



# Test-Taking Motivation in Education Students: Task Battery Order Affected Within-Test-Taker Effort and Importance

Anett Wolgast<sup>1\*</sup>, Nico Schmidt<sup>2</sup> and Jochen Ranger<sup>2</sup>

<sup>1</sup> Department of Psychology, University of Applied Sciences Hannover, Hanover, Germany, <sup>2</sup> Institute of Psychology, Martin-Luther-University Halle-Wittenberg, Halle, Germany

## OPEN ACCESS

### Edited by:

Olga Zlatkin-Troitschanskaia,  
Johannes Gutenberg University  
Mainz, Germany

### Reviewed by:

Marion Händel,  
University of Erlangen Nuremberg,  
Germany  
Martin Hecht,  
Humboldt University of Berlin,  
Germany

### \*Correspondence:

Anett Wolgast  
anett.wolgast@gmail.com

### Specialty section:

This article was submitted to  
Educational Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 06 May 2020

**Accepted:** 30 October 2020

**Published:** 25 November 2020

### Citation:

Wolgast A, Schmidt N and  
Ranger J (2020) Test-Taking  
Motivation in Education Students:  
Task Battery Order Affected  
Within-Test-Taker Effort  
and Importance.  
Front. Psychol. 11:559683.  
doi: 10.3389/fpsyg.2020.559683

Different types of tasks exist, including tasks for research purposes or exams assessing knowledge. According to expectation-value theory, tests are related to different levels of effort and importance within a test taker. Test-taking effort and importance in students decreased over the course of high-stakes tests or low-stakes-tests in research on test-taking motivation. However, whether test-order changes affect effort, importance, and response processes of education students have seldomly been experimentally examined. We aimed to examine changes in effort and importance resulting from variations in test battery order and their relations to response processes. We employed an experimental design assessing  $N = 320$  education students' test-taking effort and importance three times as well as their performance on cognitive ability tasks and a mock exam. Further relevant covariates were assessed once such as expectancies, test anxiety, and concentration. We randomly varied the order of the cognitive ability test and mock exam. The assumption of intraindividual changes in education students' effort and importance over the course of test taking was tested by one latent growth curve that separated data for each condition. In contrast to previous studies, responses and test response times were included in diffusion models for examining education students' response processes within the test-taking context. The results indicated intraindividual changes in education students' effort or importance depending on test order but similar mock-exam response processes. In particular effort did not decrease, when the cognitive ability test came first and the mock exam subsequently but significantly decreased, when the mock exam came first and the cognitive ability test subsequently. Diffusion modeling suggested differences in response processes (separation boundaries and estimated latent trait) on cognitive ability tasks suggesting higher motivational levels when the cognitive ability test came first than vice versa. The response processes on the mock exam tasks did not relate to condition.

**Keywords:** expectation-value theory, diffusion modeling, latent growth curve modeling, perspective-taking, exam test-taking motivation

## INTRODUCTION

Researchers analyzing data from the Programme for International Student Assessment (PISA) concerning the relations between motivation and test-taking achievement in mathematics reported that motivation explained 1–29% of the variance in achievement-test results (Kriegbaum et al., 2014). Further findings suggested item position effects on test performance (Weirich et al., 2016; Nagy et al., 2018, 2019; Rose et al., 2019; Liu and Hau, 2020). A problem found was decreased test performance over the course of taking a computer-assisted achievement test (List et al., 2017). That raised the question if motivation similarly decreased over the course of taking a computer-assisted achievement test. Researchers found low test-taking effort related to low test performance, discussed and tested several strategies for test takers' high effort levels, for example, incentives, integration into grading systems, or explaining test takers the relevance and importance of PISA test results (Baumert and Demmrich, 2001; Finn, 2015; Schüttelz-Brauns et al., 2020). Without applying any strategy to increase test-takers' effort, researchers found decreased intraindividual effort over the course of taking a test, this time effort of apprentices (technicians, clerks, and lab assistants, Lindner et al., 2018).

Decreasing test-taking effort is a serious problem since (computer-assisted) test-taking performance reflects an unknown amount of the tested ability in this case and threatens validity for the examined sample (Penk and Richter, 2017; Nagy et al., 2018). Furthermore, decreasing test-taking effort in a mock exam might affect achievement related choices (e.g., respond vs. not respond on a computer-assisted task) and the subsequent learning behavior in preparation of the exam. Achievement related choices in computer-assisted tasks regard test-takers' information processing. Undergraduate students in higher education often have the choice, if they respond on a computer-assisted task in a mock exam, and how much effort they spend on different types of task.

Effort is usually described as a component of achievement motivation (Eccles et al., 1984; Eccles and Wigfield, 1995; Wigfield and Eccles, 2000) or test-taking motivation (Wise and DeMars, 2005; Knekta and Eklöf, 2015; Knekta, 2017). Test-taking motivation is the engagement and effort that a person applies to a goal in order to achieve the best possible result in an achievement test (Wise and DeMars, 2005). Invested effort is conceptualized as relevant predictor on test performance according to the expectancy-value model applied to a test situation (Knekta and Eklöf, 2015, p. 663).

Knekta and Eklöf (2015) investigated adolescents' test-taking motivation and academic achievement, alongside further motivational components such as test-taking importance, expectancies, anxiety, and interest. A great deal of evidence supports the relevance of these motivational components for test performance (Eklöf and Hopfenbeck, 2018) and academic achievement (e.g., Nagy et al., 2010). For example, self-reported expectancies, test-taking effort, and test-taking importance have been found to determine performance in high-stakes tests (e.g., Knekta, 2017; Eklöf and Hopfenbeck, 2018); as one would

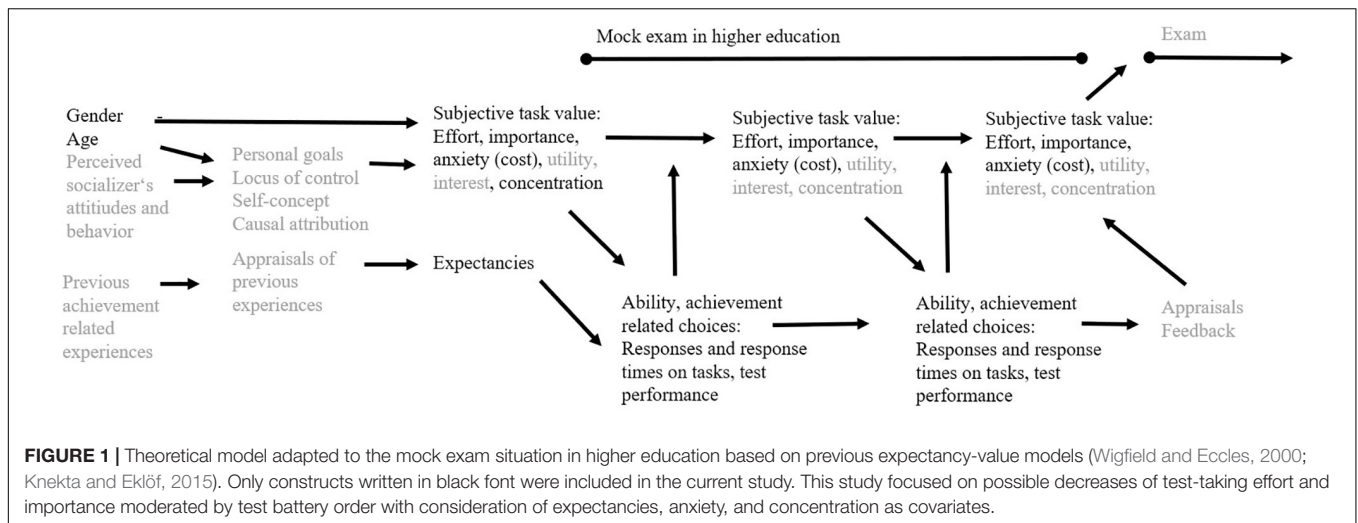
expect, high levels of these motivation components predicted higher performance levels. In low-stakes achievement tests, the relations between test-taking effort or test-taking importance (assessed by self-reports) and performance were inconsistently at zero (Sundre and Kitsantas, 2004; Penk and Richter, 2017) to low levels (Knekta, 2017; Eklöf and Hopfenbeck, 2018; Myers and Finney, 2019).

High-stakes achievement tests usually refer to ability assessments for selection purposes (e.g., enrollment in a type of school, internship, study program, or exams to complete a study program, e.g., Knekta, 2017). Low-stakes tests considered in previous research have included tests to practice high stakes-tests (e.g., mock exams), tests to evaluate educational programs (Brunner et al., 2007; Butler and Adams, 2007), tests to develop or update standardized achievement test inventories (e.g., standardization in a representative sample), and cognitive ability tests for research purposes (McHugh et al., 2004; Erle and Topolinski, 2015; Gorges et al., 2016; Goldhammer et al., 2017). Cognitive ability tests for research purposes and tests of subject content knowledge are often used in international large-scale assessments including students in school (Baumert and Demmrich, 2001; Butler and Adams, 2007; Kunina-Habenicht and Goldhammer, 2020) or standardized achievement tests in higher education in the United States (Silm et al., 2020).

The motivation at the end of a computer-assisted mock exam possibly determines undergraduate students' exam preparation. The theoretical model in **Figure 1** describes both test-taking effort and importance as significant determinants of the upcoming exam. The current study focused on test-taking effort and importance of *education students in higher education in Germany* at three measurement points *during a computer-assisted mock exam moderated by the experimentally varied order of a cognitive ability test and a battery of mock exam tasks*. Hence, the current study aimed to (1) examine whether education students' test-taking effort and importance decrease over the course of a computer-assisted cognitive ability test and subsequent computer-assisted mock-exam tasks, or vice versa, mock-exam tasks and a subsequent cognitive ability test considering further motivational components as covariates and (2) analyze differences in education students' information processing and response processes for these two task types depending on their order. The theoretical background of the model in **Figure 1** is outlined in the next section.

## THEORY AND ASSESSMENT OF TEST-TAKING EFFORT AND IMPORTANCE AS MOTIVATIONAL COMPONENTS

One way to disentangle the contributions of ability and motivation is to measure motivation in addition to the ability being tested (e.g., Baumert and Demmrich, 2001; Bensley et al., 2016), based on expectancy-value theory that has its roots in part in motive, expectancy, and incentive as determinants of "aroused motivation to achieve" proposed by Atkinson (1957, p. 362).



## Expectancy-Value Theory

Expectancy-value theory (Atkinson, 1957; Wigfield and Eccles, 2000) summarizes the relations among a number of individual background variables, that are in brief: gender, age, aptitudes, perceived socializer's beliefs and behavior, subjective appraisal of previous achievement related experiences, affective memories, and self-concept which help explain variance in learners' achievement related choices. Most of these background variables are out of the current study's scope except gender and age since Knekta and Eklöf (2015) presented their further developed expectancy-value model applied to a test situation (Knekta and Eklöf, 2015, p. 663). Both theoretical approaches (Wigfield and Eccles, 2000; Knekta and Eklöf, 2015) posit that a person's achievement related choices are in part explained by gender, age, and the subjective task value which includes a number of motivational components, namely effort, importance, expectancies, and anxiety. Expectancy of success, and subjective task value (i.e., incentive and attainment value, utility, interest).

## Evidence for the Expectancy-Value Theory

A large body of evidence supports the assumptions made in expectancy-value theory. For example, gender consistently explained variance in test-taking effort with females having an advantage over males in a review of literature (DeMars and Bashkov, 2013). The size of the gender gap seems to vary over age groups (DeMars and Bashkov, 2013). Other review results suggested that test-taking effort decreased with increasing years of age (Silm et al., 2020). Another study included undergraduates from 18 to 69 years of age with 56% being 35 years of age or less for investigating their test-taking behavior (Rios and Liu, 2017). The authors discussed the results and limitations of their study as follows: "we evaluated the comparability of proctored groups by gender, ethnicity, language, age, and GPA [grade point average]. We found no significant group differences across all variables except gender and age" (Rios and Liu, 2017, p. 11).

## Assessment of Test-Takers' Motivation by a Questionnaire

To assess levels of motivational components, researchers typically use well-established motivation inventories which were developed to measure motivation as a trait (Midgley et al., 1998; Simzar et al., 2015) or state (Vollmeyer and Rheinberg, 2006; Freund and Holling, 2011; Freund et al., 2011). Simzar et al. (2015) and other researchers (Arvey et al., 1990) have found inconsistent relations between trait motivation and test performance (Sundre and Kitsantas, 2004). Indeed, motivation while taking an achievement test is also conceptionally related to a person's motivational state in that situation. One questionnaire measuring current motivational state is the Questionnaire on Current Motivation (QCM) (e.g., Vollmeyer and Rheinberg, 2006), which several studies (e.g., Penk and Richter, 2017) have used to disentangle the relationship between current motivation, including the dimensions of anxiety, challenge, interest, and probability, and test performance (Freund et al., 2011). Findings from studies using the QCM indicate relations at moderate levels between interest and test scores (Freund et al., 2011). However, one at least partial limitation is that the QCM asks about current motivational state in a general manner. A measurement method closer to the test situation is to ask test takers how they estimate their current motivation before and after taking an achievement test (e.g., Baumert and Demmrich, 2001; Eklöf, 2006).

Eklöf (2006) developed the Test-Taking Motivation Questionnaire (TTMQ), which includes motivational components in line with the expectancy-value theory of achievement motivation (e.g., Wigfield and Eccles, 2000). The relations between these components and test performance were at low to moderate levels, indicating inconsistent findings (Wise and DeMars, 2005; Knekta and Eklöf, 2015; Penk and Schipolowski, 2015; Penk and Richter, 2017; Stenlund et al., 2018). Moreover, Penk and Richter (2017) identified changes in the motivational component of test-taking effort during test taking. Test takers' self-reported test-taking effort decreased from the beginning to the end of the test in this study (Penk and Richter, 2017) and in other studies (Attali, 2016;

Lindner et al., 2018). Test takers may easily recognize that the TTMQ items are intended to capture their motivational state and might thus respond in socially desirable ways. Hence, it is valuable to increase the validity of the TTMQ results by employing less subjective measures (AERA et al., 2014).

### Time on Task and Response Times as Indicators of Test-Takers' Motivation

Test-taking effort has been investigated by different measures, for example, response times (Wise and Kong, 2005), time on task (Attali, 2016), or self-reports (Knekta and Eklöf, 2015). A study compared test-taking effort (measured by time on task) and performance in a high-stakes achievement test vs. subsequent low-stakes achievement test with the result that the majority of test takers replicated their high stakes performance in the low-stakes condition with little effort (Attali, 2016).

Some researchers have used response times to test the assumption of low test-taking motivation reflected in low effort (Wise and Kong, 2005; Hartig and Buchholz, 2012; Debeer et al., 2014; Rios et al., 2014), examining persistence levels in terms of response times on puzzle tasks, or response times on anagram tasks (e.g., Gignac and Wong, 2018). Other studies included changes in response times over the course of an achievement test as indicators for test-taking motivation (Hartig and Buchholz, 2012; Goldhammer et al., 2017). Meta-analytic results suggested higher correlations between test-taking response time effort and test performance than self-reported effort assessed mainly by the Student Opinion Scale (Sundre and Moore, 2002) and test performance (Silm et al., 2020). Test-taking effort estimated using response times decreased over the course of test taking in these studies (Hartig and Buchholz, 2012; Debeer et al., 2014).

In summary, changes in self-reported effort over the course of test taking suggest decreased effort, which raises the question of potential strategies for keeping test-taking effort levels. The TTMQ, based on expectancy-value theory, captures current test-taking motivation (state), and is a widely used measure in large-scale assessments including students at school. Researchers examined and proposed strategies with the intention to increase German school students' test-taking motivation but examined relatively seldom changes in test-taking motivation or strategies to keep the level of test-taking motivation in education students in Germany (Silm et al., 2020). Based on expectancy-value theory and above-mentioned evidence, we focused on two motivational components among test takers: (1) the test-taking effort invested and (2) the subjective test-taking importance of the respective task (value component), while also considering the other components that are expectancies, concentration, and anxiety, as well as gender and age, as described below in the method section. Test-taking effort and importance are probably at higher levels when test takers are working on mock exam tasks than on cognitive ability tasks.

## THE PRESENT RESEARCH

We aimed to extend the findings on changes in test-taking motivation presented in the previous section (Baumert and

Demmrich, 2001; Debeer et al., 2014; Bensley et al., 2016; Knekta, 2017; Penk and Richter, 2017) by employing a computer-assisted experimental design with repeated motivational measures (test-taking effort, test-taking importance) in order to examine changes in these motivational components over experimental variations in task type order, and whether achievement related choices, information processing and response processes are affected by the electronically varied task type order. The purpose was to obtain new insights into possible changes in test-taking effort and test-taking importance across variations in task type order. Test-taking effort and importance were assessed before and after a computer-assisted cognitive ability test and mock exam to obtain insights into intraindividual changes in effort over test taking in a new context (i.e., education students in a computer-assisted environment in higher education) using different measures than in previous studies (e.g., Freund et al., 2011). Moreover, finding different levels of test-taking effort and importance in these conditions would conceptually replicate findings from previous studies on test-taking motivation in other contexts (e.g., Eklöf, 2006; Knekta, 2017). This would extend the validity of test-taking effort and/or importance scores to further test conditions and samples (Knekta and Eklöf, 2015; Penk and Schipolowski, 2015; Knekta, 2017).

Our hypotheses were as follows: (1) Test-taking effort and test-taking importance decrease across three measurement points during the test situation moderated by task type order (first cognitive ability tasks, second mock exam tasks vs. first mock exam tasks, second cognitive ability tasks) and with consideration of the five relevant covariates test expectancies, test anxiety, concentration, gender, and age. (2) Response processes on the ability tests differ depending on the task type order (first cognitive ability tasks, second mock exam tasks vs. first mock exam tasks, second cognitive ability tasks). We included the five relevant covariates in the analyses with regard to Hypothesis 1 since they are considered in the theoretical model (see **Figure 1**), previous research suggested them as relevant covariates as introduced above (DeMars and Bashkov, 2013; Rios and Liu, 2017; Silm et al., 2020), and covariates are commonly included into experimental designs to reduce variance for increasing statistical power.

To examine our assumptions, we adapted and used measures from previous research (Arvey et al., 1990; Butler and Adams, 2007; Erle and Topolinski, 2015; Knekta and Eklöf, 2015), with the exception of the mock exam tasks. We used items from the Test-Taking Motivation Questionnaire (TTMQ) developed by Eklöf (2006) that has previously been employed in large-scale surveys (e.g., Knekta and Eklöf, 2015), cognitive ability tasks (e.g., Erle and Topolinski, 2015; 10 further tasks for other research purposes, McHugh et al., 2004), and mock exam tasks in the two test order conditions. Similar to other researchers, we analyzed the changes in test-taking motivation over the course of an exam by structural equations, in particular, latent growth curve modeling (e.g., Penk and Richter, 2017). The term "latent" refers to constructs or processes which are not observable. The advantage of measuring factors and their relationships at a latent level is that measurement errors have been separated out.

We additionally analyzed the responses and the response times of the test takers in the cognitive ability tasks and the



mock exam tasks with a psychometric diffusion model. The psychometric diffusion model is capable to separate motivational parts from achievement parts of a test taker's test performance. This provides a more objective basis for the analysis of test takers' motivation. Psychometric diffusion modeling for these tasks has not yet been undertaken in previously published work. Hence, the current study, with its experimental design, extends previous research on test-taking effort with regard to response processes for different task types.

## MATERIALS AND METHODS

### Participants

The current study involved  $N = 320$  undergraduate education students (77% female,  $M_{\text{age}} = 21$ ,  $SD_{\text{age}} = 3.13$  at T1, seven missing values in gender, one missing value in age) who voluntarily attended an electronic mock exam at a German University. The sample size is sufficient for detecting moderate group differences and changes using latent growth curve modeling as simulation studies suggested (Fan, 2003). The electronic mock exam included questions concerning test-taking motivation (presented up to three times), cognitive ability tasks (less personally important tasks), mock exam tasks (personally important tasks), and demographic questions. The mock exam was computerized using the software package PsychoPy (Peirce et al., 2019) and presented on laptops in an e-exam hall. Each undergraduate student used one laptop on a desk with sight protection.

After welcoming, one of three supervisors read a standardized oral instruction in German aloud for the participants. For example, the instruction involved (in English for the current purposes): "We offer the mock exam for the first time and would like to know how you like it. Therefore, other tasks and a few questions about your motivation are included in addition to the mock exam tasks. Please answer all tasks and questions conscientiously so that the results are meaningful." The participants further received the information that they may expect 20 mock exam tasks. They could individually decide when to finish a mock exam task and proceed with the next one. There was no time limit. **Figure 2** presents the study design. All measures, task descriptions, and tasks were implemented in the programmed experiment using PsychoPy.

The data collection was completely anonymized by assigning the participants electronically generated IDs. There was no deception. All steps of the study followed international ethical standards (AERA et al., 2014). Data from 11 participants were invalid due to technical problems, such as system aborts, and had to be excluded. Thirty-four undergraduate students participated in interventions for other research purposes than presented here, leaving data from  $n = 275$  participants remaining for current analyses.

### Measures

The motivational measures employed had already been used in international large-scale surveys (e.g., Arvey et al., 1990; Knekta and Eklöf, 2015). To test the theoretical model introduced (see **Figure 1**) with the focus on education students' test-taking

effort and importance, we adapted some items to the current study as detailed below. Test-takers' expectancies, test anxiety, and concentration are included as covariates and assessed once (interest for other research purposes than presented here). Test-taking effort and importance are assessed three times (see Motivation T1, T2, and T3, in **Figure 2**). Measures only available in English were translated into German using standard cross-translation procedures. All items were presented in German during the mock exam but example items will be translated to English here. Participants indicated their concentration, expectancies, test anxiety, test-taking effort, and importance on rating scales ranging from  $-1.5$  (*strongly disagree*) to  $1.5$  (*strongly agree*).

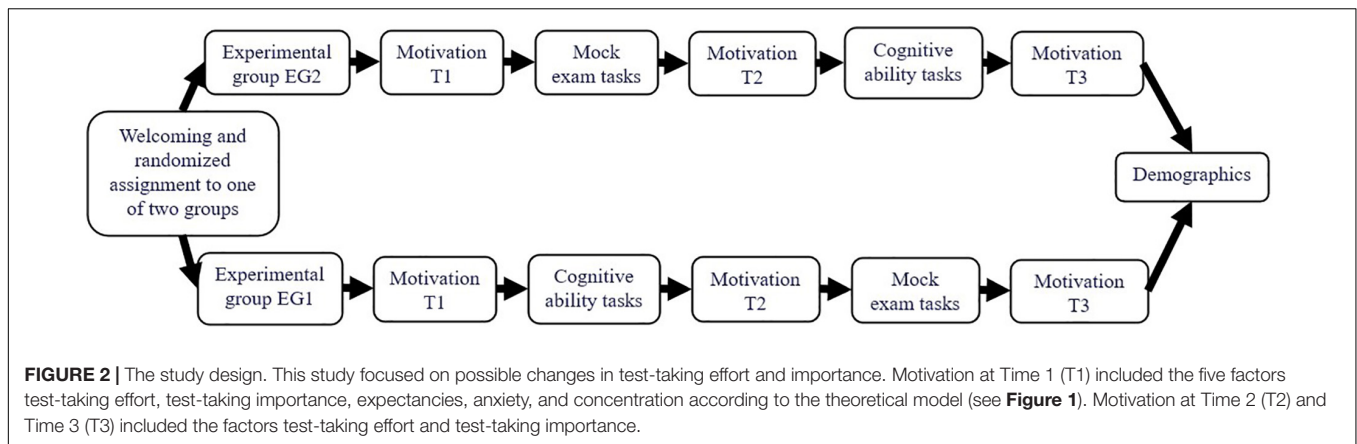
We used McDonald's  $\omega$ , instead of Cronbach's  $\alpha$ , to estimate the internal consistency of test-taking motivation and each of its dimensions test-taking effort, test-taking importance, expectancies, anxiety, and concentration simultaneously (Dunn et al., 2014). For example, Dunn et al. (2014) argued for McDonald's  $\omega$  since it is a point estimate that makes few and realistic assumptions, requires congeneric variables rather than tau-equivalent variables (Zinbarg, 2006; Revelle and Zinbarg, 2008; Hayes et al., 2020). Furthermore, inflation and attenuation of internal consistency estimation are less likely (see Dunn et al., 2014, for further advantages over Cronbach's  $\alpha$ ). McDonald's coefficient can be calculated within the R environment (R Development Core Team, 2009) using the R package *psych* (Revelle, 2019) and interpreted by the same levels as Cronbach's  $\alpha$  (Schweizer, 2011). Note the increasing number of publications about Cronbach's  $\alpha$  vs. McDonald's  $\omega$  which consistently suggest McDonald's  $\omega$  (Zinbarg, 2006; Revelle and Zinbarg, 2008; Hayes et al., 2020).

### Motivational Factors

*Test-taking effort* (Knekta and Eklöf, 2015) with regard to the current test situation was measured three times (T1–3) during the mock exam with five items: in the baseline assessment (T1), after the first task battery, and after the second task battery (T3). An example item is "I am doing my best on these tasks." McDonald's  $\omega_{\text{total}} = 0.95$  suggested good internal consistency for the three-factor solution and each factor (T1  $\omega = 0.87$ , T2  $\omega = 0.89$ , T3  $\omega = 0.89$ ). Subsequently, items assessing test-taking importance were presented.

*Test-taking importance* (Knekta and Eklöf, 2015) was measured three times (T1–3) with the same three items: in the baseline assessment (T1), after the first task battery and test-taking effort items as well as after the second task battery and effort items (T3). An example item is "The tasks are important to me." McDonald's  $\omega_{\text{total}} = 0.93$  suggested good internal consistency for the three-factor solution as well as the factors test-taking importance at T1 and T3 each except T2 with only acceptable internal consistency (T1  $\omega = 0.87$ , T2  $\omega = 0.60$ , T3  $\omega = 0.97$ ).

Moreover, McDonald's  $\omega_{\text{total}} = 0.87$  suggested good internal consistency for the motivational five-factors solution incl. expectancies ( $\omega = 0.63$ ), anxiety ( $\omega = 0.74$ ), concentration ( $\omega = 0.69$ ), test-taking effort ( $\omega = 0.87$ ), test-taking importance ( $\omega = 0.87$ ) at T1 and acceptable internal consistency of these



factors each. These five motivational variables were included according to the introduced theoretical model (see **Figure 1**). Subsequent to the test-taking importance items, *expectancies* (Knehta and Eklöf, 2015) were assessed with three items adapted to the current study (T1). An example item is “Compared with other students, I think I am doing well on the tasks.” *Test anxiety* was assessed with three items and presented before the first set of tasks. An example item is “I am so nervous when I take the tasks that I forget things I usually know” (adapted from Knehta and Eklöf, 2015, p. 666). *Concentration* (Arvey et al., 1990) was assessed with four items at the end of the baseline assessment (T1). An example item is “It is hard to keep my mind on this test.” Expectancies, anxiety, and concentration were included as manifest covariates only to consider their effects on the criterion variables test-taking effort and test-taking importance at T3 since the theoretical model and previous findings suggested such relations (Knehta and Eklöf, 2015; Silm et al., 2020).

### Cognitive Ability Tasks and Mock Exam Tasks

Pioneers of psychology already tested and described cognitive abilities such as perception (James, 1884), reasoning (Piaget, 1928), and visuo-spatial perspective-taking (Flavell et al., 1978). Since perspective-taking is highly important for education students’ social interactions with children, adolescents, and adults (Wolgast et al., 2019), we chose proven cognitive ability tasks as typical tasks for research purposes in psychology. These *cognitive ability tasks* were considered as personally less important low-stakes tasks because they were not part of the lecture or module curriculum and irrelevant for the exam the students had to take in order to finish the course. Sixteen tasks assessed the cognitive ability *visuo-spatial social perspective-taking* that is seeing what another person sees by putting oneself mentally in the target’s spatial position (Kessler and Thomson, 2010; Erle and Topolinski, 2015).

Erle and Topolinski (2015) used the visuo-spatial perspective-taking paradigm developed by Kessler and Thomson (2010). Each of the first 16 tasks involved a photograph (with friendly permission from Thorsten M. Erle for using the photographs in further research). The photograph showed a female or male target person sitting at a round table (arms on the table) from a bird’s-eye perspective. A book and a banana lay on the table

close to the person’s left arm and right arm, respectively, or vice versa. The person’s position at the table rotated from photograph to photograph between 120, 160, 200, and 240° from the participant’s point of view. Previous research has found perspective-taking to be difficult at these angles (Janczyk, 2013). Each position was presented with a female target person in eight photographs and a male target person in further eight photographs (16 tasks). Participants indicated whether the book was lying closer to the target person’s left or right arm by pressing “n” (*left*) or “m” (*right*) on the keyboard (“n” is located to the left of “m” on German keyboards). There was no time limit. All cognitive ability tasks were presented in German. McDonald’s  $\omega = 0.98$  suggested almost perfect internal consistency.

Twenty single-choice *mock exam tasks* were developed to coincide with a lecture for undergraduate education students entitled “Educational Psychology.” McDonald’s  $\omega = 0.61$  suggested acceptable internal consistency for these tasks. The mock exam tasks were considered as individually important low-stakes achievement test because the students’ upcoming module exam consisted of tasks of this type with similar content. Hence, the undergraduates had the opportunity to practice this type of task in order to be well prepared for the module exam. An example mock exam task is “Which phenomenon related to a child’s reasoning did Piaget and colleagues investigate with the three-mountain task? (a) object permanence, (b) centering, (c) egocentrism, (d) logical contradictions.” The tasks were presented in German; the example has been translated into English for current purposes.

### Procedure

Participants were randomly assigned to two conditions: EG1 responded first to the cognitive ability tasks and then to the mock exam tasks, while the order was vice versa for EG2 (first mock exam tasks, then cognitive ability tasks). All participants had the opportunity to take the mock exam tasks and subsequently receive automatically generated feedback on how many tasks they solved. The respondents participated voluntarily and gave consent to analyze their data, which was anonymously collected. Taking the tests lasted less than 1 h in total, including initial instruction.

## Statistical Analyses

### Latent Growth Curve Modeling

We used latent growth curve modeling (R Development Core Team, 2009; Rosseel, 2010) and weighted least squares with mean and variance adjustment estimation (WLSMV) (Rosseel, 2019) to test for within-test-takers' changes and differences in responses on motivational items depending on condition. We set the significance-level at  $\alpha = 0.05$ . The variables included in the modeling were grand mean centered.

First, we conducted a confirmatory factor analysis (CFA) and tested the theoretical six-factor model (test-taking effort and importance at T1, T2, T3) by the data. The unstandardized effort factor-loading of the fourth item (Item "E4," Knektá and Eklöf, 2015, p. 666, adapted to university: *I spent more effort on this test than I do on other tests we have in university.*) was  $\lambda = 0.21$  and statistically not significant with  $p = 0.15$  in the EG1. Consequently, we excluded Item E4 from further analyses. The final two factor CFA model included the latent factor *test-taking effort* measured by the respective four items and their residuals at T1, T2, and T3, and the latent factor *test-taking importance* measured by the respective three items and their residuals over the three measurement points. Scalar measurement invariance is a prerequisite for latent growth curve modeling. Measurement invariance was tested using the two factor CFA model in a multi-group analysis across groups and time points. This CFA model including constrained factor loadings suggested scalar invariance, Delta Comparative Fit Index ( $\Delta$  CFI) = 0.004; Delta Root Mean Square Error of Approximation ( $\Delta$  RMSEA) = 0.003, according to recommended cutoffs (Hu and Bentler, 1999; Svetina and Rutkowski, 2014). The factor structure and intercepts found for EG1's data were equivalent to the factor structure and intercepts found for the EG2's data and at T1, T2, and T3. This CFA model is depicted in the **Supplementary Figure S1**. A simplified version of the second order latent growth curve model for the analysis of individual within-test-takers' change in test-taking effort and importance is depicted in **Figure 3** (see **Supplementary Figure S2** for a technical version). The second order latent growth curve model was specified including random intercepts and random slopes by extending the CFA model as follows: At first order latent level, the variance of the factors *test-taking effort* and *importance* each at T1, T2, and T3 has been constrained to the same value for compound symmetry covariance structure (Rosseel, 2019, 2020). At second order latent level, *test-taking effort intercept* has been specified with the three latent factors *test-taking effort* at T1, T2, and T3 with each path fixed to one. The latent factor *test-taking effort slope* has been specified with these three factors and the paths fixed to 0, 1, 2 respectively. The means of expectancies, anxiety, and concentration from the baseline assessment as well as gender and age were included to predict the factor *effort intercept* because these covariates should explain different *effort intercepts* between the education students. The covariates' category each existed before the study such as gender and age or were assessed before assessing test-taking effort and importance. Gender and age are included as covariates in SEM (Mutz and Pemantle, 2015) since previous findings consistently suggested their relations to test-taking motivation in educational contexts (e.g., DeMars and Bashkov, 2013; Silm et al., 2020). The

covariates are included in SEM to consider their anticipated effects on the criterion variables test-taking effort and importance at T3 (see Mutz and Pemantle, 2015, for standards in experimental research).

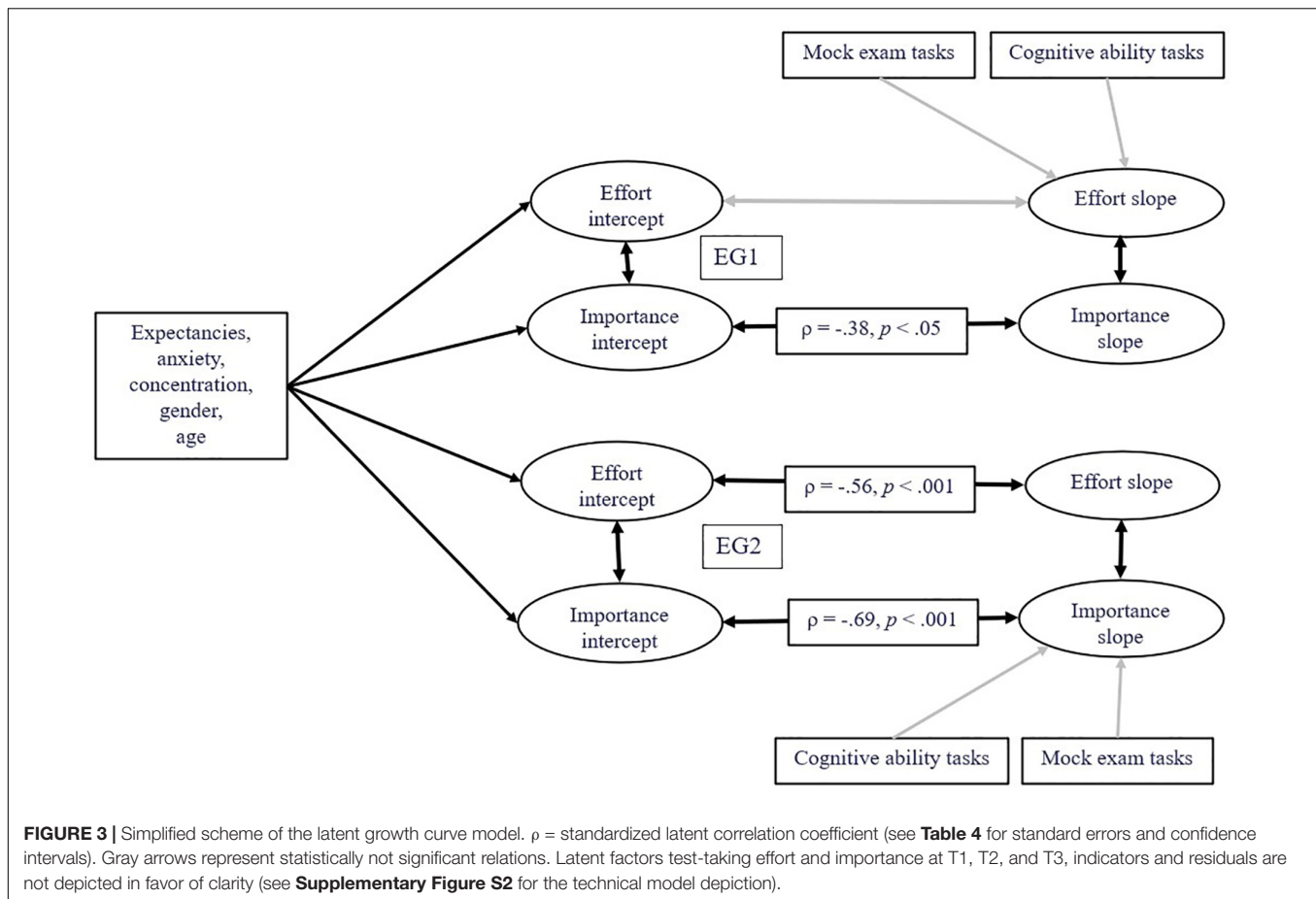
"A mean structure is automatically assumed, and the observed intercepts are fixed to zero by default, while the latent variable intercepts/means are freely estimated" (Rosseel, 2020, p. 28). The sum scores of the cognitive ability tasks and exam tasks each were included to predict *the test-taking effort slope* (instead of *test-taking effort intercept*) because we experimentally varied the exam tasks' position. We used the means and sum scores of predictor variables instead of measuring latent factors to keep the number of parameters as low as possible. Random intercepts and slopes for each latent factor were specified with correlations to themselves and to each other. The second-order latent factors *test-taking importance intercept* and *test-taking importance slope* were analogously specified and the same covariates included. Correlations between some test-taking effort indicators were allowed according to modification indices (see **Supplementary Figure S2**).

The model specification considered EG1's and EG2's data separately, so EG1's data were analyzed without EG2's data and vice versa. Criterion variables were *latent test-taking effort intercept* and *slope* as well *latent test-taking importance intercept* and *slope*.

### Latent Diffusion Modeling

To analyze participants' response processes, we included the responses and response times in a latent trait diffusion model. The diffusion model allows researchers to examine response process components in binary decision tasks (Voss et al., 2013). Binary decision tasks are, for example, the presented cognitive ability tasks where test takers have to choose one of two response options or the questions in the mock exam where the test takers have to decide between the correct and the incorrect response (Molenaar et al., 2015). The diffusion model is based on the assumption that test takers continuously accumulate evidence for the two response options. A momentary preference is formed by weighting the evidence for the two response options against each other. As soon as the momentary preference exceeds a critical level, the test taker responds by selecting the more preferred option.

Diffusion modeling involves defining three process parameters. (1) Information accumulation is a measure of one's mental simulation of two possible outcomes using the available information (*drift rate*,  $\nu$ ). (2) The amount of information required for a response is reflected in the decision threshold (*boundary separation parameter*,  $a$ ). (3) The response time includes time for reactions (e.g., moving one's finger to the keyboard in a computer-based task) and/or other sensory, mental or motor responses aside from the time needed to make a decision (*non-decision time parameter*,  $t_{er}$ ) (Ratcliff and McKoon, 2008; Voss et al., 2013). The drift rate ( $\nu$ ) provides insights into information uptake latency, with high uptake speed reflecting high performance, and is a manifestation of a test taker's capability. The lower the drift rate, the more difficult the task is in relation to a given individual latent trait (e.g., ability



or attitude, see Voss et al., 2013). Low drift rates are reflected in low response accuracy and long response times. The boundary separation reflects the response caution. It is assumed to be a manifestation of a test taker's effort or importance, reflecting their carefulness when responding. Low levels of the boundary separation are reflected in low response accuracy and short response times, two typical signs of rapid guessing.

In addition to the parameters in manifest diffusion modeling, the psychometric diffusion modeling under the item response theory allows to estimate the latent person contribution and task contribution to the response process. The person contribution refers to information processing (latent trait  $\theta$ ) and response caution (latent trait  $\omega$ ) of a person as well as investigate relationships between these latent traits ( $\theta$ ,  $\omega$ ) and constructs such as test-taking effort or importance. The task contribution refers to the task difficulty.

Two model types are distinguished, the D-diffusion (Tuerlinckx and De Boeck, 2005) and the Q-diffusion model (Van der Maas et al., 2011). They basically differ in their parameterization. In the D-diffusion model, the effective drift rate is the difference of the latent trait of a test taker and the corresponding intercept. This parameterization allows task probabilities from zero to one. In the Q-diffusion model, the effective drift rate is the quotient of latent ability and

the corresponding intercept. The Q-diffusion model requires task solving probabilities of at least 50% for calculating the diffusion parameters. In case of solving probabilities lower than 50%, the D-diffusion model can be used with consideration of its predictions.

We applied diffusion modeling within the R environment (R Development Core Team, 2009) using the R package *diffIRT* (Molenaar et al., 2015). The non-decision time (*ter*) was constrained to control delays resulting from the different laptops we used on the non-response times. The R code can be obtained from the authors.

## RESULTS

Descriptive results are summarized in **Tables 1A,B**. Product-moment correlations between the variables used are detailed in **Table 2**. The correlation coefficients suggest zero to low not significant correlations of test-taking effort (T1, T2, T3), test-taking importance (T1, T2, T3), expectancies, anxiety, and concentration with the cognitive ability tasks, and mock exam tasks. Means and standard errors of test takers' responses on test-taking effort and importance items are depicted in **Figure 4**. These line diagrams suggested changes in education students' test-taking effort and importance. For examining these changes at



latent level and with consideration of the covariates *expectancies*, *anxiety*, and *concentration*, we investigated within test-taker effects by structural equations and diffusion modeling.

## Within Test-Taker Effects

First, we employed latent growth curve modeling to disclose changes in education students' test taking *effort* and *importance* over T1, T2, and T3 moderated by condition (EG1: cognitive ability tasks first vs. EG2: mock exam tasks first). The simplified model structure is depicted in **Figure 3** (without depiction of residuals and indicators). The goodness of fit between the theoretical model and data was good (Hu and Bentler, 1999; Svetina and Rutkowski, 2014),  $\chi^2(641) = 731.04$ ,  $p = 0.008$ , CFI = 0.99, RMSEA = 0.033, 95% CI [0.018, 0.044], SRMR = 0.078 (WLSMV-estimation).

**Figure 3** provides results such as standardized latent correlation coefficients and statistical significance levels.

In **Table 3**, these standardized latent correlation coefficients are presented with standard errors, significance levels and confidence intervals each. **Table 3** provides furthermore variances of effort intercept, effort slope, importance intercept, and importance slope. The results suggested significantly decreased test-taking effort ( $\rho = -0.56$ ,  $p < 0.001$ ) and importance ( $\rho = -0.69$ ,  $p < 0.001$ ) in EG2 over test-taking time supporting Hypothesis 1.

However, the EG1's test-taking effort did not decrease over time ( $\rho = 0.63$ ,  $p = 0.27$ ), only their test-taking importance decreased ( $\rho = -0.38$ ,  $p < 0.05$ ) but less than the EG2's test-taking importance. No significant relations existed between cognitive

ability task performance or mock exam task performance and the factors *test-taking effort slope*, and *test-taking importance slope*.

This model explained 28% of variance in the latent factor effort intercept, 30% in the latent factor effort slope, 83% of variance in the latent factor test-taking effort at T3, 22% of variance in the latent factor importance intercept, 1% in the latent factor importance slope, and 83% in the latent factor importance at T3 in the EG1. In the EG2, this model explained 19% of variance in the latent factor effort intercept, 0.4% in the latent factor effort slope, 86% of variance in the latent factor test-taking effort at T3, 13% of variance in the latent factor importance intercept, 0.2% in the latent factor importance slope, and 84% in the latent factor importance at T3.

The theoretical model adapted to the current study (see **Figure 1**) implies that achievement related choices can involve decisions in response processes on tasks. We examined the EG1's vs. EG2's responses and response times on the cognitive ability tasks and mock exam tasks by diffusion modeling as described next.

## Education Students' Response Processes for the Tasks

The responses for the cognitive ability tasks as well as response times were included in a Q-diffusion model to analyze the achievement related choices and response processes according to the theoretical model in **Figure 1**. We investigated the goodness of fit between the theoretical and observed response time distribution with QQ-plots which suggested good fit for both groups (see **Supplementary Figure S3** for examples). The average intercept of the boundary separation and the average intercept of the drift rate over the items are summarized in **Table 4** for the EG1 and the EG2. A Wald test for the equivalence of the boundary intercepts in the two groups was significant ( $X^2 = 42.00$ ,  $df = 16$ ,  $p < 0.01$ ). A *post hoc* comparison of the intercepts in the single items revealed that parameters deviated in two of the 16 items on  $\alpha = 0.05$ . The corresponding Wald test for the equivalence of the drift intercepts was also significant ( $X^2 = 37.83$ ,  $df = 16$ ,  $p < 0.01$ ). The drift rates differed in three of the 16 items on  $\alpha = 0.05$ . This implies that neither the average response caution, usually considered as a motivational aspect of the response process, nor the average rate of information accumulation, usually considered as an aspect of a test taker's performance, differed between the groups in most items.

Investigating the D-diffusion model response fit using QQ-plots for the 20 exam tasks suggested acceptable fit between expected and observed distributions in both groups (see **Supplementary Figure S3** for examples). A Wald test for the equivalence of the boundary intercepts in the two groups was significant ( $X^2 = 33.90$ ,  $df = 20$ ,  $p = 0.03$ ). A *post hoc* comparison of the intercepts in the single items revealed that parameters deviated in 5 of the 20 items on  $\alpha = 0.05$ . The corresponding Wald test for the equivalence of the drift intercepts was insignificant ( $X^2 = 22.81$ ,  $df = 20$ ,  $p = 0.29$ ).

We included the latent information processing  $\omega$  (speed) and response caution  $\theta$  (trait) from diffusion modeling as criterion variables in general linear modeling to examine whether they

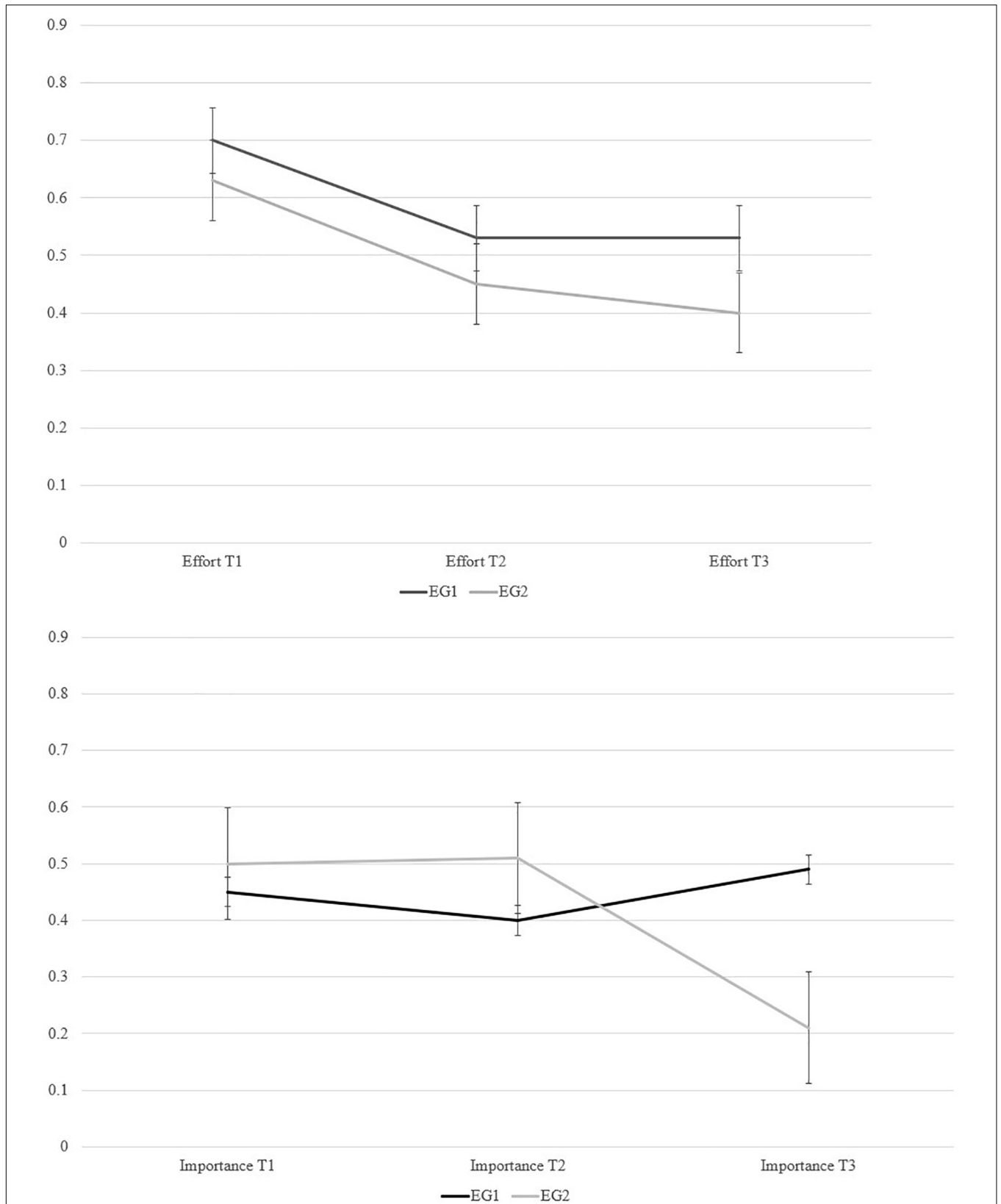
**TABLE 1A |** Means and standard deviations of the motivational components test-taking effort and test-taking importance in the EG1 ( $n = 125$ ) and EG2 ( $n = 150$ ) at T1, T2, and T3.

	Time 1		Time 2		Time 3	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>EG1</b>						
Effort <sup>a</sup>	0.70	0.38	0.53	0.45	0.53	0.40
Importance <sup>a</sup>	0.45	0.49	0.40	0.54	0.49	0.44
<b>EG2</b>						
Effort <sup>a</sup>	0.63	0.37	0.45	0.42	0.16	0.57
Importance <sup>a</sup>	0.50	0.47	0.51	0.44	0.21	0.58

<sup>a</sup>Test-taking effort and test-taking importance means of test takers' responses ( $-1.5 =$  strongly disagree,  $-0.5 =$  disagree,  $0.5 =$  agree,  $1.5 =$  strongly agree).

**TABLE 1B |** Mean accuracy and response times on cognitive ability tasks and mock-exam tasks in Experimental Group EG1 ( $n = 125$ ) vs. EG2 ( $n = 150$ ).

	Cognitive ability tasks		Mock exam tasks	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>EG1</b>				
Mean accuracy	94%	17%	57%	88%
Mean rt (s)	33.63	13.62	607.65	163.53
<b>EG2</b>				
Mean accuracy	92%	19%	54%	91%
Mean rt (s)	33.97	15.88	692.37	207.04



**FIGURE 4 |** Line diagrams of changes in test-taking effort (above) and test-taking importance (below), means of test takers' responses (-1.5 = *strongly disagree*, -0.5 = *disagree*, 0.5 = *agree*, 1.5 = *strongly agree*, error bars represent standard errors).

**TABLE 2 |** Correlations among test-taking effort at T1–3, test-taking importance at T1–3, expectancies, test-taking anxiety, concentration, cognitive ability tasks, mock exam tasks, and age in Experimental Group EG1 vs. EG2.

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1 Effort T1		<b>0.89</b>	<b>0.56</b>	<b>0.79</b>	<b>0.77</b>	0.48	<b>-0.64</b>	0.08	0.13	<0.01	0.04	-0.21
2 Effort T2	<b>0.73</b>		<b>0.68</b>	<b>0.80</b>	<b>0.86</b>	<b>0.59</b>	<b>-0.79</b>	0.02	0.13	-0.02	0.05	-0.22
3 Effort T3	<b>0.82</b>	<b>0.83</b>		0.52	<b>0.59</b>	<b>0.91</b>	<b>-0.64</b>	-0.27	0.27	-0.16	-0.28	-0.25
4 Importance T1	<b>0.70</b>	<b>0.84</b>	<b>0.81</b>		<b>0.88</b>	0.54	<b>-0.64</b>	-0.04	0.07	0.04	0.07	-0.05
5 Importance T2	0.54	<b>0.83</b>	<b>0.78</b>	<b>0.94</b>		<b>0.65</b>	<b>-0.71</b>	0.01	0.13	0.07	-0.02	-0.25
6 Importance T3	<b>0.65</b>	<b>0.79</b>	<b>0.88</b>	<b>0.92</b>	<b>0.93</b>		<b>-0.64</b>	-0.18	0.14	-0.14	-0.22	-0.26
7 Concentration	<b>-0.68</b>	<b>-0.58</b>	<b>-0.74</b>	<b>-0.58</b>	-0.48	<b>-0.61</b>		-0.11	0.04	-0.15	-0.18	0.19
8 Expectancies	0.18	0.15	0.06	0.17	0.11	0.12	-0.23		-0.41	-0.20	0.28	0.10
9 Anxiety	-0.02	0.12	0.07	0.31	0.32	0.26	0.15	-0.47		-0.09	-0.48	-0.33
10 Cogn. ability	-0.31	-0.35	-0.19	-0.44	-0.37	-0.35	-0.03	-0.39	-0.35		-0.19	-0.26
11 Mock exam	0.09	-0.12	-0.04	-0.10	-0.22	-0.16	-0.30	0.43	-0.47	-0.18		-0.14
12 Age	-0.09	-0.25	-0.13	-0.25	-0.31	-0.17	-0.26	-0.14	-0.08	0.09	0.06	

EG1 (n = 125) below the diagonal EG2 (n = 150) above the diagonal, p < 0.05 in bold. Cogn. Ability, cognitive ability tasks; T1, Time 1.

related to the condition (i.e., cognitive ability tasks or mock exam tasks first), baseline variables (i.e., expectancies, anxiety, concentration, test-taking effort, test-taking importance) gender, and age. For the cognitive ability tasks, the latent information processing  $\omega$  and response caution  $\theta$  did not relate to the baseline variables expectancies, anxiety, concentration, test-taking effort, test-taking importance, gender, and age with all eight regression coefficients close to zero ( $B_\omega = -0.04$  to  $0.01$ ,  $B_\theta = -0.01$  to  $0.02$ ) except condition that related to response caution ( $B_\theta = 0.15$ ,  $p = 0.05$ ) with higher response caution in the EG1 than EG2.

Including latent information processing  $\omega$  and response caution  $\theta$  with the baseline variables into the model to predict test-taking effort and importance at T3 yielded significant relations between concentration ( $B_{ei} = -0.32$ ,  $p < 0.001$ ), test-taking effort ( $B_{ei} = 0.48$ ,  $p < 0.001$ ), and importance ( $B_{ei} = 0.74$ ,  $p < 0.001$ ) at T1 as well as condition ( $B_{ei} = 2.80$ ,  $p < 0.001$ ) to test-taking effort and importance at T3 with EG1 having an advantage. While concentration levels were higher in the EG2 than EG1, test-taking effort and importance levels were higher in the EG1 than EG2. This model explained 36% of variance in test-taking effort and importance,  $F(10, 252) = 15.82$ ,  $p < 0.001$ .

**TABLE 3A |** Standardized latent regression coefficients and correlation coefficients, standard errors, confidence intervals, from latent growth curve modeling with intercepts and slopes of test-taking effort and test-taking importance in Experimental Group EG1 and EG2.

EG1, n = 125	$\beta$	SE	p	CI <sub>lower</sub>	CI <sub>upper</sub>
Effort slope regressed on					
Cognitive ability tasks	0.53	0.38	0.86	-0.51	0.57
Exam tasks	-0.10	0.66	0.88	-0.39	0.19
Importance slope regressed on					
Cognitive ability tasks	0.07	0.08	0.44	-0.10	0.23
Exam tasks	-0.05	0.14	0.72	-0.33	0.23
<b>EG2, n = 150</b>					
Effort slope regressed on					
Cognitive ability tasks	0.03	0.07	0.68	-0.11	0.17
Exam tasks	-0.05	0.09	0.54	-0.23	0.12
Importance slope regressed on					
Cognitive ability tasks	0.03	0.07	0.69	-0.11	0.16
Exam tasks	-0.03	0.09	0.76	-0.20	0.14
<b>EG1</b>	$\rho$	SE	p	CI <sub>lower</sub>	CI <sub>upper</sub>
Effort intercept~slope	0.63	0.57	0.27	-0.48	1.74
Importance intercept~slope	-0.38	0.18	0.03	-0.73	-0.04
<b>EG2</b>					
Effort intercept~slope	-0.56	0.12	<0.001	-0.79	-0.33
Importance intercept~slope	-0.69	0.12	<0.001	-0.91	-0.46

**TABLE 3B |** Effort intercept, effort slope, importance intercept, and importance slope: Variances at latent level, and explained variances from latent growth modeling.

EG1	Va	SE	p
Effort intercept variance	0.27	0.14	0.046
Effort slope variance	0.01	0.08	0.95
Importance intercept variance	0.41	0.10	<0.001
Importance slope variance	0.07	0.03	0.04
<b>EG2</b>			
Effort intercept variance	0.39	0.10	<0.001
Effort slope variance	0.12	0.03	<0.001
Importance intercept variance	0.54	0.12	<0.001
Importance slope variance	0.12	0.04	0.001
<b>EG1</b>	<b>R<sup>2</sup></b>		
Effort intercept	0.28		
Effort slope	0.30		
Importance intercept	0.22		
Importance slope	0.01		
<b>EG2</b>			
Effort intercept	0.19		
Effort slope	0.004		
Importance intercept	0.13		
Importance slope	0.002		

**TABLE 4 |** Mean boundary separations ( $a$ ), mean drift rates ( $v$ ), and standard deviations from diffusion modeling including the cognitive ability tasks or mock exam tasks in Experimental Group EG1 vs. EG2.

Tasks	Parameter	EG1		EG2	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cognitive ability	$a$	0.27	0.03	0.25	0.02
	$v$	0.95	0.14	1.04	0.15
Exam	$a$	0.10	0.02	0.09	0.02
	$v$	-0.04	0.10	-0.03	0.09

$a$ , boundary separation,  $v$ , drift rate,  $t_{er}$  was constrained to same values.  $Q$ -diffusion model has been used for the cognitive ability tasks,  $D$ -diffusion model for the exam tasks due to solving probabilities below 50% which the diffIRT-function cannot handle.

For the mock exam tasks, the latent information processing  $\omega$  and response caution  $\theta$  did not relate to the baseline variables expectancies, anxiety, concentration, test-taking effort, test-taking importance, gender, and age with all eight regression coefficients close to zero ( $B_{\omega} = -0.02$  to  $0.01$ ,  $B_{\theta} = -0.05$  to  $0.01$ ). Including latent information processing  $\omega$  and response caution  $\theta$  with the baseline variables into the model to predict test-taking effort and importance at T3 yielded significant relations between concentration ( $B_{ei} = -0.33$ ,  $p = 0.002$ ), test-taking effort ( $B_{ei} = 0.47$ ,  $p < 0.001$ ), and importance ( $B_{ei} = 0.74$ ,  $p < 0.001$ ) at T1 as well as condition ( $B_{ei} = 2.73$ ,  $p < 0.001$ ) to test-taking effort and importance at T3 with EG1 having an advantage. This model explained 36% of variance in test-taking effort and importance,  $F(10, 252) = 15.63$ ,  $p < 0.001$ . While concentration levels were higher in the EG2 than EG1, test-taking effort and importance levels were higher in the EG1 than EG2. This model explained 36% of variance in test-taking effort and importance at T3,  $F(10, 252) = 15.63$ ,  $p < 0.001$ .

## DISCUSSION AND CONCLUSION

Based on expectancy-value theory (Wigfield and Eccles, 2000; Knekt and Eklöf, 2015), the current study sought to examine (1) whether test-taking effort and test-taking importance decrease across three measurement points during the computer-assisted test situation moderated by test-battery order and with consideration of the five covariates test expectancies, test anxiety, concentration, gender, and age and (2) whether response processes on the ability tests and mock exam differ depending on the task type order (EG1: first cognitive ability tasks, second mock exam tasks vs. EG2: first mock exam tasks, second cognitive ability tasks). The response processes refer to the achievement related choices depicted in **Figure 1** and regard information processing. Thus, both hypotheses focus on the education students' test-taking behavior in a computer-assisted environment. Self-reported test-taking effort and importance provide subjective information about test-taking behavior in the computer-assisted environment. Information processing and response processes in tasks involve responses and response

times that provide rather objective information about test-taking behavior than self-reported test-taking effort and importance.

The results from latent growth curve modeling suggested that test-taking effort and importance in EG2 significantly decreased among the education students over different task type orders (first mock exam tasks, then cognitive ability tasks) in the computer-assisted environment. Test-taking effort significantly decreased almost linearly from T1 over the mock exam tasks and cognitive ability tasks to T3 in EG2. These declines are in accordance with Hypothesis 1. Test-taking importance significantly decreased in EG1 (moderate effect) and EG2 (strong effect) (Cohen, 1988). Previous findings suggested that test-taking effort and importance changed even over the course of low-stakes testing (e.g., Penk and Richter, 2017). The decline in test-taking effort and test-taking importance in EG2 conceptually replicates similar findings from previous studies including students in school and different tasks (e.g., Knekt, 2017; Penk and Richter, 2017).

However, test-taking effort in the current study did not decrease when cognitive ability tasks were presented first (EG1) in the computer-assisted environment. The results support Hypothesis 1 in part since test-taking effort and importance significantly decreased in EG2 with higher levels when working on mock exam tasks than on the subsequent cognitive ability tasks. Test-taking importance also decreased in EG1 but without an advantage for the mock exam tasks. EG1's not decreased test-taking effort contradicts Hypothesis 1. Note that test-taking effort and importance were assessed by computer-assisted self-report measures. Diffusion modeling allows more objective insights into response processes in the sense of achievement related choices and information processing while working on tasks.

For the mock exam tasks, boundary intercepts suggesting response caution  $\theta$  (latent trait and motivational aspect) were similar in both groups. This result implies similar motivational levels in both groups while working on the mock exam tasks. For the cognitive ability tasks, boundary intercepts suggesting response caution  $\theta$  significantly differed between the groups. Response caution  $\theta$  related to the condition with higher response caution in EG1 than EG2. Thus, EG1's motivational levels were higher than EG2's motivational levels from a more objective point of view than self-reports. This difference is in accordance with Hypothesis 2 and with the result that test-taking effort did not decrease in EG1.

The new finding here is that the education students invested similar test-taking effort in the cognitive ability tasks as in the subsequent mock exam tasks (EG1). In EG1, test-taking effort did not decrease over the task types. Latent diffusion modeling (Van der Maas et al., 2011; Voss et al., 2013) suggested similar response processes on mock exam tasks but differences in the response processes (boundary intercepts) on cognitive ability tasks suggesting higher objective motivational levels in EG1 than EG2.

Latent diffusion modeling has not been undertaken in previous research on test-taking motivation in low-stakes tests. The high accuracy on the computer-assisted cognitive ability tasks might be one explanation approach for the similar response processes on the computer-assisted mock exam tasks between conditions. Another explanation might be that the mock



exam tasks predominantly required recalling subject content knowledge of educational psychology and rarely knowledge transfer to educational practice. Changes in education students' test-taking effort might affect tasks in other computer-assisted environments than presented here which require knowledge transfer to contexts in practice because such transfer is known as cognitively difficult. Alternatively, the testing time of about 30 min was too short for a test-taking effort decline related to a cognitive performance decline.

We concluded from the results, the education students were able to keep their self-reported test-taking effort levels during computer-assisted cognitive tasks and subsequent computer-assisted mock exam tasks. Diffusion modeling suggested objectively measured higher motivational levels during the cognitive tasks when they were presented first (EG1) than when mock exam tasks were presented first (EG2). The education students were not able to keep their test-taking effort during the computer-assisted cognitive ability tasks following the computer-assisted mock exam tasks.

Weak or non-existing relations between the motivational components (assessed by self-reports) and performance in low-stakes achievement tests are already known from other studies that presented relations inconsistently at zero (Sundre and Kitsantas, 2004; Penk and Richter, 2017) to low levels (Knekt, 2017; Eklöf and Hopfenbeck, 2018; Myers and Finney, 2019). The weak or non-existing relations between the motivational components on the cognitive ability tasks and mock exam tasks might have resulted from low to moderate task difficulties which not require to be motivated for performing equally in both conditions.

## Limitations and Implications for Future Research

Participants in the current study were education students and a self-selected sample tested in an e-exam hall (one person per laptop); however, each participant was randomly assigned to one of two experimental conditions. Gender was not equally distributed in the study. The computer-assisted cognitive ability tasks were tasks for research purposes rather than widely used standardized inventories. The computer-assisted mock exam was developed based on the participating students' educational psychology curriculum, including somewhat broad fundamentals of cognitive psychology, developmental psychology, and social psychology. Consequently, the relatively low internal consistency measured by McDonald's  $\omega$  might reflect the curriculum's broad content. Despite these limitations, however, the present study contributes to the understanding of motivation-performance patterns during computer-assisted test taking in higher education. This finding might be also relevant for motivation, information processing and responses on online exam tasks and online self-assessment tasks. The described differences between the conditions might be considered in the development of new computer-assisted (online) task batteries for exams or self-assessments, especially their order.

Future research might include a within-subject design and computer-assisted (online) tests accompanied by measures

assessing test-taking effort and importance before and after the respective test (e.g., effort with regard to Test 1 assessed before Test 1 and subsequently to Test 1, then effort with regard to Test 2 assessed before Test 2 and subsequently to Test 2). It is important for further studies to examine test-taking effort and importance during different ability tasks, because responses on tasks other than those presented here may differently stimulate information processing and differently relate to test-taking effort and importance. Hence, future research should examine the relations between test-taking effort, test-taking importance, and responses on different computer-assisted (online) ability tasks to increase the validity of the presented results (AERA et al., 2014). The current study increased its validity by using test-taking-effort and importance measures, because the changes found support the hypothesis that states should be measured rather than traits (Eklöf, 2006; AERA et al., 2014).

The main contribution of the empirical work presented here is that test-taking effort and importance were assessed three times over an experimentally varied task battery order considering information processing and response processes in a computer-assisted environment in higher education. Roughly 20 years ago, Baumert and Demmrich (2001) presented insignificant findings on strategies to increase students' test-taking motivation in PISA. However, from this study's perspective, the more important question is how to keep test-taking effort and importance relatively stable and avoid declines, rather than discussing how to increase test-taking motivation, as has been the case in previous research (e.g., Baumert and Demmrich, 2001).

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: [https://osf.io/nmxq7/?view\\_only=f150703b9a664d648c772055aa3335b3](https://osf.io/nmxq7/?view_only=f150703b9a664d648c772055aa3335b3).

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AW, NS, and JR contributed in collaboration to the design, conception, and data collection of the study. NS wrote a test documentation including sample and measure descriptions as well as descriptive results presented in the current report. AW analyzed the data by structural equations and latent diffusion models, and wrote the first draft of the manuscript. JR reviewed the manuscript contributing significantly to improve it, in particular by also calculating diffusion models and writing parts of the

section about diffusion modeling. All authors contributed to the manuscript revision, reread and approved the current version.

## FUNDING

This research was funded by the German Federal Ministry of Education (Grant No. 01PL17065) including open access publication fees.

## ACKNOWLEDGMENTS

We thank Vivien Eichhoff for her support when collecting the data.

## REFERENCES

- AERA, APA, and NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arvey, R. D., Strickland, W., Drauden, G., and Martin, C. (1990). Motivational components of test taking. *Person. Psychol.* 43, 695–716. doi: 10.1111/j.1744-6570.1990.tb00679.x
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychol. Rev.* 64, 22–32. doi: 10.1037/h0043445
- Attali, Y. (2016). Effort in low-stakes assessments: What does it take to perform as well as in a high-stakes setting? *Edu. Psychol. Measur.* 76, 1045–1058. doi: 10.1177/0013164416634789
- Baumert, J., and Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *Eur. J. Psychol. Edu.* 16, 441–462. doi: 10.1007/BF03173192
- Bensley, D. A., Rainey, C., Murtagh, M. P., Flinn, J. A., Maschiochi, C., Bernhardt, P. C., et al. (2016). Closing the assessment loop on critical thinking: The challenges of multidimensional testing and low test-taking motivation. *Think. Skills Creat.* 21, 158–168. doi: 10.1016/j.tsc.2016.06.006
- Brunner, M., Artelt, C., Krauss, S., and Baumert, J. (2007). Coaching for the PISA test. *Lear. Instruct.* 17, 111–122. doi: 10.1016/j.learninstruc.2007.01.002
- Butler, J., and Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *J. Appl. Measur.* 8, 279–304.
- Cohen, J. (1988). Set correlation and contingency tables. *Appl. Psychol. Measur.* 12, 425–434. doi: 10.1177/014662168801200410
- Debeer, D., Buchholz, J., Hartig, J., and Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *J. Edu. Behav. Stat.* 39, 502–523. doi: 10.3102/1076998614558485
- DeMars, C. E., and Bashkov, B. M. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Res. Pract. Asses.* 8, 69–82.
- Dunn, T. J., Baguley, T., and Brunson, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046
- Eccles, J. S., and Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personal. Soc. Psychol. Bull.* 21, 215–225. doi: 10.1177/0146167295213003
- Eccles, J. S., Midgley, C., and Adler, T. (1984). Grade-related changes in the school environment. *Develop. Achiev. Motiv.* 3, 238–331.
- Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Edu. Psychol. Measur.* 66, 643–656. doi: 10.1177/0013164405278574
- Eklöf, H., and Hopfenbeck, T. N. (2018). "Self-reported effort and motivation in the PISA test," in *International Large-Scale Assessments in Education: Insider Research Perspectives*, ed. B. Maddox (London: Bloomsbury), 121–136.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.559683/full#supplementary-material>

**Supplementary Figure 1** | CFA-model including test-taking effort and test-taking importance at T1, T2, and T3.

**Supplementary Figure 2** | Latent growth curve model.

**Supplementary Figure 3** | QQ-plots to investigate goodness of fit in diffusion modeling including (A) cognitive ability tasks and (B) mock exam tasks.

**Supplementary Table 1** | Standardized regression coefficients, standard errors, confidence intervals, and correlations from latent growth curve modeling with intercepts and slopes of test-taking effort and test-taking importance, and covariates in both groups.

- Erle, T. M., and Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Soc. Cog.* 33, 187–210. doi: 10.1521/soco.2015.33.3.187
- Fan, X. (2003). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Struct. Equ. Modeling* 10, 380–400. doi: 10.1207/S15328007SEM1003
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Res. Rep. Ser.* 2015, 1–17. doi: 10.1002/ets2.12067
- Flavell, J. H., Omanson, R. C., and Latham, C. (1978). Solving spatial perspective-taking problems by rule versus computation: A developmental study. *Dev. Psychol.* 14, 462–473. doi: 10.1037/0012-1649.14.5.462
- Freund, P. A., and Holling, H. (2011). Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. *Personal. Individ. Differ.* 50, 723–728. doi: 10.1016/j.paid.2010.12.025
- Freund, P. A., Kuhn, J.-T., and Holling, H. (2011). Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personal. Individ. Differ.* 51, 629–634. doi: 10.1016/j.paid.2011.05.033
- Gignac, G. E., and Wong, K. K. (2018). A psychometric examination of the anagram persistence task: More than two unsolvable anagrams may not be better. *Assessment* 27, 1198–1212. doi: 10.1177/1073191118789260
- Goldhammer, F., Martens, T., and Lüdtke, O. (2017). Conditioning factors of test-taking engagement in PIAAC: an exploratory IRT modelling approach considering person and item characteristics. *Large Scale Asses. Edu.* 5:18. doi: 10.1186/s40536-017-0051-9
- Gorges, J., Maehler, D. B., Koch, T., and Offerhaus, J. (2016). Who likes to learn new things: measuring adult motivation to learn with PIAAC data from 21 countries. *Large Scale Assess. Educ.* 4:9. doi: 10.1186/s40536-016-0024-4
- Hartig, J., and Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychol. Test Assess. Modeling* 54, 418–431.
- Hayes, A. F., Coutts, J. J., and But, R. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *Commun. Methods Measures* 1, 1–24. doi: 10.1080/19312458.2020.1718629
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118
- James, W. (1884). On some omissions of introspective psychology. *Mind* 33, 1–26.
- Janczyk, M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. *J. Exp. Psychol.* 39, 1878–1887. doi: 10.1037/a0033336
- Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015
- 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, 95–111. doi: 10.1080/00313831.2015.1119723
- Knekta, E., and Eklöf, H. (2015). Modeling the test-taking motivation construct through investigation of psychometric properties of an expectancy-value-based questionnaire. *J. Psychoeduc. Assess.* 33, 662–673. doi: 10.1177/0734282914551956
- Kriegbaum, K., Jansen, M., and Spinath, B. (2014). Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learn. Individ. Diff.* 43, 140–148. doi: 10.1016/j.lindif.2015.08.026
- Kunina-Habenicht, O., and Goldhammer, F. (2020). ICT Engagement: a new construct and its assessment in PISA 2015. *Large Scale Assess. Educ.* 8:6. doi: 10.1186/s40536-020-00084-z
- Lindner, C., Nagy, G., and Retelsdorf, J. (2018). The need for self-control in achievement tests: Changes in students' state self-control capacity and effort investment. *Soc. Psychol. Educ.* 21, 1113–1131. doi: 10.1007/s11218-018-9455-9
- List, M. K., Robitzsch, A., Lüdtke, O., Köller, O., and Nagy, G. (2017). Performance decline in low-stakes educational assessments: different mixture modeling approaches. *Large Scale Assess. Educ.* 5:15. doi: 10.1186/s40536-017-0049-3
- Liu, Y., and Hau, K.-T. (2020). Measuring motivation to take low-stakes large-scale test: New model based on analyses of "Participant-Owned-Defined" missingness. *Educ. Psychol. Measur.* 2020:0013164420911972. doi: 10.1177/0013164420911972
- McHugh, L., Barnes-Holmes, Y., and Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *Psychol. Rec.* 54, 115–144. doi: 10.1007/BF03395465
- Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., et al. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contempor. Educ. Psychol.* 23, 113–131. doi: 10.1006/ceps.1998.0965
- Molenaar, D., Tuerlinckx, F., and Van der Maas, H. L. J. (2015). Package diffIRT. *J. Statistic. Software* 4:66.
- Mutz, D. C., and Pemantle, R. (2015). Standards for experimental research: Encouraging a better understanding of experimental methods. *J. Exp. Politic. Sci.* 2, 192–215. doi: 10.1017/XPS.2015.4
- Myers, A. J., and Finney, S. J. (2019). Change in self-reported motivation before to after test completion: Relation with performance. *J. Exp. Educ.* 1, 1–21. doi: 10.1080/00220973.2019.1680942
- Nagy, G., Nagengast, B., Becker, M., Rose, N., and Frey, A. (2018). Item position effects in a reading comprehension test: An IRT study of individual differences and individual correlates. *Psychol. Test Assess. Modeling* 60, 165–187.
- Nagy, G., Nagengast, B., Frey, A., Becker, M., and Rose, N. (2019). A multilevel study of position effects in PISA achievement tests: student- and school-level predictors in the German tracked school system. *Assess. Educ.* 26, 422–443. doi: 10.1080/0969594X.2018.1449100
- Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., and Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *J. Res. Adol.* 20, 482–506. doi: 10.1111/j.1532-7795.2010.00644.x
- Pearce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203.
- Penk, C., and 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. doi: 10.1007/s11092-016-9248-7
- Penk, C., and Schipolowski, S. (2015). Is it all about value? Bringing back the expectancy component to the assessment of test-taking motivation. *Learn. Individ. Differ.* 42, 27–35. doi: 10.1016/j.lindif.2015.08.002
- Piaget, J. (1928). La causalité chez l'enfant. *Br. J. Psychol.* 18, 276–301. doi: 10.1111/j.2044-8295.1928.tb00466.x
- R Development Core Team. (2009). *R: A language and environment for statistical computing [Computer software manual]*. Vienna: R Development Core Team.
- Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. doi: 10.1162/neco.2008.12-06-420
- Revelle, W. (2019). *Using R and the psych package to find  $\omega$* . Version 2.0.9. 1–20.
- Revelle, W., and Zinbarg, R. E. (2008). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika* 1, 1–14.
- Rios, J. A., and Liu, O. L. (2017). Online proctored versus unproctored low-stakes internet test administration: Is there differential test-taking behavior and performance? *Am. J. Dis. Educ.* 2017:1258628. doi: 10.1080/08923647.2017.1258628
- Rios, J. A., Liu, O. L., and Bridgeman, B. (2014). Identifying low-effort examinees on student learning outcomes assessment: A comparison of two approaches. *New Direct. Institut. Res.* 161, 69–82.
- Rose, N., Nagy, G., Nagengast, B., Frey, A., and Becker, M. (2019). Modeling multiple item context effects with generalized linear mixed models. *Front. Psychol.* 10:248. doi: 10.3389/fpsyg.2019.00248
- Rosseel, Y. (2010). lavaan: an R package for structural equation modeling and more. *J. Statistic. software* 48, 1–36.
- Rosseel, Y. (2019). *Structural equation modeling with lavaan*. Available online at: [https://personality-project.org/r/tutorials/summerschool.14/rosseel\\_sem\\_intro.pdf](https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_intro.pdf).
- Rosseel, Y. (2020). *The lavaan tutorial*, Department of Data Analysis. Belgium: Ghent University.
- Schüttpelz-Brauns, K., Hecht, M., Hardt, K., Karay, Y., Zupanec, M., and Kämmer, J. E. (2020). Institutional strategies related to test-taking behavior in low stakes assessment. *Adv. Health Sci. Educ.* 25, 321–335. doi: 10.1007/s10459-019-09928-y
- Schweizer, K. (2011). On the changing role of cronbach's  $\alpha$  in the evaluation of the quality of a measure. *Eur. J. Psychol. Assess.* 27, 143–144. doi: 10.1027/1015-5759/a000069
- Silm, G., Pedaste, M., and Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educ. Res. Rev.* 31:335. doi: 10.1016/j.edurev.2020.100335
- Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., and Conley, A. M. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high- and low-stakes test achievement. *Learn. Individ. Differ.* 39, 49–63. doi: 10.1016/j.lindif.2015.03.002
- Stenlund, T., Lyrén, P. E., and Eklöf, H. (2018). The successful test taker: exploring test-taking behavior profiles through cluster analysis. *Eur. J. Psychol. Educ.* 33, 403–417. doi: 10.1007/s10212-017-0332-2
- Sundre, D. L., and 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? *Contempor. Educ. Psychol.* 29, 6–26. doi: 10.1016/S0361-476X(02)00063-2
- Sundre, D. L., and Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assess. Update* 14, 8–9.
- Svetina, D., and Rutkowski, L. (2014). Detecting differential item functioning using generalized logistic regression in the context of large-scale assessments. *Meredith* 1993, 1–17.
- Tuerlinckx, F., and De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika* 70, 629–650. doi: 10.1007/s11336-000-0810-3
- Van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., and Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychol. Rev.* 118, 339–356. doi: 10.1037/a0022749
- Vollmeyer, R., and Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educ. Psychol. Rev.* 18, 239–253. doi: 10.1007/s10648-006-9017-0
- Voss, A., Nagler, M., and Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Exp. Psychol.* 60, 385–402. doi: 10.1027/1618-3169/a000218
- Weirich, S., Hecht, M., Penk, C., Roppelt, A., and Böhme, K. (2016). Item position effects are moderated by changes in test-taking effort. *Appl. Psychol. Measur.* 41, 115–129. doi: 10.1177/0146621616676791
- Wigfield, A., and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contempor. Educ. Psychol.* 25, 68–81. doi: 10.1006/ceps.1999.1015
- Wise, S. L., and DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educ. Assess.* 10, 1–17. doi: 10.1207/s15326977ea1001\_1

- Wise, S. L., and Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Appl. Measur. Educ.* 18, 163–183.
- Wolgast, A., Tandler, N., Harrison, L., and Umlauf, S. (2019). Adults' dispositional and situational perspective-taking: a systematic review. *Educ. Psychol. Rev.* 32, 353–389. doi: 10.1007/s10648-019-09507-y
- Zinbarg, R. E. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for  $\omega_h$ . *Appl. Psychol. Measur.* 30, 121–144. doi: 10.1177/0146621605278814

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Wolgast, Schmidt and Ranger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*