Contemporary Educational Psychology 58 (2019) 58–74

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Expectancy value interactions and academic achievement: Differential relationships with achievement measures

Jennifer Meyer[a,*], Johanna Fleckenstein[a,b], Olaf Köller[a]

[a] Leibniz Institute for Science and Mathematics Education (IPN), Olshausenstr. 62, 24118 Kiel, Germany
[b] University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Basel, Hofackerstrasse 30, OG08/West, 4132 Muttenz, Switzerland

ARTICLE INFO

Keywords:
Motivation
Expectancy value theory
Achievement measures
Latent interaction modelling

ABSTRACT

Motivation predicts academic achievement beyond cognitive ability. Expectancy value theory (Eccles et al., 1983) is a widely accepted and powerful approach explaining academic achievement as well as educational choices and attainment. Recently, attention to the multiplicative term of expectancy and value beliefs has increased. Trautwein et al. (2012) reported a detrimental effect of high task value when expectancy beliefs were low. We aimed to replicate and extend their study by using a large, representative sample of students attending upper secondary school in the German federal state Schleswig-Holstein (N = 3367). Following Trautwein et al. (2012), we applied latent interaction modelling to test whether the predictive value of expectancy value interactions differs for grades, final examinations, and standardized test scores as measures of achievement in two domains. We took the multi-dimensional structure of task value into consideration, analyzing the four components (attainment, intrinsic value, utility and cost) separately. Both a verbal and a non-verbal domain (English as a foreign language and mathematics) were investigated. Overall, the results supported those of Trautwein et al. (2012). However, our findings suggested measure- and domain-specific differences when using expectancy value beliefs and their interactions to predict academic achievement. Interaction terms predicted final examination results in both English and mathematics. Further, interaction effects were significant for grades in English but not mathematics. In general, effect sizes of multiplicative terms were small, especially in contrast to expectancy beliefs. Findings are discussed regarding the practical and conceptual importance of the multiplicative term in expectancy value theory applied in an educational setting.

1. Introduction

Individual differences in cognitive ability predict educational attainment and achievement (Deary, Strand, Smith, & Fernandes, 2007; Kuncel, Hezlett, & Ones, 2004; Rohde & Thompson, 2007). However, psychological constructs capturing individual differences in motivation (e.g., motivation, interest, and self-concept) have been shown to explain an additional amount of variance in these outcomes (Chamorro-Premuzic, Harlaar, Greven, & Plomin, 2010; Duckworth & Seligman, 2005; Kuncel, Ones, & Sackett, 2010; Steinmayr & Spinath, 2009). One of the most influential frameworks to conceptualize achievement motivation is *expectancy value theory* (EVT; Eccles, 2009; Eccles et al., 1983; Wigfield & Eccles, 2000; Atkinson, 1957; for reviews, see Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). According to expectancy value theory, achievement motivation depends on two elements: (a) expectancy of success as students' beliefs of how well they will do on the activity, and (b) value beliefs describing the extent to which students' value the activity (Eccles et al., 1983; Wigfield & Eccles, 1992; Wigfield, 1994). In other words, a person who beliefs a successful result when engaging in the task is possible, but does not have a compelling reason to do so will refrain from putting in a great deal of effort. Vice versa, if the task is important but is viewed as unlikely to be accomplished, the person might choose to engage in another task with higher expectancy of success. Value beliefs can be further differentiated into four subcomponents: cost, attainment value, intrinsic value, and utility value (see Eccles et al., 1983; Wigfield & Eccles, 1992; Trautwein et al., 2012). EVT has been shown to explain students' effort (e.g., Dietrich, Viljaranta, Moeller, & Kracke, 2017), choices (e.g., Updegraff, Eccles, Barber, & O'Brien, 1996; Nagy, Trautwein, Baumert, Köller, & Garrett, 2007; Eccles & Wang, 2015), and achievement in a variety of contexts, including academic achievement on different educational levels (e.g., Denissen, Zarrett, & Eccles, 2007). Despite their clear theoretical distinction, expectancy and value beliefs are strongly correlated empirically. This means that there are few cases in which expectancy beliefs

* Corresponding author.
E-mail addresses: jmeyer@ipn.uni-kiel.de (J. Meyer), johanna.fleckenstein@fhnw.ch (J. Fleckenstein), koeller@ipn.uni-kiel.de (O. Köller).

https://doi.org/10.1016/j.cedpsych.2019.01.006

Available online 07 January 2019

differ greatly from value beliefs. In other words, students with high expectancy beliefs are likely to have high value beliefs in this subject as well. Still, the theoretical distinction could be supported by Eccles and her team as outlined by Eccles and Wigfield (2002): They have shown that expectancy beliefs are related to actual performance, whereas task values predict enrollment and course choices even after controlling for prior performance (see Eccles et al., 1983, Eccles, Midgley, & Adler, 1984, Meece, Wigfield, & Eccles, 1990). These empirical findings are in support of differentiating between the two constructs, even if they are correlated.

Further, there are intervention studies enhancing value beliefs suggesting that it might be useful to differentiate between the two constructs. Previous research has shown the success of interventions increasing students' utility value to enhance academic effort especially in students with low expectancy beliefs (e.g., Brisson et al., 2017; Hulleman & Harackiewicz, 2009; Harackiewicz & Hulleman, 2010). This also supports the idea that expectancy beliefs are a result of previous achievements, whereas task values can impact the choice to engage in tasks. Regarding these results it makes sense to consider expectancy and value beliefs as separate constructs and to investigate their interplay as two psychological constructs associated with motivation to engage in domain-specific learning activities.

Recently, this has been done in educational research as additional attention has been dedicated to the interactions of expectancy value beliefs and their predictive value for achievement related behaviors (e.g., Nagengast et al., 2011; Trautwein et al., 2012; Guo, Marsh, Parker, Morin, & Yeung, 2015; Guo, Parker, Marsh, & Morin, 2015; Guo et al., 2016). Findings indicate a multiplicative relation with a detrimental effect of high task value when expectancy beliefs are low. This pattern was replicated for both verbal and non-verbal domains (Trautwein et al., 2012). It has to be noted, however, that in view of the high correlations of expectancy and value beliefs the number of cases is rather low, as most students with high expectancy beliefs show high value beliefs as well (see also Trautwein et al., 2012). This means the detrimental effect applies only under restricted circumstances. Further, effect sizes of the multiplicative term were small, especially in comparison to the large effects of expectancy beliefs. However, as pointed out by Nagengast, Trautwein, Kelava, and Lüdtke (2013), this finding should not be taken as an argument against the theoretical importance of the interaction term in expectancy value theory. Significant interaction effects are an indication of a multiplicative relation between the two predictor variables (Arnold & Evans, 1979; Busemeyer & Jones, 1983) regardless of effect size. Thus, despite potentially small effect sizes, investigating the multiplicative term of expectancy and value beliefs can be seen an important contribution to the understanding of motivation in educational contexts.

We aim to add to this body of research by considering differential effects of expectancy value interactions on three achievement measures in two domains (mathematics and English) by using "an enriched set of outcome variables" as suggested by Trautwein et al. (2012). Further, and again following Trautwein et al., we focus on academic achievement and its relationship with expectancy value beliefs, taking into account the four subcomponents (attainment, intrinsic value, utility and cost) separately. In order to replicate the findings by Trautwein et al. we used standardized test scores as a measure of academic achievement. In addition, we used grades as the most common measure of academic achievement as well as final examinations as a crucial high-stakes but school-based achievement indicator at the end of upper secondary education. Thus, our study aims to systematically investigate these three different measures of academic achievement. Final examinations as high-stakes tests have rarely been examined in secondary education in contexts other than the US-American standardized assessment (for an exception see Meyer, Fleckenstein, Retelsdorf, & Köller, 2019). They are particularly interesting because of their high relevance to students' academic lives, taking into account that their results make up a substantial part of higher education entrance qualifications in many European contexts. Thus, they are highly relevant for university admission. By comparing three achievement measures including final exams in two domains, we aim to gain more detailed insights on the relationship of expectancy value beliefs and their interactions with different aspects of academic achievement.

1.1. Expectancy value theory

According to expectancy value theory, achievement and achievement related behavior is associated with expectancy and value beliefs (Atkinson, 1957; Eccles et al., 1983; Wigfield & Eccles, 1992; Wigfield, 1994). Both of these components originate in the individuals' perceptions of previous situations as well as socialization processes (Wigfield & Eccles, 1992). These factors lead to the development of task-specific beliefs, such as ability beliefs, the perceived difficulty of different tasks, individuals' goals, self-schema, and affective memories that, in turn, influence the formation of expectancy and value beliefs (see Eccles et al., 1983; Eccles, Wigfield, & Schiefele, 1998). Expectancy beliefs and self-concept are theoretically distinct, however, they are not separable empirically (see Eccles & Wigfield, 2002). This is why in many studies expectancy of success is measured with a self-concept instrument (e.g., Trautwein et al., 2012; Guo, Marsh, Parker, Morin, & Dicke, 2017).

Evidence of the empirical separability of the four value components has been gathered in several studies (e.g., Trautwein et al., 2012; Luttrell et al., 2009; Conley, 2012; Guo et al., 2015; Gaspard et al., 2015; see also Flake, Barron, Hulleman, McCoach, & Welsh, 2015). According to Eccles et al. (1983), *cost* describes perceived negative consequences of engaging in a task, for example, performance anxiety amounting to emotional stress or opportunity costs of choosing this option (e.g., fewer time resources can be allocated to spending time with friends). Cost also includes the amount of perceived effort that has to be exerted in order to succeed. *Attainment value* is defined as the personal importance ascribed to succeeding in a task. According to Eccles and Wigfield (2002) attainment value can also be related to a person's self-schema, confirming relevant aspects of their identity by performing well in a task. *Intrinsic value* refers to the subjective interest in a task/domain or the enjoyment a person gains from engaging in a task. In the case of high intrinsic value, positive psychological outcomes are the reward. Thus, the intrinsic value component can be viewed as similar to intrinsic motivation as described in self-determination theory (Deci & Ryan, 1985; see also Eccles & Wigfield, 2002). In contrast, *utility value* describes the perceived individual usefulness of engagement and achievement in a task or domain. Because of the stronger importance of extrinsic performance rewards, utility value can be conceptualized as similar to the extrinsic motivation component of self-determination theory (Eccles, 2009; see also Trautwein et al., 2012). For example, utility value includes the perceived extent of how succeeding in this task can impact a student's future life. In general, previous research supports the four-dimensional structure of value beliefs. However, inter-correlations of the components are strong, especially for attainment and intrinsic value (e.g., Trautwein et al., 2012).

Despite the huge amount of previous research on expectancy value beliefs, only few studies have differentiated between the four components of task value (see Trautwein et al., 2012) – and even fewer took into account the multiplicative relationship of expectancy and value beliefs.

1.2. The interaction term

The original EVT model emphasizes the multiplicative relationship of expectancy and value beliefs (Feather & Newton, 1982; Feather, 1982, 1992). Motivation can only be strong if both expectancy and value beliefs are high. This means that low task value cannot be compensated by high expectancy beliefs and vice versa. But in modern conceptualizations of EVT, and especially in field research in educational contexts, the interaction term has largely been neglected.

Detecting multiplicative effects in these kinds of settings is challenging for different reasons (for a detailed discussion see Nagengast et al., 2011; Nagengast et al., 2013; Trautwein et al., 2012; Guo et al., 2017): first, the rate of extreme cases (e.g., very low expectancy beliefs but very high task value and vice versa) is very small in field research as indicated by the strong correlations of task value and expectancy beliefs (ranging from $r = 0.45$ to 0.80; e.g., Guo et al., 2016; Trautwein et al., 2012). Second, most studies did not have enough statistical power due to small sample sizes and methodological difficulties. Especially the inability to correct for measurement error was an issue, because when multiplying two unreliable factors measurement error multiplies as well (Busemeyer & Jones, 1983; Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007).

Recently, studies addressing these issues have been conducted, Nagengast et al. (2011) being the first to “put the X back into expectancy value theory”. They found significant interactions of expectancy and value beliefs predicting achievement-related behaviors such as engagement and academic choices in the science domain, following a similar pattern across 57 countries in the PISA study. Building on this work, Trautwein et al. (2012) were the first to consider achievement tests in two domains (English and mathematics) and investigate the different subcomponents of value beliefs separately. Their results showed significant multiplicative relations of expectancy value beliefs for all four subcomponents of task value in both English and mathematics. In order to use these insights in learning and education it is vital to consider the direction of the obtained multiplicative associations. Some studies suggest a synergistic effect of expectancy and value beliefs on homework behaviour (see Nagengast et al., 2013) and achievement-related choices (Guo et al., 2015; Nagengast et al., 2011). In contrast, results considering achievement tests suggest a detrimental effect of high task value when expectancy beliefs are low (Durik, Shechter, Noh, Rozek, & Harackiewicz, 2015; Trautwein et al., 2012). This means that students who see high value in a certain subject but do not have confidence in their abilities perform worse than students with the same expectancy of success but lower value beliefs. While this detrimental effect seems counterintuitive at first, Trautwein et al. explain it with mental contrasting effects (see Oettingen & Gollwitzer, 2010) which lead to frustration accompanied by a decreased goal-related activity. This is the case if the desired outcome is contrasted against a negative reality, while the importance of actively engaging in a goal pursuing behavior is highlighted. This finding is important because it indicates that a high task value is not always beneficial. This could easily be overlooked when investigating expectancy value models without the interaction term.

However, studies modeling latent interactions in the framework of the expectancy value model are scarce and still limited in terms of using different achievement measures. Especially with respect to the difficulty detecting interaction effects, replications are necessary in this area of research. To address this research gap and taking a more detailed approach we investigate differential relationships of expectancy value beliefs considering three important achievement measures. Because of the strong domain specificity of expectancy value theory, findings from one domain cannot just be transferred onto other domains. If results can be replicated in both domains, a strong case can be made for the importance of the multiplicative term when it comes to predicting academic achievement in the framework of EVT. Thus, similar to Trautwein et al. (2012) we investigate a verbal and a nonverbal domain: English as a foreign language and mathematics.

1.3. Expectancy value beliefs and achievement measures

Expectancy value beliefs have been shown to predict academic achievement in various domains and contexts (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Trautwein & Lüdtke, 2007; Wigfield & Guthrie, 1997). Both expectancy beliefs and task values have been associated with academic engagement: students with low expectancy beliefs are more likely to procrastinate academic tasks (Wu & Fan, 2017) and students who have higher academic self-concepts tend to invest more effort into their academic work (Levpušček, Zupančič, & Sočan, 2013; Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009). Similar relationships of task value and academic effort have been shown in previous studies, with higher task values being associated with higher academic effort (e.g., Trautwein, Lüdtke, Schnyder, & Niggli, 2006). It can be argued that the extent of these associations varies with characteristics of the utilized achievement measure. For example, interest – as a construct very similar to intrinsic value – is stronger in predicting grades than standardized test scores (Jansen, Lüdtke, & Schroeders, 2016). These results suggest that the association of expectancy value beliefs (and their interactions) with academic achievement may vary with the operationalization of achievement, as different measures capture different aspects of achievement (Willingham, Pollack, & Lewis, 2002). In the following, we summarize characteristics of three important achievement measures, deriving hypotheses on differential associations with expectancy and value beliefs and their interactions, building on previous research.

Standardized tests can be differentiated in low- and high-stakes settings. We focus on low-stakes settings, because standardized tests in German schools are commonly conducted for research and accountability purposes and thus entail low stakes for the students themselves. Even though most tests are conceptualized close to curricular demands, content and form can differ from classroom learning and assessment (Willingham et al., 2002). This is why the capacity to solve novel problems (i.e., fluid reasoning) is highly important and related to test performance, resulting in a strong association with cognitive ability. On the other hand, academic effort and related motivational aspects are less predictive of standardized tests (Borghans, Golsteyn, Heckman, & Humphries, 2011, 2016; Lechner, Danner, & Rammstedt, 2017; O'Connell, 2018; Willingham et al., 2002). So far, test scores are the achievement measure most thoroughly researched when investigating expectancy value interactions and their predictive value for academic achievement (e.g., Trautwein et al., 2012; Guo et al., 2015; Guo et al., 2016), with all of these studies obtaining significant results. Previous research shows expectancy value beliefs to be associated with learning behavior (Nagengast et al., 2013). This means for report card grades and final examinations effect sizes should be stronger as these measures are more strongly associated with learning effort and high-stakes for the individual student, as illustrated in the following.

Report card grades are the accumulation of different achievement indicators over an entire school term, containing classroom assessments such as written class examinations and evaluations of oral class participation. The amount of variance explained by cognitive ability is lower compared to standardized tests (e.g., $R^2 = 0.10$ vs. 0.58; Lechner et al., 2017). A growing number of studies provide evidence on the incremental value of motivational constructs such as conscientiousness and interest in predicting grades (e.g., Poropat, 2009; Spengler, Lüdtke, Martin, & Brunner, 2013; Steinmayr & Spinath, 2009). Furthermore, motivational factors are closely related to consistent study behaviors and school engagement (Wang & Eccles, 2013), which in turn can influence grading for three reasons: first, effort regulation and persistent studying have positive effects on the amount of learned materials, contributing to academic success (Bidjerano & Dai, 2007; Noftle & Robins, 2007; Steel, 2007). Second, teachers' expectations and judgments can influence grading (e.g., self-fulfilling prophecies and perceptual biases; Jussim & Harber, 2005). Third, grades are multi-criterial; teachers are encouraged to evaluate classroom behavior and participation during class and incorporate these into grading. Thus, if a student appears motivated in the classroom, this is likely to influence the teacher's judgment. In summary, the factors described are associated with academic effort (i.e., homework behavior), which has been shown to be predicted by expectancy value beliefs and their interaction (Nagengast et al., 2013). Considering these relationships, we argue that the expectancy value interactions have stronger effects on grades than standardized low stakes test scores.

Written final exams at the end of upper secondary education in Germany are characterized by their high stakes for students. They have great impact on final GPA and thus admission to university and vocational training opportunities after graduation. However, they are separate from final grades as described above. Both final grades and written final exams contribute to final GPA, but they are distinct measures. Teachers evaluate the written examinations, and students' preparation, academic effort, and learning behavior is crucial in order to succeed as many different materials are covered. Research on final examinations considering motivational constructs is scarce, except for studies on the SATs in the US-American context. Results suggest a stronger association with cognitive compared to motivational constructs (Noftle & Robins, 2007). However, little is known about the generalizability of these results onto other contexts. Taking into account the nature of final examinations in Germany, effort can be considered important in order to succeed in final examinations (similar to grades). The high-stakes testing situation emphasizes the relevance of effort and learning behavior, so it is even more salient to students compared to grades (which require consistent effort over months). Thus, expectancy value beliefs associated with success in the examinations may be more salient for the students as well. In this situation, the mental contrasting effect – which occurs when expectancy beliefs are low and value beliefs are high at the same time – may be facilitated and result in decreased learning effort (see Trautwein et al., 2012). Accordingly, the detrimental effect might even be stronger in a high-stakes tests, as the mental contrasting processes and the associated frustration can be argued to prevent students from applying the required effort and preparation time. Thus, we hypothesize that interaction effects predict final examination results even better than they predict grades.

In summary, achievement measures that take into account students' academic effort are more likely to be associated with motivational factors. However, to our knowledge no study has systematically compared different measures of academic achievement in their relationship with expectancy value interactions. We address this research desideratum in this study.

2. The present study

The present study had two major objectives: first, we aimed to replicate the results by Trautwein et al. (2012) investigating the relationship of expectancy value beliefs as well as their interaction with academic achievement in two domains (mathematics and English as a foreign language). Following Trautwein et al. (2012), we considered the predictive value of motivation in the framework of the expectancy value model beyond cognitive ability to gain further empirical evidence on the significance of motivational constructs in learning and education. Given the rarity of investigations regarding psychological interactions in educational research despite their theoretical importance, we particularly focused on the multiplicative relationship of expectancy and value beliefs when predicting academic achievement. Second, we extended Trautwein et al.'s approach by focusing on differential relationships with expectancy value beliefs depending on the utilized outcome measure. We considered three different achievement outcomes (grades, final examinations and standardized test scores). Thus, going beyond replication, our study provides a differential view on expectancy value beliefs by considering different aspects of achievement. To the best of our knowledge, no study on expectancy value interactions has compared these three different but common achievement measures so far. Especially final examinations are an interesting achievement outcome because their outcomes have high-stakes consequences for students' subsequent college or university attendance.

The first research question focuses on replication, with our hypotheses following the rationale of Trautwein et al. (2012).

H1. Intrinsic value, attainment, utility value and cost are strongly related, but separable subcomponents of task value. Thus, we expected value beliefs to have a four-dimensional structure.

H2. We considered the domain specificity of expectancy value beliefs, expecting small to moderate associations for mathematics and English.

H3. We investigated the predictive value of both expectancy and value beliefs when entered separately into the regression equation, expecting both to explain a significant amount of variance in all three achievement measures over and above cognitive ability.

H4. When both expectancy and value beliefs are used as predictors simultaneously we expected stronger effects of expectancy beliefs compared to value beliefs.

H5. We tested for multiplicative effects of expectancy and value beliefs, using the same latent approach as Trautwein et al. and providing further evidence on a multiplicative relationship of expectancy value beliefs. As argued above and in line with previous findings, we expected a detrimental effect of high task value when expectancy beliefs are low with mental contrasting effects impairing performance and goal-setting behavior (Oettingen & Gollwitzer, 2010), but a synergistic effect when both task value and expectancy beliefs are high (Trautwein et al., 2012).

Our second research question and corresponding set of hypotheses deals with the predictive value of expectancy value beliefs and their interaction when investigating different achievement measures as dependent variables.

H6. In line with previous research (e.g., Lechner et al., 2017; Jansen et al., 2016) we expected the overall predictive value of both expectancy and value beliefs as motivational constructs to be significantly stronger for report card grades as evaluations given by teachers, compared to standardized test scores as obtained from standardized achievement tests.

H7. We expected the relationship of motivation with final examinations to be similar to the relationship with grades, because similar aspects of achievement (i.e., high stakes and association with effort) are captured, as described above. (e.g., Meyer et al., 2019; Noftle & Robins, 2007).

H8. We expected the predictive value of expectancy value beliefs to be stronger on final exams compared to standardized test scores, because of the hypothesized stronger association with academic effort in a performance situation as compared to standardized testing for research purposes as described above (see Willingham et al., 2002) for differences between grades and test scores).

H9. For all interactions we expected a detrimental effect of high task value on achievement when expectancy beliefs are low, as indicated by previous research (Trautwein et al., 2012). This multiplicative relation of expectancy value beliefs when predicting academic achievement was expected to be stronger for both high-stakes measures and measures strongly linked to effort (i.e. final examinations and grades). This would be consistent with previous research showing expectancy value beliefs to predict academic effort (i.e., homework behavior; Nagengast et al., 2013). We expected the detrimental effect to be even stronger on final exams compared to grades, because of the emphasis on performance situation and the accumulation of knowledge that has to be retrieved on one single occasion.

H10. In line with Trautwein et al. (2012) similar patterns were expected for both mathematics and English.

H11. Similarly, we expected similar patterns of results for all four components of task value in line with Trautwein et al. (2012).

3. Method

3.1. Sample

The present study is based on data from the LISA6-study ($N = 3775$; see Leucht & Köller, 2016). Upper secondary education in the German federal state Schleswig-Holstein has a vocational and an academic track. The academic track represents the traditional Gymnasium with a variety of academic disciplines, whereas the vocational track includes more applied study fields (e.g., technical and economic courses) in addition to obligatory instruction (i.e., languages and mathematics). In both tracks the Abitur as a higher level education entrance certificate can be obtained as final certificate. In this study, achievement tests were obligatory for all students in grade 13 at randomly drawn academic track schools ($N = 1433$ students from 17 schools) and all vocational track schools ($N = 2342$ students from 27 schools) with consent of the Ministry of Schools and Professional Education, providing data representative for students in Schleswig-Holstein. Academic track schools were randomly drawn in a multiple step stratification cluster sampling procedure following the approaches in other educational large scale studies (see Bortz & Döring, 2006; Lohr, 1999; see also Leucht, Kampa, & Köller, 2016 for further information). Sampling was conducted by the IEA Data processing and research center in Hamburg. The lower number of vocational track schools made it possible to obtain data from nearly all vocational track schools (27 from 29; two were in private administration) in Schleswig-Holstein (see also Leucht et al., 2016).

Participation in questionnaires was voluntary as required in compliance with ethical standards. From the original sample we excluded students who were not taking part in English instruction ($N = 9$) and students differing in years of English instruction because they had focused on another first foreign language ($N = 129$), resulting in a final sample of $N = 3637$ (54.7% female; age $M = 19.92$ years; academic track $N = 1315$; vocational track $N = 2322$).

3.2. Measures

3.2.1. Expectancy value beliefs

Expectancy beliefs were measured with a self-concept scale consisting of three items (e.g., "I have always been good at mathematics/English"; Eccles & Wigfield, 2002; Trautwein et al., 2012). Task values were assessed with 14 items adapted from TOSCA-R (Trautwein et al., 2012; Trautwein, Neumann, Nagy, Lüdtke, & Maaz, 2010). Five items were used to measure intrinsic value (e.g. "If I can learn something new in mathematics/English, I'm prepared to use my free time to do so"), four items for attainment value (e.g., Mathematics/English is important to me personally), three for utility value (e.g., "Good grades in mathematics/English can be of great value to me later") and two for cost (e.g., "I'd have to invest a lot of time to get good grades in mathematics/English"). All scales exhibited acceptable to good internal consistency (mathematics: $\alpha = 0.87$–0.90; English: $\alpha = 0.67$–0.90), showing similar reliabilities as in Trautwein et al. To correct for measurement error in our analyses, we chose a latent modeling approach using multiple indicators for each construct (see Section 3.5). Because of voluntary questionnaire participation as described above data on expectancy value beliefs were available from $N = 2234$ students. Missing values from students participating in the study, but did not respond to the questionnaire were handled with FIML (see Section 3.5).

3.2.2. Academic achievement

3.2.2.1. Report card grades. Domain-specific end-of-the-year report card grades (year 13; mathematics and English) were collected via school administration lists. Report card grades at upper secondary school in Germany range from zero to 15 points. Higher values indicate better grades. Data on grades in English were available for $N = 3619$; in mathematics for $N = 3617$ students.

3.2.2.2. Standardized tests. Mathematics achievement was measured with a test from the National Educational Panel Study (NEPS; 20 items). The test assesses mathematics competencies based on the literacy concept and covers content areas both for lower and upper secondary level (Neumann et al., 2013; see also Kampa, Köller, Schmidt, & Leucht, 2016). On the day of testing N = 3171 students were present. In line with the NEPS framework, the test's major aim is to cover the literacy aspect of mathematical competence relevant for future life. Thus, tasks include mathematical concepts and procedures embedded in everyday life contexts that are typical for a particular age group over the life-span. As mathematical concepts and procedures are typically learned in school they follow a particular curriculum. According to the NEPS framework, mathematical competencies include two dimensions: (a) content areas in the field of mathematics (*quantity (4 items), change and relationship (6), space and shape (3), data and chance (7))* and (b) the cognitive component of mathematical competence, covering activities related to solving mathematical problems. The test assesses six cognitive processes: *technical abilities and skills (9 items), modeling (1), mathematical problem solving (4), using representational forms (5), mathematical communication (1)*. Quality and appropriateness of the items was ensured by extensive pilot studies conducted by the NEPS team (see Neumann et al., 2013). Reliability and validity of the test have been shown in previous studies. Results (see Kampa et al., 2016) show appropriate convergent validity with cognitive ability tests ($r = 0.64^{**}$) and differential correlations with students' study profiles, correlations being highest for students with thematic focus on mathematics/science ($r = 0.34^{**}$). Correlation with language study profiles is lower, supporting discriminant validity ($r = 0.07^{**}$).

English achievement was measured with listening and reading comprehension tests using a subset of items from the German National Assessment. The test items were designed to monitor the implementation of educational standards in Germany (see Köller, Knigge, & Tesch, 2010) and therefore represent competencies based on national curricula for the English language classroom. Competencies were measured with different testlets, each of which included a listening or reading text portraying real-life language situations and tasks with different item types (e.g., multiple choice, short written answers). Three to four tasks consisting of different items were presented in four 15-min blocks. Blocks were balanced in difficulty and rotated in eight different booklets to control for position effects and performance decline with test duration *(Multi-Matrix-Design)*. On the day of testing $N = 3191$ students were present. Reliability and validity of the test have been shown in previous studies: results can be linked to similar standardized tests such as PISA (see Fleckenstein, Leucht, Pant, & Köller, 2016). This is further supported by previous results (see Leucht et al., 2016), showing differential correlations with students' study profiles, correlations being highest for students with thematic focus on languages ($r = 0.\ 26^{**}$). Correlation with mathematics/science study profiles is slightly lower, indicating discriminant validity ($r = 0.21^{**}$).

3.2.2.3. Written final exams. We collected information on grades received in written final exams (*Abitur*) in both domains via school administration lists. Exam grades range from zero to 15 points, higher values indicating better performances. Different centralized abitur tasks are given for vocational and academic track schools by the Ministry of Schools and Professional Education. We took account of this difference by including school track as covariate in all analyses. In each domain students can choose between at least two assignments: in mathematics, competencies of different fields are captured (e.g., calculus and geometry), with coherent superordinate assignments consisting of several subtasks (Conference of Ministers of Education [KMK], 2002a). In English, assignments consist of text comprehension (e.g., fictional vs. non-fictional texts) combined with essay writing tasks (KMK, 2002b). Criteria for selection of these tasks are very standardized across Germany (see KMK, 2002a, 2002b). For English, both content and language quality need to be covered in the given tasks and evaluated separately according to specific criteria. The task requirements need to go beyond mere retrieval of learned information. Criteria that have to be considered in the evaluation are given for each domain. For example, in the evaluation of content in the English exam, teachers need to consider the independency, adequacy and complexity of the written content. This includes coherence of text, including a clear message and transparent text structure. Criteria are very specific, frame of references are given for "good" (grade 2, eleven points) and "sufficient" (grade 4, five points), indicating the expected performances. In both school types, two teachers evaluate the exams independently, using criteria issued by the ministry of education based on subject-specific demands. Standardized criteria are given. Teachers get a full solution for all tasks, including descriptions on what exactly needs to be written in order to achieve full credit. Example tasks including expectation levels for different grades and example answers are given. If the evaluation of the two teachers differs, the final exam commission decides what grade will be given, consulting a third teacher if necessary. Thus, individual teacher effects are kept at a minimum, especially in comparison with grades.

Students are able to choose which subjects they take final exams in to some extent. Thus, not all students take final exams in both mathematics and English. In English, final exam results were available for $N = 2950$ students, in mathematics for $N = 3044$ students. There are significant differences between students who chose to take final exams in mathematics/English and students who did not, for example in cognitive ability, socio economic status (SES) and gender (see Author, year). Even though this indicates data on the final examination variables not to be missing at random, we used the full information maximum likelihood approach (FIML) to estimate missing values. This seemed feasible because these covariates are accounted for in the analysis and the number of missing on these variables was low in comparison to the rest of the sample ($N = 735$ in English, $N = 642$ in mathematics). However, in the supplements we also provide analyses on the subsamples of students who participated in the respective exam along with robustness checks including the quadratic terms (see Supplements A and C, respectively).

3.3. Estimation of plausible values

Five plausible values (PVs) were estimated for each student, using IRT scaling techniques in ConQuest (Wu, Adams, & Wilson, 1998). PVs are used to gain more accurate estimates regarding associations on population level by correcting students' ability for measurement error (see Wu, 2005; Von Davier, Gonzalez, & Mislevy, 2009). They provide a database that allows unbiased estimation of the plausible range and the location of proficiency for groups of students. PVs are estimated based on student responses to achievement test items, as well as on other relevant background information (Mislevy, 1991). This background model is used to include covariates associated with ability to increase reliability of the ability scores. For the building of the background model only variables obtained from all students were included. Information on gender, age, English and mathematics course level as well as grades were used for the estimation (see Leucht et al., 2016). When drawing plausible values, students who were not present on the day of testing were included, so that proficiency scores were estimated based on the background variables (see Leucht et al., 2016). For all subsequent analysis we combined the five PVs for final estimations following Rubin (1987) using the M*plus* function *type = imputation*. PV-Reliabilities were satisfactory, ranging from 0.80 (cognitive ability), and 0.81 (English) to 0.92 (mathematics).

3.4. Covariates

Information on gender was collected from school administration. Socioeconomic status was measured by parents' occupational status obtained in a student questionnaire to compute the *Highest International Socio-Economic Index of Occupational Status* (HISEI; Ganzeboom, de Graaf, & Treiman, 1992).

General cognitive ability was assessed using the verbal and figural reasoning subscales of the Cognitive Ability Test (KFT4-12R; Heller & Perleth, 2000). Data were available from $N = 3172$ students.

As mentioned above, upper secondary education in Schleswig-Holstein includes different school tracks. Student characteristics differ systematically between tracks, with academic-track students showing higher academic achievement as well as higher SES and cognitive ability (see Leucht & Köller, 2016). We took account of these differences by using school track as a dummy-coded covariate (see Section 3.5).

3.5. Statistical analyses

We applied confirmatory factor analysis and structural equation modeling using M*plus* (Version 7.4; Muthén & Muthén, 1998, 2011). In all analyses we followed the procedure described by Trautwein et al. (2012) as closely as possible. Thus, we computed separate models for the four subcomponents of task value, performing all analyses four times, respectively. First, we aimed to replicate the structure of the instrument. Second, we considered the domain-specificity of expectancy and value beliefs for the four subcomponents. Third, we used multiple regression modeling to investigate expectancy and value beliefs as separate predictors of the three achievement measures. Fourth, we used both expectancy and value beliefs as simultaneous predictors. Fifth, we added the interaction term to each of the four regression models. In all regression analyses we used gender, cognitive ability, and SES as covariates. All analyses were repeated for English and mathematics, respectively. All multi-indicator constructs (expectancy beliefs and value beliefs) were modeled as latent variables. All indicators were standardized prior to analysis to enhance interpretability. Models were based on maximum likelihood estimation with robust standard errors using a numerical integration algorithm. Following Trautwein et al., latent interactions were modeled using the latent moderated structural equation (LMS) approach to correct for measurement error of latent constructs and provide unbiased interaction effects (Klein & Moosbrugger, 2000).

Because of the hierarchical data structure with students clustered in schools it was necessary to control for dependencies in the data. Thus, we used robust estimates of the model parameters (see Muthén & Satorra, 1995). FIML implemented in M*plus* 7 was used to deal with missing values (see Enders, 2010).

4. Results

4.1. The association of expectancy and task value beliefs with academic achievement

We start by first testing the hypotheses associated with Research Question 1, dealing with replication of the results by Trautwein et al. (2012). Second, we move on to Research Question 2, extending the study by comparing different achievement indicators.

4.1.1. Multidimensionality and inter-correlations of task values, expectancy and achievement measures

First, we tested the structure of the task value model (H1). We conducted separate confirmatory factor analyses for mathematics and English. Four latent factors representing the components of task value were specified, and each item was allowed to load on only one of these factors. Residual correlations were not allowed. The analyses yielded acceptable model fit indices in both mathematics (χ^2 [71, N = 2,014] = 954.92, $p \leq 0.001$, CFI = 0.960, RMSEA = 0.079, SRMR = 0.043) and English (χ^2 [71, N = 2,027] = 979.46 , $p \leq 0.001$, CFI = 0.913, RMSEA = 0.079, SRMR = 0.047). However, for our subsequent analyses this less-than-perfect model fit is acceptable as the latent modeling is restricted to one latent factor at a time. Overall, results support a four-factor structure of the instrument.[1] Factor loadings are provided in Supplement G.

Inter-correlations between the latent factors of task value components and bivariate correlations with covariates and achievement measures are displayed in Table 1 and Table 2 for mathematics and English, respectively. Standard errors are provided in Supplement F. Results show that most inter-correlations of the four latent factors do not exceed $r = 0.90$, and thus providing further support for the separability of the four task value components. Tables 1 and 2 further include inter-correlations of value components and expectancy beliefs in each domain, suggesting strong associations in support of our hypothesis.

Bivariate correlations of task value and expectancy beliefs with achievement measures suggest differential associations with grades and final examinations compared to test scores. It can be noted that correlations with test scores are consistently lower than with grades and final exams across the four subcomponents. This is in line with our hypotheses (H6, H7, H8). We performed a more rigorous test in subsequent analyses (see results Section 2).

4.1.2. Domain-specificity of expectancy value beliefs

In a second step we considered the domain specificity of expectancy and value beliefs in mathematics and English. In line with our hypotheses, we found negative latent correlations of expectancy beliefs ($r = -0.15$, $p \leq 0.001$), attainment value ($r = -0.05$, $p \leq 0.05$) and intrinsic value ($r = -0.10$, $p \leq 0.001$) between mathematics and English (H2). The association of cost in mathematics and English was non-significant ($r = -0.01$, $p = .607$), and there was a positive, but small correlation of utility value in the two domains ($r = 0.08$, $p \leq 0.001$).

4.1.3. Expectancy and value beliefs as separate vs. simultaneous predictors of achievement measures

Following Trautwein et al. (2012) we addressed the predictive value of expectancy and value beliefs on academic achievement in the next step, extending their study by simultaneously investigating three different achievement measures. First, we used expectancy and each of the four value components as separate predictors, controlling for gender, cognitive abilities, SES and school track. In line with our hypotheses, the results show expectancy and all task value components to significantly predict academic achievement for all three achievement measures (H3). The analyses yielded similar results for both mathematics and in English (see Tables 3 and 4, respectively).

In the next step, we used expectancy beliefs and attainment value simultaneously to predict academic achievement in mathematics and English (see Tables 5 and 6, respectively). In line with our assumptions, expectancy beliefs in this model were shown to be the stronger predictor of achievement in both mathematics and English (H4). Due to the strong association of expectancy and value beliefs, the effects of value beliefs were no longer statistically significant. In fact, some coefficients for value beliefs indicated a negative relationship, especially in the analyses considering attainment and intrinsic value. This is most likely a suppression effects due to the strong correlations with expectancy beliefs.

4.2. Predicting different achievement measures using the expectancy value interaction term

In the next step, we tested whether the interaction of expectancy and value beliefs would significantly predict academic achievement (H5). Going beyond replication in this step, our main research question focused on the differential effect of expectancy value interactions when predicting grades, final examinations, and test scores simultaneously (Research Question 2). Results suggested both measure- and domain-specific differences (see Tables 7 and 8 for mathematics and English, respectively).

To provide stronger support for our hypotheses comparing the three achievement measures we tested whether the effect of expectancy value differs significantly between measures. To do this, we conducted several Wald tests. Results suggested significant differential effects for grades and test scores (H6), as well as between final exams and test scores (H8) for all task value components (all p-values ≤ 0.001, see Supplement D).

Results on H7 varied between domains (H10) and different components of task value (H11). In mathematics the effects of expectancy value beliefs differed significantly between final exams and grades for all task value components (all p-values ≤ 0.001). In English, results differ for attainment and intrinsic value ($p \leq 0.001$, respectively), but no significant difference could be found when considering utility value and cost.

In H9 we were concerned with the multiplicative term, particularly with differential effects between measures. Domain-specific results (H10) were obtained as well as differential results regarding the components of task value (H11): In mathematics, the interaction terms significantly predicted achievement for final examinations, but not for grades and test scores in all four models. An exception is the task value component of cost: cost significantly moderated the association of expectancy beliefs and mathematics test score. In English, the interaction terms predicted both grades and final exams. In both domains, results differed from findings obtained by Trautwein et al. (2012), as the interaction of expectancy and value did not significantly predict test scores, with the exception of cost in mathematics (H5).

4.3. Robustness checks

Robustness analyses on the subsamples of students who participated in the respective exam (e.g., all students who took mathematics and all students who took the English exam; see Supplement A) yielded similar results. As a further check of robustness, we included quadratic terms into our models (Supplement C; see Ganzach, 1997). Because of high multicollinearity of expectancy and value beliefs and their quadratic effects standard errors increased exponentially, resulting in non-significant interaction terms, even though effect size did not decrease. Thus, conclusions on the stability of the interaction effects cannot be made. As a further robustness check we included results from manifest analyses (see Supplement B). In these analyses interaction terms are largely non-significant, as can be expected due to measurement errors as pointed out above, emphasizing the benefits of a latent approach.

4.4. Graphical illustration of significant interaction terms

To illustrate what the significant moderation of expectancy beliefs on the relationship between value components and measures of academic achievement entails, graphical displays are provided (see Fig. 1). It has to be kept in mind that in this kind of field study, the most extreme categories are rare (e.g. very high expectancy beliefs but very low task value). Nonetheless, the direction of results suggests that for students with high expectancy beliefs task value is not associated with performance as much as for students with low expectancy beliefs. For students with low expectancy beliefs, graphical displays of the interaction effects suggest that if intrinsic task value is high, performance in final examinations is lower, suggesting a detrimental effect of high task value if expectancy beliefs are low. This is in support of our hypotheses.

[1] Model-fit indices examining a one-factor structure of task value and a three-factor structure with attainment and intrinsic value combined to a single factor can be found in Supplement E. A four-factor structure compliant with theoretical assumptions and previous empirical results yielded the best fit.

Table 1
Bivariate correlations English (latent).

	1	2	3	4	5	6	7	8	9
Motivation									
1. Expectancy									
2. Attainment	0.72**								
3. Intrinsic value	0.76**	0.89**							
4. Utility	0.50**	0.73**	0.65**						
5. Cost	−0.71**	−0.43**	−0.48**	−0.26**					
Covariates									
1. IQ	0.09**	0.05	0.01	0.07*	−0.14**				
2. SES	0.15**	0.14**	0.11*	0.15**	−0.09**	0.14**			
Outcomes									
1. English grade	0.68**	0.51**	0.50**	0.36**	−0.51**	0.21**	0.19**		
2. English standardized test score	0.53**	0.38**	0.40**	0.30**	−0.44**	0.46**	0.21**	0.53**	
3. English final exam	0.66**	0.44**	0.42**	0.32**	−0.51**	0.27**	0.20**	0.77**	0.56**

Note. IQ = cognitive ability; SES = socioeconomic status. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*).

* $p < .05$.

** $p < .01$.

Table 2
Bivariate correlations mathematics (latent).

	1	2	3	4	5	6	7	8	9
Motivation									
1. Expectancy									
2. Attainment	0.76**								
3. Intrinsic value	0.80**	0.91**							
4. Utility	0.57**	0.79**	0.66**						
5. Cost	−0.74**	−0.49**	−0.59**	−0.32**					
Covariates									
1. IQ	0.31**	0.23**	0.24**	0.19**	−0.26**				
2. SES	0.06	−0.01	0.01	0.02	−0.05	0.14**			
Outcomes									
1. Mathematics grade	0.69**	0.47**	0.49**	0.37**	−0.49**	0.30**	0.10**		
2. Mathematics standardized test score	0.45**	0.36**	0.38**	0.34**	−0.37**	0.64**	0.17**	0.44**	
3. Mathematics final exam	0.65**	0.47**	0.50**	0.35**	−0.49**	0.37**	0.12**	0.70**	0.52**

Note. IQ = cognitive ability; SES = socioeconomic status. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*). *$p < .05$.

** $p < .01$.

5. Discussion

In this study we had two major aims. First, we intended to replicate the results by Trautwein et al. (2012), showing the multiplicative effect of expectancy and value beliefs when predicting academic achievement. Second, we extended the investigation by focusing on the differential predictive value of expectancy value beliefs and their interactions depending on the utilized achievement measure.

5.1. Replication of Trautwein et al. (2012)

Regarding the first question we replicated most of the findings by Trautwein et al. (2012). Our results provide further support for the multi-dimensionality of value beliefs, showing they can be empirically differentiated into four components (H1), even though inter-correlations are high. Moreover, our results are in line with previous findings concerning the domain-specificity of expectancy value beliefs (H2), with utility value appearing to be somewhat less domain-specific (Trautwein et al., 2012). Further, our results replicate previous research as both expectancy and value beliefs predict achievement when entered into the model separately (H3). We also found no additional predictive effect of value beliefs when entering them next to expectancy beliefs (H4). Thus, most of our findings were in line with previous research.

However, we could not replicate the multiplicative effects of expectancy and value beliefs predicting standardized test scores (H5) that were found by Trautwein et al. (2012). Potential reasons need to be discussed. First, different achievement tests were used in the studies. Trautwein et al. applied a version of the TOEFL test to measure English competencies, and a TIMSS (Third International Mathematics and Science Study; e.g., Baumert, Bos, & Lehmann, 2000) test in mathematics, whereas in our study we assessed English using items from the German National Assessment and mathematics using a NEPS test. It can be argued that these mathematics tests differ in conceptualization, with the NEPS test being designed in line with the literacy concept and thus measuring competencies, whereas TIMSS is conceptualized closer to the school curriculum, capturing knowledge. If test content is more school based in one study it could have resulted in a stronger association with grades compared with the other study, explaining a stronger effect of motivation on test scores with significant interactions in Trautwein et al. (2012). This hypothesis needs to be considered in future research. Second, different student samples were tested. Trautwein et al.'s study was carried out in the German federal state Baden-Württemberg, ours in Schleswig-Holstein. Even though both studies were conducted in upper secondary schools in year 13 and academic as well as vocational school track were included, population differences can be assumed. Previous research has shown that students in Baden-Württemberg perform better

Table 3
Mathematics linear regression results for expectancy and value beliefs entered separately, controlling for cognitive ability, gender and SES.

	Expectancy			Intrinsic			Attainment			Utility			Cost		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation															
Exp. beliefs	0.73***	0.63***	0.27***	–	–	–	–	–	–	–	–	–	–	–	–
	(0.02)	(0.03)	(0.03)												
Value beliefs	–	–	–	0.57***	0.53***	0.26***	0.50***	0.46***	0.22***	0.38***	0.32***	0.20***	−0.52***	−0.48***	−0.22***
				(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Covariates															
Gender[a]	−0.22***	−0.04	0.71***	−0.26***	−0.08	0.69***	−0.25***	−0.07	0.69***	−0.25***	−0.06	0.69***	−0.22***	−0.05	0.71***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)
SES	0.06**	0.07***	0.03	0.07***	0.08***	0.04	0.08***	0.09***	0.04	0.07**	0.08***	0.03	0.07**	0.08***	0.03
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Cognitive ability	0.09***	0.17***	0.47***	0.19***	0.24***	0.49***	0.19***	0.25***	0.50***	0.23***	0.29***	0.51***	0.19***	0.25***	0.50***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.04	0.04	0.22***	0.09	0.11	0.24***	0.10	0.12	0.24***	0.10	0.12	0.25***	0.01	0.03	0.20***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)
Residual variance	0.495	0.536	0.387	0.704	0.668	0.403	0.721	0.687	0.409	0.791	0.758	0.414	0.716	0.675	0.410
Model fit															
χ^2	200.802			247.825			190.625			66.155			4.327		
Df	14			23			33			14			6		
CFI	0.979			0.974			0.986			0.993			1.000		
RMSEA	0.061			0.052			0.036			0.032			0.000		
SRMR	0.022			0.024			0.019			0.020			0.002		

Note. Exp. beliefs = expectancy beliefs; SES = socioeconomic status; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*). $^{*}p \leq 0.05$.

** $p \leq 0.01$.

*** $p \leq 0.001$.

[a] Gender: 0 = female, 1 = male.

[b] School type: 0 = vocational track, 1 = academic track.

Table 4
English linear regression results for expectancy and value beliefs entered separately, controlling for cognitive ability, gender and SES.

	Expectancy			Intrinsic			Attainment			Utility			Cost		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation															
Exp. beliefs	0.71***	0.70***	0.52***												
	(0.02)	(0.03)	(0.02)												
Value beliefs				0.69***	0.58***	0.53***	0.57***	0.50***	0.41***	0.42***	0.38***	0.35***	−0.55***	−0.54***	−0.42***
				(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
Covariates															
Gender[a]	−0.10**	−0.03	−0.03	−0.16***	−0.11**	−0.07	−0.09*	−0.04	−0.02	−0.19***	−0.13**	−0.10*	−0.14***	−0.07	−0.05
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
SES	0.08**	0.06*	0.03	0.12***	0.10***	0.06	0.10***	0.09**	0.05	0.12***	0.10**	0.06	0.13***	0.11***	0.07*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Cognitive ability	0.14***	0.18***	0.34***	0.20***	0.24***	0.38***	0.17***	0.21***	0.36***	0.18***	0.22***	0.36***	0.14***	0.18***	0.34***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.05	0.19***	0.45***	0.09	0.22**	0.48***	0.06	0.20**	0.46***	0.09	0.23**	0.48***	0.02	0.16*	0.43***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)
Residual variance	0.501	0.512	0.504	0.698	0.750	0.592	0.700	0.745	0.609	0.814	0.826	0.654	0.697	0.694	0.594
Model fit															
χ^2	164.794			195.476			168.494			94.344			20.385		
Df	14			23			33			14			6		
CFI	0.981			0.971			0.983			0.983			0.998		
RMSEA	0.054			0.045			0.034			0.040			0.025		
SRMR	0.017			0.032			0.022			0.027			0.005		

Note. Exp. beliefs = expectancy beliefs; SES = socioeconomic status; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*).

* $p \le 0.05$.

** $p \le 0.01$.

*** $p \le 0.001$.

[a] Gender: 0 = female, 1 = male.

[b] School type: 0 = vocational track, 1 = academic track.

Table 5
Mathematics linear regression results with both value and expectancy beliefs as simultaneous predictors.

	Intrinsic			Attainment			Utility			Cost		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation												
Exp. beliefs	0.85***	0.67***	0.24***	0.83***	0.67***	0.26***	0.74***	0.64***	0.23***	0.77***	0.61***	0.25***
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Value beliefs	−0.18***	−0.05	0.05	−0.13***	−0.05	0.02	−0.01	−0.02	0.07**	0.06	−0.03	−0.03
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Covariates												
Gender[a]	−0.21***	−0.04	0.70***	−0.21***	−0.04	0.71***	−0.22***	−0.04	0.69***	−0.22***	−0.04	0.71***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
SES	0.06**	0.07***	0.03	0.05**	0.07***	0.03	0.06**	0.07***	0.03	0.06**	0.07***	0.03
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Cognitive ability	0.10***	0.17***	0.47***	0.10***	0.17***	0.47***	0.09***	0.17***	0.47***	0.09***	0.17***	0.47***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.01	0.03	0.22***	0.01	0.03	0.22***	0.04	0.04	0.23***	0.04	0.04	0.21***
	(0.05)	(0.07)	(0.06)	(0.05)	(0.07)	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)
Residual variance	0.495	0.537	0.387	0.495	0.537	0.387	0.496	0.536	0.384	0.495	0.536	0.387
Model fit												
χ^2	593.459			576.110			343.967			217.512		
Df	48			61			36			25		
CFI	0.963			0.970			0.976			0.983		
RMSEA	0.056			0.048			0.048			0.046		
SRMR	0.030			0.028			0.032			0.020		

Note. Exp. beliefs = expectancy beliefs; SES = socioeconomic status; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*). *$p \leq 0.05$.

** $p \leq 0.01$.

*** $p \leq 0.001$.

[a] Gender: 0 = female, 1 = male.

[b] School type: 0 = vocational track, 1 = academic track.

Table 6
English linear regression results with both value and expectancy beliefs as simultaneous predictors.

	Intrinsic			*Attainment*			*Utility*			*Cost*		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation												
Exp. beliefs	0.71***	0.80***	0.47***	0.68***	0.75***	0.51***	0.70***	0.71***	0.49***	0.70***	0.67***	0.47***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)
Value beliefs	−0.03	−0.23***	0.07	0.04	−0.10*	0.01	0.02	−0.04	0.06	−0.02	−0.04	−0.07
	(0.06)	(0.07)	(0.06)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Covariates												
Gender[a]	−0.10**	−0.04	−0.03	−0.10**	−0.05	−0.03	−0.10**	−0.03	−0.03	−0.10**	−0.03	−0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
SES	0.08**	0.06*	0.03	0.08**	0.06*	0.03	0.07**	0.06*	0.03	0.08**	0.06*	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Cognitive ability	0.14***	0.18***	0.34***	0.14***	0.18***	0.34***	0.14***	0.18***	0.34***	0.14***	0.18***	0.33***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.04	0.18**	0.45***	0.05	0.19***	0.45***	0.05	0.18***	0.45***	0.04	0.18***	0.44***
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)
Residual variance	0.509	0.510	0.503	0.507	0.513	0.503	0.503	0.512	0.502	0.500	0.511	0.502
Model fit												
χ^2	525.440			396.547			305.943			192.710		
Df	48			61			36			25		
CFI	0.958			0.975			0.972			0.985		
RMSEA	0.052			0.039			0.045			0.043		
SRMR	0.035			0.026			0.032			0.015		

Note. Exp. beliefs = expectancy beliefs; SES = socioeconomic status; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*).

* $p \leq 0.05$.

** $p \leq 0.01$.

*** $p \leq 0.001$.

[a] Gender: 0 = female, 1 = male.

[b] School type: 0 = vocational track, 1 = academic track.

Table 7
Results of linear regression analyses with both value and expectancy beliefs as simultaneous predictors and interaction term in mathematics.

	Intrinsic			*Attainment*			*Utility*			*Cost*		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation												
Exp. beliefs	0.85***	0.67***	0.24***	0.83***	0.66***	0.26***	0.73***	0.63***	0.23***	0.77***	0.61***	0.25***
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Value beliefs	−0.18***	−0.07	0.04	−0.13***	−0.05	0.02	−0.01	−0.02	0.07**	0.06	−0.02	−0.01
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Exp. X value beliefs	0.01	0.06*	0.03	0.01	0.06**	0.01	0.00	0.04*	−0.01	−0.02	−0.07**	−0.08*
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Covariates												
Gender[a]	−0.21***	−0.04	0.70***	−0.21***	−0.04	0.71***	−0.22***	−0.04	0.69***	−0.22***	−0.04	0.71***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
SES	0.06**	0.07***	0.03	0.05**	0.07**	0.03	0.06**	0.07***	0.03	0.06**	0.07**	0.03
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Cognitive ability	0.10***	0.17***	0.47***	0.10***	0.17***	0.47***	0.09***	0.17***	0.46***	0.09***	0.17***	0.47***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.01	0.03	0.22***	0.01	0.03	0.22***	0.04	0.04	0.23***	0.04	0.04	0.21***
	(0.05)	(0.07)	(0.06)	(0.05)	(0.07)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.07)	(0.05)
Residual variance	0.494	0.537	0.387	0.493	0.537	0.386	0.494	0.537	0.383	0.494	0.537	0.386

Note. All multi-indicator constructs were modeled as latent variables. Traditional fit indices are not available for models with latent product terms. Exp. beliefs = expectancy beliefs; exp. X value beliefs = interaction terms. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*).

* $p \le 0.05$.
** $p \le 0.01$.
*** $p \le 0.001$.
[a] Gender: 0 = female, 1 = male.
[b] School type: 0 = vocational track, 1 = academic track.

Table 8
Results of linear regression analyses with both value and expectancy beliefs as simultaneous predictors and interaction term in English.

	Intrinsic			*Attainment*			*Utility*			*Cost*		
	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score	Grades	Final Exams	Test score
Motivation												
Exp. beliefs	0.71***	0.81***	0.47***	0.66***	0.73***	0.51***	0.70***	0.70***	0.49***	0.69***	0.66***	0.47***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Value beliefs	−0.03	−0.27***	0.08	0.07	−0.07	0.01	0.04	−0.03	0.06	−0.05	−0.07	−0.08*
	(0.06)	(0.07)	(0.06)	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Exp. X value beliefs	0.08***	0.16***	−0.05	0.08***	0.13***	−0.01	0.06**	0.11***	−0.03	−0.06***	−0.08***	−0.02
	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Covariates												
Gender[a]	−0.10**	−0.03	−0.03	−0.09**	−0.04	−0.03	−0.10**	−0.03	−0.03	−0.09**	−0.03	−0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
SES	0.08**	0.06*	0.03	0.08**	0.06*	0.03	0.07**	0.06*	0.03	0.08**	0.06*	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Cognitive ability	0.14***	0.17***	0.34***	0.14***	0.18***	0.34***	0.14***	0.18***	0.34***	0.14***	0.17***	0.33***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
School track[b]	0.04	0.18***	0.45***	0.04	0.18***	0.45***	0.05	0.18***	0.45***	0.05	0.19***	0.44***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)
Residual variance	0.505	0.502	0.500	0.502	0.505	0.502	0.500	0.507	0.501	0.497	0.506	0.502

Note. All multi-indicator constructs were modeled as latent variables. Traditional fit indices are not available for models with latent product terms. Exp. beliefs = expectancy beliefs; exp. X value beliefs = interaction terms. Standardized test score refers to the results from standardized achievement tests conducted for research purposes. Final exam refers to the results obtained in written final matriculation exams at the end of upper secondary school, thus associated with high stakes for students as they influence college entrance certificates (*Abitur*).

* $p \le 0.05$.
** $p \le 0.01$.
*** $p \le 0.001$.
[a] Gender: 0 = female, 1 = male.
[b] School type: 0 = vocational track, 1 = academic track.

on achievement tests in mathematics compared to students from Hamburg (Nagy et al., 2007; Neumann, Nagy, Trautwein, & Lüdtke, 2009). Similar differences can be found for students from Schleswig-Holstein (see Stanat, Böhme, Schipolowski, & Haag, 2016; Pant, Stanat, Schroeders, Roppelt, Siegle, & Pöhlmann, 2013). If students perform better on standardized tests in general because of differing features of the school systems or varying population characteristics, it can be argued that the effect of motivation and the interaction of expectancy and value beliefs have a stronger impact on test performance. These hypotheses remain to be tested in future research. However, our results

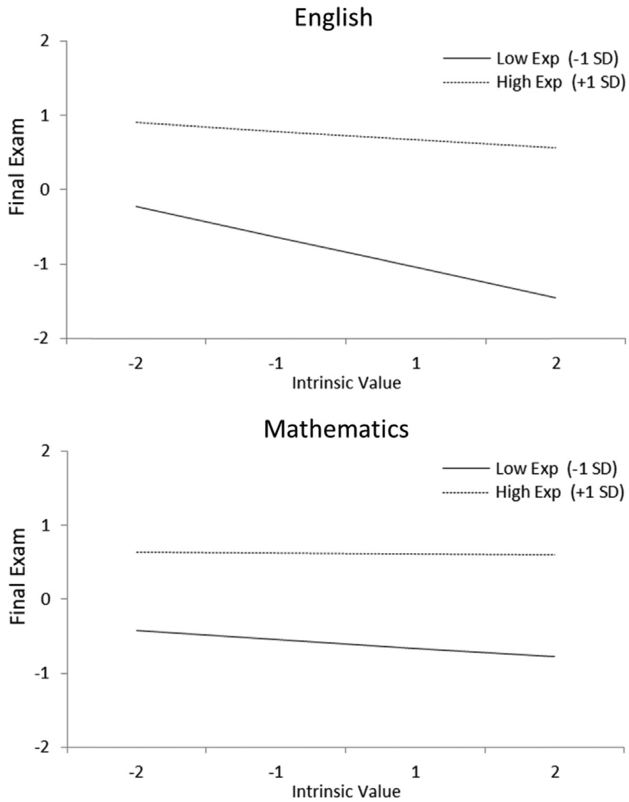

Fig. 1. Plots of the significant moderating effects of expectancy beliefs (Exp) on the relationship between value beliefs and measures of academic achievement (final examinations). Depicted are simple slopes at ± 1 standard deviation from the mean of expectancy value. The independent (value) and dependent (achievement measures) variables range from −2 to +2 standard deviations of the mean.

support these hypotheses to some degree. When considering test scores, effect sizes for expectancy and value beliefs can in general be described as smaller compared to effect sizes in Trautwein et al.' s study. For example, effect sizes for expectancy beliefs in the multiplicative model in Trautwein et al.'s study were estimated at about $\beta = 0.50–0.60$, whereas in our study they were estimated at about $\beta = 0.25–0.50$ for test scores. Third, it has to be noted that the data collections of the studies are about seven years apart; it is possible that cohort effects can explain the differential results.

5.2. Differential effects of multiplicative terms on achievement measures

Moving on to our second research question, the effects of expectancy value beliefs on test scores can be described as smaller in size compared to the effects on other achievement measures (i.e., grades and final examinations; H6, H7, H8). This was in line with our hypotheses, as motivational constructs generally do not predict as much variance in test scores as in grades (Lechner et al., 2017). By comparing the measures systematically, we attempted to gain more detailed insights concerning the relationship of expectancy value motivation with academic achievement in both verbal and non-verbal domains and potentially differential effects depending on the utilized achievement measure.

Our results indicated significant multiplicative terms of expectancy and value beliefs when predicting both grades and final exams in English (H9). However, there were differential effects for grades and final exams in mathematics (H10), with the multiplicative term predicting only final exams but not grades. To understand these differences and deduce potential explanations, we will first consider the meaning of the multiplicative relations of expectancy and value beliefs in our study.

Our interpretation of the interactions is in line with the results by Trautwein et al. (2012): A multiplicative relation implies that the effect of expectancy beliefs on achievement depends on the extent to which an individual values a given domain and vice versa. Accordingly, we find the effect of task value to be dependent on the level of expectancy of success: when expectancy beliefs are high, a high task value can have positive effects on academic achievement (i.e., a synergistic relationship). When expectancy beliefs are low, a high task value does not compensate, but instead have a detrimental effect on a students' academic achievement.

As suggested by Trautwein et al. (2012), this detrimental effect can be explained with mental contrasting processes (see Oettingen & Gollwitzer, 2010) that lead to decreased goal- related activity if the outcome is negative. Further, effort can be understood as a double edged sword (Covington & Omelich, 1979; Marsh et al., 2016; Nicholls, 1976): a situation associated with high task value but with low expectancy for success presents a threat to a students' self-concept and, thus, self-esteem: if he or she fails or expects to fail in a situation that is perceived to be relevant, negative emotions are set off. It can be argued that the negative emotions associated with failing can be prevented if effort is decreased, as the failure can then be attributed on flexible characteristics associated with the situation (e.g., lack of studying), but if effort is high and the student still fails, it would have to be attributed on stable factors associated with the individual (e.g., lack of ability). Previous studies indicated that inability attributions and negative affect were greatest when failure followed a bigger amount of effort (Covington & Omelich, 1979). Particularly if failure is a likely outcome—a possibility that is supposedly more evident for students with low self-concept— students might not try hard, because trying hard and failing would further destabilize their subsequent self-concept (Marsh et al., 2016). Our results suggest that this effect might be even stronger if task value and stakes are high, as expectancy value interactions consistently predict final exams, but not grades in mathematics. These results are in line with our expectations that the detrimental effect of low expectancy beliefs and high task value should be strongest for final examinations. As argued above, in order to succeed in final examinations it is important to put in effort. In view of the high-stakes testing situations for students the relevance of effort and learning behavior is emphasized and thus even more salient to the students than with grades. This means that expectancy value beliefs associated with success in the examinations might be more salient for the students as well. In this situation, the mental contrasting effect that occurs when expectancy beliefs are low with value beliefs high at the same time, might be facilitated and result in decreased learning efforts (see Trautwein et al., 2012). This seems be even more detrimental in high stakes performance situations such as final exams because there is no way to compensate later. However, for the interpretation of these findings it has to be noted that in view of the high correlations of expectancy and value beliefs the number of cases in which this detrimental effects applies are be rather limited (see also Trautwein et al., 2012).

The question remains why the pattern of results differs in English, where expectancy value interactions predict both final exams and grades and, further, the overall effect of expectancy value motivation does not significantly differ for the components of utility and cost. Apart from this inconsistency, most of our findings were stable for all four components of task value (H11). Explanations can be found when considering domain-specific differences considering class participation and oral performance in the classroom. With English as a foreign language representing a verbal domain, speaking is one of the core competencies in this subject. Therefore, English grades include oral participation in the classroom to a greater extent than grades in other domains. A student with both high expectancy and value beliefs in English would potentially be involved more in classroom activities, whereas a student with low expectancy but high value beliefs might be more reserved during the lessons due to the processes describes above (i.e., mental contrasting and self-esteem danger), which could result in lower grades. This process is less likely to apply to mathematics. This would explain why high task value combined with low expectancy beliefs would be detrimental for English grades, but not mathematics. Further, it can be argued that expectancy and value beliefs can differ depending on topics or emotional states that can change over the course of the school year, thus allowing for compensation in other contexts to improve the grade. This hypothesis should be considered in future research in view of domain-specific differences between mathematics and English.

With respect to the findings of previous research, there is another discrepancy that should be addressed. Nagengast et al. (2013) found synergistic effects of expectancy and value beliefs on homework behavior. Reasons for the differential interaction effects for measures of effort (i.e., homework behavior) vs. measures of achievement need to be discussed. There are several explanations for this discrepancy. First, the mental contrasting effect might not apply when it comes to homework behavior, as there are no direct consequences for students depending on their homework assignments (assuming they are not explicitly graded). In Nagengast et al. (2013) items referred to expectancy and value beliefs associated with homework in a given domain, whereas in our study (as well as in Trautwein et al., 2012) items referred to the domain in a more general way. This might indicate that expectancy and value beliefs concerning homework in a domain might be different from students' beliefs regarding the domain itself. This hypothesis needs to be considered in future research.

An alternative explanation for differences between achievement measures regarding their relationship with expectancy value motivation might be found in the types of tasks included. For example, final exam tasks in English include writing assignments requiring the production of coherent texts. This is not the case in the standardized test used in this study, which consisted of shorter answers and multiple choice questions. It can be argued that these differences can explain a stronger effect of motivation on text production as opposed to receptive competencies, such as reading and understanding texts. Because of the need for elaborate and costly evaluation when scoring longer written texts, this type of assignment has rarely been used in large scale research. Thus, investigating the differential effect of motivation on different English competencies remains an interesting question for future studies.

Further, the effect size of interactions needs to be discussed in relation to the main effects of expectancy and value. Similar to Nagengast et al. (2013) and Trautwein et al. (2012), we found small effect sizes. Especially in comparison with the strong effects of expectancy interaction effects the additional amount of explained variance can be seen as minor. However, as pointed out by Nagengast et al. (2013), this finding should not be taken as an argument against the theoretical importance of the interaction term in expectancy value theory. Significant interaction effects are an indication of a multiplicative relation between the two predictor variables (Arnold & Evans, 1979; Busemeyer & Jones, 1983) regardless of effect size. The small size of interaction effects also warrants applying latent interaction modeling because in manifest models effect sizes would be even smaller, potentially resulting in the unjustified rejection of multiplicative relations. Furthermore, following the argumentation by Nagengast et al. (2013) we considered the effect size of the interaction effect in comparison to that of other background variables, for example we obtained similar effect sizes for SES and school track. In summary, this suggests that small effect size does not diminish the importance of our findings.

5.3. Limitations and future research

While this study adds to the body of research on expectancy value interactions and provides new differential insights by including multiple achievement measures, some issues have to be kept in mind when interpreting our results. First, we used FIML to handle missing values. Final examinations are achievement measures associated with a certain amount of choices. Thus, these missing values cannot be assumed to be missing at random. It is known that course selection is strongly associated with expectancy value beliefs (see Nagy et al., 2007), which is why it is possible that missing values associated with final exam choice behavior may have affected our results. By conducting several robustness analyses we have tried to consider potential effects, but further research is needed.

Second, on a related note, a large amount of data was missing on the expectancy and value items. This is the result of the voluntary nature of students' questionnaire as required by law and ethics. We handled these missing data with FIML to include all accessible information. However, this might lead to decreasing validity of our findings as results might differ if all students had participated in the questionnaire. Additional analyses (see Leucht & Köller, 2016) showed that selectivity of samples depended on school track, with selectivity being more distinct in vocational track students. In tendency, students with lower final GPA ($d = -0.05/-0.11$; for academic and vocational track, respectively) and lower cognitive ability ($d = 0.09/0.04$) were less likely to respond to the questionnaire. Female students were more likely to participate ($d = 0.08/\ 0.14$). Further, as described above, we estimated plausible values for students who were not present on the day of testing. Additional analyses show that the absent students differ slightly from students who participated considering final GPA ($d = -0.16$). In consideration of these selectivity issues the found relationships might be even stronger in a more heterogeneous sample. However, as the relevant variables were included in our models it can be argued that FIML estimation provided reliable estimates for population parameters that would be comparable to parameters estimated using Multiple Imputation (see Graham, Olchowski, & Gilreath, 2007).

Third, another measurement issue concerns the limitation of variables in the background model. Because of the high missing percentage in the questionnaire data we discussed above, we included only complete variables from the school administration lists. This means that some relevant variables were not accounted for in the estimation of plausible values. This might have affected our results as expectancy and value beliefs as well as the interaction term were not included. This could have led to an underestimation of effects in our models. Thus, it can be argued if plausible values could have been estimated more accurately, effects would be stronger in our sample. However, regarding the comparison with grades and final exams, a correction for measurement error would not be possible for these measures. Thus, not accounting for these variables in the estimation of plausible values might lead to a better comparability of the measures. Still, it is possible that the nonsignificance of interaction terms when using standardized test scores as outcome are a result of measurement error in our study. This provides another explanation for the varying results when comparing with the results of Trautwein et al. (2012).

Fourth, we used items adapted from Trautwein et al. (2012) to assess expectancy and value beliefs to closely replicate their study. However, the measurement of these constructs with self-report scales can be questioned and, similar to Trautwein et al. (2012), measurement model fit was not perfect. Especially in view of the high inter-correlations of expectancy and value beliefs different instruments should be used in future research, potentially reducing common method variance as caused by similar wording. Thus, as previously discussed (see Trautwein et al., 2012) findings need to be replicated in studies using different and possibly longer instruments assessing expectancy value beliefs.

Fifth, another issue is the nonsignificance of interaction effects when controlling for quadratic effects of expectancy and value beliefs as observed in our robustness analyses. In view of the robust effect size compared to the original analyses, this can be explained by the increased standard errors resulting from high multicollinearity of predictors. This means that the robustness of interaction effects could not be shown in this analysis, which is why further replications are needed to provide evidence on the robustness of our results.

Sixth, longitudinal studies are needed to provide evidence on the long-term stability of our results as well as the temporal ordering of effects, as pointed out by Guo et al. (2016). Previous research (e.g., Marsh et al., 2016) suggests that reciprocal effects can be assumed for the relationship of self-concept and achievement in a domain, with achievement and value beliefs being viewed as both cause and effect of each other. Similar associations can be assumed for related motivational constructs such as expectancy value beliefs; however, these should be tested in future studies.

Seventh, further replications are needed to clear up differing results between studies and investigate potential explanations as well as to investigate the generalizability across other samples, age groups, and educational levels.

5.4. Implications

Our study has theoretical as well as practical implications. First, our results provide empirical evidence on the theoretical importance of the multiplicative term in expectancy value theory. Further our results indicate interaction effects depend on the nature of the applied achievement measure. This preliminary evidence for differential effects can help enhance the understanding of student motivation, as factors such as high stakes and effort might potentially impact expectancy value beliefs and their interplay on academic achievement. Of course, this needs to be tested more rigorously in future studies.

Second, taken together with previous research, our findings can offer careful suggestions for practical application in education. First of all, our replication of a multiplicative relation between expectancy and value beliefs supports the idea by Guo et al. (2015) that isolated interventions targeting at strengthening only one component might be less successful at promoting academic achievement. Of course it has to be kept in mind that cases of students with low expectancy but high task value are rare, which is consistent with research showing the success of interventions increasing students utility value to enhance academic effort especially in students with low expectancy beliefs (e.g., Brisson et al., 2017; Hulleman & Harackiewicz, 2009; Harackiewicz & Hulleman, 2010). This point is closely related to the assumed reciprocal effects between expectancy and value beliefs and achievement. It can be argued that expectancy beliefs are a result of previous achievements, whereas task values can impact the choice to engage in tasks. This would mean if task values could be enhanced and result in stronger academic efforts as shown in these previous studies; this could foster academic achievement and in turn have positive impact on expectancy beliefs. However, this cannot be tested with our data as longitudinal studies are needed to investigate these reciprocal relationships rigorously. Still, we think this idea can support the necessity of distinguishing between expectancy and value beliefs as well as considering their interactions when predicting academic achievement.

On the other hand, in line with findings by Guo et al. (2015) and Durik et al. (2015) our results indicate the importance of fostering both expectancy and value beliefs at the same time and that just increasing value beliefs while expectancy beliefs stay low could in fact be detrimental to academic achievement for these students. Rather, interventions targeting the promotion of educational outcomes should seek to enhance both expectancy and value beliefs. Also, researchers developing new interventions should take into account the different aspects of achievement depending on the utilized measure when evaluating their findings.

5.5. Conclusion

In summary, our results highlight the theoretical and practical importance of the interaction term for expectancy value theory and its application in educational context. Furthermore, they emphasize the importance of future replications using the same methodology in similar student populations.

6. Declarations of interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2019.01.006.

References

Arnold, H. J., & Evans, M. G. (1979). Testing multiplicative models does not require ratio scales. *Organizational Behavior and Human Performance, 24*(1), 41–59. https://doi.org/10.1016/0030-5073(79)90014-X.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6, Pt.1), 359–372. https://doi.org/10.1037/h0043445.

Baumert, J., Bos, W., & Lehmann, R. (Eds.) (2000). TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Vol.2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe [Third international mathematics and science study. Mathematics and science education at the end of upper secondary school]. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences, 17*(1), 69–81. https://doi.org/10.1016/j.lindif.2007.02.001.

Borghans, L., Golsteyn, B. H. H., Heckman, J., & Humphries, J. E. (2011). Identification problems in personality psychology. *Personality and Individual Differences, 51*(3), 315–320. https://doi.org/10.1016/j.paid.2011.03.029.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences of the United States of America, 113*(47), 13354–13359. https://doi.org/10.1073/pnas.1601135113.

Bortz, J., & Döring, N. (2006). Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler [Methods and Evaluation for the Humanities and Social Sciences], 4th ed. Berlin: Springer.

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal, 54*(6), 1048–1078. https://doi.org/10.3102/0002831217716084.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93*(3), 549–562. https://doi.org/10.1037//0033-2909.93.3.549.

Chamorro-Premuzic, T., Harlaar, N., Greven, C. U., & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence, 38*(4), 385–392. https://doi.org/10.1016/j.intell.2010.05.002.

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy value perspectives. *Journal of Educational Psychology, 104*(1), 32–47. https://doi.org/10.1037/a0026042.

Covington, M. V., & Omelich, C. L. (1979). Effort: The double-edged sword in school achievement. *Journal of Educational Psychology, 71*(2), 169–182. https://doi.org/10.1037/0022-0663.71.2.169.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19*(2), 109–134. https://doi.org/10.1016/0092-6566(85)90023-6.

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development, 78*(2), 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x.

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction, 47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009.

Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology, 3*(3), 100–114. https://doi.org/10.1027/1614-2241.3.3.100.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939–944. https://doi.org/10.1111/j.1467-9280.2005.01641.x.

Durik, A. M., Shechter, O. G., Noh, M., Rozek, C. S., & Harackiewicz, J. M. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion, 39*(1), 104–118. https://doi.org/10.1007/s11031-014-9419-0.

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. https://doi.org/10.1080/00461520902832368.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.). *Achievement and achievement motivation. Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Eccles, J. S., Midgley, C., & Adler, T. (1984). Grade-related changes in the school environment. *The Development of Achievement Motivation,* 238–331.

Eccles, J. S., & Wang, M.-T. (2015). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development, 40*(2), 100–106. https://doi.org/10.1177/0165025415616201.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*, 830–847. https://doi.org/10.1111/j.1467-8624.1993.tb02946.x.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon, & N. Eisenberg (Eds.). *Handbook of child psychology: Social, emotional, and personality development* (pp. 1017–1095). Hoboken, NJ, US: John Wiley & Sons.

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press.

Feather, N. T. (1982). Expectancy–value approaches: Present status and future directions. In Author (Ed.). *Expectations and actions: Expectancy–value models in psychology* (pp. 395–420). Hillsdale, NJ: Erlbaum.

Feather, N. T. (1992). Values, valences, expectations, and actions. *Journal of Social Issues, 48*(2), 109–124. https://doi.org/10.1111/j.1540-4560.1992.tb00887.x.

Feather, N. T., & Newton, J. W. (1982). Values, expectations, and the prediction of social action: An expectancy valence analysis. *Motivation and Emotion, 6*(3), 217–244. https://doi.org/10.1007/BF00992246.

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary educational psychology, 41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002.

Fleckenstein, J., Leucht, M., Pant, H. A., & Köller, O. (2016). Proficient beyond borders: Assessing non-native speakers in a native speakers' framework. *Large-scale Assessments in Education, 4*(1), 19.

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods, 2*(3), 235–247. https://doi.org/10.1037//1082-989X.2.3.235.

Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, 107*, 663–677. https://doi.org/10.1037/edu0000003.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science: The Official Journal of the Society for Prevention Research, 8*(3), 206–213. https://doi.org/10.1007/s11121-007-0070-9.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy by-value interactions. *Learning and Instruction, 49*, 81–91. https://doi.org/10.1016/j.learninstruc.2016.12.007.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences, 37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008.

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., ... Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open, 2*(1), https://doi.org/10.1177/2332858415626884.

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology, 51*(8), 1163–1176. https://doi.org/10.1037/a0039440.

Harackiewicz, J. M., & Hulleman, C. S. (2010). The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass, 4*(1), 42–52. https://doi.org/10.1111/j.1751-9004.2009.00207.x.

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+ R. Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (3., revidierte Aufl. des KFT 4-13+) [Cognitive ability test for class level 4 to 12, revised version].* Göttingen, Germany: Beltz.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science), 326*(5958), 1410–1412. https://doi.org/10.1126/science.1177067.

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology, 46*, 116–127. https://doi.org/10.1016/j.cedpsych.2016.05.004.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3.

Kampa, N., Köller, O., Schmidt, F. T. C., & Leucht, M. (2016). Kompetenzen im Fach Mathematik. In M. Leucht, N. Kampa, & O. Köller (Eds.), Fachleistungen beim Abitur. LISA 6 : Vergleich allgemeinbildender und beruflicher Gymnasien in Schleswig-Holstein [An empirical study on the educational outcomes of students from vocational and academic upper secondary schools] (pp. 119–143).

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*(4), 457–474. https://doi.org/10.1007/BF02296338.

KMK = Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland [Conference of Ministers of Education]. (2002a). Einheitliche Prüfungsanforderungen in der Abiturprüfung Englisch [Standardized requirements for final examinations in English at the end of upper secondary school]. (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 24.05.2002). Bonn, Germany: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Retrieved from http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Englisch.pdf.

KMK = Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland [Conference of Ministers of Education]. (2002b). Einheitliche Prüfungsanforderungen in der Abiturprüfung Mathematik [Standardized requirements for final examinations in mathematics at the end of upper secondary school]. (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 24.05.2002). Bonn, Germany: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Retrieved from http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Mathe.pdf.

Köller, O., Knigge, M., & Tesch, B. (Eds.). (2010). *Sprachliche Kompetenzen im Ländervergleich [Assessment of language proficiency in the German National Assessment of Educational Progress: Monitoring the achievement of standards for German and the first foreign language at the end of grade 9.].* Münster: Waxmann.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161. https://doi.org/10.1037/0022-3514.86.1.148.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences, 49*(4), 331–336. https://doi.org/10.1016/j.paid.2010.03.042.

Lechner, C., Danner, D., & Rammstedt, B. (2017). How is personality related to intelligence and achievement? A replication and extension of Borghans et al. and Salkever. *Personality and Individual Differences, 111*, 86–91. https://doi.org/10.1016/j.paid.2017.01.040.

Leucht, M., & Köller, O. (2016). Anlage und Durchführung der Studie. In M. Leucht, N. Kampa, & O. Köller (Eds.), Fachleistungen beim Abitur. LISA 6 : Vergleich allgemeinbildender und beruflicher Gymnasien in Schleswig-Holstein [An empirical study on the educational outcomes of students from vocational and academic upper secondary schools] (pp. 79–98).

Leucht, M., Kampa, N., & Köller, O. (Eds.). (2016). Fachleistungen beim Abitur: LISA 6 : Vergleich allgemeinbildender und beruflicher Gymnasien in Schleswig-Holstein [An empirical study on the educational outcomes of students from vocational and academic upper secondary schools].

Levpušček, M. P., Zupančič, M., & Sočan, G. (2013). Predicting achievement in mathematics in adolescent students. *The Journal of Early Adolescence, 33*(4), 523–551. https://doi.org/10.1177/0272431612450949.

Lohr, S. (1999). *Sampling and design.* Pacific Grove, CA: Duxbury.

Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2009). The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement, 70*(1), 142–160. https://doi.org/10.1177/0013164409344526.

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology, 52*(8), 1273–1290. https://doi.org/10.1037/dev0000146.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60–70. https://doi.org/10.1037//0022-0663.82.1.60.

Meyer, J., Fleckenstein, J., Retelsdorf, J., & Köller, O. (2019). The relationship of personality traits and different measures of domain-specific achievement in upper secondary education. *Learning and Individual Differences, 69*, 45–59. https://doi.org/10.1016/j.lindif.2018.11.005.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*(2), 177–196. https://doi.org/10.1007/BF02294457.

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267. https://doi.org/10.2307/271070.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "x" out of expectancy value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science, 22*(8), 1058–1066. https://doi.org/10.1177/0956797611415540.

Nagengast, B., Trautwein, U., Kelava, A., & Lüdtke, O. (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research, 48*, 428–460. https://doi.org/10.1080/00273171.2013.775060.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2007). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*(4), 323–345. https://doi.org/10.1080/13803610600765687.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., & Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal For Educational Research Online/Journal FüR Bildungsforschung Online, 5*(2), 80–109.

Neumann, M., Nagy, G., Trautwein, U., & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen. *Zeitschrift für Erziehungswissenschaft, 12*(4), 691–714. https://doi.org/10.1007/s11618-009-0099-6.

Nicholls, J. G. (1976). Effort is virtuous, but it's better to have ability: Evaluative responses to perceptions of effort and ability. *Journal of Research in Personality, 10*(3), 306–315. https://doi.org/10.1016/0092-6566(76)90020-9.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big

five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116–130. https://doi.org/10.1037/0022-3514.93.1.116.
O'Connell, M. (2018). The power of cognitive ability in explaining educational test performance, relative to other ostensible contenders. *Intelligence, 66*, 122–127. https://doi.org/10.1016/j.intell.2017.11.011.
Oettingen, G., & Gollwitzer, P. (2010). Strategies of setting and implementing goals: Mental contrasting and implementation intentions. In J. E. Maddux, & J. P. Tangney (Eds.). *Social psychological foundations of clinical psychology*. New York, NY: Guilford Press.
Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., & Pöhlmann, C. (2013). IQB-Ländervergleich 2012: mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I. [IQB – Comparison of federal states 2012: Competencies in mathematics and science at the end of lower secondary education]. Münster/New York/München/Berlin: Waxmann.
Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338. https://doi.org/10.1037/a0014996.
Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004.
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* Hoboken, NJ: John Wiley & Sons.
Spengler, M., Lüdtke, O., Martin, R., & Brunner, M. (2013). Personality is related to educational outcomes in late adolescence: Evidence from two large-scale achievement studies. *Journal of Research in Personality, 47*(5), 613–625. https://doi.org/10.1016/j.jrp.2013.05.008.
Stanat, P., Böhme, K., Schipolowski, S., & Haag, N. (Eds.). (2016). IQB-Bildungstrend 2015: Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich. [IQB – trends in education 2015: language competencies at the end of lower secondary education]. Waxmann.
Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65.
Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*(1), 80–90. https://doi.org/10.1016/j.lindif.2008.05.004.
Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., & Maaz, K. (Eds.). (2010). Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand [School achievement of high school students: Testing the new upper school structure]. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.
Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology, 99*(2), 432–444. https://doi.org/10.1037/0022-0663.99.2.432.
Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97*(6), 1115–1128. https://doi.org/10.1037/a0017048.
Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98*(2), 438–456. https://doi.org/10.1037/0022-0663.98.2.438.
Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104*(3), 763–777. https://doi.org/10.1037/a0027470.
Updegraff, K. A., Eccles, J. S., Barber, B. L., & O'Brien, K. M. (1996). Course enrollment as self-regulatory behavior: Who takes optional high-school math courses? *Learning and Individual Differences, 8*(3), https://doi.org/10.1016/S1041-6080(96)90016-3.
Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series, 2*, 9–36.
Wang, M.-T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002.
Wigfield, A. (1994). Expectancy value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49–78. https://doi.org/10.1007/BF02209024.
Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*(3), 265–310. https://doi.org/10.1016/0273-2297(92)90011-P.
Wigfield, A., & Eccles, J. S. (2000). Expectancy value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015.
Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology, 89*(3), 420–432. https://doi.org/10.1037/0022-0663.89.3.420.
Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*(1), 1–37. https://doi.org/10.1111/j.1745-3984.2002.tb01133.x.
Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2–3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005.
Wu, M. L., Adams, R. J., & Wilson, M. (1998). *ACER ConQuest: Generalised item response modelling software.* Melbourne, Australia: Acer Press.
Wu, F., & Fan, W. (2017). Academic procrastination in linking motivation and achievement-related behaviours: A perspective of expectancy value theory. *Educational Psychology, 37*(6), 695–711. https://doi.org/10.1080/01443410.2016.1202901.