

RESEARCH

Open Access



The correlation between vocational school students' test motivation and the performance in a standardized test of economic knowledge: using direct and indirect indicators of test motivation

Roland Happ^{1*} and Manuel Förster²

*Correspondence:

roland.happ@uni-mainz.de

¹ Department of Law,
Business & Economics, Chair
of Business and Economics
Education, Johannes
Gutenberg University
Mainz, Jakob-Welder-Weg 9,
55128 Mainz, Germany
Full list of author information
is available at the end of the
article

Abstract

Background: In this study, the correlation between test motivation and performance on an economic knowledge test was investigated. To this end, the economic knowledge of 1018 students in vocational schools in Germany was assessed using a standardized test, and their self-reported test motivation and interest in receiving feedback on their performance on the test functioned as indicators of test motivation. The expectancy-value model served as the theoretical basis, while this paper followed Knekta and Eklöf (*J Psychoeduc Assess* 33(7):662–673, 2015) and focused on invested effort in particular. Further, the number of missing values on the test was examined as a potential external criterion for test motivation. The correlation between gender and test motivation being a subject of frequent discussion in the literature, gender was incorporated as a control variable into the modeling.

Methods and results: Three structural equation models of the results show that self-reported test motivation (direct indicator) and the number of missing values on the test (external criterion) correlated significantly with the economic knowledge test score achieved. Interest in receiving feedback (indirect indicator) had no significant correlation with the economic knowledge test score. However, there is a positive correlation between interest in receiving feedback and self-reported test motivation. The analyses taking into account gender show that there is no correlation between gender and interest in receiving feedback or gender and self-reported test motivation. There are, however, correlations between gender and the economic knowledge test score as well as between gender and the number of missing values.

Conclusions: The findings underline the importance of a differentiated view on the assessment of correlations between test motivation and test performance. Thereby, dividing indicators of test motivation into direct (self-reported test motivation) and indirect (interest in receiving feedback) indicators, as well as taking into account the external criterion (number of missing values), in particular, were seen as the value added by this study.

Keywords: Test motivation, Invested effort, Low-stakes test, Feedback, Missing values, Gender, Test of economic literacy, Vocational school sector

Relevance

Although a test may be designed to measure knowledge, performance on that test may not be an accurate reflection of knowledge of the test construct. Test-takers' motivation to try their best to answer all the questions correctly and thoroughly may influence their scores on the test. This is especially the case with low-stakes tests (see Kane 2006; Thelk et al. 2009; Wise and DeMars 2005), which usually are administered to measure academic achievement, identify learning problems, or inform instructional adjustments, and performance on them has no direct short- or long-term consequence for test-takers (Penk and Richter 2017, p. 55).¹ As interpretation of test scores alone is valid to a limited degree only (Asseburg and Frey 2013, p. 92; Eklöf 2010a; Wise 2009), other important influences on performance on knowledge tests such as test motivation should be considered when interpreting results.

Although numerous studies have been conducted of regular secondary school students' performance on large-scale assessments such as PISA² (Asseburg and Frey 2013) and TIMSS³ (Eklöf 2008), very few have been conducted on vocational school students' performance on such tests (e.g., Baethge et al. 2006; Winther and Achtenhagen 2009). This is not surprising considering the difficulty in producing large-scale assessments for the diverse types of vocational school and their heterogeneous student bodies (see Billet 2011; Jacob and Solga 2015). While some vocational schools accommodate lower-ability students (e.g., within the transition system⁴), others cater to high-performance students (e.g., vocational secondary schools "berufliches Gymnasium") (cf. Beicht and Walden 2016; Protsch and Solga 2015). The correlation between the ability of the test-takers and test motivation has been demonstrated by several studies (cf. Eklöf and Nyroos 2013; Penk and Richter 2017; Wise and DeMars 2005). In doing so, test motivation was conceptualized by means of students' self-reported data on the importance of the test, motivation, invested effort, and test anxiety (Eklöf and Nyroos 2013; Wise and DeMars 2005) and on the expectancy and value components (e.g. perceived probability of success; Penk and Richter 2017), respectively. This data was collected in the course of an achievement test. Moreover, very few studies have been conducted of the influence of vocational school students' test motivation on their test scores.⁵ The main aim of this study is to examine the correlation between vocational school students' test motivation and their performance on an economic knowledge test.

1018 students from vocational secondary schools (berufliches Gymnasium), commercial vocational schools (Berufsausbildung) and upper vocational schools (Berufsoberschule) were surveyed during the second term of the 2015/2016 school year. The purpose of this project was to measure economic knowledge of students in vocational schools. To this end, a test instrument in a paper-and-pencil format was used (Happ et al. 2016),

¹ Short- or long-term consequences for students result from high-stakes tests (Finn 2015).

² Program for International Student Assessment.

³ Trends in International Mathematics and Science Study.

⁴ The transition system ("Übergangssystem") offers a temporary solution to students (mostly with a low school-leaving grade or none at all) who were unable to begin vocational training after finishing school (Solga et al. 2014). Most courses in this transition system are neither standardized nor certified.

⁵ The recent studies of vocational education conducted by Rausch and Kögler (2016) and Kögler and Rausch (2017) may be referred to at this point.

composed of 45 items in a multiple-choice format. For the assessment of test motivation, two indicators were applied in this paper (detailed in “[Theoretical basis and state of research](#)” section). Both direct assessment of test motivation through test-takers’ self-evaluation (1) and indirect assessment operationalized through the test-takers’ interest in receiving feedback on their test performance (2) were performed. Furthermore, the number of missing values on the economic knowledge test was incorporated into the modeling as external criterion that should correlate with test motivation.⁶ In addition to investigating the correlation between direct and indirect indicators of test motivation, another aim of this paper is to examine the correlation between both of the direct and indirect indicators for test motivation and the number of missing values (external criterion) and the scores on the economic knowledge test. Furthermore, the state of research regarding test motivation indicates differences between the two genders (on the theoretical backgrounds see “[Theoretical basis and state of research](#)” section) (cf. Butler and Adams 2007; Kornhauser et al. 2014; O’Neil et al. 2005). For this reason, the test-takers’ gender was also integrated into the analyses of this paper.

In “[Theoretical basis and state of research](#)” section, the construct of test motivation and indicators thereof are examined, with the expectancy-value model serving as a basis and the paper focusing on invested effort. Also, hypotheses were formulated in “[Theoretical basis and state of research](#)” section. In “[Measurement instrument and sample](#)” section, information on the sample and details about the instruments used to collect data are provided. In “[Statistical approach](#)” section, the statistical approach is presented in three structural equation models. Results of latent variable modeling are presented in “[Results](#)” section and discussed in “[Discussion](#)” section. In “[Conclusions and limitations](#)” section, the conclusions and limitations of the study are presented.

Theoretical basis and state of research

Test motivation and invested effort

Baumert and Demmrich (2001, p. 441) define test motivation as “the willingness to engage in working on test items and to invest effort and persistence in this undertaking.” A similar understanding of test motivation can be found in Wise and DeMars (2005, p. 2): “We define test-taking effort as a student’s engagement and expenditure of energy toward the goal of attaining the highest possible score on the test.” Similarly in this study, test motivation is understood as the willingness to invest the greatest possible effort to perform well on a test. This means that the goal of a motivated test-taker is to obtain the highest possible score on a test. The expectancy-value model of achievement motivation by Wigfield and Eccles (2000) has proven to be helpful as a theoretical framework for defining test motivation. In this model, test motivation is a function of expectancy of success and relevance of a test to the test-taker, which together influence effort to complete a test. In a simplified schematic picture, Knekta and Eklöf (2015) demonstrate the application of the expectancy-value theory in a test situation. The expectancies and

⁶ The number of missing values is not a direct indicator of test motivation (see “[Theoretical basis and state of research](#)” section). Rather, it can be assumed that a low test motivation results in the abortion of the item or of the entire questionnaire, thus producing a higher number of missing values. However, a high number of missing values may also be explained by other causes and through other constructs (such as lack of knowledge or slow response behavior). Therefore, the missing values in this paper were incorporated as an external criterion that should correlate with test motivation.

the task value are significant predictors of effort in this case. Effort, in turn, predicts significantly the performance in a test situation. Accordingly, both authors regard effort as an aspect of test motivation which results from expectancies as well as the task value and precedes action.⁷ The above definitions clarify that most test motivation definitions focus on invested effort. This paper focuses on the actual willingness to make an effort as a part of test motivation, not on explaining it by means of expectancy and value components. Consequently, only invested effort is taken into account (also cf. the selection of items in Table 3 in [Appendix](#)).

Test-takers with a high level of test motivation exhibit more effort to answer questions correctly and thoroughly and demonstrate more commitment to completing all test items than test-takers with lower levels of test motivation (Baumert and Demmrich 2001). Consequently, ambiguities arise in the interpretation of test scores: It is unclear whether performance on a test is the result of having the necessary ability and/or the result of (missing) test motivation. Hence, misinterpretation of test scores may occur when influence factors are not considered or misidentified (Wise and DeMars 2005). For example, it is difficult to determine whether test items are ignored due to a low level of test motivation alone or whether other factors are at play and whether responses to test items are mere guesses without any significant effort being made (Knekta 2017).

A covariate repeatedly discussed in research on test motivation is gender (cf. Finn 2015; Eklöf 2007; King 2016). Several studies indicate that women have a higher test motivation than men (cf. DeMars et al. 2013; Eklöf 2007; Marrs and Sigler 2012). Current research approaches explain this with specific characteristics associated with both of the genders, also in test situations. Thus King (2016, p. 62) characterizes women as having a higher level of self-discipline and a lower tendency to avoid work. According to DeMars et al. (2013), men take the line of least resistance and avoid unnecessary work. Men are thus characterized as work-avoidant. When the indicators for the test motivation are examined and further correlations are highlighted in this paper, the findings resulting from gender-specific differences in test motivation demonstrate that the gender should be controlled in these analyses. In addition, the results on standardized economic tests show, that men perform better on them than women (cf. Asarta et al. 2014; Beck et al. 1998; Happ et al. 2016; Walstad et al. 2013). Consequently, this paper can also examine if a part of this effect can be explained with varying test motivation or if the difference may even be larger when taking into account test motivation.

To assess test motivation, various indicators can be analyzed. In this study, test-takers' self-reported motivation and interest shown in receiving feedback on their performance on an economic knowledge test are explored as indicators of test motivation ("[Indicators of test motivation](#)" section). Also the number of missing values on the test is examined as an external criterion.

⁷ In addition to effort, Knekta and Eklöf (2015) refer to the following four additional aspects: expectancies, importance, interest, and test anxiety.

Indicators of test motivation

Self-reported motivation (direct indicator)

Studies of test motivation traditionally have been based on test-takers' self-reports (Liu et al. 2015, p. 80; Swerdzewski et al. 2011). Thereby, test-takers assess their motivation to work on the test instrument using Likert scales in most studies. The assessment of self-reported motivation can take place either prior to, during, or after the performance test (the economic knowledge test in the context of this paper). In the literature no clear suggestions have been made as to when test motivation should be assessed as a control variable: In some studies test motivation has been assessed immediately upon completion of a knowledge test (Knekta 2017; Ortner et al. 2014); in others, an intermediate questionnaire has been completed at various points throughout the test (Stenlund et al. 2018). Repeated assessment of test motivation certainly would make sense from a theoretical point of view because motivation may change during the test. Due to time constraints, however, in most studies test motivation usually is assessed as a control variable at one point in time only. This is also the case in the present paper (see "[Economic knowledge and test motivation](#)" section).⁸

It can be assumed that test-takers with a higher self-reported test motivation have invested, or are investing, more effort into completing the items (Barry and Finney 2016; Knekta 2017). Consequently, there should be a positive correlation between self-reported test motivation and test performance (Knekta and Eklöf 2015). On the basis of these theoretical considerations the first hypothesis of this paper can be formulated as follows:

H1 Self-reported test motivation correlates positively with test performance.

Some scholars question whether test-takers give truthful answers when reporting on their test motivation (Wise and Kong 2005). It is possible that responses are biased and self-knowledge related to test motivation as reported on in self-evaluations and assessed on Likert scales (Ziegler et al. 2011) is inaccurate. Also, giving socially desirable responses to questions about test motivation may be a way for test-takers to avoid potential negative consequences of expressing lack of motivation and consequently may skew interpretation of results (Wise and Kong 2005). Lastly, when a test-taker has low level of test motivation he or she may lack willingness to make an effort to report on test motivation (Finn 2015). Furthermore, if test motivation is self-reported upon completion of a test, it may be influenced by the expected test score: test-takers may, consciously or unconsciously, attribute their expected test scores to a low or high level of test motivation. Due to the unreliability of self-reported data alone, other indicators of test motivation should be investigated.

⁸ Results of studies of change in test motivation are heterogeneous: while some indicate an increase in test motivation (e.g., Barry and Finney 2016), others indicate a decrease over the duration of the test (e.g., Penk and Richter 2017; Sundre and Kitsantas 2004; Wise et al. 2009). This heterogeneity in findings may be because both personal characteristics of the test-takers (e.g. ability) and item characteristics (e.g., difficulty) are responsible for changes in test motivation.

Interest in receiving feedback (indirect indicator)

Interest in receiving feedback on performance on a test may indicate test motivation (Kong et al. 2006; Liu et al. 2015, p. 81). Furthermore, Kong et al. (2006, p. 3) note that if test-takers know before starting a test that they will receive individual feedback on their performance, they are less likely to make guesses or finish the test hastily instead of thinking carefully about how to respond best to test items. In large-scale student assessments such as PISA students have been promised individual feedback (Baumert and Demmrich 2001, p. 448), which may influence their test motivation. Also, there might be more pressure to perform well on large-scale studies such as PISA due to the international visibility of results and ranking of countries. This type of group feedback may boost test motivation. The theoretical state of research regarding interest in receiving feedback indicates that promising feedback may result in a positive effect on invested effort and thus on test performance. Therefore, the indicator “interest in receiving feedback” should be taken into account not only due to the potential biases in self-reported measures. From a theoretical point of view, also, an effect resulting from “interest in receiving feedback” on invested effort can be expected. Hence, this indirect indicator of test motivation is also taken into account in this paper.

Two assumptions about the motivational effects of receiving feedback can be distinguished. One is that obligatory feedback⁹ increases test motivation because the test is perceived by test-takers as more personally meaningful (Huffman et al. 2011; Zilberberg et al. 2013).¹⁰ The second is that when receiving feedback is a choice and test-takers actively seek feedback, their level of test motivation is higher than those who do not. In this paper test-takers’ conscious decision to obtain feedback on their performance on a knowledge test is an indicator of their test motivation. Theoretically, the task value component (Knekta and Eklöf 2015) of test motivation (see “[Test motivation and invested effort](#)” section), in particular, should be more pronounced among these test-takers (see Wigfield and Eccles 2000). With the schematic picture of Knekta and Eklöf (2015) as the basis, the higher importance of the task value, in turn, can be expected to result in a higher invested effort, which indicates a positive correlation with test performance, according to the model. To determine the effects of feedback on test motivation in numerous studies with experimental designs test-takers have been categorized in two groups: one that is offered the option to receive feedback and the other not (cf. Baumert and Demmrich 2001, p. 448; Wise 2004). Categorizing test-takers into a feedback group and a no-feedback group was not an option for the present study. The individual benefit for each test-taker was paramount to school supervising authorities; therefore, all students were given the option to receive feedback even though not all students (see “[Sample and descriptive statistics](#)” section) made use of it. Therefore the feedback in this study is not obligatory. Based on the considerations regarding interest in feedback and test performance, the following second hypothesis of the present paper can be formulated:

⁹ Obligatory feedback means that the test-takers get a feedback regardless whether they ask for it or not.

¹⁰ The state of research is not entirely clear as several studies have not provided evidence that promising feedback to students on their subsequent test performance impacts their test motivation (Wise 2004). However, studies revealing a significant correlation between feedback and test motivation predominate.

H2 Interest in receiving feedback correlates positively with test performance.

The literature indicates that test-takers who demonstrate a higher test motivation are more interested in their test results (cf. Knekta and Eklöf 2015). These test-takers show a more pronounced importance component within the expectancy-value model, which should result in a higher invested effort and a higher test performance. For this reason, both indicators should correlate positively with each other, which results in the following Hypothesis:

H3 Self-reported test motivation correlates positively with the interest in receiving feedback.

The number of missing values as external criterion

An external criterion that should be associated with test motivation is the number of missing values (Finn 2015, p. 11; Stocking et al. 2001). If test-takers leave a large number of questions unanswered, this may be due to a low level of test motivation (Boe et al. 2002; Musekamp and Pearce 2016; Wise and Kong 2005) or it may be because they do not know the right answer (Shoemaker et al. 2000) or they are too slow in answering the test items so they cannot work on all the items. Thus, a low level of test motivation cannot automatically be assumed when there are many missing values on knowledge tests. In several studies, Musekamp et al. (2014) arranged test items in order of decreasing difficulty. The assumption was that due to the items becoming easier to answer for the test-takers, the number of missing values would decrease toward the end of the test. However, Musekamp and colleagues observed the opposite effect: the number of missing values increased toward the end of the test. This result can be attributed, inter alia, to the decreasing level of test motivation toward the end of the test.

In this paper, the number of missing values is examined as an external criterion. The number of missing values is regarded not as an indicator of test motivation, but rather as an external criterion. It cannot be clearly established, from a theoretical point of view, if the missing values result from a low test motivation or from lack of knowledge. A combination of both reasons is also possible. It is clear that the number of missing values should correlate with test performance. Ultimately, missing values are frequently coded as wrong answers on performance tests, resulting in a lower test performance (cf. Baker and Seock-Ho 2004; Coates 2004; Schwab and Helm 2015).

The literature on test motivation indicates that a lower test motivation correlates with a lower invested effort (Finn 2015, p. 11; Knekta and Eklöf 2015). Test-takers with a lower test motivation complete the questionnaire with less diligence, or cancel it earlier (cf. DeMars et al. 2013, p. 70–71; Eklöf 2010b, p. 8; Lee and Chen 2011, p. 361). This results in a higher number of missing values. Therefore, both test-takers' self-reported test motivation and their interest in receiving feedback on their performance on the test should correlate with the number of missing values. Based on these considerations, the two following hypotheses can be formulated:

H4.1 Test-takers with a lower self-reported motivation demonstrate a higher number of missing values.

H4.2 Test-takers with a lower interest in receiving feedback demonstrate a higher number of missing values.

Measurement instrument and sample

Economic knowledge and test motivation

In this study, vocational education students' economic knowledge was assessed¹¹ using the German adaptation (TEL4-G) of the fourth edition of the American test of economic literacy (TEL4; Walstad et al. 2013). The original TEL4 was adapted for use in German-speaking areas in 2014 during a 6-month adaptation and validation process (see Happ et al. 2016 and Förster et al. 2017). The TEL4 was developed for secondary school students in upper-grade levels. Curricular analysis indicate that the content-related validity of TEL4 is also applicable to the German-speaking area (cf. Förster et al. 2017; Happ et al. 2016). Students have 40 min to complete the items, which makes it possible to complete the test during one school lesson. The test is available in two similar versions (A and B) with 45 items each linked to each other via 10 identical anchor items.¹² During administration of both the TEL4 (Walstad et al. 2013) and the TEL4-G (Förster et al. 2017), it became apparent that item difficulty varied between Versions A and B. Because all students should retain the possibility of viewing their individual results from the test, the decision was made not to use both the A as well as the B versions of the TEL4-G. Ultimately, the results may differ slightly depending on the test version, thus complicating the interpretation of the achieved scores for the students. Because both versions measure the same construct, only version A was deployed for these reasons, ensuring all test-takers work on the same items. This should prevent the differences in the test versions from causing distortions in the result feedback.¹³

Sociodemographic variables of the participants such as type of school and gender were assessed in addition to their results on the TEL4-G. Upon completion of the TEL4-G, participants rated their test motivation on a scale (Knekta 2017; Ortner et al. 2014) containing four items (see Table 3 in Appendix) from the Test Attitude Survey by Arvey et al. (1990, p. 714) and the Student Opinion Scale of Sundre (2007, p. 4).¹⁴ When translating the items into German, Giermann (2012, p. 41) was followed. Participants rated their test motivation on a five-point Likert scale ranging from “strongly disagree” (–) to “strongly agree” (++) . The items focus mostly on “invested effort” as an aspect of test motivation (see “Test motivation and invested effort” section).

¹¹ Because the test administered in this study was in paper-and-pencil format, the average time spent on each item could not be determined (for analyses, see Liu et al. 2015). In future studies, tests could be administered on computers so that time spent on items can be recorded and modeled.

¹² See Förster et al. (2017) for a detailed analysis regarding both versions of the TEL4-G.

¹³ For copyright reasons, the German edition of the test cannot be presented. The items of the original American edition of the TEL4 are available on the following website: http://www.c3teachers.org/wp-content/uploads/2016/09/Walstad_Rebeck.pdf.

¹⁴ The Student Opinion Scale is an instrument for assessing test motivation frequently employed in the USA (Sundre 1997; Sundre and Moore 2002; Sundre and Kitsantas 2004). This instrument measures interviewees' self-assessed effort on tests as well as general attitudes toward taking tests (see theory-based model by Wigfield and Eccles 2000).

After rating their test motivation, the participants completed a feedback sheet on which they could generate an imaginary word¹⁵ to use four to six weeks later to obtain confidential feedback on their performance on the test. Both participation in the assessment and completion of the feedback sheet were voluntary and no direct (negative or positive) consequence of taking the test was evident; therefore, the assessment could be perceived as low-stakes (see Finn 2015). Accordingly, the participants who provided an imaginary word were assumed to have a high level of interest in receiving feedback on their performance on the test.

Sample and descriptive statistics

The 1018 participants in the sample under investigation were students in 66 classes at seven vocational schools in the federal state of Rhineland-Palatinate, Germany. With 625 students (61.8%), the vocational secondary school (berufliches Gymnasium) had the largest share. There were 357 students (32.1%) who were in training at a commercial vocational school (kaufmännisch-verwaltende Berufsausbildung). With 32 students (3.2%), the upper vocational school (Berufsoberschule) had only a very small share in the sample. However, the correlation between test motivation and performance on an economic knowledge test being the main focus of this paper (and not the level of economic knowledge), the students of the upper vocational school were retained in the modeling. In Table 1 an overview is given of the descriptive distribution of several sociodemographic characteristics of participants in the sample. Missing values also are presented in the table.

Gender distribution was more or less balanced, with a slightly higher percentage of female students (51.4%). A total of 28.6% of the test-takers did not provide an imaginary word on the feedback sheet and therefore did not receive feedback on their performance. Missing values on the imaginary word (2.3%) represent those participants who returned the test without the feedback sheet. It is not clear whether these test-takers accidentally took the sheet with them despite having generated an imaginary word or did not generate an imaginary word. Hence, the 23 questionnaires without the attached feedback sheet were considered as missing values. Participants rated on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree) statements regarding their test motivation. The average value of 2.33 indicates moderate test motivation. With a score of roughly 23.56 of 45 points, the participants completed on average slightly more than half of the items on the TEL4-G correctly.¹⁶ In Table 4 in Appendix, a correlation matrix for the correlations between the variables involved can be found.

¹⁵ For data protection reasons, obtaining information on the generation process of an individual code was not possible. In other studies, such codes may be generated by taking the third and fourth letters of the student's place of birth or the second and third letters of his or her mother's first name. After consulting with the data protection officer, the decision was made to generate an imaginary word (prominent athletes or politicians) despite overlaps that may occur.

¹⁶ Although no comparison was made in this study between regular and vocational secondary school students' performance on the TEL4-G, Happ et al. (Happ et al. 2018) found the average score on the TEL4-G of first-semester bachelor students at universities in Germany was 27.55 points, which is better than the score of the participants in this study.

Table 1 Sample statistics

Variable	N	%
Type of school		
Commercial vocational school (Wirtschaftsgymnasium)	629	61.8
Vocational training (Berufsausbildung)	357	35.1
Upper vocational school (Berufsoberschule)	32	3.1
Missing values	–	–
Gender		
Male	492	48.3
Female	523	51.4
Missing values	3	0.3
Interest in feedback		
Yes	704	69.2
No	291	28.6
Missing values	23	2.3
Age (min = 15; max = 36)	19.26 (2.618)	
Test scores		
Mean value test motivation (SD)	2.33 (.905)	
Mean value TEL4-G (SD)	23.56 (7.1)	

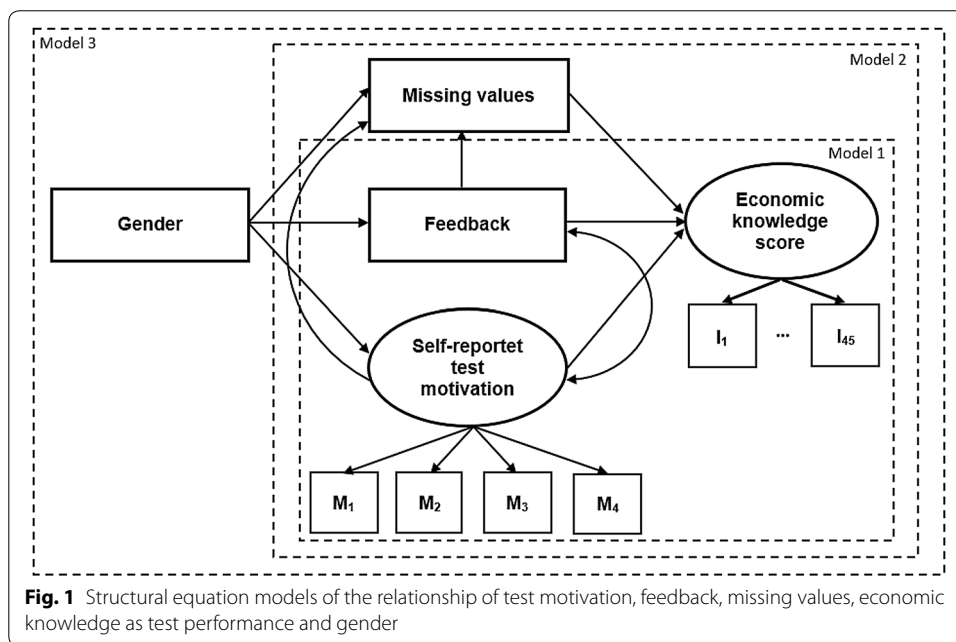


Fig. 1 Structural equation models of the relationship of test motivation, feedback, missing values, economic knowledge as test performance and gender

Statistical approach

To test the hypotheses, three structural equation models were specified in Mplus software version 7.3 (Muthén and Muthén 1998–2012; Byrne 2012; Kline 2016). These are illustrated in Fig. 1.

To assess each of the three models, the usual criteria were applied: Chi-squared (χ^2) test, root mean square error of approximation (RMSEA), comparative fit index (CFI), and the Tucker-Lewis index (TLI). The following cut-off criteria were taken as a basis for the interpretation of the goodness-of-fit indices: RMSEA: $\leq .06$, CFI: $\geq .95$, TLI: $\geq .97$

Table 2 Unstandardized and standardized path coefficients in the three models

	Model 1		Model 2		Model 3	
	B unst.	B stand.	B unst.	B stand.	B unst.	B stand.
Regression on TEL4-G score	R ² = .110		R ² = .226		R ² = .258	
Self-reported test motivation	.126***	.329***	.114***	.296***	.123***	.318***
Feedback	.007	.012	-.012	-.021	-.011	-.044
Number of missing values			-.021***	-.344***	-.023***	-.378***
Feedback with self-reported test motivation	.052***	.171***	.052***	.171***	.153***	.228***
Regression on number of missing values			R ² = .022		R ² = .055	
Self-reported test motivation			-.625	-.097	-.347	-.054
Feedback			-.926**	-.098**	-.610**	-.142**
Gender					-1.418***	-.164***
Regression on self-reported test motivation					R ² = .003	
Gender					.075	.056
Regression on feedback					R ² = .000	
Gender					.002	.001

* $p < .05$, ** $p < .01$, *** $p < .001$

(Hu and Bentler 1999; Schermelleh-Engel and Moosbrugger 2003). The measurement models of the two variables, self-reported test motivation and test score, were interpreted as good. We used Full-Information-Maximum-Likelihood (FIML) in order to consider missing values. The congeneric reliability of the measurement model in TEL4-G amounted to .888, and the one for the self-reported test motivation to .856.¹⁷

Results

In Table 2 an overview is given of the unstandardized and standardized path coefficients of the three models.

In accordance with the theoretical assumptions discussed in “[Theoretical basis and state of research](#)” section, self-reported test motivation, interest in receiving feedback, and test scores were analyzed together in the first model (Model 1), allowing the correlation between the two indicators of test motivation to be examined. With the reported criteria ($\chi^2 = 1491.3$, $df = 1173$, $p < .000$; number of estimated model parameters = 115; RMSEA = .016 (90% CI .014/.019); CFI = .968; TLI = .967), Model 1 was interpreted as suitable and showing a good fit. Self-reported test motivation was found to have a highly significant influence on the test score of .329 ($p < .001$), and a highly significant correlation was found between self-reported test motivation and interest in receiving feedback [.171; ($p < .001$)]. This indicates that participants with a higher level of self-reported test motivation achieved higher scores on the test (H1). Furthermore, it shows that participants with a generally higher level of self-reported test motivation were more likely to express interest in receiving feedback on their performance on the test (H3). The correlation between the two indicators of test motivation was weak. Because interest in receiving feedback was coded dichotomously, little variance was found.¹⁸ The very weak

¹⁷ Congeneric reliability is an ordinary reliability index for Structural Equation Modeling (SEM) and can be compared to Cronbach's Alpha.

¹⁸ Converting the correlation of 0.171 into an effect size resulted in a value corresponding to Cohen's d of 0.347, which is a small effect size (Cohen 1988; Rosenthal 1994).

correlation between interest in receiving feedback and test scores (.012; $p = .716$) was not significant (H2).

In the second model (Model 2), the number of missing values on the test was added to the structures of Model 1. With the criteria ($\chi^2 = 1573.704$, $df = 1220$, $p < .000$; number of estimated model parameters = 120; RMSEA = .017 (90% CI .014/.019); CFI = .966; TLI = .964) Model 2 also was considered as suitable and having a good fit. The influence of self-reported test motivation on test scores decreased slightly, indicating a value of .296. The correlation, however, remained significant ($p = .000$) (H1). The newly added number of missing values had a significant negative influence ($-.344$; $p < .001$) on performance on the test. This is because missing values generally were interpreted as resulting from lack of knowledge and, thus, as incorrect responses to test items (Baker and Seock-Ho 2004; Coates 2004; Schwab and Helm 2015). Therefore, the two variables were not independent of each other, which was considered to be unproblematic in the present study. Furthermore, students who showed interest in receiving feedback by creating an imaginary word had fewer missing values on the test ($-.098$; $p = .006$) (H4.2). The influence of self-reported test motivation on the number of missing values ($p = .052$) was significant only marginally (H4.1). However, the influence of interest in receiving feedback on their performance on the test ($p = .52$), as in Model 1, was not significant.

In the third model (Model 3) gender was taken into account in addition to the indicators of test motivation (DeMars et al. 2013). The model exhibited a good fit and quality ($\chi^2 = 1675.368$, $df = 1268$, $p < .000$; number of estimated model parameters = 122; RMSEA = .018 (90% CI .015/.020); CFI = .96; TLI = .958).¹⁹ Both self-reported test motivation (.32; $p < .001$) and the number of missing values ($-.38$; $p < .001$) correlated significantly with test scores (H1). Furthermore, the correlation between self-reported test motivation and interest in receiving feedback (.23; $p < .001$) as well as the influence of interest in receiving feedback on the missing values ($-.142$; $p < .01$) were maintained (H3 and H4.2). Gender correlated significantly with the number of missing values ($-.164$; $p < .001$) only. This means that male participants had fewer missing values than female participants. After incorporating gender into Model 3, the total variance explained in test scores was slightly greater than in Model 2 (R^2 -Model 2 = .226; R^2 -Model 3 = .26).

Discussion

In the analyses, self-reported test motivation had a significant influence on performance on the TEL4-G (H1). Surprisingly, no correlation was found between self-reported test motivation and the number of missing values on the domain-specific test (external criterion) (H4.1). A possible interpretation of this finding might be that the participants' perception of their performance on the test influenced their subsequent self-reported test motivation, resulting in a greater correlation between test motivation and test scores instead of between test motivation and the number of missing values. This (potential) link between test motivation self-reported after completion of the test and perception of performance on the test might result in the interpretation of the correlation with the test

¹⁹ While we had 3 missing values for the gender variable the sample size for model 3 is 1015 instead of 1018 in the models 1 and 2.

score as a maximum correlation between test motivation and test score. Another reason might be that test scores are already controlled for student test motivation and thus all motivational parts of test scores are already accounted for by the test motivation scale so that “nothing is left” for missing values. In further studies, test motivation should be assessed before and during the test in order to estimate more accurately the extent to which it is potentially influenced by performance on the test. As described above, test motivation was interpreted as the result of perceived item difficulty, students’ individual skill (expectation component), and the relevance of the test to the test-taker (Wigfield and Eccles 2000) and, thus, is not assumed to be constant. Therefore, one can argue that perceived item difficulty and self-estimated performance were relevant for self-reported test motivation.

At the same time, there was a moderate correlation between interest in receiving feedback and self-reported test motivation (H3). Interest in receiving feedback also correlated with the number of missing values (H4.2) but not with the latent test score. Interest in receiving feedback in particular can be seen as an indicator of relevance of the test for an individual while self-reported test motivation indicates the willingness to make an effort to complete the test, and thus can be interpreted as resulting from expectations and values (see “[Test motivation and invested effort](#)” section). The assessment of test motivation in this paper had some limitations. Self-efficacy as a relevant expectation component was not assessed. The focus of further studies could be on the development of value-and-expectation constructs over time and their influence on performance on tests. The findings in this study reveal that there are other indicators of test motivation in addition to self-reported test motivation that should be integrated into analyses for the purposes of incremental validation for assessment of test motivation (AERA et al. 2014). Moreover, it became clear that 11% of the test scores could be explained primarily by self-reported test motivation. This emphasizes the importance of taking test motivation into account when interpreting test scores.

Conclusions and limitations

Over the past two decades, administration of low-stakes tests has increased in schools in Germany. Simultaneously, more research has been done on the correlation between test motivation and performance on tests. Results of studies conducted in Germany (cf. for example Penk et al. 2014) and of international studies (cf. for example Eklöf and Nyroos 2013; Wise and DeMars 2005) confirm the correlation between a low level of test motivation and test scores, particularly on low-stakes tests. Similar research in the field of vocational education is scarce (see for example Kögler and Rausch 2017). The present paper demonstrates that in vocational education, too, a significant correlation must be postulated between test motivation and scores on a standardized test. Two indicators of the operationalization of test motivation were investigated in this study. This procedure distinguishes the present study from numerous other studies on test motivation. Incorporating interest in receiving feedback in particular can be regarded as innovative for the vocational education sector.

Economic knowledge ought to be of particular interest to vocational school students in Germany because economic content often is incorporated into the curriculum.²⁰ Before doing the test, the participants were given the opportunity to receive feedback on their performance on the test as an incentive to complete the test as accurately and thoroughly as possible. In this light, the overall moderate self-reported test motivation is considered to be critical. Due to the limited number of participants in this sample, particularly from upper-vocational schools, a comparison of self-reported test motivation was not performed. Considering the pronounced heterogeneity in student bodies within and across the vocational schools where data were collected, in future studies differentiated analyses should be performed according to the type of school.²¹ With the vocational education student body in mind, samples should include more low-ability students (e.g., from the transition system). This would not only increase the representativeness of the survey but also enable exciting comparative analyses of the students from the diverse school types (see “[Relevance](#)” section).

There are some limitations to this study which should be mentioned. First, the sample included students from vocational schools in one federal state only. Therefore, the results of this study are not representative of all students in vocational schools throughout Germany. Second, the selection of schools in Rhineland-Palatinate to participate in this study was not random. Generally, willingness to cooperate and interest in the study were criteria needed to gain access to students of a vocational school.

Neither self-reported test motivation nor interest in receiving feedback were significantly influenced by the gender of the test-takers. It should be noted that this differs from the current findings of related research (DeMars et al. 2013; Finn 2015; Marrs and Sigler 2012). However, the highly significant correlation between gender and the number of missing values on the economic knowledge test is remarkable, with female students showing more missing values. At this point, the ongoing discussion that was started several years ago about the extent to which the multiple-choice (MC) format influences the results in a knowledge test (Biggs 1999; Bridgeman and Lewis 1994; Ferber et al. 1983; Lumsden and Scott 1987) should also be referenced. MC items are seldom employed in the German educational system (Wuttke 2007). Therefore, students have a relatively low testwiseness in dealing with MC items, regardless of gender. In this study female students left more items unanswered, which could be because females take fewer risks on MC tests than males (on the subject of differences in risk aversion between the genders, see Powell and Ansic 1997; Byrnes et al. 1999; Weber et al. 2002). However, further research is needed to substantiate this conclusion. The low level of risk tolerance possibly resulted in less willingness to guess an answer in the case of uncertainty, which subsequently resulted in a higher number of missing values for female test-takers. The majority of missing values, however, could be attributed to a potential difference in knowledge between male students and female students.²² Because the male test-takers

²⁰ This field is different from general secondary schools in Germany because economic contents are not extensively incorporated into the curricula in all federal states (Förster et al. 2017).

²¹ A first look at the data from the present study showed that self-reported test-taking motivation amounted to 2.36 (SD .904) for students of the vocational secondary school (berufliches Gymnasium), 2.31 (SD .874) for students of the commercial vocational school, and 1.93 (SD 1.147) for students of the upper vocational school (Berufsoberschule). Thus, further studies should pursue a school-type-specific modeling—in doing which a significantly greater number of cases is an essential prerequisite, particularly for the upper vocational school.

²² There are studies in the literature reporting, from an international comparative perspective, both variations in the number of missing values on the domain-specific test in a MC format as well as variations in the effect size of the gender gap difference between male and female test-takers in economic knowledge (see Brückner et al. 2015; Förster et al. 2015). Here, starting points for further research can be found.

achieved better test scores, it is possible that they were able to respond correctly to more items and leave fewer unanswered. This surely is where more in-depth studies are needed of the influence of item formats and missing values. In this study no gender differences were found in self-reported test motivation or interest in receiving feedback indicators of test-taking motivation, no differences between male and female students were found.

Abbreviations

CEE: council for economic education; CFI: comparative fit index; df: degrees of freedom; FIML: Full-Information-Maximum-Likelihood; H: hypothesis; MC: multiple-choice; N: sample; PISA: Program for International Student Assessment; RMSEA: root mean square error of approximation; SD: standard deviation; SEM: Structural Equation Modeling; TEL4: test of economic literacy 4th version; TEL4-G: test of economic literacy German version; TIMSS: Trends in International Mathematics and Science Study; TL: Tucker-Lewis index; χ^2 : Chi-squared.

Authors' contributions

Both authors contributed substantially to this work. RH and MF developed the theoretical framework of the paper. Data analysis for this paper was conducted by MF and RH. Both authors discussed the manuscript at all stages. Both authors read and approved the final manuscript.

Author details

¹ Department of Law, Business & Economics, Chair of Business and Economics Education, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 9, 55128 Mainz, Germany. ² Faculty of Social Sciences, Economics, and Business Administration, Chair for Business and Economics Education, Otto Friedrich University of Bamberg, Kärntenstraße 7, 96052 Bamberg, Germany.

Acknowledgements

We particularly thank the two anonymous reviewers who provided very detailed, constructive feedback and helpful guidance during the revision of this paper.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable. There is a copyright on the test instrument by the Council for Economic Education (CEE; US) and the data set cannot be shared.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Taking part in the research project was entirely voluntary. All students consented to participate. A ethic committee was established at the end of 2014 at the department of business and economics at the University in Mainz. Until that point the design and the questionnaire was finished. We have still talked to a member of the ethics committee and he remarked no critical points in the questionnaire.

Funding

Not applicable.

Appendix

See Tables 3 and 4.

Table 3 Items for assessing test motivation

Original English version (Arvey et al. 1990, p. 714; Sundre 2007, p. 4)	German translation and adaptation (Giermann 2012, p. 185)
I engaged in good effort throughout this test	Ich habe in diesem Test durchgehend eine hohe Anstrengung aufgebracht
I did not give this test my full attention while completing it	Während des Tests war ich vollkommen aufmerksam
While taking this test or tests, I concentrated and tried to do well	Als ich diesen Test bearbeitet habe, war ich konzentriert und habe versucht, alle Aufgaben richtig zu beantworten
Doing well on this test (or these tests) is important to me	In diesem Test gut zu sein, war mir persönlich sehr wichtig

Table 4 Estimated correlation matrix for the latent variables

	1	2	3	4	5
1. Economic knowledge score	1				
2. Self-reported test motivation	.331	1			
3. Interest in receiving feedback	.091	.224	1		
4. Missing score	-.375	-.114	-.139	1	
5. Gender	.172	-.002	.031	-.016	1

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 June 2018 Accepted: 1 October 2018

Published online: 15 October 2018

References

- AERA (American Educational Research Association), APA (American Psychological Association), NCME (National Council on Measurement in Education) (2014) The standards for educational and psychological testing. AERA, Washington, DC
- Arvey RD, Strickland W, Drauden G, Martin C (1990) Motivational components of test taking. *Pers Psychol* 43(4):695–716
- Asarta CJ, Butters RB, Thompson E (2014) The gender question in economic education: is it the teacher or the test? *Perspect Econ Educ Res* 9(1):1–19
- Asseburg R, Frey A (2013) Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychol Test Assess Model* 55(1):92–104
- Baethge M, Achtenhagen F, Arends L, Babic E, Baethge-Kinsky V, Weber S (2006) PISA-VET—a feasibility-study. Steiner, Stuttgart
- Baker FB, Seock-Ho K (2004) Item response theory: parameter estimation techniques, 2nd edn. Dekker, New York
- Barry CL, Finney SJ (2016) Modeling change in effort across a low-stakes testing session: a latent growth curve modeling approach. *J Appl Meas* 29(1):46–64. <https://doi.org/10.1080/08957347.2015.1102914>
- Baumert J, Demmrich A (2001) Test motivation in the assessment of student skills: the effects of incentives on motivation and performance. *Eur J Psychol Educ* 16(3):441–462
- Beck K, Krumm V, Dubs R (1998) Wirtschaftskundlicher Bildungs-Test (WBT) [Test of economic literacy]. Hogrefe, Göttingen
- Beicht U, Walden G (2016) Transitions into vocational education and training by lower and intermediate secondary school leavers. Can male adolescents compensate for their school-based educational disadvantage in comparison with female adolescents? *Empir Res Vocat Educ Train* 8:11. <https://doi.org/10.1186/s40461-016-0037-9>
- Biggs J (1999) Teaching for quality learning at University. Society for Research into Higher Education and Open University, Buckingham
- Billet S (2011) Vocational education: purposes, traditions and prospects. Springer, Dordrecht
- Boe EE, May H, Boruch RF (2002) Student task persistence in the Third International Mathematics and Science Study: a major source of achievement differences at the national, classroom, and student levels. University of Pennsylvania, Center for Research and Evaluation in Social Policy, Philadelphia
- Bridgeman B, Lewis C (1994) The relationship of essay and multiple-choice scores with grades in college courses. *J Educ Meas* 31(1):37–50
- Brückner S, Förster M, Zlatkin-Troitschanskaia O, Happ R, Walstad WB, Yamaoka M, Asano T (2015) Gender effects in assessment of economic knowledge and understanding: differences among undergraduate business and economics students in Germany, Japan, and the United States. *Paebody J Educ* 90(4):503–518
- Butler J, Adams RJ (2007) The impact of differential investment of student effort on the outcomes of international studies. *J Appl Meas* 8:279–304
- Byrne BM (2012) Structural equation modeling with Mplus. Routledge, New York
- Byrnes JP, Miller DC, Schafer WD (1999) Gender difference in risk taking: a meta-analysis. *Psychol Bull* 125:367–383
- Coates H (2004) Treating test item nonresponse. *J Appl Meas* 5(1):1–25
- Cohen J (1988) Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale
- DeMars CE, Bashkov BM, Socha AB (2013) The role of gender in test-taking motivation under low-stakes conditions. *Res Pract Assess* 8:69–82
- Eklöf H (2007) Test-taking motivation and mathematics performance in TIMSS 2003. *Int J Test* 7:311–326. <https://doi.org/10.1080/15305050701438074>
- Eklöf H (2008) Test-taking motivation on low-stakes tests: a Swedish TIMSS 2003 example. *IERI Monogr Ser Issues Methodol Large Scale Assess* 1:9–21
- Eklöf H (2010a) Skill and will: test-taking motivation and assessment quality. *Assess Educ Princ Policy Pract* 17:345–356
- Eklöf H (2010b) Student motivation and effort in the Swedish TIMSS Advanced Field Study. In: Paper presented at the meeting of the 4th IEA international research conference, Gothenburg

- Eklöf H, Nyroos M (2013) Pupil perceptions of national tests in science: perceived importance, invested effort, and test anxiety. *Eur J Psychol Educ* 28(2):497–510. <https://doi.org/10.1007/s10212-012-0125-6>
- Ferber MA, Birnbaum BG, Green CA (1983) Gender differences in economic knowledge: a re-evaluation of the evidence. *J Econ Educ* 14(2):24–37
- Finn B (2015) Measuring motivation in low-stakes assessments. ETS Research Report RR-15-19. Educational Testing Service, Princeton. <https://doi.org/10.1002/ets2.12067>
- Förster M, Zlatkin-Troitschanskaia O, Brückner S, Happ R, Hambleton RK, Walstad WB, Asano T, Yamaoka M (2015) Validating test score interpretations by cross-national comparison: comparing the results of students from Japan and Germany on an American test of economic knowledge in higher education. *Z Psychol* 223(1):14–23
- Förster M, Brückner S, Happ R, Beck K, Zlatkin-Troitschanskaia O (2017) Strukturanalyse eines kognitiven Messinstruments im Multiple Choice-Format. Das Beispiel des Test of Economic Literacy (TEL4-G) [Structural analysis of a cognitive multiple choice measuring instrument—exemplified by the test of economic literacy (TEL4-G)]. *Zeitschrift für Berufs- und Wirtschaftspädagogik [J Vocat Bus Educ]* 113(3):366–396
- Giermann I (2012) Der Einfluss von Testmotivation auf die Leistung in einem Leistungstest [The impact of test motivation on performance in a performance test]. Diploma Thesis, University of Regensburg
- Happ R, Förster M, Zlatkin-Troitschanskaia O, Carstensen V (2016) Assessing the previous economic knowledge of beginning students in Germany—implications for teaching economics in basic courses. *Citizsh Soc Econ Educ* 15(1):45–57
- Happ R, Förster M, Beck K (2018) Eingangsvoraussetzungen von Studierenden der Wirtschaftswissenschaften mit und ohne Migrationshintergrund [Migration background and economic knowledge of beginning university students]. *Zeitschrift für empirische Hochschulforschung [J Empir Res High Educ]* 2(1):5–22. <https://doi.org/10.3224/zehf.v2i1.01>
- Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
- Huffman L, Adamopoulos A, Murdock G, Cole A, McDermid R (2011) Strategies to motivate students for program assessment. *Educ Assess* 16:90–103
- Jacob M, Solga H (2015) Germany's vocational education and training system in transformation: changes in the participation of low- and high-achieving youth over time. *Eur Sociol Rev* 31(2):161–171
- Kane M (2006) Content-related validity evidence in test development. In: Haladyna TM, Downing SM (eds) *Handbook of test development*, vol 1. Lawrence Erlbaum, Mahwah, pp 131–153
- King RB (2016) Gender differences in motivation, engagement and achievement are related to students' perceptions of peer—but not of parent or teacher—attitudes toward school. *Learn Individ Differ* 52:60–71. <https://doi.org/10.1016/j.lindif.2016.10.006>
- Kline RB (2016) *Principles and practice of structural equation modeling*, 4th edn. The Guildford Press, New York
- Kornhauser Z, Minahan J, Siedlecki K, Steedle JT (2014) A strategy for increasing student motivation on low-stakes assessments. In: Paper presented at the Annual Meeting of the American Educational Research Association, Philadelphia.
- Knekta E (2017) Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes. *Scand J Educ Res* 61(1):95–111. <https://doi.org/10.1080/00313831.2015.1119723>
- Knekta E, Eklöf H (2015) Modeling the test-taking motivation construct through investigation of psychometric properties of an expectancy-value-based questionnaire. *J Psychoeduc Assess* 33(7):662–673. <https://doi.org/10.1177/0734282914551956>
- Kögler K, Rausch A (2017) Bedingungsfaktoren und Veränderlichkeit der Testmotivation im domänenspezifischen Problemlösen [Determinants and changeability of test-taking motivation in domain-specific problem solving]. In: Beitrag im Symposium auf der Jahrestagung der Sektion Berufswirtschaftspädagogik am 16. September [Contribution to the symposium at the annual meeting of the section for vocational and business education on 16 September], Stuttgart
- Kong XJ, Wise SL, Harnes JC, Yang S (2006) Motivational effects of praise in response-time based feedback: a follow-up study of the effort-monitoring CBT. In: Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA
- Lee YH, Chen H (2011) A review of recent response-time analyses in educational testing. *Psychol Test Assess Model* 53(3):359–379
- Liu OL, Rios JA, Borden V (2015) The effects of motivational instruction on college students' performance on low-stakes assessment. *Educ Assess* 20:79–94. <https://doi.org/10.1080/10627197.2015.1028618>
- Lumsden KG, Scott A (1987) The economics student re-examined: male–female differences in comprehension. *J Econ Educ* 18(4):365–375
- Marrs H, Sigler EA (2012) Male academic performance in college: the possible role of study strategies. *Psychol Men Masc* 13:227–241. <https://doi.org/10.1037/a0022247>
- Musekamp F, Pearce J (2016) Student motivation in low-stakes assessment contexts: an exploratory analysis in engineering mechanics. *Assess Eval High Educ* 41(5):750–769. <https://doi.org/10.1080/02602938.2016.1167164>
- Musekamp F, Spöttl G, Mehrafza M, Heine J-H, Heene M (2014) Modeling of competences for students of engineering mechanics. *Int J Eng Pedagog* 4(1):4–12. <https://doi.org/10.3991/ijep.v4i1.2917>
- Muthén LK, Muthén BO (1998–2012) *Mplus user's guide*, 7th ed. Muthén & Muthén, Los Angeles
- O'Neil HF, Abedi J, Miyoshi J, Mastergeorge A (2005) Monetary incentives for low-stakes tests. *Educ Assess* 10:185–208
- Ortner TO, Weißkopf E, Koch T (2014) I will probably fail. Higher ability students' motivational experiences during adaptive achievement testing. *Eur J Psychol Assess* 30(1):48–56. <https://doi.org/10.1027/1015-5759/a000168>
- Penk C, Richter D (2017) Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educ Assess Eval Account* 29:55–79. <https://doi.org/10.1007/s11092-016-9248-7>
- Penk C, Pöhlmann C, Roppelt A (2014) The role of test-taking motivation for students' performance in low-stakes assessments: an investigation of school-track-specific differences. *Large Scale Assess Educ* 2:5. <https://doi.org/10.1186/s40536-014-0005-4>

- Powell M, Ansic D (1997) Gender differences in risk behavior in financial decision-making: an experimental analysis. *J Econ Psychol* 18(6):605–628
- Protsch P, Solga H (2015) The social stratification of the German VET system. *J Educ Work* 29(5):637–661. <https://doi.org/10.1080/13639080.2015.1024643>
- Rausch A, Kögler K (2016) Authenticity and efficiency in assessing domain-specific problem-solving competence: conflicting goals in large-scale assessments? Discussion Paper, International conference on competence theory, research and practice, Wageningen, NL
- Rosenthal R (1994) Parametric measures of effect size. In: Cooper H, Hedges LV (eds) *The handbook of research synthesis*. Sage, New York, pp 231–244
- Schermelleh-Engel K, Moosbrugger H (2003) Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online* 8(2):23–74
- Schwab S, Helm C (2015) Überprüfung von Messinvarianz mittels CFA und DIF-Analysen [Testing for measurement invariance in students with and without special educational needs]. *Empirische Sonderpädagogik [Empir Special Educ]* 3:175–193
- Shoemaker J, Eichholz M, Skewes EA (2000) Item response: distinguishing between don't know and refuse. *Int J Public Opin Res* 14(2):193–201
- Solga H, Protsch P, Ebner C, Crzinsky-Fay C (2014) The German vocational education and training system: Its institutional configuration, strengths, and challenges. Discussion Paper. WZB Berlin Social Science Center, Berlin. <https://bibliothek.wzb.eu/pdf/2014/i14-502.pdf>. Accessed 04 June 2018
- Stenlund T, Lyrén P-E, Eklöf H (2018) The successful test taker: exploring test-taking behavior profiles through cluster analysis. *Eur J Psychol Educ* 33:403–417. <https://doi.org/10.1007/s10212-017-0332-2>
- Stocking ML, Steffen MS, Eignor DR (2001) A method for building a realistic model of test taker behavior for computerized adaptive testing (RR-01-22). Educational Testing Service, Princeton
- Sundre DL (1997) Differential examinee motivation and validity: a dangerous combination. In: Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL
- Sundre DL (2007) The Student Opinion Scale (SOS). A measure of examinee motivation. Test manual. The Center for Assessment & Research Studies, Harrisonburg
- Sundre DL, Kitsantas A (2004) An exploration of the psychology of the examinee: can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemp Educ Psychol* 29:6–26
- Sundre DL, Moore DL (2002) The students opinion scale: a measurement of examinee motivation. *Assess Update* 14(1):8–9
- Swerdzewski PJ, Harnes JC, Finney SJ (2011) Two approaches for identifying low-motivated students in a low-stakes assessment context. *Appl Meas Educ* 24(2):162–188
- Thelk AD, Sundre DL, Horst SJ, Finney SJ (2009) Motivation matters: using the Student Opinion Scale to make valid inferences about student performance. *J Gen Educ* 58(3):129–151. <https://doi.org/10.1353/jge.0.0047>
- Walstad WB, Rebeck K, Butters RB (2013) Test of economic literacy: development and results. *J Econ Educ* 44(3):298–309
- Weber EU, Blais AR, Betz NE (2002) A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *J Behav Decis Making* 5:1–28
- Wigfield A, Eccles JS (2000) Expectancy-value theory of achievement motivation. *Contemp Educ Psychol* 25:68–81
- Winther E, Achtenhagen F (2009) Measurement of vocational competencies—a contribution to an international large-scale assessment on vocational education and training. *Empir Res Vocat Educ Train* 1:85–108
- Wise VL (2004) The effects of the promise of test feedback on examinee performance and motivation under low-stakes testing conditions. University of Nebraska, Lincoln
- Wise SL (2009) Strategies for managing the problem of unmotivated examinees in lowstakes testing programs. *J Gen Educ* 58:152–166
- Wise SL, DeMars CE (2005) Low examinee effort in low-stakes assessment: problems and potential solutions. *Educ Assess* 10:1–17
- Wise SL, Kong X (2005) Response time effort: a new measure of examinee motivation in computer-based tests. *Appl Meas Educ* 18(2):163–183. https://doi.org/10.1207/s15324818ame1802_2
- Wise SL, Pastor DA, Kong XJ (2009) Correlates of rapid-guessing behavior in low-stakes testing: implications for test development and measurement practice. *Appl Meas Educ* 22(2):185–205. <https://doi.org/10.1080/08957340902754650>
- Wuttke J (2007) Die Insignifikanz signifikanter Unterschiede: Der Genauigkeitsanspruch von PISA ist illusorisch [The insignificance of significant differences: PISA's claim to accuracy is illusory]. In: Jahnke T, Meyerhöfer W (eds) *Kritik eines Programms [Criticism of a program]*, 2nd edn. PISA & Co., Franzbecker, Berlin, pp 99–246
- Ziegler M, MacCann C, Roberts R (eds) (2011) *New perspectives on faking in personality assessment*. Oxford University Press, Oxford
- Zilberberg A, Anderson RD, Finney SJ, Marsh KR (2013) American college students' attitudes toward institutional accountability testing: developing measures. *Educ Assess* 18(3):208–234. <https://doi.org/10.1080/10627197.2013.817153>