today, students take more external tests than they did at the beginning of this century. Thus, after more than a decade of intense testing, it seems likely that test-taking motivation in low-stakes assessments has developed an influence on effort and performance. Just how motivated students remain throughout the testing session is another area of particular interest. An analysis of this kind would rely on more than just two measurements (i.e., before and after the test). Once such data are to hand, the course of students' test-taking motivation during testing sessions can be more robustly examined.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CPe analyzed the data and wrote the manuscript. CPo supported in drafting the manuscript. ARO supported the performance of the statistical analysis. All authors read and approved the final manuscript.

Received: 2 September 2013 Accepted: 19 June 2014

Published online: 19 August 2014

References

- Artelt, C, Demmrich, A, & Baumert, J. (2001). Selbstreguliertes Lernen: Motivation und Strategien in den Ländern der Bundesrepublik Deutschland [Self-regulated learning: Motivation and strategies in the German federal states]. In Deutsches PISA-Konsortium (Ed.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 271–298). Opladen: Leske + Budrich.
- Barry, CL, & Finney, SJ. (2009). *Exploring change in test-taking motivation*. Rocky Hill, CT: Northeastern Educational Research Association.
- Baumert, J, & Demmrich, A. (2001). Test motivation in the assessment of student skills: the effects of incentives on motivation and performance. *European Journal of Psychology of Education*, *16*(3), 441–462.
- Baumert, J, Trautwein, U, & Artelt, C. (2003). Schulumwelten: institutionelle Bedingungen des Lehrens und Lernens [School environments: Institutional conditions of teaching and learning]. In Deutsches PISA-Konsortium (Ed.), *PISA 2000: ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 261–331). Opladen: Leske + Budrich.
- Baumert, J, Stanat, P, & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [School structure and the emergence of differential learning and developing milieus]. In J Baumert, P Stanat, & R Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften/GWW Fachverlage GmbH, Wiesbaden.
- Boekaerts, M, & Otten, R. (1993). Handlungskontrolle und Lerneranstrengung im Schulunterricht [Action control and learning-related effort in the classroom]. *Zeitschrift für Pädagogische Psychologie*, *7*(2/3), 109–116.
- Brunner, M, Keller, U, Hornung, C, Reichert, M, & Martin, R. (2009). The cross-cultural generalizability of a new structural model of academic self-concepts. *Learning Individual Differences*, *19*(4), 387–403. doi:10.1016/j.lindif.2008.11.008.
- Chen, S-K, Yeh, Y-C, Hwang, F-M, & Lin, S-S. (2013). The relationship between academic self-concept and achievement: a multicohort–multioccasion study. *Learning Individual Differences*, *23*, 172–178. doi:10.1016/j.lindif.2012.07.021.
- Cole, JS, Bergin, DA, & Whittaker, TA. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, *33*(4), 609–624.
- Deutsches PISA-Konsortium. (2003). *PISA 2000: Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland [PISA 2000: A differentiated view of the German federal states]*. Opladen: Leske + Budrich.
- Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing*, *7*(3), 311–326.
- Eklöf, H. (2008). Test-taking motivation on low-stakes tests: A Swedish TIMSS 2003 example. In *Issues and methodologies in large-scale assessments: IERI monograph series* (1st ed., pp. 9–21). Hamburg: IEA-ETS Research Institute.
- Eklöf, H, & Nyroos, M. (2013). Pupil perceptions of national tests in science: perceived importance, invested effort, and test anxiety. *European Journal of Psychology of Education*, *28*(2), 497–510. doi:10.1007/s10212-012-0125-6.
- Hodapp, V, Laux, L, & Spielberger, CD. (1982). Theorie und Messung der emotionalen und kognitiven Komponente der Prüfungsangst [Theory and measurement of emotional and cognitive component of test anxiety]. *Zeitschrift für Pädagogische Psychologie*, *3*(3), 169–184.
- Köller, O, & Baumert, J. (2001). Leistungsgruppierungen in der Sekundarstufe I [Performance grouping in secondary education]. *Zeitschrift für Pädagogische Psychologie*, *15*(2), 99–110. doi:10.1024/1010-0652.15.2.99.
- Köller, O, & Baumert, J. (2012). Schulische Leistung und ihre Messung [School achievement and its measurement]. In W Schneider & U Lindenberger (Eds.), *Entwicklungspsychologie* (Vol. 7, pp. 645–661). Weinheim: Beltz/PVU.
- Kunter, M, Schümer, G, Artelt, C, Baumert, J, Klieme, E, Neubrand, M, Prenzel, M, Schiefele, U, Schneider, W, Stanat, P, Tillmann, K-J, & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente (Bd. 72) [PISA 2000: Documentation of the survey instruments]*. Berlin: Max-Planck-Inst. für Bildungsforschung.
- Lau, AR, Swerdzewski, PJ, Jones, AT, Anderson, RD, & Markle, RE. (2009). Proctors matter: strategies for increasing examinee effort on general education program assessments. *The Journal of General Education*, *58*(3), 196–217. doi:10.1353/jge.0.0045.
- Marsh, HW. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Education & Psychology*, *79*(3), 280–295. doi:10.1037/0022-0663.79.3.280.
- Marsh, HW. (1990). *Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. San Antonio, TX: The Psychological Corporation.
- Muthén, LK, & Muthén, BO. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nie, Y, Lau, S, & Liao, AK. (2011). Role of academic self-efficacy in moderating the relation between task importance and test anxiety. *Learning and Individual Differences*, *21*(6), 736–741. doi:10.1016/j.lindif.2011.09.005.
- O'Neil, HF, Sugrue, B, & Baker, EL. (1995). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, *3*(2), 135–157.
- O'Neil, HF, Abedi, J, Miyoshi, J, & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment*, *10*(3), 185–208. doi:10.1207/s15326977ea1003_3.
- OECD & UNESCO Institute for Statistics. (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publishing.
- Putwain, DW, & Daniels, RA. (2010). Is the relationship between competence beliefs and test anxiety influenced by goal orientation? *Learning and Individual Differences*, *20*(1), 8–13. doi:10.1016/j.lindif.2009.10.006.
- Schunk, DH, Pintrich, PR, & Meece, JL. (2008). *Motivation in education: Theory, research, and applications* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Thelk, AD, Sundre, DL, Horst, SJ, & Finney, SJ. (2009). Motivation matters: using the student opinion scale to make valid inferences about student performance. *The Journal of General Education*, *58*(3), 129–151. doi:10.1353/jge.0.0047.
- Trautwein, U, Lüdtke, O, Marsh, HW, Köller, O, & Baumert, J. (2006). Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Education & Psychology*, *98*(4), 788–806. doi:10.1037/0022-0663.98.4.788.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.
- Wise, SL, & DeMars, CE. (2005). Low examinee effort in low-stakes assessment: problems and potential solutions. *Educational Assessment*, *10*(1), 1–17. doi:10.1207/s15326977ea1001_1.

- Wolf, LF, & Smith, JK. (1995). The consequence of consequence: motivation, anxiety, and test performance. *Applied Measurement Education*, 8(3), 227–242. doi:10.1207/s15324818ame0803_3.
- Wu, ML, Adams, RJ, Wilson, MR, & Haldane, SA. (2007). *ACER ConQuest Version 2.0: Generalised item response modelling software*. Camberwell, VIC: ACER Press.
- Yen, WM, & Fitzpatrick, AR. (2006). Item response theory. In RL Brennan (Ed.), *Educational measurement* (4th ed., pp. 111–153). Westport, CT: Praeger Publishers.

doi:10.1186/s40536-014-0005-4

Cite this article as: Penk et al.: The role of test-taking motivation for students' performance in low-stakes assessments: an investigation of school-track-specific differences. *Large-scale Assessments in Education* 2014 2:5.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
