Promoting Value Beliefs in Mathematics: A Multidimensional Perspective and the Role of Gender

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Dipl.-Psych. Hanna Gaspard

aus Karlsruhe

Tübingen

2015

Tag der mündlichen Prüfung:	30.03.2015
Dekan:	Prof. Dr. rer. soc. Josef Schmid
1. Gutachter:	Prof. Dr. Benjamin Nagengast
2. Gutachter:	Prof. Dr. Ulrich Trautwein

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Prof. Dr. Benjamin Nagengast and Prof. Dr. Ulrich Trautwein for their guidance throughout this dissertation project. I greatly appreciate the opportunity to work and learn in an extremely inspiring research environment and the high level of feedback, support, and challenge that I was provided with in every phase of this dissertation.

I am also deeply grateful to Prof. Jacque Eccles for the opportunity to spend a very interesting research stay in Irvine and for all her advice, encouragement, and support before, during, and after this stay.

Special thanks go to Anna-Lena Dicke who closely accompanied each part of this dissertation. Without her invaluable assistance, feedback, and ideas, this dissertation would not have been the same. I would also like to say a big thank you to Barbara Flunger for all her moral and professional support, especially—but not only—with statistics.

The completion of the project “Motivation in Mathematics” (MoMa), which is the basis for the present dissertation, was only possible as a team effort. I am very happy that I have been able to work in a team with Brigitte Brisson and Isabelle Häfner. Thank you both for your company, the collaboration, and our discussions during the last three years! Many other people helped in successfully conducting this project. Particular thanks are due to Katharina Allgaier and Evelin Herbein as well as all our student assistants.

Furthermore, I would like to thank all my colleagues at the Hector Research Institute for Education Sciences, the LEAD Graduate School, and the Cooperative Research Training Group. The shared experiences made the last three years much more enjoyable than they would have been without you. I am especially thankful to Norman Rose for his statistical support as well as to Jenna Cambria, Sara Dallinger, and Cora Parrisius for their help in the last stages of this dissertation.

Lastly, I owe many thanks to my family and friends who were always there for me when I needed them. Thank you all for believing in me, for your understanding, and for taking my mind off things from time to time!

ABSTRACT

Students' value beliefs for subjects such as mathematics are important predictors for their course and career choices (Wigfield, Tonks, & Klauda, 2009). Fostering students' value beliefs for mathematics in school is one way to address the leaking pipeline towards science, technology, engineering, and mathematics (STEM) because mathematics can be seen as a gatekeeper for STEM careers. As females are underrepresented in many STEM disciplines compared to males and tend to show lower motivation in related subjects, they are a particular target group for motivational interventions in STEM. Previous intervention studies have shown that helping students understand the value of the course material can be an effective tool to promote motivation, achievement, and course choices in mathematics and sciences (e.g., Hulleman & Harackiewicz, 2009). However, a number of questions with regards to the effects of such interventions still remain unanswered. The three empirical studies that were conducted within this dissertation address some of these questions. Specifically, the dissertation closely examines (a) the effectiveness of value interventions from a multidimensional perspective on value beliefs and (b) the role of gender for value beliefs in mathematics and how gender differences in these beliefs are affected through value interventions.

Each of the three studies used data from an intervention project that aimed to promote ninth grade students' value beliefs for mathematics. Within a cluster randomized trial, 82 classrooms were randomly assigned to one of two experimental conditions or a waiting control condition. The intervention in both experimental conditions consisted of a 90-minutes session on the relevance of mathematics. Two tasks designed to foster the perceived relevance of mathematics were compared: evaluating quotations and writing a text. To evaluate the effects of the intervention, students' motivation was assessed via self-reports before the intervention as well as six weeks and five months after the intervention.

Using pretest data, Study 1 examined the dimensionality of value beliefs for mathematics and gender differences in these beliefs. Students' responses to a newly developed measure of value beliefs consisting of 37 items were used to investigate how many subfacets of value beliefs could be distinguished empirically and whether gender differences could be found on these facets. Confirmatory factor analyses supported the

differentiation of value beliefs into a total of eleven value facets. Whereas the factor structure was invariant across gender, considerable differences in mean levels favoring boys were found on some but not all value facets.

Applying the newly developed instrument, Study 2 assessed the effects of the intervention on students’ value beliefs in mathematics and the moderating role of gender. The results suggested that both intervention conditions fostered more positive value beliefs in mathematics up until five months after the intervention. Comparing the two intervention conditions, the quotations condition showed stronger and more comprehensive effects on students’ value beliefs than the text condition. Stronger intervention effects were found for those value facets that were targeted in the intervention. When assessing intervention effects separately by gender, evidence for stronger effects for females than for males was found.

Exploring the multidimensional perspective on value more broadly, Study 3 investigated side effects of the intervention in math on motivation in German and English as two verbal subjects. To examine the breadth of effects, students’ self-concept and effort were considered as outcomes in addition to value beliefs. Negative effects on value for German, but not for English, were found five months after the intervention. Additionally, this study took an intraindividual difference perspective in examining intervention effects on differences between these constructs in math, on the one hand, and the two verbal subjects, on the other hand. It was shown that the long-term effects on the difference between math and German value were larger than on math value alone. However, the effects did not generalize to students’ self-concepts and effort in the three subjects considered.

The findings of the three studies are summarized and discussed in light of the broader research context. Implications for future research and educational practice are derived.

ZUSAMMENFASSUNG

Wertüberzeugungen von Schülerinnen und Schülern für Fächer wie Mathematik sind wichtige Prädiktoren für ihre Kurs- und Karriereentscheidungen (Wigfield, Tonks, & Klauda, 2009). Die Förderung von Wertüberzeugungen in Mathematik ist ein möglicher Ansatz, um dem Fachkräftemangel in Mathematik, Informatik, Naturwissenschaften und Technik (MINT) entgegenzuwirken, da Mathematik eine Schlüsselqualifikation für MINT-Karrieren darstellt. Da Frauen in vielen MINT-Disziplinen im Vergleich zu Männern unterrepräsentiert sind und tendenziell eine niedrigere Motivation in entsprechenden Fächern aufweisen, stellen sie eine spezielle Zielgruppe von Motivationsinterventionen im MINT-Bereich dar. Bisherige Interventionsstudien haben gezeigt, dass Motivation, Leistung und Kurswahlen in Mathematik und Naturwissenschaften gefördert werden können, indem Schülerinnen und Schülern der Wert dieser Fächer aufgezeigt wird (z.B. Hulleman & Harackiewicz, 2009). Dennoch bleibt eine Reihe an Fragen hinsichtlich der Effekte solcher Interventionen unbeantwortet. Die drei empirischen Studien, die innerhalb dieser Dissertation durchgeführt wurden, beschäftigen sich mit einem Teil dieser Fragen. Die Dissertation erforscht insbesondere (a) die Effektivität von Interventionen zur Steigerung der Wertüberzeugungen auf diese Überzeugungen aus einer mehrdimensionalen Perspektive und (b) die Rolle des Geschlechts für Wertüberzeugungen in Mathematik sowie die Beeinflussbarkeit von Geschlechtsunterschieden durch entsprechende Interventionen.

Alle drei Studien nutzten Daten eines Interventionsprojekts, das die Förderung der Wertüberzeugungen von Neuntklässlerinnen und Neuntklässlern im Fach Mathematik zum Ziel hatte. Innerhalb einer Cluster-randomisierten Studie wurden 82 Klassen zufällig einer von zwei Experimentalbedingungen oder einer Wartekontrollbedingung zugewiesen. Die Intervention bestand in beiden Experimentalbedingungen aus einer 90-minütigen Unterrichtseinheit zur Relevanz der Mathematik. Zwei Arten von Aufgaben zur Förderung der wahrgenommenen Relevanz der Mathematik wurden verglichen: die Beurteilung von Zitaten und das Schreiben eines Textes. Um die Effekte der Intervention zu evaluieren, wurde die Motivation der Schülerinnen und Schüler vor der Intervention sowie sechs Wochen und fünf Monate nach der Intervention durch Selbstberichte erfasst.

Studie 1 untersuchte anhand von Prätestdaten die Dimensionalität von Wertüberzeugungen für Mathematik und Geschlechtsunterschiede in diesen Überzeugungen. Die Antworten der Schülerinnen und Schüler zu einem neu entwickelten Fragebogen bestehend aus 37 Items wurden genutzt, um zu erforschen, wie viele Subfacetten von Wertüberzeugungen sich empirisch unterscheiden lassen und ob sich Geschlechtsunterschiede in diesen Facetten finden. Konfirmatorische Faktorenanalysen bestätigten die Aufteilung von Wertüberzeugungen in insgesamt elf Facetten. Während die Faktorstruktur sich als über das Geschlecht hinweg invariant erwies, fanden sich bedeutsame Geschlechtsunterschiede zugunsten der Jungen auf einigen, aber nicht allen Facetten.

Studie 2 wandte das neu entwickelte Instrument an, um die Effekte der Intervention auf die Wertüberzeugungen der Schülerinnen und Schüler in Mathematik und das Geschlecht als Moderator dieser Effekte zu untersuchen. Beide Interventionsbedingungen führten bis zu fünf Monate nach der Intervention zu positiveren Wertüberzeugungen in Mathematik. Im Vergleich der beiden Bedingungen zeigten sich in der Zitatebedingung stärkere und umfassendere Effekte auf die Wertüberzeugungen als in der Textbedingung. Besonders starke Interventionseffekte zeigten sich für die Wertfacetten, die im Fokus der Intervention standen. Bei einer getrennten Betrachtung der Interventionseffekte für Mädchen und Jungen zeigten sich tendenziell stärkere Effekte für Mädchen als für Jungen.

Studie 3 erweiterte die mehrdimensionale Perspektive auf Wertüberzeugungen um andere Fächer und untersuchte Nebenwirkungen der Intervention in Mathematik auf die Motivation für Deutsch und Englisch. Um die Breite der Effekte zu überprüfen wurden neben den Wertüberzeugungen das Selbstkonzept und die Anstrengungsbereitschaft als weitere Outcomes in den Blick genommen. Negative Effekte auf die Wertüberzeugungen fünf Monate nach der Intervention wurden für Deutsch, nicht jedoch für Englisch gefunden. Weiterhin wurden in dieser Studie Interventionseffekte im Hinblick auf intraindividuelle Unterschiede zwischen diesen Konstrukten in Mathematik auf der einen Seite und den beiden sprachlichen Fächern auf der anderen Seite betrachtet. Es wurde gezeigt, dass die langfristigen Effekte auf den Unterschied zwischen Mathematik und Deutsch größer als die Effekte auf Mathematik alleine waren. Jedoch zeigten sich die entsprechenden Effekte nicht für das Selbstkonzept und die Anstrengungsbereitschaft in den drei Fächern.

Die Ergebnisse der drei Studien werden im Hinblick auf ihren breiteren Forschungskontext zusammengefasst und diskutiert. Implikationen für die zukünftige Forschung und die Praxis werden abgeleitet.

CONTENT

1

Introduction and Theoretical Framework

1 Introduction and Theoretical Framework

Governments and economies in many industrialized countries have raised concerns about a lack of qualified personnel in science, technology, engineering, and mathematics (STEM) and have called for initiatives to foster achievement and motivation in these domains (e.g., National Science Board, 2007; Plünnecke & Klös, 2009). Compared to males, females are underrepresented in many—but not all—disciplines within STEM and are therefore seen as a particular target group of such initiatives (National Academy of Sciences, 2006; OECD, 2004; Schoon & Eccles, 2014; Watt & Eccles, 2008). In the last years, researchers have developed a number of interventions to foster motivation in mathematics and sciences (for an overview, see Karabenick & Urdan, 2014) and to reduce gender gaps in motivation for these subjects (e.g., Häussler & Hoffmann, 2002; Kerger, Martin, & Brunner, 2011).

One prominent framework to explain academic choices in general as well as gender differences in particular is the expectancy-value theory of achievement related choices by Eccles and colleagues (1983). According to this theory, subjective beliefs about the success expectancy and the value related to a task or activity are the most direct factors influencing achievement-related choices. During the last decades, expectancy-value theory has stimulated an enormous number of empirical studies supporting its basic assumptions: Expectancy and value beliefs have been shown to be important predictors of achievement and achievement-related behaviors in various school subjects as well as course and career choices (for reviews, see Wang & Degol, 2013; Wigfield, Tonks, & Klauda, 2009). Fostering students' expectancy and value beliefs about subjects such as math, thus, seems to be one possible way to address the leaking pipeline towards STEM-related careers at an early stage. Only recently, the results from correlational research using the expectancy-value framework have been translated into interventions in the classroom, which try to help students to understand the value of the course material. Previous studies have shown that such interventions can be an effective tool to promote motivation and performance in STEM courses as well as STEM course choices in high school (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009).

The present dissertation deals with the central question of how student motivation in mathematics as one important precursor of STEM-related careers can be fostered by value interventions. To this end, the dissertation addresses a number of questions that are not only of high relevance for intervention research, but also for motivational research in general. First of all, to be able to assess intervention effects, students' value beliefs need to be measured in an adequate way. A differentiated measurement of value beliefs is proposed in this dissertation to learn more about the structure of value beliefs in general and to better understand existing gender differences in value beliefs, which can be helpful in order to develop targeted interventions. Next, the newly developed instrument was applied in examining the effects of an intervention in mathematics classrooms on students' value beliefs about mathematics. More needs to be known about the strategies that are most successful in fostering value, about the complexity of effects on students' value beliefs and about the possibility to reduce gender differences by such interventions. Last, it is argued that a comprehensive consideration of the effects of interventions in STEM subjects needs to consider motivation for other—particularly verbal—subjects as well. As choices are always made considering different options available, it is not only the motivation for STEM subjects that affects the likelihood that students pursue a STEM-related career, but also the motivation for non-STEM subjects. Therefore, side effects of motivational interventions on non-targeted subjects should be taken into account.

The present dissertation is structured in the following way: The introductory chapter presents the theoretical background for the three empirical studies and aims at situating these studies within their broader research context. In the first section of the introduction, the expectancy-value theory of achievement-related choices will be explained further. The conceptualization as well as the operationalization of value beliefs will be discussed in-depth, including the distinction from other motivational constructs. The most important research findings on the development of expectancy and value within the school context as well as the role of expectancy and value for students' academic development will then be presented. In the second section, possible explanatory factors for gender differences in choices will be investigated and previous findings on gender differences in expectancies and value beliefs will be described. In the next part, previous intervention studies focusing on value will be reviewed and challenges for intervention studies in the classroom setting will be illustrated. The first chapter will conclude by introducing the research questions guiding the three empirical studies. The following chapters will present the three empirical studies realized within this dissertation: The first study presents a differentiated approach to measure value beliefs in math and examines gender differences in these beliefs. The second study assesses how such beliefs are affected by a motivational intervention and whether intervention effects are moderated by student gender. The third study deals with side effects of a motivational intervention in math on motivation in the verbal domain. In the final chapter of this dissertation, the findings of the three empirical studies will be summarized and integrated into a broader conceptual framework. The dissertation closes with a discussion of implications for future research and educational practice.

1.1. Expectancy-Value Theory of Achievement-Related Choices

Expectancy-value theories have a long tradition in motivation research. Starting with Atkinson (1957, 1964), these theories assume that task performance is most directly linked to subjective beliefs about the expectancy and the value related to a task. Whereas early expectancy-value theories tried to explain behavior in the laboratory applying experimental paradigms, modern expectancy-value theories aim at explaining behavior in real-world contexts, mostly using correlational approaches (Trautwein et al., 2013; Wigfield et al., 2009). The most influential modern expectancy-value theory in educational research is the Eccles et al. model of achievement-related choices (1983), which was developed to explain gender differences in the choice of mathematics courses and majors. In comparison to Atkinson’s expectancy-value theory, the expectancy and value component in this model are elaborated further and are connected with a variety of psychological, social, and cultural determinants. Another major difference between this modern approach and traditional expectancy-value theories is the assumed relation between expectancies and values: Eccles et al. assume

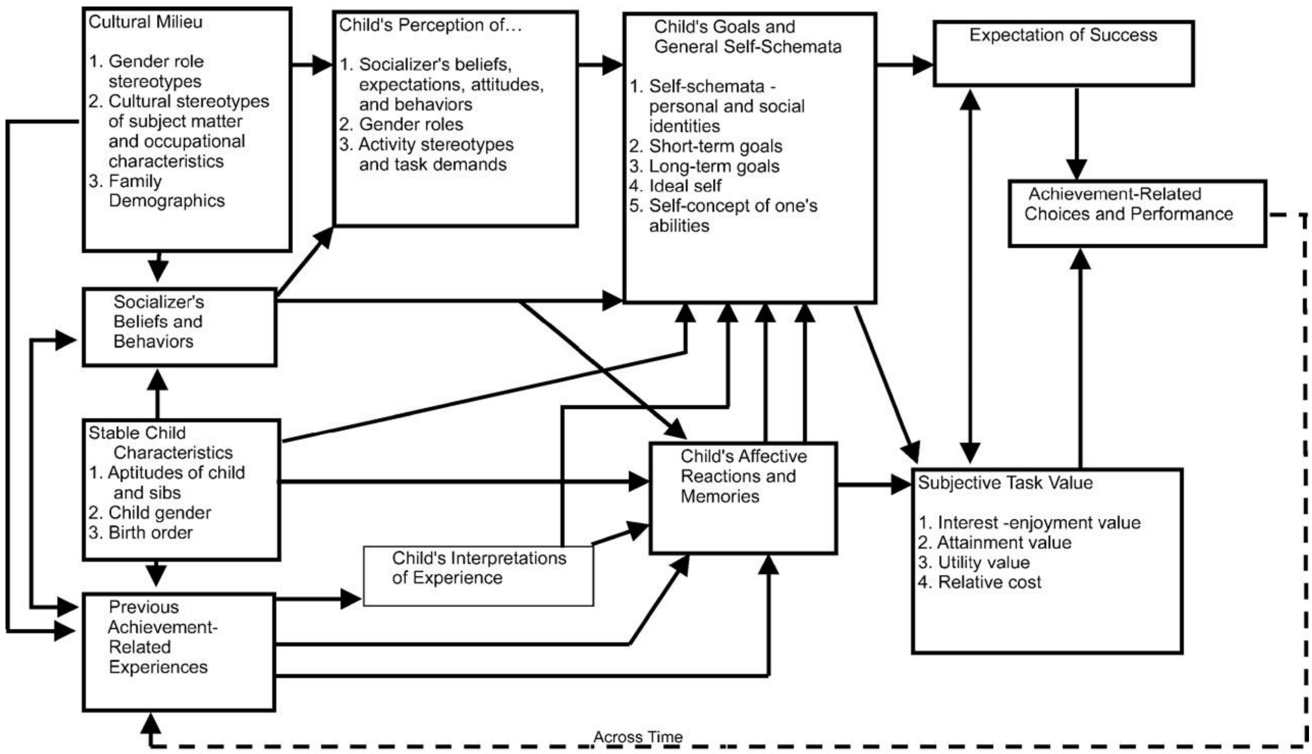

Figure 1.1.1. Eccles et al. expectancy-value theory of achievement-related choices (from Eccles & Wigfield, 2002)

expectancies and values to be positively associated, whereas Atkinson proposed a negative association between them.

Figure 1.1.1 depicts the latest version of the Eccles et al. expectancy-value theory of achievement-related choices (Eccles & Wigfield, 2002; Wigfield et al., 2009). Moving in the model from right to left, expectancies and values are assumed to influence achievement-related choices and performance directly. Expectancies and values themselves are affected by individuals' goals and self-schemata such as competence perceptions as well as by individuals' affective memories for achievement-related events. These beliefs, goals, and affective memories are in turn influenced by individuals' perceptions of the expectations and attitudes of their environment and by their interpretation of previous achievement experiences. Individuals' perceptions and interpretations are finally influenced by a number of social and cultural factors, including the cultural milieu in which they live and the beliefs and behaviors from important socializers such as parents and teachers, as well as by individual aptitudes and experiences. Finally, the model includes a feedback loop across time from achievement-related performance and choices to the experiences an individual makes. The model thus considers a wide range of possible influences on achievement-related choices, including individual as well as structural factors. Effects of the context on students' motivation are mediated by the individual's perception of the environment and the interpretation of experiences. The model thus takes a social-cognitive perspective on student motivation (Pintrich, 2003).

This dissertation focuses largely on the right part of the model; that is, on students' task-specific beliefs. In the following, expectancies and values will therefore be further defined. As the dissertation has its particular focus on value beliefs, the conceptualization of these beliefs will be discussed in more detail along with their empirical operationalization in previous research.

1.1.1. Theoretical conceptualization of expectancy and value beliefs

A number of social-cognitive theories of motivation are concerned with the beliefs that motivate students to learn. According to Pintrich, Marx, and Boyle (1993), these beliefs can be grouped into two broad categories: beliefs about capabilities to do a task and beliefs about reasons for engaging in a task. Expectancy-value theories integrate these two perspectives: Whereas expectancies are linked to the question "Can I do this task?", values have to do with the question "Why should I do this task?". In line with other social-cognitive motivation theories, Eccles et al. (1983) consider both expectancies and values as beliefs about specific tasks, typically referring to different school subjects. Supporting this assumption, it has been shown that both expectancy and value beliefs are highly domain-specific; that is, beliefs in different subjects show relatively low correlations (Bong, 2001a; Eccles, Wigfield, Harold, & Blumenfeld, 1993).

Expectancies for success, on the one hand, are conceptualized as individuals' beliefs about how well they will do on a task in the immediate or long-term future (Eccles & Wigfield, 2002). Expectancies are conceptually related to other constructs referring to self-evaluations of abilities, such as academic self-concept (cf., Marsh & Craven, 1997) and self-efficacy (cf., Bandura, 1997; Pajares, 1996). Within the expectancy-value model, expectancies are determined by ability beliefs, which are defined as individuals' evaluations of their competence in a given domain. Theoretically, ability beliefs (as well as academic self-concepts) refer to beliefs about competencies in broader domains, whereas expectancies for success (as well as self-efficacy) refer to more specific upcoming tasks. However, ability beliefs and expectancies for success have been shown to be highly correlated in empirical studies, and therefore research using the expectancy-value framework typically either collapsed these constructs or used them interchangeably (e.g., Eccles & Wigfield, 1995; Eccles, Wigfield, et al., 1993). Based on their research, Eccles and Wigfield (2002) concluded that these two constructs were empirically indistinguishable in real-world achievement situations. Other authors have, however, shown that ability beliefs and expectancies for success can be differentiated—at least under certain conditions (see Bong & Clark, 1999; Bong & Skaalvik, 2003). As this differentiation is beyond the scope of the present dissertation, the term expectancies is used throughout to refer to all competence-related beliefs.

On the other hand, values are defined as the relative worth of an object or activity and come along with the psychological experience of attraction or repulsion by this object or activity (Higgins, 2007). As Atkinson (1957, 1964) defined value as being directly determined by success expectancies (i.e., as its inverse), values did not have a unique contribution to achievement choices in his model. In contrast, the Eccles et al. expectancy-value theory provides a more elaborate conceptualization of value beliefs that builds upon the work on broad human values by Rokeach (1979), the work on more task-specific values by Battle (1965, 1966) and the integration of these two perspectives by Feather (1982, 1988). Eccles and her colleagues define values as subjective beliefs about specific tasks that affect the individual's desire to engage in these tasks. They distinguish four value components: intrinsic value, attainment value, utility value, and cost (Eccles, 2005; Eccles et al., 1983; Wigfield & Eccles, 1992). *Intrinsic value* is defined as the enjoyment a person derives from engaging in an activity and can therefore be seen as an affective value component. *Attainment value* indicates the personal importance of doing well on a given task and has been linked to identity-related questions such as confirming important aspects of the self. *Utility value* refers to the perceived usefulness of engaging in a task for achieving short- as well as long-term future goals. Finally, *cost* describes all the perceived negative consequences of engaging in a task, including effort and negative emotions associated with the activity itself as well as opportunity costs of choosing one option over another.

The broad conceptualization of value beliefs including multiple components is one of the particular strengths of the Eccles et al. model. This is also one of the major differences between the Eccles et al. model and other modern expectancy-value models, such as Pekrun's control-value theory (2006), which do not provide such a full elaboration of value beliefs. The differentiation between several aspects makes it possible to relate value beliefs to various other motivational theories that also focus on reasons for engagement[1]. As the conceptualization of value beliefs is a central topic of the present dissertation, the dimensionality of value beliefs will be elaborated further in the following. The different value components will first be distinguished from other motivational constructs, after which their measurement in previous research will be described.

1.1.2. Value beliefs and related motivational constructs

Besides the expectancy-value theory, there are several other motivational theories that also focus on different reasons for students' engagement. These

[1] Within motivational research, student engagement is often used as an overarching framework referring to students' schoolwork-related thoughts, behaviors, and feelings (Fredricks, Blumenfeld, & Paris, 2004). Although there is no consensus on the conceptualization of engagement, it can be seen as an outcome of motivation (Martin, 2012).

motivational theories comprise self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000), interest theory (Renninger, Hidi, & Krapp, 1992), goal orientation theory (Dweck & Leggett, 1988; Nicholls, 1984) and future time perspective (Husman & Lens, 1999). The conceptual similarities and differences between value beliefs and the constructs defined by these other motivational theories will be discussed in the following sections.

Value beliefs and intrinsic vs. extrinsic motivation

Self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000) distinguishes between two basic types of motivation: intrinsic and extrinsic motivation. *Intrinsic motivation* refers to doing something for the inherent enjoyment of the activity, whereas *extrinsic motivation* refers to doing an activity for some kind of external reinforcement, such as receiving positive feedback. In the educational setting, it has been found that intrinsic motivation is positively related to learning and performance (for a review, see Ryan & Deci, 2009). However, when examined critically, most behaviors are not purely intrinsically motivated as they occur within a social context, which comes with expectations from others. Self-determination theory therefore assumes a continuum of extrinsic motivation that varies from external to integrated regulation, depending on the degree to which the value and regulation of a behavior have been internalized and—as a further step—integrated in the own self (Ryan & Deci, 2000).

The types of value described in the Eccles et al. expectancy-value theory include intrinsic as well as more extrinsic aspects of motivation (Trautwein et al., 2013). Intrinsic value is similar in certain aspects to intrinsic motivation, whereas utility value shares some characteristics with extrinsic motivation (Wigfield & Cambria, 2010b). When an activity is intrinsically valued, it is done for the experience of doing it; the activity is thus an end to itself. When an activity is performed out of its usefulness, the activity serves for achieving future goals; it is thus a means to an end. However, utility value can also be tied to important personal goals such as attaining a certain occupation (Eccles, 2005). More specifically, Eccles (2005) related the different value components to the types of behavioral regulation as proposed by Ryan and Deci (2000). According to her, intrinsic value comes closest to internalized regulation with engagement in the activity being fully self-determined, attainment value comes closest to integrated regulation with the activity being important for the self, and utility value comes closest to identified regulation because of the link to personal goals. Intrinsic, attainment, and utility value can thus be arranged on a continuum from more internal to more external behavioral regulation. However, it should be acknowledged that these constructs come from different theoretical frameworks which differ with regards to their underlying assumptions (Eccles, 2005; Wigfield & Cambria, 2010b). Whereas self-determination theory juxtaposes intrinsic and extrinsic motivation as two opposite ends of a motivational continuum, expectancy-value theory assumes that different types of value jointly contribute to the total value of a task. Self-determination is more concerned with the quality of student motivation, whereas expectancy-value theory assumes that student engagement is determined by the overall value or the quantity of student motivation (Vansteenkiste, Lens, & Deci, 2006).

Value beliefs and interest

Another influential construct in motivation research that describes students' engagement with different learning materials is interest. In his person-object theory of interest, Krapp (2002) defined interest as a relationship between a person and an object; interest is thus of relational nature and always refers to a specific object, topic or subject area. Two major types of interest can be distinguished: situational and individual interest (Schiefele, 2009). *Situational interest* is conceptualized as a temporary state of focused attention triggered by specific features of a situation, task, or object. *Individual interest* describes a relatively stable tendency to engage with an object of interest. Individual interest is further composed by feeling-related and value-related valences (Krapp, 2002). Feeling-related valences refer to positive emotions that are associated with an object or activity; value-related valences refer to the personal importance attached to an object or activity. Both types of valences are directly related to the object of interest rather than to the relation of this object to other objects and are thus intrinsic in nature. For instance, the personal importance is not based on the significance of the object of interest (e.g., mathematics) for achieving certain outcomes (e.g., good grades). Feeling- and value-related valences have been shown to be highly correlated (Schiefele, 2009). It still seems useful to differentiate between these types of valences, as some individual interests might be based more strongly either on experienced feelings or on personal importance (Schiefele, 2009; Wigfield & Eccles, 1992).

Intrinsic value has often been associated with interest (e.g., Pintrich, 2003; Wigfield & Cambria, 2010b). Within the literature on expectancy-value theory, intrinsic value is often even labeled interest (see Wigfield & Cambria, 2010b). However, as outlined above, the theoretical conceptualization of interest is more complex, including not only affective, but also cognitive components. Using the concept of feeling-related and value-related valences, intrinsic value comes close to feeling-related valences (Schiefele, 2009). Value-related valences, on the other hand, can be related to attainment value—or more precisely, to identity-related notions, where the importance of a task is based on personal reasons (cf., Eccles, 2005). Additionally, value beliefs as conceptualized in expectancy-value theory contain situational as well as more stable aspects (Wigfield & Cambria, 2010b). Values are assumed to vary across specific tasks and situations, but have also been shown to predict long-term engagement and persistence in a given domain. Therefore, intrinsic value can be linked to situational as well as to individual interest. Several scholars have argued that values are beliefs, which can lead to developing interest over time, and are thus seen as antecedents of interest (Hidi & Renninger, 2006; Hulleman, Durik, Schweigert, & Harackiewicz, 2008).

Value beliefs and goals

Students' motivation can also be described in terms of their goals. Goals refer to broader approaches that students show in their achievement-related behaviors and the types of purposes or reasons that direct these behaviors (Elliot, 2005). Achievement goal orientation theory is the most popular form of goal theory in educational research (for a review, see Maehr & Zusho, 2009). It focuses on two types of goals: mastery or task involved goals and performance or ego involved goals (Dweck & Leggett, 1988; Nicholls, 1984). Students who endorse *mastery goals* aim at improving their competence, mastering the material, and understanding the topic; students who endorse *performance goals* strive at demonstrating high achievement to others (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). Within a traditional perspective, students were assumed to either adopt mastery or performance goals with mastery goals being more favorable for students' learning (Dweck & Leggett, 1988; Nicholls, 1984). Current research, however, favors a multiple goals perspective: It has been shown that students can pursue multiple goals simultaneously and that performance goals are not always detrimental to learning (Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002). To explain inconsistent findings with regards to the role of performance goals, it seems important to further distinguish between goals focusing on approaching success and goals focusing on avoiding failure (Elliot, 1999). This distinction has been made with regards to performance as well as mastery goals, but has received more attention regarding performance goals (Hulleman, Schrager, et al., 2010). The adoption of performance-approach goals has been shown to predict higher achievement, whereas the pursuit of performance-avoidance goals has been found to be detrimental for students' learning (Maehr & Zusho, 2009).

Broader perspectives on students' goals proposed by other researchers (Ford, 1992; Wentzel, 1991) also offer valuable insights into the range of goals that might affect task specific values. Ford (1992) described an extensive taxonomy of goals comprising desired intrapersonal outcomes, which have to do with the person him-/herself, and desired person-environment outcomes, which concern the interaction between the person and his/her environment. Wentzel (1991) examined the multiple goals that students pursue in the achievement setting. Her work demonstrates that social as well as academic goals predict students' performance and behavior (see Wentzel, 2005).

All these types of goals can be seen as determinants of students' task-specific values. Goal orientations are conceptualized as broader beliefs, which can influence how students approach specific tasks and the value they perceive in these tasks (Eccles, 2005; Hulleman et al., 2008; Maehr & Zusho, 2009). Wigfield and Eccles (1992) discuss how pursuing mastery vs. performance goals might be connected with intrinsic value on the one hand and attainment and utility value on the other hand. In addition, different tasks provide students with opportunities to demonstrate and fulfill their personal goals (Eccles, 2005). An understanding of students' goals in terms of their content is therefore especially important for Eccles' notion of attainment and utility value. As soon as multiple goals conflict with each other in the classroom setting, cost also comes into play (cf., Boekaerts, 2009).

Value beliefs and future time perspective

Another motivational construct related to value beliefs is future time perspective (Husman & Lens, 1999; Nuttin & Lens, 1985). The authors who advanced this theory have emphasized the role of future for student motivation: As students know that education should primarily prepare them for the future, they are more likely to be motivated if they perceive their current educational engagement as useful to them in the longer term. More precisely, future time perspective has been defined as “the degree to which and the way in which the chronological future is integrated into the present life-space of an individual through motivational goal-setting processes” (Husman & Lens, 1999, p. 114). Individuals are assumed to vary in their future time perspectives, which can be characterized by their extension (i.e., the time span for which goals are set), their density (i.e., the number of goals), and their realism (i.e., the degree to which these goals are realistic). A number of studies have shown that perceived instrumentality (i.e., the value of present activities to the future) predicts a more positive motivation, self-regulated behavior, and higher achievement (Husman & Lens, 1999). However, it has been noted that the relation between instrumentality and motivation is complex and depends on the type of instrumentality: If students understand that gaining competence on certain tasks is useful for reaching their personal goals (e.g., working in a specific job) and not only for overcoming obstacles (e.g., getting into college), this can promote more intrinsic motivation (Husman, Derryberry, Crowson, & Lomax, 2004; Husman & Lens, 1999; Simons, Vansteenkiste, Lens, & Lacante, 2004). Furthermore, students’ future time perspectives can be distinguished in terms of the domains they refer to. Peetsma and van der Veen (2011) established future time perspectives for the life domains school and professional career, social relations, and leisure time. Whereas students with a long-term perspective in leisure time showed negative trajectories in their investment in learning, students with long-term perspectives in school and professional career as well as social relations showed positive trajectories.

With respect to the aspects of value defined by Eccles and colleagues, there seems to be most overlap between future time perspective and utility value (Husman et al., 2004; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). However, future time perspective makes a clearer distinction between different time orientations (see Husman et al., 2004). Although utility value has been defined as the usefulness of engagement in a task for short- as well as long-term goals, these time orientations have been either neglected or mixed in its operationalization. The distinction between perceived instrumentality and valuing specific tasks has also been discussed more generally (Miller & Brickman, 2004; Wigfield & Cambria, 2010a). Instrumentality is seen as a broader belief affecting how specific tasks are valued and is more directly tied to future goals.

Conclusion

The four value components are, thus, not only separately defined in the expectancy-value model by Eccles and colleagues (1983), but can also be related to different constructs defined in other motivational theories. This broad spectrum of value beliefs included in expectancy-value theory contributes to a high power in predicting academic choices (see Wigfield et al., 2009). The inclusion of cost as one crucial determinant of choices (Eccles et al., 1983) seems to be rather unique within motivational research and can be seen as a further strength of the model.

1.1.3. Empirical measurement of value beliefs

The measurement of value beliefs in empirical research has not fully covered this theoretical richness of value beliefs. Few studies using the Eccles' et al. model as theoretical framework have incorporated separate scales for all four components (for exceptions, see Conley, 2012; Trautwein et al., 2012). Instead, several alternative strategies have been adopted: Many studies combined the positive value aspects (i.e., intrinsic, attainment, and utility value) into one general value scale consisting of a small number of items (e.g., Eccles, Wigfield, et al., 1993; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002); other studies collapsed two of the value components (often importance as a combination of attainment and utility value; e.g., Durik, Vida, & Eccles, 2006; Watt et al., 2012) or used only one component as an indicator of task value (e.g., intrinsic value; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006). Although the importance of cost for choices is emphasized in expectancy-value theory, this component has been included less frequently in the measurement of value beliefs. In their review on the perspectives for research based on expectancy-value theory, Wigfield and Cambria (2010a) suggested a further exploration of this component as one major line for future research.

Partly, previous studies were simply not able to separate the value components (e.g., Eccles, Adler, & Meece, 1984; Eccles, Wigfield, et al., 1993). Although various factors might contribute to the empirical separability of different value aspects (e.g., students' age), the set of items that is used to assess value beliefs is certainly one important factor. The items that were used in different studies were not always consistent. Table 1.1.1 presents an exemplary set of measures for value beliefs that were used in different studies with sample items (for a review on measures of task value, see also Wigfield & Cambria, 2010b). Eccles and colleagues developed measures of intrinsic, attainment, and utility value with two to three items each (for the full set of items, see Wigfield et al., 1997). These items refer to the value of domains such as math, reading, sports, or music and have been used in most of their research in this or an adapted form. Conley (2012) combined this set of items with newly developed ones to explicitly cover all four value components, including two items referring to cost. Trautwein et al. (2012) also included separate scales for all four value components in their study, using a total of twelve items on value beliefs. Both studies by Conley (2012) and Trautwein et al. (2012) were able to separate the four value components empirically. The most broadly applied questionnaire to assess motivational beliefs as well as self-regulated learning strategies in various setting is the Motivated Strategies for Learning Questionnaire (MSLQ), which was developed by Pintrich and colleagues (Pintrich & De Groot, 1990; Pintrich, Smith, Garcia, & McKeachie, 1993). The questionnaire is designed to refer to a specific class and includes a scale on task value with the items referring to intrinsic, attainment, and utility value aspects. Similarly, Hulleman et al. (2008) used measures of intrinsic and utility value that referred to a specific course and separated the value of this course from interest and goals referring to psychology as a domain. Recently, several researchers (A. Battle & Wigfield, 2003; Luttrell et al., 2010; Perez, Cromley, & Kaplan, 2014) developed elaborate measures of cost, particularly for university students. Perez et al. (2014) explicitly distinguished between three types of cost: effort cost, opportunity cost, and psychological cost.

Table 1.1.1

Measures of task value with sample items

Authors and scale title	Sample item
Wigfield et al. (1997)	
Interest value	How much do you like doing math?
Attainment value	For me, being good in math is (not at all important, very important).
Utility value	In general, how useful is what you learn in math?
Conley (2012)	
Interest value	Math is exciting to me.
Attainment value	Thinking mathematically is an important part of who I am.
Utility value	Math concepts are valuable because they will help me in the future.
Cost value	I have to give up a lot to do well in math.
Trautwein et al. (2012)	
Intrinsic value	I enjoy puzzling over mathematics problems.
Attainment value	Mathematics is important to me personally.
Utility value	I'll need good mathematics skills for my later life (training, studies, work).
Cost value	I'd have to sacrifice a lot of free time to be good at mathematics.
Pintrich & De Groot (1990)	
Task value	It is important for me to learn what is being taught in this class.
Hulleman et al. (2008)	
Intrinsic value	Lectures in this class are entertaining.
Utility value	What I am learning in this class is relevant to my life.
Battle & Wigfield (2003)	
Intrinsic-attainment value	I'm excited about the idea of going to graduate school.
Utility value	I don't think a graduate degree will be very useful for what I want to do in the future.
Cost	Getting a graduate degree sounds like it really requires more effort than I'm willing to put into it.
Perez et al. (2014)	
Effort cost	When I think about the hard work needed to get through my science major, I am not sure that getting a science degree is going to be worth it in the end.
Opportunity cost	I'm concerned my science major may cost me some treasured friendships.
Psychological cost	I'm concerned that I won't be able to handle the stress that goes along with my science major.

Several aspects can be noted when comparing the measures of task value across studies and research groups. First, the various measures imply different levels of task-specificity. Whereas expectancy-value theory conceptualizes values as task-specific beliefs, most research uses value measures referring to broader subjects (e.g., math) as domains. Some studies also assessed the value of specific courses (e.g., Hulleman et al., 2008; Pintrich & De Groot, 1990) or particular education forms (e.g., graduate school; A. Battle & Wigfield, 2003). Second, different measures of the same construct only partially overlap in the value aspects that are tapped by the wording of the items. The items could sometimes also be used as measures of related motivational constructs such as interest (e.g., "Mathematics is important to me personally."; Trautwein et al., 2012), whereas other items seem to capture qualitatively different aspects (e.g., "For me, being good at math is [not at all important, very important]"; Wigfield et al., 1997). Such inconsistencies in the operationalization of constructs have been described as jingle-jangle-fallacies (Marsh, Craven, Hinkley, & Debus, 2003): Scales with the same name do not always reflect the same construct, and scales with different names do not always reflect different constructs. In their review on the measurement of task value, interest, and goal orientations, Wigfield and Cambria (2010b) note similar problems and call for more research on the empirical distinctiveness of motivational constructs.

1.1.4. Development of expectancy and value beliefs

The state of research with regards to the development of expectancy and value beliefs can be summarized along the following questions (cf., Wigfield et al., 2009): How does the structure of expectancy and value change across age? How does the level of these beliefs change with students' age? Which factors influence the development of these beliefs? In line with the focus of the dissertation, the summary of the empirical evidence with regards to these questions focuses on value beliefs. However, expectancy and value beliefs are assumed to be shaped through the same processes so that they do not develop independent from each other. Therefore, research on the development of expectancy and value beliefs is presented jointly in this paragraph.

Concerning the development of the structure of expectancy and value beliefs, Eccles, Wigfield, and colleagues (Eccles, Wigfield, et al., 1993; Eccles & Wigfield, 1995) conducted factor analyses on responses to items assessing expectancy and value beliefs for students of different ages. They found that students distinguish between expectancy and value within one domain (e.g., mathematics) as well as between these beliefs across domains (e.g., value for mathematics vs. reading) from the beginning of elementary school on. Whereas the beliefs across subjects tend to become more distinct over time, the association between expectancy and value beliefs within one domain seems to increase with students' age (Denissen, Zarrett, & Eccles, 2007; Wigfield et al., 1997). This increasing association has been explained as an effect of students coming "to value what they are good at" (Wigfield et al., 2009, p. 61). Effects in the other direction (i.e., value beliefs affecting expectancies) are also plausible as choices based on values can lead to higher achievement and expectancies (Wigfield & Eccles, 1992). However, more support in empirical research has been found for effects of expectancies on the development of value beliefs (Jacobs et al., 2002; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). So far, less work has examined the development of the structure underlying multiple value components. From their factor-analytic studies with elementary and secondary school students (Eccles, Wigfield, et al., 1993; Eccles & Wigfield, 1995), Wigfield et al. (2009) conclude that the value components can be separated from fifth grade on. On a theoretical basis, it has been discussed how the concept of task value might change as students get older: Whereas subjective task value for younger students might be centered heavily around intrinsic enjoyment, older students might also consider the aspect of usefulness for future goals (Wigfield & Eccles, 1992; Wigfield, 1994).

How do mean levels of expectancy and value beliefs develop? Longitudinal studies with samples from different countries (such as the United States, Australia and Germany) show a consistent pattern in different academic subjects: Expectancy and value-related beliefs decrease from elementary school years onwards (Wigfield et al., 1997), and this downward-trend continues into secondary school (Fredricks & Eccles, 2002; Frenzel, Goetz, Pekrun, & Watt, 2010; Jacobs et al., 2002; Watt, 2004). Several possible explanations have been offered for these findings (for an overview, see Wigfield et al., 2006). Focusing on the development of expectancies, children seem to have overly optimistic beliefs about their levels of competence when they are young and become much more realistic in evaluating their own achievement when they grow up (Dweck & Elliott, 1983; Stipek & Mac Iver, 1989). This more realistic competence appraisal will lead to a decrease in students' expectancies and, as a consequence, in values. Interest theory further suggests that interests differentiate naturally through identity formation processes as children get older (Krapp, 2002). Intraindividual differences in interests thus become more pronounced with some interests remaining high and others going down, and this can explain the negative development in the average level of subject specific interests. Furthermore, social interests seem to increasingly compete with students' school-related interests during adolescence (Hidi & Ainley, 2002). In addition to such natural developmental factors, some researchers have argued that the lack of fit between students' developmental needs and the school environment contributes to a decline in expectancies and values (Eccles, Midgley, et al., 1993). The interpretation of mean-level changes in expectancies and values, however, relies on the assumption that measures of these beliefs assess the same underlying constructs across time. This assumption is not always tested in empirical studies; there is, however, some support for structural changes in motivational constructs (Frenzel, Pekrun, Dicke, & Goetz, 2012).

Which psychological and social factors influence the development of expectancy and value beliefs? Although there is more literature available on the factors involved in the development of expectancy beliefs, the same factors are assumed to also affect how students' value beliefs develop, and much can also be learned from the work on the development of interest (see Wigfield et al., 2006, 2009). Generally, expectancies and values are developed through experiences with different tasks, which can be made in various contexts. Students use the feedback provided by important socializers such as parents and teachers to build their beliefs about the expectancies and values of different tasks (Eccles, 2007). More broadly, expectancies and values are also influenced by cultural norms (Eccles, 2005). All these experiences with different tasks provide students with a set of different comparisons that they can use as sources of information (see Marsh, 1986; Möller & Marsh, 2013). First, students engage in social or external comparisons; that is, they compare their own abilities, and probably also their interests, with those of others. Second, students rely on dimensional or internal comparisons; that is, they compare their own ability or interest in one domain (e.g., math) with their ability or interest in another domain (e.g., language arts). Although these comparison processes have been mainly investigated for students' expectancies, there is also some support for their role in the development of values and related constructs (Goetz, Frenzel, Hall, & Pekrun, 2008; Nagy et al., 2006; Schurtz, Pfost, Nagengast, & Artelt, 2014). In particular, support for dimensional comparison effects has been found between mathematical subjects on the one hand and verbal subjects on the other hand (Marsh, 1986; Möller & Marsh, 2013).

1.1.5. Educational relevance of value beliefs

Value beliefs—as well as expectancies—are of high educational relevance as there is ample evidence that they predict important student outcomes, such as effort, persistence, and achievement, in various school subjects (e.g., Marsh et al., 2005; Nagengast, Trautwein, Lüdtke, & Kelava, 2013; Trautwein & Lüdtke, 2007) as well as academic choices (e.g., Durik et al., 2006; Nagy et al., 2006; Simpkins, Davis-Kean, & Eccles, 2006). Although expectancies and values are both associated with achievement and choices, a differential pattern emerges when their unique predictive effects are examined. Whereas expectancies are stronger predictors for achievement (Marsh et al., 2005; Meece, Wigfield, & Eccles, 1990; Trautwein et al., 2012), value beliefs are especially important for predicting academic choices (Bong, 2001b; Meece et al., 1990). The predictive power of these beliefs also holds in longitudinal studies. Value beliefs are not only associated with concurrent enrollment intentions and career aspirations, but value beliefs also predict actual choices at later time points, such as course enrollment in high school (Durik et al., 2006; Nagy et al., 2006).

It has been argued that the predictive power of value beliefs could even be increased by including separate measures for all value components because these components might differentially predict outcomes such as persistence and choice and predictive patterns might also vary across age and context (Trautwein et al., 2013; Wigfield & Eccles, 1992). As only few studies have included separate components so far, the empirical evidence for such a differential predictive validity is limited, but all value components have been associated with students’ choices. Intrinsic value has been found to predict related extracurricular activities (Durik et al., 2006; Nagengast et al., 2011). Furthermore, intrinsic, attainment, and utility value have all been associated with course choices and career aspirations (Durik et al., 2006; Nagy et al., 2006; Watt, 2006). Watt et al. (2012) found that the specific role of expectancies and different types of values for predicting high school math participation, educational aspirations, and career plans varied to some extent with the context (Australia, Canada, or the US) and student gender. Some studies also assessed the predictive validity of cost. Battle and Wigfield (2003) assessed female college students’ intentions to attend graduate school and found a negative effect of cost together with positive effects of the other value components. In another study with college students, Perez et al. (2014) could show that perceptions of cost predicted students' intentions to leave their STEM major over and above other aspects of value. Their results further supported the differential validity of different kinds of cost with effort cost showing stronger effects than opportunity cost and psychological cost.

Most of this research has investigated the effects of domain-specific expectancies and values on achievement or choices in matching domains. However, in her expectancy-value theory of achievement-related choices, Eccles (2009, 2011) emphasizes that choices are always made considering the options that are available. Therefore, the expectancies and values for all of these options come into play. The rank ordering of expectancies and values or intraindividual hierarchies of these beliefs across subjects are crucial for understanding why students choose one educational option instead of another. This assumption has been supported by several recent studies. For instance, Chow, Eccles, and Salmela-Aro (2012) used latent profile analyses to show that priority patterns of task values in several subjects predicted aspirations towards physical and information technology related sciences for samples in the US as well as in Finland. Using a different analytical strategy, Parker et al. (2012) could show that choosing a math-intensive major was predicted positively by math self-concept and negatively by verbal self-concept (and vice versa for verbal-intensive majors). Similarly, Nagy et al. (2006) found that self-concept and intrinsic value for math and biology were positively associated with choosing advanced courses in the same domain, but negatively associated with choosing courses in the respective other domain.

1.2. Using the Expectancy-Value Theory to Explain Gender Differences in Choices

Across a wide range of countries, males and females differ notably in how often they pursue degrees and careers in certain STEM fields (National Science Foundation, 2011; OECD, 2004). Precursors for such gendered career pathways can be found in secondary school, where males and females already differ in their career aspirations and—given that they have the choice—in the math and science courses they take (Schoon & Eccles, 2014; Watt & Eccles, 2008; Watt et al., 2012). To describe this phenomenon, several authors have used the metaphor of a leaky pipeline towards STEM (e.g., Watt & Eccles, 2008): People drop out of a pathway towards STEM-related careers at various time points and females are generally more likely to drop out than males. Mathematics is seen as the critical filter in this pipeline because math courses in secondary school affect the career options that one has at a later time point (Watt & Eccles, 2008).

The Eccles et al. expectancy-value theory (1983) is a very prominent approach to explain such gender differences in choices (e.g., Chow & Salmela-Aro, 2011; Nagy et al., 2008; Watt et al., 2012). The most proximal factors that are assumed to explain gender differences in academic choices are expectancy and value beliefs for related subjects—with a particular focus on value beliefs as the driving force of choices (Eccles, 2005, 2009, 2011). These beliefs, in turn, are supposed to be affected through gendered socialization processes. Males and females make different experiences related to their gender role. As long as this gender role is a central part of their identity, these experiences will lead to males and females having different expectancies and values. For instance, the cultural definition of gender roles can affect the priorities in the long-term goals that males and females pursue. Males' and females' expectancies and values are also supposed to be shaped by their experiences with parents, teachers, and peers who might provide them with different feedback on their opportunities.

Whereas expectancy-value theory focuses on socialization processes affecting motivational beliefs in explaining gender differences in choices, several alternative explanatory factors have been proposed (for reviews, see Ceci, Williams, & Barnett, 2009; Wang & Degol, 2013). Ceci et al. (2009) provide a comprehensive review on the factors driving the underrepresentation of women in math-intensive fields, considering biological as well as sociocultural causes. Based on the evidence considered in this review, they suggest that women's preferences are the most powerful explanatory factor. In addition, they consider gender differences on gatekeeper tests such as the mathematics section of the Scholastic Assessment Test (SAT) in the US (especially at the right tail of the distribution) as one important factor. Based on the empirical evidence, they conclude that such gender differences in achievement are more likely caused from sociocultural than from biological factors. Gender differences in mathematics achievement (at the mean level as well as in the distribution), however, can only partially explain the female underrepresentation in STEM-related careers (Ceci et al., 2009; Wang & Degol, 2013). In another review on the factors explaining gendered career choices, Wang and Degol (2013) stress the role of occupational preferences and lifestyle values for women's underrepresentation in STEM and illustrate the sociocultural influences on STEM choices. Expectancy-value theory seems to be an especially powerful explanatory framework for gender differences in choices as it considers a wide range of contributing factors.

According to expectancy-value theory, gendered socialization should result in boys reporting more favorable expectancy and value beliefs in male-typed domains such as mathematics and girls reporting more favorable expectancy and value beliefs in female-typed domains such as languages. The empirical evidence mostly confirms this pattern of gender differences in expectancy and value beliefs. With regards to expectancies, it has been consistently found across diverse samples that boys rate their expectancies in mathematics higher than girls—regardless of their abilities (e.g., Jacobs et al., 2002; Marsh et al., 2005; Nagy et al., 2010). Although similar gender differences can be found across different Western cultures (Nagy et al., 2010), this gender effect seems to be culturally shaped as it varies considerably across nations all over the world (Else-Quest, Hyde, & Linn, 2010). Females, on the other hand, tend to report higher expectancies for more female-typed domains such as language arts and foreign languages, although this female advantage has not always been found (e.g., Durik et al., 2006; Jacobs et al., 2002; Nagy et al., 2008; Watt, 2004). Again, it needs to be noted that such gender stereotypic differences in subjective beliefs cannot fully be explained by boys' and girls' achievement. In fact, for grades—which seem to more important for students' expectancies than achievement tests (see Marsh et al., 2005)—females tend to earn higher grades in almost all school subjects. A recent meta-analysis (Voyer &

Voyer, 2014) reported that this female advantage was largest in language arts (*d* = 0.374) and smallest for math (*d* = 0.069). To explain this somewhat paradoxical pattern, it has been argued that females tend to underestimate their abilities in comparison to males (Stetsenko, Little, Gordeeva, Grasshof, & Oettingen, 2000).

For value beliefs, the pattern of results is not straightforward for math. Whereas some studies reported higher values in mathematics for boys (e.g., Marsh et al., 2005; Steinmayr & Spinath, 2008), others reported no difference between boys and girls (e.g., Jacobs et al., 2002; Meece et al., 1990; Wigfield et al., 1997). These inconsistencies can partly be explained by the operationalization of value beliefs and differences in the value dimensions incorporated (i.e., one general value scale or separate value components). Studies that considered separate value components found that male adolescents reported higher scores on measures of math interest and intrinsic value (Frenzel et al., 2010; Frenzel, Pekrun, & Goetz, 2007; Marsh et al., 2005; Watt, 2004), whereas gender differences in attainment and utility value are more inconsistent and seem to depend on the specific operationalization (Frenzel et al., 2007; Meece et al., 1990; Steinmayr & Spinath, 2010; Watt, 2004; Watt et al., 2012). Few studies, though, included measures of all value components, making direct comparisons of the results across studies difficult. Other moderators such as the age group or the cultural background of the sample are also possible. For the area of languages, on the other hand, girls were consistently found to place higher values on language arts and foreign languages (Durik et al., 2006; Jacobs et al., 2002; Nagy et al., 2008; Watt, 2004).

Several studies addressed the question of how such gender differences in expectancy and value beliefs develop over the school years (Frenzel et al., 2010; Jacobs et al., 2002; Nagy et al., 2010; Watt, 2004). Stereotypic gender differences have been shown to already emerge at the beginning of elementary school and to remain relatively stable throughout the school years. Little evidence, thus, supports an intensification of gender differences as hypothesized by gender role socialization perspectives (Eccles, 1987; Hill & Lynch, 1983). However, socialization processes in the school and at home have been found to contribute to gender differences in children’s beliefs (Eccles, 2007; Jacobs, Davis-Kean, Bleeker, Eccles, & Malanchuk, 2005; Wang & Degol, 2013): Parents and teachers show gender-dependent behavior that can shape boys’ and girls’ expectancies and values; they communicate different expectations for boys and girls, provide them with gender-typed experiences, and act as role models.

To summarize, expectancy-value theory is a very powerful approach for explaining gender differences in choices. A particular emphasis in the literature has been put on expectancy and value beliefs in mathematics as a precursor of choosing a career within STEM. Although values are assumed to be the driving force for choices, the gender differences regarding math that have been found in previous research seemed to be more pronounced for expectancies than for value beliefs. More research is therefore needed to explain this seemingly inconsistent pattern of results.

1.3. Intervening on Students' Value Beliefs

Most evidence supporting expectancy-value theory stems from correlational research. This research has clearly demonstrated the role of expectancies and values for predicting student outcomes such as career-related choices (see chapter 1.1.5). Furthermore, it has been shown that expectancy and value beliefs are shaped through the learning environment (see chapter 1.1.4). During the last years, researchers have started to translate these research findings into targeted educational interventions (e.g., Hulleman, Godes, et al., 2010; Hulleman & Harackiewicz, 2009). Such intervention studies try to manipulate one central component of student motivation (e.g., value) and evaluate the effects on student outcomes. Intervention research testing motivation theory in the field is a prime example for use-inspired basic research as it provides valuable information for theory as well as for practice (Pintrich, 2003). From a theoretical perspective, experimental studies can provide insight into causal relationships in a way that observational research cannot (Shadish, Cook, & Campbell, 2002). From a practical perspective, intervention research develops and tests teaching practices that can help in fostering student outcomes (Hulleman & Barron, in press).

Previous interventions grounded in expectancy-value theory aimed at helping students to understand the value of the course material for their lives. These interventions focused on the value rather than the expectancy part of the model as it has been argued that value beliefs might be more amenable to external interventions than expectancies (Harackiewicz, Tibbetts, Canning, & Hyde, 2014; for a meta-analysis on the effects of self-concept interventions, see O'Mara, Marsh, Craven, & Debus, 2006). Specifically, previous value interventions focused on utility value as one of the value components. This component is more extrinsic in nature than the other components (Eccles & Wigfield, 2002), and therefore seems to be more easily influenced from the outside. Directly triggering enjoyment or identification with a task may be difficult as these value aspects seem to depend strongly on characteristics and preferences of the individual student (Eccles, 2005). However, it might be easier for educators to foster reflections on the usefulness and relevance of the learning material for students' lives (see also Brophy, 1999). Several authors have argued that initial stimulation of a more extrinsic motivation should lead to enhanced engagement and can therefore also promote students' intrinsic motivation and interest development over time (Hidi & Harackiewicz, 2000; Hidi & Renninger, 2006).

1.3.1. Previous value intervention studies

Recently, a number of value interventions have been tested in the laboratory as well as in the classroom (for reviews, see Durik, Hulleman, & Harackiewicz, in press; Harackiewicz et al., 2014). These interventions applied different strategies to promote students' perceptions of utility value. One strategy was to provide students with information on the usefulness of the learning material, whereas the other strategy was to encourage students to generate arguments for the utility of the learning material themselves. Providing students with information on the utility of a task seems to be what happens naturally in the classroom, when teachers discuss the relevance of a particular topic. This strategy has been applied in a number of laboratory studies with college students (Durik & Harackiewicz, 2007, study 2; Durik, Shechter, Noh, Rozek, & Harackiewicz, 2015, study 1; Shechter, Durik, Miyamoto, & Harackiewicz, 2011). In these studies, students in both the experimental and the control group learned a new math technique using standardized instructional material. In the experimental group, this instructional material contained information about the usefulness of this math technique for achieving goals in the short term (i.e., for everyday life situations) or in the long term (i.e., for future career opportunities). The effects of this utility information depended on students' initial motivation: The utility value information promoted interest and achievement for students with high initial motivation, whereas it tended to undermine interest for students with low initial motivation. This pattern of results was consistently found in previous studies, although several moderators were examined as indicators of initial motivation (i.e., interest or success expectancies). Furthermore, Shechter et al. (2011) assessed how the effects of different types of utility information (i.e., short- vs. long-term) depended on students' cultural background. Western students benefited more from information about the usefulness in the short term, whereas East Asian students benefited more from long-term utility value information. The effects of utility information on students' interest in the task were shown to be mediated by competence valuation, task involvement, and perceived competence (Durik & Harackiewicz, 2007).

The second intervention strategy, encouraging students to self-generate utility value, was applied in the laboratory as well as in the classroom (Hulleman, Godes, et al., 2010; Hulleman & Harackiewicz, 2009). In these studies, students were randomly assigned to a relevance or a control condition. Students in the relevance condition were asked to make connections between the learning material and their lives, whereas students in the control condition were asked to simply summarize the learning material. Again, various specific techniques were used to generate connections. In a study with high school science students (Hulleman & Harackiewicz, 2009), students in the relevance condition wrote a total of eight essays during the semester about the meaning of the course material to their lives. Hulleman, Godes, et al. (2010) implemented similar interventions in two studies with college students. In a first study in the laboratory, participants were asked to write an essay on the relevance of a math technique to their lives. In a second study in the psychology classroom, two types of each relevance and control conditions were implemented which all included writing two essays on current course topics at different times during the semester. In the relevance conditions, students either wrote a letter to another person describing the relevance of this topic or wrote an essay about the relevance of a media report to the topic covered in class. In the control conditions, students either wrote a summary of the topic or discussed two study abstracts related to the topic. In all of these studies, the relevance conditions showed positive effects on student outcomes compared to the control conditions for students with low success expectancies (Hulleman, Godes, et al., 2010, study 1; Hulleman & Harackiewicz, 2009) or initial performance (Hulleman, Godes, et al., 2010, study 2). Effects were shown on utility value, interest, and sometimes also on course grades. The relevance interventions were, however, not effective for students with high expectancies or high initial performance.

These different intervention strategies, thus, yielded contrary results regarding the group of students for which they worked. Whereas the first strategy (i.e., providing utility value information) was only successful for students with high initial motivation, the second strategy (i.e., self-generating utility value) was more beneficial for students with low expectancies. However, different intervention strategies were not systematically compared to date. As both strategies worked for some groups of students, one might argue for combining different strategies. However, detrimental effects that were found for lowly motivated students when they were provided with utility value information should be taken into account. It seems as if utility information is threatening for students who do not expect to do well (Durik et al., in press). Durik et al. (2015, study 2) recently tested a confidence boost manipulation in a laboratory study to buffer against such detrimental effects. Before learning a math technique, participants in this condition were told that they had good potential to learn the technique. This encouraging feedback combined with the information on utility led to increased competence valuation, task involvement, and situational interest for students with low expectancies.

In addition to these interventions which directly targeted students' utility value beliefs, parents' utility value beliefs have been addressed as a more indirect way to foster students' motivation. Harackiewicz et al. (2012) designed an intervention aiming at helping parents to motivate their children for mathematics and science courses in high school. The intervention consisted of information material on the usefulness of STEM courses (i.e., two brochures and a web site) that was sent to parents. Students whose parents were in the intervention condition took more STEM courses in the last two years of high school than students in the control condition. This effect on STEM course-taking was mediated through changes in mothers' and students' STEM utility value. Rozek, Hyde, Svoboda, Hulleman, and Harackiewicz (2015) further investigated the factors that moderate this intervention effect. They found that the intervention was most beneficial for high-achieving girls and low-achieving boys. The intervention, however, did not affect high-achieving boys and tended to negatively affect low-achieving girls. To avoid negative effects for low-achieving female students, who may have particularly low self-concepts, Rozek and colleagues call for taking the role of expectancies into account when designing future interventions.

To conclude, previous studies have shown that utility value interventions can be effective in increasing student motivation and performance in the targeted domains. However, the pattern of effects is inconsistent with regards to the groups of students which can be reached by these interventions. More research is therefore needed that systematically compares different intervention strategies.

1.3.2. Designing motivational intervention studies in the classroom

Several aspects need to be considered when designing intervention studies in the classroom setting. One important consideration for interventions in the classroom setting is how they can be designed to be maximally effective. Generally, one can assume that interventions in the classroom setting will mostly show small effects as the outcomes of such interventions (i.e., student achievement or motivation) are multiply determined (cf., Hattie, 2008). One possibility to get larger intervention effects might be to increase the intervention strength in terms of dosage or frequency. At the same time, it seems important from a practical perspective to develop short, yet effective interventions. Psychological interventions often rely on affecting recursive processes in educational settings so that even short interventions can show lasting effects (Yeager & Walton, 2011). As previous value intervention studies varied with regards to the frequency of exposure to the intervention, it is not clear how much it takes to promote students' value beliefs in the classroom setting. In comparison to targeted motivational interventions such as the value interventions described here, motivation researchers have also developed multi-component interventions (e.g., Concept-Oriented Reading Instruction; Guthrie et al., 2004). Whereas interventions targeting multiple motivational constructs may be more effective, it is harder to determine the factors that are primarily responsible for the effects for these types of interventions (Hulleman & Barron, in press).

Another important question is the level at which the intervention is assigned (Plewis & Hurry, 1998). Students are nested within classrooms within schools, and interventions can be assigned at all of these levels (i.e., student, class or school). The level of assignment has consequences for the intervention designs that can be used as well as for the evaluation of these interventions. For randomized experiments, the basic distinction from a multilevel perspective is whether the randomization is done at the individual or the cluster level. Previous value intervention studies in the classroom randomized at the individual level (Hulleman, Godes, et al., 2010; Hulleman & Harackiewicz, 2009). In such a design, students within one class are in different experimental conditions. As this might increase the likelihood that students in the intervention and the control condition interact with each other, diffusion effects (i.e., the control group is affected by the intervention) can occur more easily (Craven, Marsh, Debus, & Jayasinghe, 2001; Plewis & Hurry, 1998). Such diffusion effects represent a threat to the internal validity of a study and lead to biased estimates of intervention effects (see also Shadish et al., 2002). It is, however, difficult to predict the direction and size of such biases as they depend on how students deal with being in different experimental conditions. Plewis and Hurry (1998) additionally illustrated how the integrity of the intervention can be threatened in such within-classroom designs. The attempt to separate between students in the intervention and the control condition enforces an unnatural situation which can lead to the intervention being delivered in an unauthentic way. Cluster randomized trials, on the other hand, assign and implement interventions at the cluster level; that is, all students within one class are in the same condition. This is one way to reduce diffusion effects within classrooms. Also, some kinds of interventions (e.g., interventions including group discussions) can only be meaningfully implemented at the cluster level. However, cluster randomized trials require relatively large sample sizes to have an adequate power. Moerbeek, van Breukelen, and Berger (2000) as well as Raudenbush (1997) describe how the level of randomization can be optimally chosen and how resources can be optimally allocated to achieve an adequate power in multilevel randomized experiments.

The design of intervention studies in the classroom setting, thus, needs to take its multilevel structure into account. This multilevel structure should be considered with regards to the study design and the analysis. To detect small effects of motivational interventions, adequate sample sizes are necessary. In order to develop effective interventions in the classroom, researchers need to carefully consider the intervention dosage that is necessary as well as the number of intervention elements that need to be implemented.

1.4. Research questions of the present dissertation

The present dissertation investigates how students' value beliefs for mathematics can be fostered in the classroom setting. This seems to be an important endeavor especially in light of the widespread call for more young professionals in the STEM field. One particular target group of initiatives to foster achievement and motivation within STEM are females as they are still underrepresented in many STEM disciplines compared to males. This dissertation therefore specifically addresses the role of gender within mathematics.

The empirical studies that were conducted within this dissertation all use the expectancy-value theory of achievement-related choices (Eccles et al., 1983) as their theoretical framework. Therefore, the dissertation builds on a large body of research which has supported the assumptions of this theory empirically. Whereas this research has clearly demonstrated the usefulness of this model for predicting student outcomes such as academic choices (see section 1.1.5) and for explaining gender differences in choices (see section 1.2), there still remain important questions to be solved, particularly with regards to value beliefs (for a similar conclusion, see also Wigfield et al., 2009). Previous value intervention studies (Hulleman, Godes, et al., 2010; Hulleman & Harackiewicz, 2009) have provided first valuable insights into how these research findings can be applied in the classroom. However, it is still unclear how such interventions should be designed to be most effective in the classroom context.

The present dissertation aims at extending previous research on expectancy-value theory with regards to several points that are of particular relevance for developing and evaluating value interventions. In doing so, the dissertation takes a multidimensional perspective on value beliefs in two senses. First, it proposes a multidimensional measurement of value beliefs within one domain (i.e., the target subject of the intervention). This seems important for overcoming inconsistencies in the operationalization of value beliefs in previous research (see section 1.1.3). Based on the broad definition of value beliefs in expectancy-value theory, it is assumed that most of the value components include multiple facets. A differentiated measurement considering these subfacets can help in understanding existing gender differences in value beliefs as well as in developing and evaluating targeted interventions. Second, the dissertation calls for taking value in multiple subjects into account when evaluating the effects of motivational interventions. Value beliefs are highly domain-specific and choices are affected by intraindividual hierarchies in value beliefs. Therefore, it seems important to investigate possible side effects of motivational interventions, which have been neglected in previous research.

All three empirical studies of this dissertation use data from a large cluster randomized trial in which a utility value intervention was implemented in ninth grade math classrooms. Math was chosen as the target of the intervention as math competencies are an important prerequisite for future careers in many different fields, including all STEM disciplines. The target group of the intervention—ninth grade students—was chosen for two reasons: First, value beliefs, particularly in mathematics, have been found to decrease during secondary school (see section 1.1.4). Second, ninth grade students are at an age where they are supposed to begin reflecting about their future careers. Utility value interventions should therefore be within their “motivational zone of proximal development” (Brophy, 1999). Building upon the results of previous utility value interventions, the study aimed at extending the knowledge about the effectiveness of such interventions in the classroom. To be able to evaluate intervention effects, the study design contained randomized assignment at the class level with an adequate sample size based on a prior power analysis (see section 1.3.2): A total of 82 ninth grade classes were randomly assigned to one of two intervention conditions or a control condition. All classes in the intervention conditions received a 90-minute intervention on the usefulness of mathematics that consisted of a presentation as well as tasks for individual students. Two types of tasks to foster perceived utility were systematically compared. To evaluate effects of the intervention and to control for initial differences between the three experimental conditions, students’ motivation was assessed via self-reports before the intervention as well as six weeks and five months after the intervention. A specifically developed instrument was used to assess students’ value beliefs for mathematics in a comprehensive way. Additionally, students were asked to report on their value beliefs for two other subjects: German and English. These two verbal subjects were expected to be the most susceptible targets for dimensional comparisons between math and other subjects. In the following, the specific research questions of the three empirical studies will be elaborated.

Study 1 (*More Value Through Greater Differentiation: Gender Differences in Value Beliefs About Math*) examined the dimensionality of value beliefs for mathematics and gender differences in these beliefs. This study followed up on inconsistencies in previous findings of gender differences in math values potentially due to differences in the operationalization between studies. Based on the broad definition of value beliefs in expectancy-value-theory, several subfacets for attainment value, utility value, and cost were assumed. Study 1 used data from the pretest of the intervention study to investigate whether these subfacets could be distinguished empirically and whether gender differences could be found on these facets.

Study 2 (*Fostering Adolescents' Value Beliefs for Mathematics with a Relevance Intervention in the Classroom*) assessed effects of the intervention on students' value beliefs in mathematics. Specifically, this study addressed three open questions with regards to the effectiveness of value interventions in the classroom setting. First, as previous studies using different intervention strategies found mixed patterns of results, more research is needed that systematically examines which strategies are most effective. This study compared the effects of a previously used strategy to foster utility value (i.e., writing essays on the usefulness of math) to the effects of a newly developed strategy (i.e., evaluating interview quotations on the usefulness of math). Second, more needs to be known about the complexity of the effects of utility value interventions on value dimensions other than utility. Therefore, this study assessed and compared effects of the two intervention conditions on different value facets. Third, as boys and girls tend to differ in their career aspirations, interventions that highlight the usefulness of math for various careers might be a way to reduce gender differences in math values. The study therefore explored the role of gender as a moderator of the intervention effects. Study 2 used data on students' value beliefs in mathematics from all three measurement times.

Study 3 (*Adverse or Desired Side Effects of STEM Interventions? Effects of a Motivational Math Intervention on Motivation in Verbal Domains*) focused on side effects of the intervention in math on German and English as two verbal subjects. Considering that choices are influenced by intraindividual hierarchies in expectancies and values, it seems important to explore how such intraindividual hierarchies are affected through motivational interventions in one subject. Previous research, however, typically only addressed motivation in the target subject as an outcome. Based on the literature on dimensional comparisons, side effects of the intervention were assessed on motivation in German and English. To investigate the breadth of these effects, students' self-concept and effort were considered as outcomes in addition to value beliefs. Taking an intraindividual-difference perspective, intervention effects on differences between these constructs in math, on the one hand, and the two verbal subjects, on the other hand, were considered. Study 3 used data on students' value, self-concept, and effort in math, German, and English from all three measurement times.

References

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Battle, A., & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior*, *62,* 56–75. doi:10.1016/S0001-8791(02)00037-4

Battle, E. (1965). Motivational determinants of academic task persistence. *Journal of Personality and Social Psychology, 2,* 209–218. doi:10.1037/h0022442

Battle, E. (1966). Motivational determinants of academic competence. *Journal of Personality and Social Psychology*, *4*, 634–642. doi:10.1037/h0024028

Boekaerts, M. (2009). Goal-directed behavior in the classroom. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 105–122). New York, NY: Routledge.

Bong, M. (2001a). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task-value, and achievement goals. *Journal of Educational Psychology*, *93,* 23–34. doi:10.I037//0022-0663.93.1.23

Bong, M. (2001b). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology*, *26*, 553–570. doi:10.1006/ceps.2000.1048

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, *34*, 139–153. doi:10.1207/s15326985ep3403_1

Bong, M., & Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*, 1–40. doi:10.1023/A:1021302408382

Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist*, *34*, 75–85. doi:10.1207/s15326985ep3402_1

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218–61. doi:10.1037/a0014412

Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology*, *48*, 1612–1628. doi:10.1037/a0030194

Chow, A., & Salmela-Aro, K. (2011). Task-values across subject domains: A gender comparison using a person-centered approach. *International Journal of Behavioral Development*, *35*(3), 202–209. doi:10.1177/0165025411398184

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

Craven, R. G., Marsh, H. W., Debus, R. L., & Jayasinghe, U. (2001). Diffusion effects: Control group contamination threats to the validity of teacher-administered interventions. *Journal of Educational Psychology*, *93*, 639–645. doi:10.1037//0022-0663.93.3.639

Deci, E. L., & Ryan, A. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Publishing Co.

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*, 430–447. doi:10.1111/j.1467-8624.2007.01007.x

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, *99*, 597–610. doi:10.1037/0022-0663.99.3.597

Durik, A. M., Hulleman, C. S., & Harackiewicz, J. M. (in press). One size fits some: Instructional enhancements to promote interest. In K. A. Renninger & N. M. (Eds.), *Interest, the self, and K-16 mathematics and science learning*. Washington, DC: American Educational Research Association.

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*, 104–118. doi:10.1007/s11031-014-9419-0

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*, 382–393. doi:10.1037/0022-0663.98.2.382

Dweck, C. S., & Elliott, E. S. (1983). Achievement motivation. In P. Mussen & E. M. Heatherington (Eds.), *Handbook of child psychology* (pp. 643–691). New York, NY: Wiley.

Dweck, C. S., & Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. *Psychological Review, 95*, 256–273. doi:10.1037/0033-295X.95.2.256

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*, 135–172. doi:10.1111/j.1471-6402.1987.tb00781.x

Eccles, J. S. (2005). Subjective task values and the Eccles et al. model of achievement related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105–121). New York, NY: Guilford.

Eccles, J. S. (2007). Families, schools, and developing achievement-related motivations and engagement. In J. E. Grusec & P. J. Hastings (Eds.), *Handbook of socialization* (pp. 665–691). New York, NY: Guilford.

Eccles, J. S. (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*, 78–89. doi:10.1080/00461520902832368

Eccles, J. S. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development*, *35*, 195–201. doi:10.1177/0165025411398185

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74–146). San Francisco, CA: W. H. Freeman.

Eccles, J. S., Adler, T. F., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology*, *46*, 26–43. doi:10.1037/0022-3514.46.1.26

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence—The impact of stage-environment fit on young adolescents experiences in schools and in families. *American Psychologist*, *48*, 90–101. doi:10.1037//0003-066x.48.2.90

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*, 215–225. doi:10.1177/0146167295213003

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

Eccles, J. S., 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. doi:10.1111/j.1467-8624.1993.tb02946.x

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*, 169–189. doi:10.1207/s15326985ep3403_3

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52–72). New York, NY: Guilford.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. doi:10.1037/a0018053

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

Feather, N. T. (1988). Values, valences, and course enrollment: Testing the role of personal values within an expectancy-value framework. *Journal of Educational Psychology*, *80*, 381–391. doi:10.1037/0022-0663.80.3.381

Ford, M. E. (1992). *Human motivation: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109. doi: 10.3102/00346543074001059

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38,* 519–533. doi:10.1037//0012-1649.38.4.519

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20,* 507–537. doi:10.1111/j.1532-7795.2010.00645.x

Frenzel, A. C., Pekrun, R., Dicke, A.-L., & Goetz, T. (2012). Beyond quantitative decline: Conceptual shifts in adolescents' development of interest in mathematics. *Developmental Psychology*, *48,* 1069–1082. doi:10.1037/a0026895

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics - A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *12,* 497–514. doi:10.1007/BF03173468

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33,* 9–33. doi:10.1016/j.cedpsych.2006.12.002

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., … Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96,* 403–423. doi:10.1037/0022-0663.96.3.403

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, *94*, 638–645. doi:10.1037/0022-0663.94.3.638

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, *23*, 899–906. doi:10.1177/0956797611435530

Harackiewicz, J. M., Tibbetts, Y., Canning, E., & Hyde, J. S. (2014). Harnessing values to promote motivation in education. In S. A. Karabenick & T. Urdan (Eds.), *Motivational interventions (Advances in motivation and achievement, Volume 18)* (pp. 71–105). Bingley, UK: Emerald Group Publishing Limited.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest , self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, *39*, 870–888. doi:10.1002/tea.10048

Hidi, S., & Ainley, M. (2002). Interest and adolescence. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (pp. 247–275). Greenwich, CT: Information Age Publishing.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, *70*, 151–179. doi:10.3102/00346543070002151

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. doi:10.1207/s15326985ep4102_4

Higgins, E. T. (2007). Value. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 454–472). New York, NY: Guilford.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. C. Petersen (Eds.), *Girls at puberty* (pp. 201–228). New York, NY: Plenum Publishing Co.

Hulleman, C. S., & Barron, K. E. (in press). Motivation interventions in education: Bridging theory, research, and practice. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (3rd ed.). New York, NY: Routledge, Taylor and Francies.

Hulleman, C. S., Durik, A. M., Schweigert, S. B., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology*, *100*, 398–416. doi:10.1037/0022-0663.100.2.398

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*, 880–895. doi:10.1037/a0019506

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

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*, 422–449. doi:10.1037/a0018947

Husman, J., Derryberry, P. W., Crowson, H. M., & Lomax, R. (2004). Instrumentality, task value, and intrinsic motivation: Making sense of their independent interdependence. *Contemporary Educational Psychology*, *29*, 63–76. doi:10.1016/S0361-476X(03)00019-5

Husman, J., & Lens, W. (1999). The role of the future in student motivation. *Educational Psychologist*, *34*(2), 113–125.

Jacobs, J. E., Davis-Kean, P. E., Bleeker, M., Eccles, J. S., & Malanchuk, O. (2005). "I can, but I don't want to": The impact of parents, interest, and activities on gender differences in math. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 246–263). New York, NY: Cambridge University Press.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527. doi:10.1111/1467-8624.00421

Karabenick, S. A., & Urdan, T. (Eds.). (2014). *Motivational interventions (Advances in motivation and achievement, Volume 18)*. Bingley, UK: Emerald Group Publishing Limited.

Kerger, S., Martin, R., & Brunner, M. (2011). How can we enhance girls’ interest in scientific topics? *British Journal of Educational Psychology*, *81*, 606–628. doi:10.1111/j.2044-8279.2011.02019.x

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*, 383–409. doi:10.1016/S0959-4752(01)00011-1

Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement*, *70*, 142–160. doi:10.1177/0013164409344526

Maehr, M. L., & Zusho, A. (2009). Achievement goal theory: The past, present, and future. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 77–104). New York, NY: Francis.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research*, *23*, 129–149. doi:10.3102/00028312023001129

Marsh, H. W., & Craven, R. G. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment. Educational psychology series* (pp. 131–198). San Diego, CA: Academic Press.

Marsh, H. W., Craven, R. G., Hinkley, J. W., & Debus, R. L. (2003). Evaluation of the big-two-factor theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research*, *38*, 189–224. doi:10.1207/s15327906mbr3802_3

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416. doi:10.1111/j.1467-8624.2005.00853.x

Martin, A. J. (2012). Motivation and engagement: Conceptual, operational, and empirical clarity. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 303-311). New York, NY: Springer.

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*, 60–70. doi:10.1037//0022-0663.82.1.60

Miller, R. B., & Brickman, S. J. (2004). A model of future-oriented motivation and self-regulation. *Educational Psychology Review*, *16*, 9–33. doi:10.1023/B:EDPR.0000012343.96370.39

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, *25*, 271–284. doi:10.3102/10769986025003271

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*, 544–560. doi:10.1037/a0032459

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T. 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*, 1058–1066. doi:10.1177/0956797611415540

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

Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115–143). Washington, DC: American Psychological Association.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, *12*, 323–345. doi:10.1080/13803610600765687

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence*, *20*, 482–506. doi:10.1111/j.1532-7795.2010.00644.x

National Academy of Sciences. (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: National Academies Press.

National Science Board. (2007). *A national action plan for addressing the critical needs of the U.S. science, technology, engineering, and mathematics education system*. Arlington, VA: National Science Foundation.

National Science Foundation. (2011). *Women, minorities, and persons with disabilities in science and engineering*. Arlington, VA: National Science Foundation.

Nicholls, J. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*, 328–346. doi:10.1016/1041-6080(89)90010-1

Nuttin, J. R., & Lens, W. (1985). *Future time perspective and motivation: Theory and research method*. Leuven, Belgium: Leuven University Press.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*, 181–206. doi:10.1207/s15326985ep4103_4

OECD. (2004). *Education at a Glance 2004: OECD Indicators*. Paris, France: OECD Publishing. doi:10.1787/eag-2004-en

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*, 543–578. doi:10.2307/1170653

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology*, *48*, 1629–1642. doi:10.1037/a0029167

Peetsma, T., & van der Veen, I. (2011). Relations between the development of future time perspective in three life domains, investment in learning, and academic achievement. *Learning and Instruction*, *21*, 481–494. doi:10.1016/j.learninstruc.2010.08.001

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, *106*, 315–329. doi:10.1037/a0034027

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*, 667–686. doi:10.1037/0022-0663.95.4.667

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33–40. doi:10.1037//0022-0663.82.1.33

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change, *Review of Educational Research*, *63,* 167–199. doi:10.3102/00346543063002167

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*, 801–813. doi:10.1177/0013164493053003024

Plewis, I., & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation*, *4*, 13–26. doi:10.1076/edre.4.1.13.13014

Plünnecke, A., & Klös, H.-P. (2009). *MINT-Meter–MINT-Lücke in Deutschland und Indikatoren im internationalen Vergleich [STEM meter–STEM gap in Germany and indicators in international comparison]*. Köln, Germany: Institut der deutschen Wirtschaft Köln.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*, 173–185. doi:10.1037/1082-989X.2.2.173

Renninger, K. A., Hidi, S., & Krapp, A. (Eds.). (1992). *The role of interest in learning and development*. Hillsdale, NJ: Erlbaum.

Rokeach, M. (1979). From individual to institutional values with special references to the values of science. In M. Rokeach (Ed.), *Understanding human values* (pp. 47–70). New York, NY: Free Press.

Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. *Journal of Educational Psychology, 107,* 195-206. doi:10.1037/a0036981

Ryan, R. M., & Deci, E. I. (2009). Promoting self-determined school engagement: Motivation, learning and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171–195). New York, NY: Routledge.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78. doi:10.1037/0003-066x.55.1.68

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197–222). New York, NY: Routledge.

Schoon, I., & Eccles, J. S. (Eds.). (2014). *Gender differences in aspirations and attainment*. Cambridge, UK: Cambridge University Press.

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. doi:10.1016/j.learninstruc.2014.08.001

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shechter, O. G., Durik, A. M., Miyamoto, Y., & Harackiewicz, J. M. (2011). The role of utility value in achievement behavior: The importance of culture. *Personality and Social Psychology Bulletin*, *37*, 303–317. doi:10.1177/0146167210396380

Simons, J., Vansteenkiste, M., Lens, W., & Lacante, M. (2004). Placing motivation and future time perspective theory in a temporal perspective. *Educational Psychology Review*, *16*, 121–139. doi:10.1023/B:EDPR.0000026609.94841.2f

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, *42,* 70–83. doi:10.1037/0012-1649.42.1.70

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality*, *22*, 185–209. doi:10.1002/per.676

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and validation of a scale for the assessment of school-related values]. *Diagnostica*, *56*, 195–211. doi:10.1026/0012-1924/a000023

Stetsenko, A., Little, T. D., Gordeeva, T., Grasshof, M., & Oettingen, G. (2000). Gender effects in children's beliefs about school performance: A cross-cultural study. *Child Development*, *71*, 517–527. doi:10.1111/1467-8624.00161

Stipek, D. J., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*, 521–538. doi:10.1111/1467-8624.ep7252701

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*, 432–444. doi:10.1037/0022-0663.99.2.432

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*, 763–777. doi:10.1037/a0027470

Trautwein, U., Nagengast, B., Marsh, H. W., Gaspard, H., Dicke, A.-L., Lüdtke, O., & Jonkmann, K. (2013). Expectancy-value theory revisited. From expectancy-value theory to expectancy-valueS theory? In D. M. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self-processes and human development* (pp. 233–249). Charlotte, NC: Information Age Publishing.

Vansteenkiste, Lens, & Deci (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*, 19-31. doi: 10.1207/s15326985ep4101_4

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174-1204. doi:10.1037/a0036620

Wang, M.-T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*, 304–340. doi:10.1016/j.dr.2013.08.001

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*, 1556–1574. doi:10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G. (2006). The role of motivation in gendered educational and occupational trajectories related to maths. *Educational Research and Evaluation*, *12*, 305–322. doi:10.1080/13803610600765562

Watt, H. M. G., & Eccles, J. S. (Eds.). (2008). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences*. Washington, DC: American Psychological Association.

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*, 1594–1611. doi:10.1037/a0027838

Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research*, *61*, 1–24. doi:10.3102/00346543061001001

Wentzel, K. R. (2005). Peer relationships, motivation, and academic performance at school. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 279–296). New York, NY: Guilford.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*, 49–78. doi:10.1007/BF02209024

Wigfield, A., & Cambria, J. (2010a). Expectancy-value theory: Retrospective and prospective. In T. C. Urdan & S. A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement. Volume 16 Part A* (pp. 35–70). London, UK: Emerald Group Publishing Limited.

Wigfield, A., & Cambria, J. (2010b). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*, 1–35. doi:10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*, 265–310. doi:10.1016/0273-2297(92)90011-P

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In N. Eisenberg (Ed.), *Handbook of child psychology*. *Volume 3: Social, emotional, and personality development* (pp. 933–1002). Hoboken, NJ: Wiley.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*, 451–469. doi:10.1037/0022-0663.89.3.451

Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). New York, NY: Routledge.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*, 267–301. doi:10.3102/0034654311405999

2

More Value Through Greater Differentiation: Gender Differences in Value Beliefs about Math

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U. & Nagengast, B. (2014). More Value Through Greater Differentiation: Gender Differences in Value Beliefs about Math. *Journal of Educational Psychology*. Online first publication. doi:10.1037/edu0000003

Abstract

Expectancy-value-theory (EVT, Eccles et al., 1983) is a prominent approach to explaining gender differences in math-related academic choices, with value beliefs acting as one important explanatory factor. EVT defines four value components: intrinsic value, attainment value, utility value, and cost. The present study followed-up on inconsistencies in research findings on gender differences in math values that might partially be due to differences in the operationalization of the value construct. To this end, we examined if sub-facets of the four value components can be established empirically and if gender differences can be found on these facets. A total of 1868 ninth grade students completed a set of 37 items assessing their value beliefs in mathematics. Confirmatory factor analyses supported the conceptual differentiation of value beliefs into a total of eleven value facets. Whereas the factor structure was invariant across gender, there were considerable differences in mean levels favoring boys on some, but not all value facets. These gender differences depended not only on the value component, but also on the specific facet under consideration.

Keywords: expectancy-value theory, gender differences, mathematics, task value

More Value Through Greater Differentiation: Gender Differences in Value Beliefs about Math

Males and females are not equally represented in math-related domains, and this is only partially explainable by gender differences in achievement (Else-Quest, Hyde, & Linn, 2010; OECD, 2004; Watt & Eccles, 2008). Expectancy-value-theory (EVT; Eccles et al., 1983) is a widely used explanatory framework for the overrepresentation of males in math, and research based on EVT has proved highly effective in explaining gender differences and, more generally, achievement-related outcomes (for a review, see Wigfield, Tonks, & Klauda, 2009). According to EVT, value beliefs are the central factor in explaining gender differences in academic choices (Eccles, 2005, 2009). Nevertheless, previous research on gender differences in values of mathematics yielded somewhat inconsistent findings: Some studies reported higher values for boys (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Steinmayr & Spinath, 2008; Watt, 2004), whereas other studies reported no differences between boys and girls (e.g., Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Meece, Wigfield, & Eccles, 1990; Wigfield et al., 1997).

Two factors seem to contribute to this lack of coherence. First, there have been vast differences in the operationalization of value beliefs. In fact, the value measures used in previous studies often have little overlap. Such phenomena, i.e., scales assuming to measure the same construct by different operationalizations, are not unusual in motivational science and hinder integration of research results (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010; Marsh, Craven, Hinkley, & Debus, 2003). Second, studies differ considerably in how much emphasis was put on a general value scale or separate value components. The EVT model by Eccles and colleagues differentiates between four value components: attainment value, intrinsic value, utility value, and cost (Wigfield & Eccles, 1992). However, although theoretically separable, previous research has typically found high correlations among the value components and many studies have collapsed them into a single, more general value scale (e.g., Eccles, Wigfield, Harold, & Blumenfeld, 1993; Jacobs et al., 2002). More recent studies were able to separate the four components using confirmatory factor analysis with items that explicitly tapped all of them (Conley, 2012; Trautwein et al., 2012).

In this study, we argue that a careful operationalization and statistical modeling of value beliefs is needed to fully understand gender differences in these beliefs. To this end, the present study proposes to measure value beliefs at a facet level with two major goals. First, we examined the dimensionality of domain-specific value beliefs. Starting from both theoretical considerations as well as previous operationalizations, we utilized a large set of items that did not only measure all four value components, but also covered different facets of these components. We examined whether the proposed facets could be separated empirically using confirmatory factor analyses. The second aim of the present study was an investigation of gender differences in value beliefs in mathematics. We expected to find differences between males and females on many – but not all – value facets. By taking a closer look at differences in task value at a facet level, the present study aimed at solving some of the inconsistencies found in previous research on gender differences.

Theoretical Conceptualization of Subjective Task Value

In the EVT model by Eccles and colleagues (1983), the most immediate predictors of academic performance and choice are two kinds of beliefs: expectancies, i.e., the perceived ability to succeed on a task, and value beliefs, i.e., reasons for engaging in a task. Value beliefs are conceptualized as task-specific; that is, they are shaped by qualities of different tasks that influence the probability that an individual will engage in them (Eccles, 2005; Eccles et al., 1983; Wigfield & Eccles, 1992). Although these beliefs can refer to specific tasks and situations, research based on EVT mostly refers to more global beliefs regarding different school subjects that have been shown to predict performance and academic choices (e.g., Meece et al., 1990; Nagengast et al., 2011; Trautwein et al., 2012). Value beliefs are, however, highly domain-specific with beliefs in different school subjects such as mathematics and English showing only low or even negative correlations (e.g., Bong, 2001; Eccles, Wigfield, et al., 1993; Trautwein et al., 2012). Further, task value is conceptualized as being subjective, hence the term *value beliefs* that we use in the present study.

Eccles et al. (1983) proposed four major components of task value: intrinsic value, attainment value, utility value, and cost (for a detailed discussion of these components, see Eccles, 2005; Eccles & Wigfield, 2002; Wigfield & Eccles, 1992). *Intrinsic value* is defined as the enjoyment a person derives from doing a task. The task is, thus, an end to itself, which is similar to the concept of intrinsic motivation by Ryan and Deci (2009). Moreover, intrinsic value has been linked to the construct of individual interest (e.g., Pintrich, 2003; Wigfield & Cambria, 2010). *Attainment value* refers to the importance that individuals attach to doing well on or being competent at a given task. Eccles (2009) has linked it to the relevance of a task to one's personal and social identities. Attainment value as well as intrinsic value can be related to different parts of interest according to the person-object theory of interest (Krapp, 2002). Whereas intrinsic value is closer to feeling-related valences, attainment value can be linked to value-related valences, i.e., the personal importance of an interest object. *Utility value* indicates the perceived individual usefulness of engagement and achievement in a certain domain for short- and long-term goals. When a task is done for its utility value, the task is a means to an end rather than an end in itself. Utility value is, thus, instrumental in nature and closer to extrinsic motivation (Ryan & Deci, 2009). Finally, *cost* describes the perceived negative consequences of engaging in a task. These include the opportunity cost of choosing one option over another, but also the amount of effort required to succeed and negative emotions that are associated with this engagement (see Eccles et al., 1983 for a detailed discussion).

The concept of task value as it is defined in EVT by Eccles et al. (1983) is, thus, very broad. It covers conceptually different beliefs that can be differentiated in terms of their intrinsic versus extrinsic nature with intrinsic value and attainment value being more intrinsic motivational factors and utility value being a more extrinsic motivational factor (Trautwein et al., 2013). The broad spectrum of value components that are included in EVT can be seen as one of its strengths because it offers the possibility of relating task value to several other motivational theories (Eccles, 2005; Eccles & Wigfield, 2002; Wigfield & Cambria, 2010) and it leads to a high power in predicting academic choices (Wigfield et al., 2009). These strengths are clearly reflected in the previous empirical research (for an overview, see Wigfield & Cambria, 2010; Wigfield et al., 2009).

Measurement and Dimensionality in Previous Research

The measurement of value beliefs in research on EVT, however, has only partially incorporated this multidimensionality of value beliefs. Few studies have used separate measures for all four value components. Many studies have incorporated the value components, measured by a small number of items, into a single, more general value scale (e.g., Eccles, Wigfield, et al., 1993; Jacobs et al., 2002). Other studies have measured value beliefs by only one of these components (e.g., Nagengast et al., 2011) or have used a combination of two or more components (often importance value as a combination of utility and attainment; e.g., Durik, Vida, & Eccles, 2006; Watt et al., 2012). In spite of the theoretical importance of cost in EVT, it has been incorporated less frequently.

Regarding the empirical separability of the value components, earlier studies were often not able to separate them into four factors ending up with one general factor (Eccles, Adler, & Meece, 1984; Eccles, Wigfield, et al., 1993). However, there is more recent empirical support for four distinct components using confirmatory factor analyses on items tapping all of these components (Conley, 2012; Trautwein et al., 2012). Assessing all components separately increased the explanatory power of value beliefs: Conley found different clusters of motivational beliefs with clearly distinct patterns of task value; Trautwein et al. (2012) found synergistic effects of expectancy and value in predicting achievement for all four value components. Both Conley (2012) and Trautwein et al. (2012) reported a similar correlation pattern among the value components with the highest correlations between intrinsic and attainment value. Nevertheless, each of those studies used different sets of items. The specific aspects of value beliefs indicated by the items might have affected the correlations obtained between different value components as well as the correlations with other constructs. Based on their own results and a critical assessment of the available evidence, Trautwein et al. (2012; 2013) argued for a differentiated measurement including all four value components in future research and suggested using a larger set of more diverse items to test empirically if the value components could be broken down into facets.

Value Facets

Although EVT describes four value components in depth, the definition of some components has been rather broad. Their typical descriptions and operationalizations point to the possibility that some value components could be further differentiated into multiple facets (Trautwein et al., 2013). In fact, differentiating between facets of value beliefs may further increase the predictive power of EVT and resolve some of the discrepancies concerning gender differences. Whereas the definition and the nature of intrinsic value is clear-cut by focusing on positive feelings associated with engagement in a task, the other value components could be differentiated further into facets.

First, attainment value is defined as the importance of doing well on a given task, but has also been linked to the relevance of a task to the individual's identity (Eccles, 2005, 2009). Whereas the first notion clearly refers to the importance of high performance, the second notion goes beyond this by relating the task to central aspects of the self. That is, a task can be important because one wants to get good grades—for whatever reason—or because it is central to one's identity and one wants to master the task itself. Both notions have been incorporated in previous research to various degrees. Whereas many studies have assessed attainment value with items referring to high performance (e.g., "For me, being good at math is (not at all important, very important)", Eccles, Wigfield, et al., 1993), other studies also included items referring to personal importance (e.g., "Thinking mathematically is an important part of who I am", Conley, 2012). The importance of high achievement has also been assessed under the name of achievement value (Frenzel, Pekrun, & Goetz, 2007) or need for high achievement (Luttrell et al., 2010), whereas personal importance has often been assessed as one part of individual interest (i.e., value-related valences) (e.g., Marsh et al., 2005; Ramm et al., 2006). Although these aspects are both congruent with the definition of attainment value (Eccles, 2009; Eccles et al., 1983; Eccles & Wigfield, 2002), they have not been differentiated systematically in previous empirical studies.

Second, the definition of *utility value* in EVT relates to short- as well as long-term goals (Eccles & Wigfield, 2002), but is relatively open concerning the content of such goals. Depending on students' age and their developmental stage students' goals can vary with regards to different life domains. Research on future time perspectives has shown that students differentiate between life domains such as school and professional career, social relations, and leisure time (Peetsma & van der Veen, 2011). When adolescents think about the usefulness of a subject within a more short- as well as a more long-term perspective, they can refer to their career on the one hand (school in the short term and job in the long term) and life outside of school on the other hand (i.e., leisure time as well as social relations). In line with the broad definition in EVT, the items used to indicate utility value in previous research referred to different life domains. Whereas many items were rather general (e.g., "Math will be useful for me later in life", Conley, 2012), some items also referred to utility for job and future education (e.g., "Being good at math will be important when I get a job or go to college", Conley, 2012), and utility for daily life (e.g., "I do not need math in my everyday life", Luttrell et al., 2010). However, previous research did not differentiate between those life domains systematically, but either focused on one life domain or included very few items for each life domain. It is, thus, not clear if students differentiate between life domains when evaluating the usefulness of a task. Moreover, utility for social goals was not included in previous research, but may be of particular relevance to adolescents (cf. Boekaerts, 2009). During adolescence, acceptance by peers becomes an increasingly important goal and youths also tend to select friends who have similar values (Eccles, Midgley, et al., 1993; Juvonen, Espinoza, & Knifsend, 2012; Wentzel, 2005).

Finally, *cost* includes multiple facets in its original definition, such as the amount of time and energy lost for other activities, i.e., the opportunity cost of engagement in a task, as well as anticipated negative emotions and effort required to succeed (Eccles et al., 1983). Regarding the operationalization of cost, studies on value beliefs within secondary school mostly focused on opportunity cost and did not consider other aspects (e.g., Conley, 2012; Trautwein et al., 2012). More elaborated measures of cost that also include effort required and emotional cost have been developed specifically for college students (Battle & Wigfield, 2003; Luttrell et al., 2010; Perez, Cromley, & Kaplan, 2014). Perez et al. (2014) assessed cost in terms of three dimensions, which were separable in an exploratory factor analysis, namely effort cost, opportunity cost, and psychological cost. They found some support for differential contributions of these types of cost for predicting choices.

In sum, as is the case with many other motivational constructs, the items that have been used to measure value beliefs are only partially consistent across studies or research groups (for a discussion of jingle-jangle-fallacies, see Marsh et al., 2003; for a review on the operationalization of value beliefs, see Wigfield & Cambria, 2010). The scales measuring separate components share some overlap, while also capturing qualitatively different aspects. Such differences in operationalizations may produce inconsistent research results and point to an underlying problem: There is more than one facet to most of the value components. As previous research has not taken into account all facets simultaneously, it remains to be investigated whether these facets can be differentiated empirically and whether differences in operationalization affect substantive conclusions, such as findings on gender differences.

Gender Differences in Value Beliefs

According to EVT, expectancy and value beliefs are shaped by gender norms and roles through socialization processes affecting identity formation (Eccles, 2009). In consequence, boys develop more favorable beliefs in male-typed domains like mathematics and girls develop more favorable beliefs in female-typed domains like English. Results of previous research on the existence of gender differences in math values are, however, not straightforward. Differences in the number of value dimensions incorporated in previous studies and their operationalizations seem to contribute to some inconsistencies in these findings.

Studies that examined composite scores for math value yielded inconsistent results: Whereas Jacobs et al. (2002) found no gender differences in math values in an US sample tracked from grade 1-12, Steinmayr and Spinath (2008) found higher values for males in a German sample from grade 11. Studies that examined separate components of math value found gender differences depending on the value component. The overall pattern of research seems to indicate that girls are aware of the importance of attaining good grades in mathematics, while they perceive it as a rather unattractive subject. In German as well as Australian samples, males have been found to report higher intrinsic value in math than females in secondary school (Frenzel et al., 2007; Watt, 2004; Watt et al., 2012). Gender differences favoring males were also found in a German study in grade 7 and 10 for interest measured as a combination of both positive affect and personal importance (Marsh et al., 2005; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006), whereas no difference was found for importance of achievement in samples from different nations and grades (Frenzel et al., 2007; Meece et al., 1990; Steinmayr & Spinath, 2010). Regarding utility value, Watt (2004) found no gender differences for Australian students from grade 7 to 11, whereas Steinmayr and Spinath (2010) found differences favoring males for German 11th graders. Studies using importance measures combining attainment and utility value found no differences between males and females in grades 9 and 10 in Australia, Canada, and the US (Watt, Eccles, & Durik, 2006; Watt et al., 2012). Regarding cost, Australian female adolescents tend to perceive math as requiring more effort than their male counterparts (Watt, 2004). Moreover, in German as well as in US secondary school samples, girls reported less intense positive emotions such as enjoyment and pride and more negative emotions such as anxiety, hopelessness, and shame in math compared to boys (Frenzel et al., 2007; Meece et al., 1990).

To our knowledge no study has examined mean gender differences in secondary school on all four value components within one sample (for gender differences in motivational profiles, see Conley, 2012). Gender differences that were found in previous studies may depend on the components that were examined, on the items that were used as well as on the characteristics of the specific sample (such as age group or cultural background). As all of these factors varied across studies, it is difficult to conclude how they contributed to the inconsistent pattern of research results. Valid comparisons of gender differences across age groups (e.g., Frenzel, Goetz, Pekrun, & Watt, 2010; Jacobs et al., 2002; Watt, 2004) or cultures (e.g., Watt et al., 2012) would require using the same instruments across those groups. Additionally, analyses of gender differences based on scale scores, as they are reported in many studies (e.g., Frenzel et al., 2007), place high demands on the measurement quality of instruments that are often neglected in applications. Specifically, such analyses rely on strict measurement invariance (Meredith, 1993). Compared to analyses with scale scores, latent variable modeling allows for controlling such assumptions and assesses differences at the construct level, thereby relying on fewer assumptions. To get a closer look at gender differences in value beliefs, a broader measurement and an adequate statistical modeling within a large sample is needed.

The Present Investigation

In the present study, we examined the structure of value beliefs and gender differences in these beliefs in mathematics. A large sample of ninth grade students completed 37 items assessing all four value components of the EVT by Eccles et al. (1983). Building on the broad definition of the value components in EVT and differences in their previous operationalization, we expanded this model by assuming several facets within the four value components. This systematic differentiation between value facets aims at solving some inconsistencies in previous research on gender differences.

An overview of the facets assessed in our study is presented in Figure 1. Our measures of intrinsic value focused on positive affect related to the task. Attainment value was divided into *importance of achievement*, focusing on high performance, and *personal importance*, focusing on mastering the content and its relation to one's identity. Utility value was differentiated into five facets that referred to short- as well as long term goals and different life domains within these time perspectives. Utility for short-term goals included *utility for school*, relating to the usefulness for present and future education; *utility for daily life,* relating to daily routines and leisure time activities; and *social utility*, referring to the usefulness of subject knowledge for being accepted by peers. Utility for long-term goals was assessed in terms of *utility for job*, referring to future career opportunities, and *general utility for future life*, relating to unspecified future life activities. Whereas one could possibly think of even more utility facets in terms of different life domains, we considered those aspects as particularly important for young adolescents, given the research on personal goals (Boekaerts, 2009; Salmela-Aro, 2009) and future time perspectives (Peetsma & van der Veen, 2011) as well as previous operationalizations of utility value. Cost was divided into several possible negative effects of engaging in a task; in line with theoretical assumptions (Eccles et al., 1983) and previous assessments with college students (Perez et al., 2014), we distinguished between *opportunity cost* (i.e. time lost for other activities), *effort required* (i.e., perceived exhaustion) and *emotional cost* (i.e., associated negative emotions).

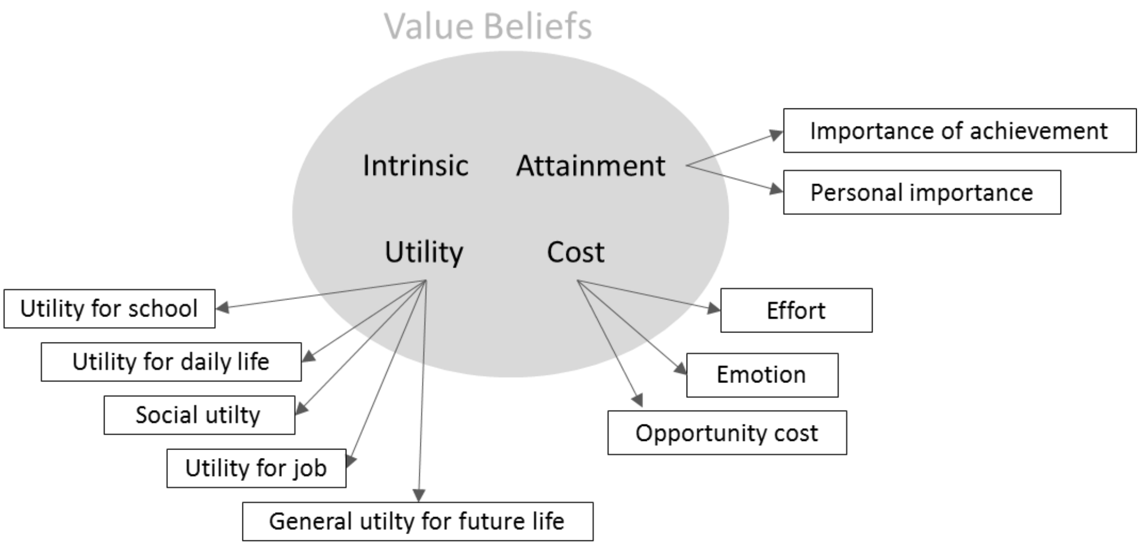

Figure 1. Conceptual representation of value facets assessed in our study.

We had two major research questions. First, we tested if the theoretical differentiation of value components into further sub-facets holds empirically. We assessed this question in several steps beginning with each value component separately to see if attainment value, utility value, and cost can be split into the assumed facets. Then, we tested if those facets can also be separated within a complete model. At last, we tested whether the dimensionality of value beliefs can be generalized across gender. In addition to testing measurement invariance as a prerequisite of comparing means, latent variable modeling also allows to test if the structure of the underlying constructs is stable across groups.

Second, we examined gender differences in mean levels of value facets. We expected that gender differences would favor male adolescents where they occurred. Based on previous research (Frenzel et al., 2007; Watt, 2004; Watt et al., 2006), we hypothesized to find gender differences in intrinsic value and some aspects of cost, i.e., emotional cost and effort required. For attainment and utility value, results on gender differences in previous research were less consistent (Frenzel et al., 2007; Meece et al., 1990; Steinmayr & Spinath, 2010; Watt, 2004; Watt et al., 2006; Watt et al., 2012); suggesting that gender differences might depend on the facet level. We, thus, expected to find a differentiated pattern of gender differences that would favor boys on some, but not on all value facets, thereby helping to understand inconsistencies in previous research.

Methods

Sample

Data were collected as part of a larger study on motivation in mathematics. Participation was voluntary and only students with active parental consent participated in the study (96% participation rate). Of the 1978 students who had parental consent, 111 students were absent at the time of data collection for this study. The current study was, thus, based on data from 1867 ninth-grade students who were enrolled in 82 classes. The data were collected in 25 academic-track secondary schools (*Gymnasium* schools) in the German state of Baden-Württemberg in 2012. On average, students were 14.62 years old (*SD* = 0.46). In line with the typical composition of academic track schools, girls (53.3 % of the sample) were slightly overrepresented in this study. Data collection took place at the beginning of the school year. The questionnaires were administered by trained research assistants.

Measures

We assessed value beliefs in the domain of mathematics with 37 items tapping all four components of the EVT by Eccles et al. (1983). Moreover, we differentiated between multiple facets of attainment value, utility value, and cost. All items are reported in Table 1. Items from previous national and international large-scale studies were included (e.g., Conley, 2012; Rakoczy, Buff, & Lipowsky, 2005; Steinmayr & Spinath, 2010; Trautwein et al., 2012; see Appendix for more information on specific facets). As not all value facets that we thought of importance for adolescents were adequately represented in previous research, we developed additional items in order to measure value beliefs very broadly. While developing the instrument, we tried to find items that could clearly be assigned to one of the facets. Whereas we assessed a number of facets of attainment value, utility value, and cost, we tried to keep the overall balance between the length of the instrument and a broad measurement of value beliefs.

To assess the internal psychometric properties of the resulting scales, we estimated the scale reliability ρ, an estimator for reliability in latent variable modeling (Bollen, 1989; Raykov, 2009) and an alternative to Cronbach's alpha (Sijtsma, 2009). These estimates were based on models with correlated factors representing the respective value facets computed separately for each value component (see Results for more information on these models). Intrinsic value was assessed by four items (ρ = .94). Attainment value was assessed by ten items covering the facets importance of achievement (four items, ρ = .88) as well as personal importance (six items, ρ = .83). Utility value was assessed by twelve items focusing on different life domains, namely utility for school (two items, ρ = .52), utility for daily life (three items, ρ = .83), social utility (three items, ρ = .76), utility for job (two items, ρ = .68), and general utility for future life (two items, ρ = .79). Cost was assessed by eleven items that were subdivided into the facets of opportunity cost (four items, ρ = .83), effort required (four items, ρ = .90), and emotional cost (four items, ρ = .87). All items were measured with a four-point Likert scale ranging from *completely disagree* to *completely agree*.

Table 1

Standardized Factor Loadings (in the Eleven-Factor Model), Sample, Means, and Standard Deviation for All Items

Item	Factor loadings	*N*	*M*	*SD*
Intrinsic				
Math is fun to me.	.91	1844	2.26	.93
I like doing math.	.92	1853	2.36	.93
I simply like math.	.90	1850	2.24	.97
I enjoy dealing with mathematical topics.	.81	1843	2.16	.89
Importance of achievement				
It is important to me to be good at math.	.78	1851	3.06	.74
Being good at math means a lot to me.	.82	1850	2.76	.79
Performing well in math is important to me.	.83	1844	2.94	.72
Good grades in math are very important to me.	.79	1850	3.03	.76
Personal importance				
I care a lot about remembering the things we learn in math.	.68	1852	2.74	.74
Math is not meaningful to me.	.70	1853	3.10	.87
I‘m really keen on learning a lot in math.	.66	1843	2.41	.77
Math is very important to me personally.	.79	1840	2.33	.84
To be honest, I don't care about math.	.59	1813	2.81	.92
It is important to me to know a lot of math.	.76	1845	2.62	.79
Utility for school				
It is worth making an effort in math, because it will save me a lot of trouble at school in the next years.	.58	1854	3.10	.75
Being good at math pays off, because it is simply needed at school.	.60	1853	3.13	.69
Utility for daily life				
Understanding math has many benefits in my daily life.	.78	1852	2.49	.90
Math comes in handy in everyday life and leisure time.	.80	1851	2.28	.81
Math is directly applicable in everyday life.	.80	1835	2.45	.80

(continued)

Item	Factor loadings	*N*	*M*	*SD*
Social utility				
Being well versed in math will go down well with my classmates.	.69	1824	1.73	.78
I can impress others with intimate knowledge in math.	.65	1841	1.78	.81
If I know a lot in math, I will leave a good impression on my classmates.	.81	1830	1.72	.73
Utility for job				
Good grades in math can be of great value to me later on.	.71	1855	3.14	.79
Learning math is worthwhile, because it improves my job and career chances.	.72	1850	3.03	.83
General utility for future life				
Math contents will help me in my life.	.81	1845	2.67	.83
I will often need math in my life.	.81	1842	2.74	.80
Effort required				
Doing math is exhausting to me.	.78	1857	2.57	.93
I often feel completely drained after doing math.	.81	1835	2.48	.96
Dealing with math drains a lot of my energy.	.86	1851	2.42	.95
Learning math exhausts me.	.89	1845	2.43	.97
Emotional cost				
I'd rather not do math, because it only worries me.	.76	1841	2.01	.89
When I deal with math, I get annoyed.	.78	1849	2.18	.96
Math is a real burden to me.	.87	1847	2.05	.99
Doing math makes me really nervous.	.75	1844	1.85	.85
Opportunity cost				
I have to give up other activities that I like to be successful at math.	.70	1854	1.62	.82
I have to give up a lot to do well in math.	.82	1848	1.66	.79
I'd have to sacrifice a lot of free time to be good at math.	.83	1851	1.88	.92

Statistical Analysis

The data were analyzed using confirmatory factor analysis (CFA) with Mplus (Muthén & Muthén, 1998-2012). All analyses used the robust maximum likelihood estimator and the design-based correction of standard errors and model-fit statistics to account for nonnormality of the indicator variables and the nesting of students within classes within schools (Muthén & Muthén, 1998-2012). We assessed the model fit of the CFAs using the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). In addition, we report the Yuan-Bentler (1998) robust χ^2 test statistics and carefully inspected the parameter estimates. CFI and TLI values (on a scale from 0 to 1) greater than .90 and .95 are typically taken to reflect acceptable and excellent fit to the data, respectively (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). RMSEA of less than .06 are typically taken to reflect a reasonable fit, although there is no golden rule (F. Chen, Curran, Bollen, Kirby, & Paxton, 2008; Hu & Bentler, 1999; Marsh et al., 2004). For SRMR, values of less than .08 are considered to indicate a good model fit (Hu & Bentler, 1999).

Test of measurement and structural invariance. We pursued a two-group analysis to test for the invariance of the resulting model over gender. Tests of measurement invariance examine if comparisons across groups are valid or if they are biased due to a different conceptual understanding of the items for these groups. Additionally, tests of structural invariance refer to tests concerning the latent variables themselves, such as differences in factor variances, covariances, and means (Vandenberg & Lance, 2000). Such tests include a series of nested models in which the least restrictive model has no invariance constraints and the most restrictive model constrains all parameters to be equal across groups (Meredith, 1993). If the introduction of increasingly stringent invariance constraints results in little or no change in goodness of fit, this is considered as support of measurement invariance.

In total, six models were tested for invariance analyses following the steps outlined by Meredith (1993), Widaman and Reise (1997), and Vandenberg and Lance (2000), among others. Model 1 included measurement models with identical loading patterns, but no invariance of any parameters, examining configural invariance. Model 2 constrained the factor loadings to be equal across groups, testing for weak measurement invariance. Model 3 required the factor loadings and, additionally, the item intercepts to be invariant over groups. Such a model assesses strong measurement invariance, which is required when comparisons in latent means are conducted. Model 4 required invariance of item uniquenesses (in addition to invariant factor loadings and intercepts). This model tested strict measurement invariance, which justifies the comparison of manifest means over groups. Model 5 constrained factor variances and covariances (in addition to the constraints in model 4) to be equal across groups. Finally, model 6 additionally constrained latent means to be equal over responses by males and females. In this set of nested models, models 1 to 4 assessed measurement invariance, whereas models 5 and 6 assessed structural invariance. For tests of measurement invariance, we used the guidelines provided by Chen (2007) and Cheung and Rensvold (2002): Their simulations suggest that a decrease of less than .01 in the fit of the more parsimonious model on incremental fit indices such as the CFI should be treated as support for that model. Chen (2007) suggested that an increase of less than .015 in the RMSEA and an increase in the SRMR of .030 for invariance of factor loadings and .015 for invariance of intercepts and residual variances indicate support for the more constrained model. There are, however, no clear guidelines for detecting mean structure invariance (F. F. Chen, Sousa, & West, 2005; Fan & Sivo, 2009).

Missing data. For the variables considered here, the percentage of missing data was low (2.9 % at maximum, see Table 1 for the exact number of missing values for each variable). We used the full information maximum likelihood approach implemented in Mplus to deal with missing values. This approach takes all available information into account when estimating the model parameters (Schafer & Graham, 2002).

Results

Facets of Task Value

Separate analyses for each value component. Our first research question concerned the separability of value facets. As a first step to this end, we conducted separate CFAs for each value component. These analyses assessed our predictions concerning the dimensionality of each of these components as we wanted to assess the separability of facets within components before testing all components together in the next step (see Figure 1 for an overview of components and facets). We expected intrinsic value to be one-dimensional. For attainment value, we assumed two separable dimensions: importance of achievement and personal importance. For utility value, we assumed five facets: utility for school, utility for daily life, social utility, utility for job, and general utility for future life. For cost, we expected to find three dimensions: effort required, emotional cost, and opportunity cost. By comparing the fit of a one-factor model to a model with correlated factors representing the respective facets assumed, we assessed the usefulness of differentiating between facets to explain the underlying structure of these value beliefs. In all analyses, each item was allowed to load on only one factor. Residual correlations were allowed for two negatively worded items of attainment value (items 4 and 6). Table 2 reports the fit statistics for these CFAs.

Table 2

Model Fit Statistics for Confirmatory Factor Analyses Assessing Each Component Separately and All Components Together

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Intrinsic						
One factor	3.41	2	1.000	.999	.019	.003
Attainment						
One factor	404.55	34	.951	.935	.076	.038
Two factors	140.63	33	.986	.981	.042	.021
Utility						
One factor	1830.35	54	.748	.692	.133	.095
Five factors	174.34	44	.982	.972	.040	.026
Cost						
One factor	1301.96	44	.869	.836	.124	.071
Three factors	259.85	41	.977	.969	.053	.034
All						
Four factors	6117.08	622	.849	.838	.069	.078
Eleven factors	2087.50	573	.958	.952	.038	.048
2^{nd} order	2827.14	613	.939	.934	.044	.065

Note. χ^2 is the Yuan-Bentler robust test statistics; *df* = degrees of freedom*;* CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

For each component, the models yielded a good fit for the assumed factor structure (see Table 2). For intrinsic value, a one-factor structure of the instrument was supported (χ^2 = 3.41, *df* = 2, CFI = 1, TLI = .999, RMSEA = .019, SRMR = .003). For attainment value, differentiating between two dimensions (χ^2 = 140.63, *df* = 33, CFI =

.986, TLI = .981, RMSEA = .042, SRMR = .021) considerably improved the fit in comparison to a one-factor solution (χ^2 = 404.55, *df* = 34, CFI = .951, TLI = .935, RMSEA = .076, SRMR = .038). For utility value, separating between five facets lead to a good fit to the data (χ^2 = 174.34, *df* = 44, CFI = .982, TLI = .972, RMSEA = .040, SRMR = .026), whereas a one-factor model did not yield an adequate fit (χ^2 = 1830.35, *df* = 54, CFI = .748, TLI = .692, RMSEA = .133, SRMR = .095).[1] For cost, the theoretical three-facet model exhibited a good fit (χ^2 = 259.85, *df* = 41, CFI = .977, TLI = .969, RMSEA = .053, SRMR = .034), whereas a one-factor model did not exhibit a reasonable fit (χ^2 = 1301.96, *df* = 44, CFI = .869, TLI = .836, RMSEA = .124, SRMR = .071).[2] In sum, these results support the dimensionality that was assumed for each component and point to the usefulness of differentiating between multiple facets of attainment value, utility value, and cost.

Comprehensive model including all components. In the next step, we conducted CFAs based on all items to further assess the dimensionality of value beliefs when put into a complete model and examined the associations between the different facets. To assess the separability of the value facets, we compared a four-factor model, in which the factors represented the value components intrinsic, attainment, utility, and cost, to an eleven-factor model, in which the factors represented the value facets.

The fit statistics for the CFAs including all 37 items are also reported in Table 2. A four-factor model, in which the factors represented the value components did not yield a reasonable fit (χ^2 = 6117.08, *df* = 622, CFI = .849, TLI = .838, RMSEA = .069, SRMR = .078). This four-factor model was compared against an eleven-factor model that resulted from the conjunction of the differentiated models described above. This model, in which the factors represented the theoretically assumed facets, fitted the data well (χ^2 = 2087.50, *df* = 573, CFI = .958, TLI = .952, RMSEA = .038, SRMR = .048). These results again support the differentiation of value components into distinct facets.

---

[1] As general utility for future life and utility for daily life were highly correlated (*r*= .86), we also tested a four-factor model in which those facets were combined into one factor. This four-factor model, however, showed a significantly lower fit than the five-factor model ($\Delta\chi^2$ (4) = 270.42, *p* < .001, ΔCFI = -.039, ΔTLI = -.050, ΔRMSEA = .027, ΔSRMR = .012).

[2] As we found a high intercorrelation between effort required and emotional cost (r= .90), we also compared the three-factor model to a two-factor model in which those facets were combined into one factor. However, this alternative model yielded a considerably weaker fit ($\Delta\chi^2$ (2) = 213.75, *p* < .001, ΔCFI = -.028, ΔTLI = -.034, ΔRMSEA = .025, ΔSRMR = .007).

We also specified a second-order model, in which the first-order factors representing the facets of attainment value, utility value, and cost were set to load on a second-order factor representing the correspondent value component. This second-order model assessed the relations between intrinsic value, attainment value, utility value, and cost as the major components of task value. This model yielded an acceptable fit (χ^2 = 2827.14, *df* = 613, CFI = .939, TLI = .934, RMSEA = .044, SRMR = .065).[3] Table 3 reports the factor loadings of the first-order factors and the intercorrelations between the higher-order value components resulting from this second-order-model. The factor loadings varied from .41 (social utility) to 1 (personal importance). There was, thus, some variance in how well the value facets represented the respective higher-order factor. The correlations between the higher-order value components ranged from |.41| (utility value and cost) to |.76| (intrinsic value and cost).

Table 3

Second-Order Confirmatory Factor Analysis: Factor Loadings of the First-Order Factors and Intercorrelations Between Value Components

Factor	Intrinsic	Attainment	Utility	Cost
Importance of achievement		.83		
Personal importance		1.00		
Utility for school			.65	
Utility for daily life			.83	
Social utility			.41	
Utility for job			.76	
General utility for future life			.95	
Effort required				.91
Emotional cost				.99
Opportunity cost				.68
Intrinsic	-			
Attainment	.72	-		
Utility	.57	.75	-	
Cost	-.76	-.50	-.41	-

Note. The residual variance of the latent variable representing personal importance was fixed to 0 to identify the model. All correlations were significant at $p < .001$.

[3] When this model was computed without further constraints, a negative residual variance for personal importance was estimated. The residual variance for personal importance was then fixed to 0. This resulting model did not lead to any further estimation problems.

Table 4

Intercorrelations (Corrected For Measurement Error) Among Value Facets

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Intrinsic value	-										
(2)	Importance of achievement	.50	-									
(3)	Personal importance	.76	.86	-								
(4)	Utility for school	.38	.68	.69	-							
(5)	Utility for daily life	.49	.43	.61	.47	-						
(6)	Social utility	.40	.39	.46	.30	.36	-					
(7)	Utility for job	.39	.55	.64	.66	.51	.27	-				
(8)	General utility for future life	.53	.48	.69	.53	.86	.32	.76	-			
(9)	Effort required	-.69	-.25	-.45	-.12	-.25	-.20	-.21	-.33	-		
(10)	Emotional cost	-.76	-.34	-.56	-.32	-.33	-.19	-.31	-.44	.90	-	
(11)	Opportunity cost	-.47	-.17	-.26	*-.10*	-.16	(-.04)	-.13	-.23	.64	.67	-

Note. Correlations in parentheses were not statistically significant, correlations printed in italics were significant at $p < .01$, all other correlations reported were significant at $p < .001$.

Compared to the eleven-factor model, the second-order-model had a somewhat weaker fit, indicating that the correlations among the value facets were not perfectly represented by this higher-order-structure. We, thus, examined the correlation pattern of the value facets in more detail. Table 4 reports the intercorrelations among the value facets that resulted from the eleven-factor model. Relations within one component and between aspects of intrinsic, attainment, and utility value were positive, whereas correlations between cost and the other value components were negative. In line with a higher-order-structure, we found rather high correlations among the facets within the major task value components of attainment value, utility value, and cost that ranged from .27 to .90. However, some of the correlations among facets within one component were lower than correlations between components. For example, social utility was related more closely to aspects of attainment value (.39 - .46) than to other utility aspects (.27 - .36). These correlations seem to have resulted in a weaker fit of the second-order model relative to the first order models described above. Moreover, intrinsic value was highly correlated with some aspects of attainment value and cost, i.e., personal importance, effort required, and emotional cost (|.69| - |.76|), and personal importance also showed high correlations with facets of utility value. As we were interested in examining gender differences at a facet-level and the eleven-factor model yielded the best fit of the comprehensive models tested, we continued our analyses concerning structural invariance across gender with this eleven-factor model.

Measurement and structural invariance across gender. To address the generalizability of value facets across gender, we tested the invariance of the full factor model including all value facets. We conducted a two-group analysis in which we constrained various sets of parameter estimates to be invariant over gender to test not only for measurement invariance, but also for structural invariance. We compared fit indices for models with different sets of invariance constraints ranging from model 1 (no invariance constraints for any parameter estimates) to the most restrictive model 6 (all parameter estimates—factor loadings, item intercepts, item uniquenesses, factor variances, covariances, and means—constrained to be the same in solutions for males and females). Model fit indices for this set of nested models are reported in Table 5.

Table 5

Model Fit Statistics for Two-Group Confirmatory Factor Analysis Models

Model	Hypothesis	χ^2	*df*	CFI	TLI	RMSEA	SRMR
1	Configural invariance	2700.18	1147	.957	.950	.038	.051
2	Weak measurement invariance	2697.82	1172	.958	.952	.037	.050
3	Strong measurement invariance	2815.61	1198	.955	.950	.038	.051
4	Strict measurement invariance	2871.49	1235	.955	.951	.038	.051
5	Factor variances/covariances	2986.71	1301	.954	.952	.037	.060
6	Factor means	3143.73	1312	.950	.949	.039	.061

Note. χ^2 is the Yuan-Bentler robust test statistics; *df* = degrees of freedom*;* CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

The initial model (model 1), which included eleven correlated factors for both groups, fitted well (χ2 = 2700.18, df = 1147, CFI = .957, TLI = .950, RMSEA = .038, SRMR = .051). To ensure measurement invariance, we consecutively compared the model fit from models 1 (configural invariance) to 4 (strict factorial invariance). As the decreases in model fit were in an acceptable range according to the criteria by Cheung and Rensvold (2002) and Chen (2007), strict factorial invariance of the eleven-factor model across gender can be accepted. The measurement properties of the instrument were, thus, reasonably equal for males and females. Models 5 and 6 assessed invariance of the structure of latent constructs across gender. The low decrease in model fit from model 4 to 5 (Δ χ2 (55) = 72.13, p = .060, ΔCFI = -.001, ΔTLI = .001, ΔRMSEA = -.001, ΔSRMR = .009) provide support for structural invariance in terms of factor variances and covariances. The results provide, thus, strong support for the generalizability of the structure of value beliefs over gender. Importantly, the support for measurement invariance ensured comparability of latent constructs and allowed us to examine gender differences in latent means. When comparing a model in which latent means were constrained to be equal across gender (model 6) to a model without this constraint (model 5), we found a significant decrease in model fit, but small changes in terms of goodness-of-fit-indexes. (Δ χ2 (11) = 154.35, p < .001, ΔCFI = -.004, ΔTLI = -.003, ΔRMSEA = .002, ΔSRMR = .001) As goodness-of-fit-indexes have been developed to assess model fit in terms of covariance structure and may be insensitive to differences in mean structure (F. F. Chen et al., 2005; Fan & Sivo, 2009), we continued with inspecting possible differences in latent means in the next step.

Gender Differences in Value Facets

To evaluate the nature of latent mean differences, we compared responses for males and females on the eleven constructs based on model 5 (with factor loadings, item intercepts, item uniquenesses, factor variances, and covariances invariant over solutions for males and females, but factor means freely estimated). Females were used as a reference group in which the latent mean values were constrained to be zero. The latent means for the male group represent, thus, latent mean differences between males and females. Moreover, latent mean differences were standardized setting the variance of the latent factors to one. They can, thus, be interpreted as effect sizes comparable to Cohen's *d* (Cohen, 1988) according to Hancock (2001).

Table 6

Latent Mean Differences for Males and Females (Positive Values Reflect Higher Scores for Males)

Dimension	*M*	*SE*	*p*
Intrinsic	0.338	0.072	<.001
Importance of achievement	0.002	0.067	.980
Personal importance	0.326	0.072	<.001
Utility for school	-0.146	0.077	.057
Utility for daily life	0.122	0.066	.065
Social utility	0.103	0.060	.084
Utility for job	0.293	0.065	<.001
General utility for future life	0.312	0.071	<.001
Effort required	-0.358	0.059	<.001
Emotional cost	-0.295	0.059	<.001
Opportunity cost	-0.020	0.055	.718

Note. Standardized mean differences (based on model 5 with factor loadings, variable intercepts, variable uniquenesses, and factor variances/covariances all invariant over solutions for males and females, but latent factor means freely estimated) can be interpreted like effect sizes comparable to Cohen's *d*.

The resulting standardized mean differences are reported in Table 6. In general, latent mean differences favored boys. In line with previous research, males showed a higher intrinsic value (M = 0.338, p < .001). However, gender differences in the other major value components (attainment value, utility value, and cost) depended on the respective facet. For attainment value, females attached the same importance of achievement to math as males (M = 0.002, p = .980), but males attached more personal importance to it than females (M = 0.326, p < .001). Regarding utility value, gender differences favoring boys were found for general utility for future life (M = 0.312, p < .001) and utility for job (M = 0.293, p < .001), whereas no significant differences were found regarding utility for school (M = -0.146, p = .057), utility for daily life (M = 0.122, p = .066), and social utility (M = 0.103, p = .091). For cost, boys perceived less effort required (M = -0.358, p < .001) and emotional cost (M = -0.295, p < .001) than girls, whereas both genders perceived the same amount of opportunity cost (M = -0.020, p = .718).

In summary, tests of gender differences showed that the structure of value beliefs was similar for male and female students. However, we found considerable gender differences in latent means with boys having higher positive values for math and girls perceiving higher costs. Importantly, in line with our expectations, gender differences in latent means varied by the specific facet under consideration.[4]

Discussion

The present study examined two central questions: How many value dimensions are separable – just the four major value components or many more at a facet level? Can inconsistencies in the literature regarding gender differences in math value be resolved when switching to the facet level? In this study with a large sample of secondary school students, we found empirical support for the separability of value facets and for additional explanatory power of measuring value at the facet level. For attainment

[4] We also assessed gender differences in the 2nd order factor model. First, we assessed measurement invariance of this 2nd order factor model following the guidelines by Chen, Sousa, and West (2005). As for the eleven factor model, we found strong support for measurement invariance of this model across gender when considering Δgoodness-of-fit-indexes. Second, we found gender differences in second-order factor means favoring boys for all four value components (intrinsic value: M = 0.338, SE = 0.072, p < .001, attainment value: M = 0.253, SE = 0.071, p < .001, utility value: M = 0.253, SE = 0.070, p < .001, cost: M = -0.311, SE = 0.059, p < .001).

value, utility value, and cost, we were able to differentiate the facets that were assumed based on their definitions in EVT as well as on their previous operationalizations. The differentiated operationalization and the multidimensional statistical modeling used in the present study allowed us to take a closer look at gender differences in value beliefs about math. Whereas the structure of value beliefs was invariant across gender, we found considerable gender differences in latent means favoring boys for some factors. Importantly, these gender differences depended not only on the component, but also on the facet under consideration.

Dimensionality of Value Beliefs

The multidimensionality of value beliefs is not limited to the differentiation between the four value components defined in EVT (Eccles et al., 1983; Eccles & Wigfield, 2002). Rather, some of them include multiple aspects. In the present study, we were able to differentiate between two facets of attainment value that refer to the importance of achievement (i.e. how important it is to show high performance) and personal importance (i.e. how important it is to master a task and how central the task is for one’s identity). For utility value, it was possible to differentiate between the usefulness of math for fulfilling short- and long-term goals in various life domains. Correlations between utility facets can be interpreted as facets which share the same time perspective (e.g., general utility for future life and utility for job) or refer to similar life domains (e.g., utility for school and utility for job) being closely related. An exception was social utility which seems to be related more closely to identification with math (i.e., attainment value) than to the usefulness for other life domains. For cost, we found three facets: effort required, emotional cost, and opportunity cost.

We found mostly high correlations among the facets of one task value component. In line with the differentiation between four value components in EVT, a second-order-model (in which the facets loaded on one of the four value components) fitted the structure underlying those value facets reasonably well. However, there were several indications that the value facets and the correlations among them cannot be perfectly represented by such a higher-order-structure. Some of the correlations between facets within one component were lower than correlations of facets from different components. These correlations may partly be explainable by similarities/differences in specific item formulations. Our items assessing social utility, for instance, referred to one aspect of social goals, namely impressing others with math competencies. This can be related to confirming important aspects of the self as one notion of attainment value (cf., Eccles & Wigfield, 2002). As a consequence of these correlations, the second-order model had a considerably lower fit compared to the model that differentiated between all value facets, and the loadings in this second-order model showed some variance in how well facets represented a higher-order factor. The value facets measured in our study appear to not be strictly hierarchically ordered under the respective value components. Whereas a higher-order structure in terms of the four value components is reasonable, differentiating between a larger set of value facets provides more information. In fact, we would not have been able to detect the complex pattern of gender differences if we had assessed these differences based on the four value components.

Furthermore, the correlations between different value components depend on the specific facet. For instance, intrinsic value was correlated to a higher degree with personal importance than with the importance of achievement. As intrinsic value and personal importance can be linked to feeling- and value-related valences as parts of interest (Krapp, 2002; Trautwein et al., 2013), such a high correlation is in line with theoretical considerations and empirical research (Schiefele, 2009; Trautwein et al., 2012). Also, positive and negative feelings associated with mathematics (i.e., intrinsic value and emotional cost) were highly interrelated (see also Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). When looking at the overall correlation pattern of value facets, it seems that they assess value for mathematics at different levels of specificity. Personal importance was closely related to almost all other facets and can be seen as a more general attribution of value to mathematics that could have been influenced by other facets such as the utility for different life domains (including career opportunities, but also leisure time and social relations). However, causal-ordering of relations among value facets is not possible relying on data of the present study, but would require a longitudinal assessment of value beliefs.

Gender Differences in Math Values

Whereas the structure of value beliefs was similar for boys and girls, we found gender differences in mean levels of these constructs. In accordance with gender-linked stereotypes of mathematics, these differences generally favored boys. Although this would be expected based on EVT and previous research, the pattern of results at a facet level was more complex. In line with previous research, we found lower intrinsic value for girls than boys. In addition, girls perceived math as less personally important and less useful for their general as well as for their professional future than boys, while at the same time, they perceived higher emotional cost and effort required compared to boys. No significant gender differences were found for the importance of achievement, opportunity cost, and the usefulness for short-term goals (i.e., school, daily life, social life). The only difference that tended to favor girls was utility for school, which is in line with girls perceiving school and good grades in general as more valuable than boys (Kessels, Heyder, Latsch, & Hannover, 2014; Steinmayr & Spinath, 2010).

Altogether, the results of our study point to differences as well as similarities in girls' and boys' value beliefs about math (cf., Hyde, 2005). It appears that girls perceive math as a subject of importance in the school context, while perceiving it as personally unimportant and unrelated to their future plans. Math is, thus, perceived as a necessary evil that is not intrinsically valued. The pattern of value beliefs in math found in females could potentially lead to an impaired affective experience of math classes (e.g., Frenzel et al., 2007) and eventually result in dropping math as soon as the school systems allows for it (e.g., Watt et al., 2006). All significant differences found were small in size ($0.29 < d < 0.36$) according to the widely used guidelines by Cohen (1988). However, these differences in value beliefs are still larger than gender differences reported in meta-analyses for mathematics achievement, where close-to-zero differences have been found for standardized tests ($d < 0.15$) as well as grades ($d = -0.07$; Else-Quest et al., 2010; Voyer & Voyer, 2014). Value beliefs for math are, thus, a likely factor contributing to gender differences in choices.

The differentiated measurement of value beliefs applied in the current study extends previous findings in several ways. First, we found gender differences in value beliefs for aspects that have been less examined, such as emotional cost and effort required. Second and even more importantly, differentiating between separate value facets actually helps to explain inconsistent findings in previous research. Concerning attainment value, Marsh et al. (2005) and Nagy et al. (2006) assessed interest as a combination of what we call intrinsic value and personal importance and found gender differences, whereas other studies (Frenzel et al., 2007; Meece et al., 1990; Steinmayr & Spinath, 2010) assessed what we call importance of achievement and found no differences. Regarding utility value, Watt (2004) found no gender differences while including items tapping short- as well as long-term goals (focusing on general utility, everyday world, and the workplace), whereas Spinath and Steinmayr (2010) found gender differences for general utility for future life. That is, whereas gender differences might also vary depending on students' age and cultural background, inconsistencies in previous findings can be explained by differences in the operationalizations and lack of systematic differentiation between facets.

Implications

Our results have several implications for educational research. First, they point to the difficulties that result from differing operationalizations of motivational constructs in educational research. Items used in different studies that are assumed to measure the same construct (e.g., attainment value) seem to measure slightly different constructs (i.e., importance of achievement vs. personal importance). When it comes to the operationalization of value beliefs in future research, even more attention should be paid to the fit between the defined construct and its operationalization. We have provided one example of jingle-jangle-fallacies in motivational research concerning task value, but these problems get even more complex when comparing different motivational constructs and theories. As described in a comprehensive review by Wigfield and Cambria (2010), similarities in the measures for task values, goal orientations, and interest can obscure conceptual differences between these constructs and more research is needed to further investigate their associations. Some of the distinctions we have made between different sets of value beliefs here correspond to distinctions that other motivational theories make. For instance, the distinction between importance of achievement and personal importance for attainment value can be related to the distinction of mastery and performance goal orientation (Maehr & Zusho, 2009).

Second, our results point to taking into account multiple facets in the measurement of value beliefs and its statistical modeling. Although facets of one value component were highly correlated in the present study, we only found pronounced gender differences for individual facets. The less differentiated modeling that was used for value beliefs in many studies, often including one general value scale, potentially masked such gender differences. Differentiating between value facets can increase the explanatory power of EVT for gender differences in academic choices. The next step is to examine whether different value facets are differentially associated with related constructs such as expectancy and whether they have differential predictive validity for relevant outcomes, such as effort, persistence, and achievement. Another way of modeling interindividual differences that allows examining how individual patterns of motivation predict outcomes is the person-centered approach. This approach takes combinations of different motivational sources, as highlighted in EVT (Wigfield & Cambria, 2010), into account. To date, only few studies have used a person-centered perspective for looking at gender differences in task value (Chow, Eccles, & Salmela-Aro, 2012; Conley, 2012). It would be interesting to investigate which role value facets play in differentiating between motivational patterns and how males and females differ in those patterns.

Third, differentiating between separate types of value beliefs might help develop fruitful interventions to enhance students' motivation. Given its more extrinsic nature, utility value seems to be a value component which can be influenced by means of classroom interventions. Several studies have been successful in enhancing students' interest and grades by targeting students' utility perceptions (Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009). However, it might make a difference whether utility for job or for daily life is the aim of such an intervention. Intervention studies should carefully check for such differential effects. Interventions designed to improve attitudes towards mathematics would be specifically promising for girls, who especially differ more from boys in their perceptions of the usefulness for long-term. Knowing about the specific differences in beliefs between boys and girls can help in developing such targeted interventions.

Limitations and Future Research

Some limitations of our study should be considered when interpreting its results. First, one has to keep the specific characteristics of our sample in mind, that is, ninth graders within German academic track schools. The study should be replicated with other age groups to determine whether the structure of value beliefs is specific to our sample of ninth grade students or whether it also applies to other age groups. It is possible that younger students have less concrete ideas about their future and might, thus, have less differentiated beliefs about the usefulness of a subject. Moreover, it is unclear to what extent the gender differences we detected in our study depend on this age cohort. Previous studies investigating the development of interest for mathematics found that gender differences in these beliefs emerge relatively early and are rather stable (Frenzel et al., 2010; Jacobs et al., 2002; Nagy et al., 2010; Watt, 2004). Finally, generalizability of our results to other school types and cultures should also be investigated in future research.

Second, our items were limited to the domain of mathematics. However, other studies have shown fairly consistent correlation patterns in different domains such as mathematics and English (Trautwein et al., 2012) as well as an invariant factor structure across different subjects (Nagengast, Trautwein, Kelava, & Lüdtke, 2013). Moreover, we were especially interested in gender differences in math values for their assumed importance for explaining gender differences in educational and occupational careers. Future research should replicate our findings for different school subjects and try to map gender differences in mathematics to other domains, such as English, where females typically have more favorable beliefs than males.

Third, some of the scales at the facet level were rather short to keep the length of the questionnaire in balance. Specifically, some utility facets were measured by only two items. Using short scales can undermine reliability as well as validity. However, confirmatory factor analyses as used in the present study correct for effects of measurement errors (Bollen, 1989). Still, a scale with few items but a high reliability might represent the respective construct to a limited extent. For instance, the items we developed to assess utility for social goals referred to impressing others (i.e., classmates) with competencies in a certain domain, which might be only one aspect of how socials goals can influence value beliefs. Future research might, therefore, develop the items further to broaden the coverage of the underlying construct. Still, we were able to differentiate between utility facets and even found differentiated gender effects. Using more extensive measures might even lead to more complex patterns. Future studies on value facets might want to expand the instrument with more and refined items relating to the value aspects of interest. At the same time, our value instrument was already rather long with a total of 37 items, and using such a long questionnaire might not be feasible in many studies, especially if value beliefs are not the major focus. It seems, therefore, also important to further develop short scales that are efficient in predicting students' learning behavior and academic choices (on the advantages and disadvantages of short scales, see Crede, Harms, Niehorster, & Gaye-Valentine, 2012;

Gogol et al., 2014; Smith, McCarthy, & Anderson, 2000). Our results provide strong support for the multidimensionality of value beliefs. If future studies cannot assess value beliefs in terms of all the facets we differentiated in this study, it still seems important to carefully choose the aspects that are of interest for a particular study.

Conclusion

The present study aimed at taking a closer look at gender differences in value beliefs about math using a differentiated approach. To this end, we examined value beliefs at the facet-level within a large sample of ninth grade students. Our results showed that it is crucial to understand value beliefs as a multidimensional construct with gender differences depending on the respective value facet. Future research should examine further if different value facets have differential importance for explaining achievement-related outcomes. Furthermore, the generalizability of the structure of value beliefs to age groups beyond adolescence and to subject domains other than mathematics should be examined. Finally, the potential effectiveness of classroom interventions targeting boys' and girls' value perceptions should be a primary focus of future research.

References

Battle, A., & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior, 62,* 56-75. doi: 10.1016/s0001-8791(02)00037-4

Boekaerts, M. (2009). Goal-directed behavior in the classroom. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 105-122). New York, NY: Routledge.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task-value, and achievement goals. *Journal of Educational Psychology, 93*, 23-34. doi: 10.I037//0022-0663.93.1.23

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods and Research, 36*, 462-494. doi: 10.1177/0049124108314720

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504. doi: 10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471-492. doi: 10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi: 10.1207/S15328007SEM0902_5

Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology, 48,* 1612-1628. doi: 10.1037/a0030194

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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

Crede, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888. doi: 10.1037/a0027403

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology, 98*, 382-393. doi: 10.1037/0022-0663.98.2.382

Eccles, J. S. (2005). Subjective task values and the Eccles et al. model of achievement related choices. In A. J. Elliott & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York, NY: Guilford.

Eccles, J. S. (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*, 78-89. doi: 10.1080/00461520902832368

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74-146). San Francisco, CA: W. H. Freeman.

Eccles, J. S., Adler, T. F., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*, 26-43. doi: 10.1037/0022-3514.46.1.26

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Maciver, D. (1993). Development during adolescence—The impact of stage-environment fit on young adolescents experiences in schools and in families. *American Psychologist, 48*, 90-101. doi: 10.1037//0003-066x.48.2.90

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

Eccles, J. S., 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. doi: 10.1111/j.1467-8624.1993.tb02946.x

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103-127. doi: 10.1037/a0018053

Fan, X., & Sivo, S. A. (2009). Using Δgoodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16,* 54-69. doi: 10.1080/10705510802561311

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence, 20,* 507-537. doi: 10.1111/j.1532-7795.2010.00645.x

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics - A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 12,* 497-514. doi: 10.1007/BF03173468

Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., . . . Pekrun, R. (Eds.). (2009). *PISA 2006 Skalenhandbuch. Dokumentation der Erhebungsinstrumente [PISA 2006 handbook of scales. Documentation of assessment instruments]*. Münster, Germany: Waxmann.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., . . . Preckel, F. (2014). “My questionnaire is too long!” The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39,* 188-205. doi: 10.1016/j.cedpsych.2014.04.002

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing means on a single latent construct. *Psychometrika, 66,* 373-388. doi: 10.1007/BF02294440

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55. doi: 10.1080/10705519909540118

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102,* 880-895. doi: 10.1037/a0019506

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

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136,* 422-449. doi: 10.1037/a0018947

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581-592. doi: 10.1037/0003-066X.60.6.581

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509-527. doi: 10.1111/1467-8624.00421

Juvonen, J., Espinoza, G., & Knifsend, C. (2012). The role of peer relationships in student academic and extracurricular engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 387-401). New York, NY: Springer.

Kessels, U., Heyder, A., Latsch, M., & Hannover, B. (2014). How gender differences in academic engagement relate to students' gender identity. *Educational Research, 56*, 220-229. doi: 10.1080/00131881.2014.898916

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12*, 383-409. doi: 10.1016/S0959-4752(01)00011-1

Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The Mathematics Value Inventory for general education students: Development and initial validation. *Educational and Psychological Measurement, 70*, 142-160. doi: 10.1177/0013164409344526

Maehr, M. L., & Zusho, A. (2009). Achievement goal theory: The past, present, and future. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 77-104). New York, NY: Francis.

Marsh, H. W., Craven, R. G., Hinkley, J. W., & Debus, R. L. (2003). Evaluation of the big-two-factor theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research, 38,* 189-224. doi: 10.1207/s15327906mbr3802_3

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341. doi: 10.1207/s15328007sem1103_2

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects of causal ordering. *Child Development, 76,* 397-416. doi: 10.1111/j.1467-8624.2005.00853.x

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*, 60-70. doi: 10.1037/0022-0663.82.1.60

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543. doi: 10.1007/BF02294825

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition.* Los Angeles, CA: Muthén & Muthén.

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*, 1058-1066. doi: 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. doi: 10.1080/00273171.2013.775060

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*, 323-345. doi: 10.1080/13803610600765687

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence, 20,* 482-506. doi: 10.1111/j.1532-7795.2010.00644.x

OECD. (2004). *Education at a Glance 2004: OECD Indicators*. Paris, France: OECD Publishing.

Peetsma, T., & van der Veen, I. (2011). Relations between the development of future time perspective in three life domains, investment in learning, and academic achievement. *Learning and Instruction*, *21*, 481-494. doi: 10.1016/j.learninstruc.2010.08.001

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36-48. doi: 10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., & Perry, R. P. (2005). Academic Emotions Questionnaire (AEQ). User's manual: Department of Psychology, University of Munich.

Pekrun, R., Götz, T., S., J., Zirngibl, A., Hofe, R. v., & Blum, W. (2002). Skalenhandbuch PALMA: 1. Messzeitpunkt (5. Klassenstufe) [PALMA handbook of scales: First measurement point (5th grade)]: Department of Psychology, University of Munich.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106,* 315-329. doi: 10.1037/a0034027

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667-686. doi: 10.1037/0022-0663.95.4.667

Rakoczy, K., Buff, A., & Lipowsky, F. (2005). Befragungsinstrumente [Questionnaires]. In E. Klieme, C. Pauli, & K. Reusser (Eds.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis. Materialien zur Bildungsforschung [Technical report of the Swiss-German video study Instructional quality, learning behavior, and mathematical* *understanding. Materials for educational research]*. Frankfurt am Main, Germany: GFPF.

Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., . . . Schiefele, U. (Eds.). (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente [PISA 2003. Documentation of assessment instruments]*. Münster, Germany: Waxmann.

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42*, 223-232. doi: 10.1177/0748175609344096

Ryan, R. M., & Deci, E. I. (2009). Promoting self-determined school engagement: motivation, learning and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 171-195). New York, NY: Routledge.

Salmela-Aro, K. (2009). Personal goals and well-being during critical life transitions: The four C's - channelling, choice, co-agency and compensation. *Advances in Life Course Research, 14*, 63-73. doi: 10.1016/j.alcr.2009.03.003

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177. doi: 10.1037//1082-989X.7.2.147

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197-222). New York, NY: Routledge.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120. doi: 10.1007/s11336-008-9101-0

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111. doi: 10.1037//1040-3590.12.1.102

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality, 22*, 185-209. doi: 10.1002/per.676

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and validation of a scale for the assessment of school-related values]. *Diagnostica, 56*, 195-211. doi: 10.1026/0012-1924/a000023

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*, 763-777. doi: 10.1037/a0027470

Trautwein, U., Nagengast, B., Marsh, H. W., Gaspard, H., Dicke, A.-L., Lüdtke, O., & Jonkmann, K. (2013). Expectancy-value theory revisited. From expectancy-value theory to expectancy-valueS theory? In D. M. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self-processes and human development* (pp. 233-249). Charlotte, NC: Information Age Publishing.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. doi: 10.1177/109442810031002

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174-1204. doi: 10.1037/a0036620

Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development, 75*, 1556-1574. doi: 0.1111/j.1467-8624.2004.00757.x

Watt, H. M., Eccles, J., & Durik, A. M. (2006). The leaky mathematics pipeline for girls: A motivational analysis of high school enrolments in Australia and the USA. *Equal Opportunities International, 25*, 642-659. doi: 10.1108/02610150610719119

Watt, H. M., & Eccles, J. S. (Eds.). (2008). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences*. Washington, DC: American Psychological Association.

Watt, H. M., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Developmental Psychology, 48*, 1643-1657. doi: 10.1037/a0027838

Wentzel, K. R. (2005). Peer relationships, motivation, and academic performance at school. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 279-296). New York, NY: Guilford Press.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*, 1-35. doi: 10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values. *Developmental Review, 12*, 265-310. doi: 10.1016/0273-2297(92)90011-P

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451-469. doi: 10.1037/0022-0663.89.3.451

Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. W. Wentzel, A. (Ed.), *Handbook of motivation at school* (pp. 55-75). New York, NY: Routledge.

Yuan, K. H., & Bentler, P. M. (1998). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 51*, 63-88. doi: 10.1111/j.2044-8317.1998.tb00667.x

Appendix

Item Sources for All Value Facets

Value facets are based on items from the following sources:

1. Intrinsic: Frey et al. (2009); Pekrun, Goetz, & Perry (2005); Pekrun, Götz, Zirngibl, Hofe, & Blum (2002), Steinmayr & Spinath (2010)
2. Importance of achievement: Steinmayr & Spinath (2010) and self-developed items
3. Personal importance: Rakoczy et al. (2005); Trautwein et al. (2012), and self-developed items
4. Utility for school: self-developed items
5. Utility for daily life: self-developed items
6. Social utility: self-developed items
7. Utility for job: Ramm et al. (2006); Trautwein et al. (2012)
8. General utility for future life: Steinmayr & Spinath (2010) and self-developed items
9. Effort required: self-developed items
10. Emotional cost: self-developed items
11. Opportunity cost: Conley, (2012); Trautwein et al. (2012)

3

Fostering Adolescents' Value Beliefs for Mathematics with a Relevance Intervention in the Classroom

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (in press). Fostering Adolescents' Value Beliefs for Mathematics with a Relevance Intervention in the Classroom. *Developmental Psychology*.

Abstract

Interventions targeting students' perceived relevance of the learning content have been shown to effectively promote student motivation within science classes (e.g., Hulleman & Harackiewicz, 2009). Yet, further research is warranted to understand better how such interventions should be designed in order to be successfully implemented in the classroom setting. A cluster randomized controlled study was conducted to test whether ninth-grade students' value beliefs for mathematics (i.e., intrinsic value, attainment value, utility value, and cost) could be fostered with relevance interventions in the classroom. Eighty-two classrooms were randomly assigned to one of two experimental conditions or a waiting control condition. Both experimental groups received a 90-minute intervention within the classroom on the relevance of mathematics, consisting of a psychoeducational presentation and relevance-inducing tasks (either writing a text or evaluating interview quotations). Intervention effects were evaluated via self-reports of 1916 participating students six weeks and five months after the intervention in the classroom. Both intervention conditions fostered more positive value beliefs among students at both time points. Compared to the control condition, classes in the quotations condition reported higher utility value, attainment value, and intrinsic value, and classes in the text condition reported higher utility value. Thus, stronger effects on students' value beliefs were found for the quotations condition than for the text condition. When assessing intervention effects separately for females and males, some evidence for stronger effects for females than for males was found.

Keywords: expectancy-value theory; task value; intervention; motivation; mathematics; gender.

Fostering Adolescents' Value Beliefs for Mathematics with a Relevance Intervention in the Classroom

"Why should I learn all this stuff in mathematics?" Most students have already asked themselves this question. Students can find different answers ranging from "It's just fun" to "It will help me get my dream job" (cf., Eccles et al., 1983). Such beliefs about the value of certain subjects have been found to predict academic choices, effort, and persistence (for a review, see Wigfield, Tonks, & Klauda, 2009). Research has shown that—on average—students' value beliefs in various subjects, particularly in mathematics, decline across secondary school (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Watt, 2004). In mathematics, female students are especially at risk as they have been found to report even lower value beliefs for math than their male counterparts in secondary school (e.g., Frenzel, Pekrun, & Goetz, 2007; Watt, 2004).

Is it possible to buffer decreases in student motivation and to reduce gender differences in motivation for mathematics? Within the last few years, a number of interventions have been developed to enhance motivation in areas related to science, technology, engineering, and mathematics (STEM) (for an overview, see Karabenick & Urdan, 2014). Some of these interventions foster motivation by helping students find meaning for what they learn (Brophy, 1999; Hidi & Harackiewicz, 2000). Studies with high school (Hulleman & Harackiewicz, 2009) and college students (Hulleman, Godes, Hendricks, & Harackiewicz, 2010) have shown that such relevance interventions are a successful tool to foster motivation. In these studies, relevance interventions were administered to individual students who wrote several essays about the relevance of the learning content to their lives. In the educational context, it is of central interest whether short interventions implemented at the classroom level can be used to effectively promote students' motivational development. More research is also needed on the tasks that are most effective for inducing relevance.

In the present study, we tested whether ninth-grade students' value beliefs for mathematics would be enhanced by relevance interventions in the classroom setting. To this end, 82 classes were randomly assigned to one of two intervention conditions or a waiting control group. The intervention consisted of a 90-minute session in which a psychoeducational presentation providing information on the relevance of mathematics was combined with individual tasks triggering relevance. For the relevance-inducing tasks, we compared a previously used task (i.e., self-generation of arguments for the usefulness of mathematics) with a newly developed task (i.e., reflection on typical arguments given by young adults). We assessed effects of the two intervention conditions on all four value components (intrinsic, attainment, utility, cost) and also tested whether the intervention was equally effective for boys and girls.

Intervening on Students' Value Beliefs

Several intervention studies targeting value beliefs have recently been conducted in the lab and to some extent also in the classroom (for an overview, see Harackiewicz, Tibbetts, Canning, & Hyde, 2014). These intervention studies utilized the expectancy-value theory (EVT) by Eccles et al. (1983) as a theoretical framework, which provides an elaborate view of the role of value beliefs for academic development. The Eccles et al. (1983) EVT model conceptualizes task value in terms of four distinct value components: intrinsic, attainment, utility, and cost. Intrinsic value is defined as the enjoyment a person derives from doing a task and has been linked to individual interest. Attainment value refers to the importance that individuals attach to doing well on a given task and relates to the relevance of a task for one's identity. Utility value indicates the perceived usefulness of engagement in a task for short- as well as long-term goals. Finally, cost describes the perceived negative consequences of engaging in a task (Eccles, 2005; Eccles & Wigfield, 2002; Wigfield & Eccles, 1992). Research has supported the basic assumptions of EVT showing that value beliefs predict positive student outcomes in various school subjects (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Nagengast, Trautwein, Kelava, & Lüdtke, 2013; Trautwein & Lüdtke, 2007) as well as academic choices (e.g., Durik, Vida, & Eccles, 2006; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Simpkins, Davis-Kean, & Eccles, 2006).

Most evidence supporting EVT and the role of task value thus stems from correlational research. How can students' perceptions of task value be promoted? Triggering intrinsic or attainment value may be difficult as the enjoyment of a task and identification with it seem to depend on individual characteristics (Eccles, 2005). Elaborating on more rational reasons why a subject is relevant for a student's life, however, may be a feasible way to foster perceptions of meaningfulness. Compared to attainment and intrinsic value, utility value is more extrinsic in nature (Eccles & Wigfield, 2002) and seems to be more easily influenced from the outside. In line with these assumptions, previous intervention studies (Hulleman et al., 2010; Hulleman &

Harackiewicz, 2009) focused on utility value.

Two types of intervention approaches have been previously applied to enhance utility value in different studies. The first approach, directly communicating utility information, was applied in a number of laboratory studies with college students (Durik & Harackiewicz, 2007; Shechter, Durik, Miyamoto, & Harackiewicz, 2011). When learning a new math technique, intervention groups received information about how this technique could be useful for achieving short- or long-term goals. This information had positive effects on competence valuation, task involvement, and perceived competence as well as interest and performance for students with high initial motivation.

The second approach encouraged students to self-generate arguments for the utility of the material to their lives and was successfully applied in the lab and in the classroom (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009). Hulleman and colleagues (2010) conducted two randomized experiments using this approach: In the laboratory, participants were asked to write an essay on the relevance of a math technique to their lives. In the classroom, students in two intervention conditions completed two writing tasks each, either letters about the relevance of a topic to their lives or essays about the relevance of a media report to the topic covered in class. These writing interventions promoted utility value and interest compared to a control condition. In a similar study in high school science classrooms (Hulleman & Harackiewicz, 2009), 262 students were randomized within classrooms and students within the relevance condition wrote a total of eight essays about the meaning of the course material to their lives. This had positive effects on interest and course grades for students with low expectancies.

Remaining Questions on the Effects of Relevance Interventions

Altogether, previous studies provided valuable insights into the effects of relevance interventions on student motivation. When it comes to applying interventions theoretically grounded in EVT in the classroom, some of the most important questions are, however, still unresolved: How should interventions be designed to get an effect in real classroom situations? Which kinds of beliefs can be affected by relevance interventions—only utility value or other value beliefs as well? Are relevance interventions a way to reduce gender differences in motivation for STEM subjects? To address these questions, several factors need to be taken into account, which will be addressed in the following paragraphs.

Designing relevance interventions that are effective in the classroom setting. As described above, two kinds of interventions have been used so far to foster utility value: providing arguments for the usefulness of a topic and self-generating such arguments. Combining both approaches within one intervention might have a stronger impact on motivation. A combination of persuasive messages and writing assignments was already successfully applied in a small-scale intervention study within an undergraduate introductory statistics course (Acee & Weinstein, 2010). The intervention applied various strategies to foster self-regulated learning and to guide students in exploring the value of statistics. Studies by Hulleman and colleagues (2010; 2009) have also shown that making personal connections and triggering reflection are crucial elements of effective interventions in the classroom. These processes can be triggered in various ways and writing essays seems to be one of them. However, writing essays might be difficult, especially for younger students, if students are not provided with any background information. One way to trigger reflection processes and elicit more connections would be combining both approaches: providing some possible arguments beforehand and have students generate connections to their own lives afterwards.

Reflection and personal connections could also be promoted when students receive typical arguments for the utility of mathematics from people that they can easily connect to. Drawing on a social cognition perspective, several theories such as social learning theory (Bandura, 1977), possible-selves theory (Markus & Nurius, 1986), and identity-based motivation (Oyserman & Destin, 2010) suggest that adolescents can benefit from positive role models. Such role models can be important in terms of representing a possible future identity as well as providing information on the path to this identity. Interview quotations in which older students describe the usefulness of subject knowledge to them may be one way to give students personal and authentic information about the relevance to their future lives. Harackiewicz, Rozek, Hulleman, and Hyde (2012) implemented this idea as part of an intervention targeting parents by presenting interviews with college students referring to the usefulness of high school STEM courses on a website. These interviews were, however, part of a more comprehensive intervention, so that their effect was not directly evaluated.

In order to create interventions that are effective in real life, one also needs to consider the context: Students are nested within classrooms. Previous studies assigned individual students within classes to experimental conditions (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009). However, implementing interventions at the classroom level might be more beneficial as this comes closer to the natural learning setting in schools. The classroom setting could be utilized for providing information on the relevance of subject material for future life, engaging students in discussions and thereby also triggering active reflection. Implementing interventions at the classroom level not only has benefits for creating more powerful interventions, but can also increase the precision for evaluating effects of such interventions. In within-classroom designs, students within one class are in different experimental conditions, and interactions between students in those groups can lead to biased estimates of intervention effects (Craven, Marsh, Debus, & Jayasinghe, 2001; Plewis & Hurry, 1998). Between-classrooms designs, in which all students within one class are in the same condition, are a means to reduce the risk of diffusion effects; however, they require relatively large sample sizes to have an adequate power.

Effects of relevance interventions on subcomponents of task value. Whereas it has become clear that relevance interventions can be an effective way to foster motivation (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009), more needs to be learned about the complexity of the effects on value beliefs. Although the Eccles et al. (1983) EVT model describes four theoretically distinct components, previous research on students' value beliefs often incorporated positive value aspects (i.e., attainment, intrinsic, and utility value) into a single value scale (e.g., Bong, 2001; Jacobs et al., 2002). Recent studies, however, were able to separate four components using confirmatory factor analysis with items that explicitly tapped all of them (Conley, 2012; Trautwein et al., 2012). When assessed separately, all value components have been associated with important student outcomes: Attainment and utility value seem to be especially important for career aspirations and course choices (Durik et al., 2006; Watt et al., 2012), intrinsic value predicts leisure time activities (Durik et al., 2006; Nagengast et al., 2011), and cost adds to the predictive power of positive value beliefs for educational intentions (Battle & Wigfield, 2003; Perez, Cromley, & Kaplan, 2014). Although it seems that the four value components predict different important outcomes, more research disentangling the role of separate components is needed.

Theoretically, the four value components are assumed to be formed through different processes (cf., Eccles, 2005) and might, therefore, also be affected through interventions in different ways. In previous intervention studies, effects on students' value beliefs have been assessed in terms of utility value (Hulleman et al., 2010) and constructs related to intrinsic and attainment value such as interest (Durik & Harackiewicz, 2007; Hulleman et al., 2010; Hulleman & Harackiewicz, 2009) and competence valuation (Durik & Harackiewicz, 2007). To assess effects of interventions on students' value beliefs comprehensively, however, all components need to be taken into account simultaneously using theoretically valid and psychometrically sound instruments. Theoretically, stimulation of relevance should not only foster utility value, but also engagement and a more intrinsic motivation by eliciting positive feelings associated with a task and fostering identification, that is intrinsic and attainment value (Hidi & Harackiewicz, 2000; Hidi & Renninger, 2006).

Depending on the focus of the intervention, one might expect stronger effects for some value beliefs than for others. Relevance interventions can either focus on the usefulness for long-term goals such as career opportunities or on the usefulness for short-term goals such as solving daily life problems (cf. Shechter et al., 2011). Reflections drawing on different future time perspectives (cf., Nuttin & Lens, 1985) might affect different kinds of value beliefs. To be able to assess such effects, value beliefs need to be measured with even more differentiation. When looking more closely at the definition of the four value components, subfacets of attainment value, utility value, and cost can be distinguished (Gaspard et al., 2014; Trautwein et al., 2013). Utility value can refer to short- and long-term goals in a variety of life domains, including school, daily life, and social life in the short term and job and future life in general in the long term. Attainment value can be differentiated into a facet that focuses on performance (importance of achievement) and a facet that is more related to identity issues (personal importance). Cost can be divided into effort required, negative emotions associated with engagement in a task, and opportunity cost of choosing one option over another (Gaspard et al., 2014; Perez et al., 2014). Support for validity of these subfacets has been found in previous research in terms of a differentiated pattern of gender differences in math value beliefs (Gaspard et al., 2014) as well as differential contributions of types of cost for predicting choices (Perez et al., 2014). In intervention studies, measurement at the facet level is needed to gain insight into the kinds of beliefs that were affected. For relevance interventions, subfacets of utility value are of particular interest. Relevance interventions promote connections between the learning material and students' personal goals. These connections can refer to different life domains, such as future careers or daily life. Other life domains such as social goals might be more difficult to affect as they depend on students' context.

Gender as a potential moderator of the effects of relevance interventions. Are relevance interventions a way to reduce gender differences in motivation for mathematics? Females are underrepresented in mathematics and related careers and this cannot be explained sufficiently by gender differences in achievement (Else-Quest, Hyde, & Linn, 2010; OECD, 2004; Watt & Eccles, 2008). EVT has been applied successfully to explain such gender differences in choices by expectancy and value beliefs (e.g., Chow, Eccles, & Salmela-Aro, 2012; Nagy et al., 2006; Watt et al., 2012). Gender differences reported in value beliefs in previous studies are, however, somewhat inconsistent and seem to depend on the type of value (e.g., Frenzel et al., 2007; Gaspard et al., 2014; Marsh et al., 2005; Meece, Wigfield, & Eccles, 1990; Watt, 2004; Watt et al., 2012). The overall pattern of gender differences can be interpreted as girls seeing high performance in mathematics as important, whereas perceiving it as a rather unattractive subject. Using data from the sample participating in the present intervention study, Gaspard et al. (2014) found that boys reported higher intrinsic value, higher personal importance as one facet of attainment value, higher utility for job and general utility for future life as facets of utility value, and lower effort required and emotional cost as facets of cost before the intervention.

Drawing on these findings, how can females' motivation for mathematics be fostered? As females tend to differ from males regarding the type of career they aspire to (Eccles, 2011; Watt, 2008), they may especially benefit from information regarding the usefulness of mathematics for more female-typed domains (e.g., statistics for psychology classes in college), which might be new to them (cf. Wang, 2012). Rozek, Hyde, Svoboda, Hulleman, and Harackiewicz (2015) showed that the effects of a relevance intervention helping parents to motivate their adolescent children in STEM were moderated by gender and achievement. The intervention increased the number of STEM courses taken for high-achieving girls and low-achieving boys, whereas no effect was found for high-achieving boys and the intervention tended to have negative effects for low-achieving girls. Hulleman and Harackiewicz (2009) found no moderating effect of gender for the effects of a relevance intervention on interest and performance in high school sciences classes. There is thus no clear evidence on whether relevance interventions decrease gender differences in motivation for STEM fields, and gender effects might also depend on the type of intervention.

The Present Study

Extending previous intervention studies (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009), we conducted a cluster randomized controlled study to test whether students' value beliefs in mathematics could be promoted by relevance interventions in the classroom. Several new design features were introduced to further increase the effectiveness and practicality of these interventions. First, we used a between-classrooms design with an adequately large number of classrooms to utilize the classroom setting for triggering reflections and to reduce diffusion effects that can occur in within-classroom designs. Eighty-two ninth-grade classes were randomly assigned to one of two intervention groups or a waiting control group. Second, to make our intervention as effective as possible, we combined different approaches: background information on the utility of mathematics and relevance-inducing tasks to trigger reflections and personal connections. Both intervention conditions consisted of a 90-minute session about the relevance of mathematics and two short reinforcement exercises to be done at home. Students in both conditions first participated in a psychoeducational presentation that focused on the relevance of one's attitude for learning mathematics and the relevance of mathematics for future life providing examples from different fields. Then, students in the two intervention conditions worked on relevance-inducing tasks, where we systematically compared a new strategy, (i.e., evaluating interview quotations) to a previously used one (i.e., writing a text about the relevance of mathematics).

To evaluate the effects of the intervention conditions, we assessed students' value beliefs before and after the intervention as well as in a follow-up test. The value instrument consisted not only of measures of all four value components, but also included subfacets of attainment value (i.e., importance of achievement and personal importance), utility value (i.e., utility for school, daily life, social life, job, and future life in general), and cost (i.e., emotional cost, effort required, and opportunity cost).

Our study had three major research questions. First, we examined whether ninth-grade students' value beliefs (intrinsic, attainment, utility, and cost) could be enhanced by two different relevance interventions within mathematics classrooms. Strongest effects for both conditions were expected on utility value. However, students may draw more personal connections and involve more deeply in the task when realizing the utility of a task leading to an increase in intrinsic and attainment value (Hidi & Renninger, 2006; Shechter et al., 2011). Second, we assessed whether intervention effects differ depending on the value facet under consideration. With regards to utility value, we expected stronger effects for those life domains specifically addressed in the intervention, particularly utility for future job opportunities, but also daily life. Third, we investigated whether intervention effects differ depending on gender. Previous research has shown that girls report lower value beliefs for mathematics than boys including the sample under investigation (e.g., Frenzel et al., 2007; Gaspard et al., 2014). We, therefore, wanted to test if gender differences can be reduced by relevance interventions.

Methods

Sample and Procedure

Data for the study “Motivation in Mathematics” (MoMa) were collected in 82 ninth-grade classes in 25 academic track schools in the German state of Baden-Württemberg from September 2012 to March 2013. In Germany, mathematics is taught as one comprehensive course including different topics, such as algebra, geometry, or calculus. In the academic schools we studied, mathematics was compulsory with no level of choice regarding the amount or level of courses (i.e., all students have four compulsory mathematics lessons per week). A total of 1978 students with active parental consent participated in the study. These 1978 students are 96% of the total number of students in these 82 classes, yielding a very high participation rate. For the current study, 62 students in the two intervention conditions were excluded as they were absent during the intervention. Data analyses were, thus, based on a sample of 1916 students (mean age at the beginning of the study = 14.62, SD = 0.47, 53.5% female). The study consisted of three waves of data collection. Students were administered questionnaires by trained research assistants before the intervention (pretest), on average six weeks after the intervention (posttest), and on average five months after the intervention (follow-up).

Before recruiting the participating classes for our study, we conducted a power analysis for a multi-site cluster randomized trial with the treatment implemented at level 2 (i.e., classes within schools are randomly assigned to experimental conditions) with Optimal Design (Raudenbush et al., 2011). This power analysis indicated that we would get an acceptable power ($\beta = .73$) to detect intervention effects of $\delta = 0.20$ (comparing a single intervention condition to the control condition) for a total number of 25 schools (with one class per experimental condition and $n = 25$ students per class), under the following realistic assumptions: First, that the intra-class correlations for our outcomes were low (.05); second, that only little variance was explained by the school level (0.005); third, that 50% of the variance at level 2 was explained by a pretest measure used as a covariate. Given our resources, this set-up seemed to represent the best we could achieve balancing test power and feasibility (cf. Moerbeek, 2006; Raudenbush, 1997) and we, therefore, set out to recruit 25 schools.

We initially recruited 26 schools with a total number of 77 teachers and 87 classes (1-5 classes per school) that were willing to participate in our study. Before the first wave of data collection, within each school, the teachers (and their classes) were randomly assigned to one of two intervention conditions or a waiting control group. After randomization and before the first wave of data collection, four teachers from two different schools dropped out of the study (quotations condition: 1 class, text condition: 3 classes, waiting control condition: 1 class) due to organizational reasons. The remaining classes (quotations condition: 25 classes, text condition: 30 classes, waiting control condition: 27 classes) participated in all waves of data collection. Unequal class sample sizes for different conditions resulted from the fact that nine teachers participated with two classes, which had been intentionally assigned to the same condition. The classes in the three intervention conditions did not differ significantly in their class size, teachers' age, teachers' teaching experience, teachers' gender or the relevance of math instruction reported by teachers (all p's $\geq .101$).

Relevance Intervention

From October to November 2012, the intervention was implemented in all classes in the two experimental conditions by five trained female doctoral students. All doctoral students carried out 8-13 interventions in total, roughly equally distributed between the two experimental conditions. The intervention consisted of a 90-minute lesson on the relevance of mathematics which included a psychoeducational presentation for the whole class and relevance-inducing tasks for individual students.

The psychoeducational presentation had two main components. First, research results on the importance of effort and self-concept for math achievement were presented and students were told about frame of reference effects that can occur within the classroom. This part aimed at inoculating students against potential negative effects of highlighting the importance of a subject which might be anxiety-inducing if students judge their own achievement in this subject as low (cf., Durik, Shechter, Noh, Rozek, & Harackiewicz, 2014). Second, to prepare students for their individual tasks, they were provided with various examples on the utility of mathematics for future education, career opportunities, and leisure time activities including female- and male-typed careers. This presentation was identical for both intervention conditions.

After this presentation, students worked on relevance-inducing tasks which differed between the two conditions. In the quotations condition, students were asked to read a total of six interview quotations of young adults describing situations in which mathematics was useful to them and to evaluate these quotations based on their personal relevance. In the text condition, students were asked to make a list of arguments for the personal relevance of mathematics to their current and future lives and to write an essay explaining these arguments. Thus, in both conditions, the students had to apply the relevance of mathematics to their lives, whereas the two conditions differed in the specific structure of the task and the extent to which arguments had to be self-generated.

Additionally, each intervention group received two reinforcements that were embedded into a homework diary, which was filled out by all classes for four weeks after the intervention. The first reinforcement was filled out one week after the intervention; students were asked to reproduce what they remembered from their individual tasks. The second reinforcement was filled out two weeks after the intervention and differed by condition. In the quotations condition, students were given the link to a webpage on the value of mathematics (www.dukannstmathe.de), where they should search for reasons why mathematics could be useful for them and report the most convincing one. In the text condition, students were asked to think of a person they knew for whom mathematics was useful and to report why mathematics was useful to this person. Those reinforcements resembled the individual tasks assigned to the students within the intervention lesson: Students in the quotations condition had to evaluate given arguments and students in the text condition had to generate arguments themselves. Classes in the waiting control condition also filled out homework diaries, but these did not include any intervention reinforcements.

Students in the waiting control condition received the intervention that was shown to be more successful after the last wave of data collection.

Measures

Value beliefs. We assessed value beliefs in the domain of mathematics with a German instrument (Gaspard et al., 2014) that was developed to capture the multidimensionality of value beliefs as described in the expectancy-value model by Eccles et al. (1983). In addition to the four value components, subscales describing multiple facets of attainment value, utility value, and cost can be differentiated. Support for the separability of these subfacets as well as a second-order model was found in a previous study (Gaspard et al., 2014). Intrinsic value was assessed by four items. Attainment value was assessed by ten items covering the facets importance of achievement as well as personal importance. Utility value was assessed by twelve items focusing on the utility for different life domains within a short-term (school, daily life, social life) as well as a long-term perspective (job, future life in general). Cost was assessed by eleven items that covered the facets opportunity cost, effort required, and emotional cost. All items were measured with a four-point Likert scale ranging from completely disagree to completely agree. Sample items and reliabilities for all measurement points are reported in Table 1. Correlations between scales are reported in Table 2 for value components and in Table 3 for subfacets.

Confirmatory factor analyses were conducted with Mplus (Muthén & Muthén, 1998-2012) using the robust maximum likelihood estimator and the design-based correction of standard errors and model-fit statistics to account for nonnormality of the indicator variables and the nested data structure. These analyses supported the differentiation of the value facets with a good fit of an eleven-factor model at all three measurement points (T1: $\chi2$ = 2098.03, df = 574, CFI = .957, TLI = .950, RMSEA = .038; T2: $\chi2$ = 1818.56, df = 574, CFI = .964, TLI = .958, RMSEA = .035; T3: $\chi2$ = 1504.24, df = 574, CFI = .970, TLI = .966, RMSEA = .031). As a prerequisite of comparing value beliefs across time and groups, preliminary analyses were conducted to assess measurement invariance for this full-factor model across the three measurement waves as well as across the three conditions (see Supplement for fit indices). These analyses supported strict measurement invariance across time as well as across groups with changes in fit indices for more restrictive models meeting recommended cutoff criteria (see Chen, 2007; Cheung & Rensvold, 2002). Second-order models with the value components intrinsic value, attainment value, utility value, and cost as higher-order factors also resulted in an acceptable fit at all three measurement points (T1: $\chi2$ = 2789.13, *df* = 614, CFI = .939, TLI = .933, RMSEA = .044 for T1; T2: $\chi2$ = 2645.43, *df* = 614, CFI = .942, TLI = .937, RMSEA = .043; T3: $\chi2$ = 2120.66, *df* = 614, CFI = .952, TLI = .948, RMSEA = .038), thereby supporting the aggregation of value facets to value components.

Covariates. Background information on students was assessed before the intervention. Teachers provided students' math grades at the end of eighth grade, student gender as well as the test scores from a state-wide standardized, curriculum-based math achievement test that was conducted at the beginning of ninth grade. Students completed a test assessing their nonverbal cognitive abilities, namely the Figure Analogies subscale (α=.79) from the Cognitive Abilities Test 4 – 12 + R (Heller & Perleth, 2000).

Table 1

Sample Items, Reliabilities and Intraclass Correlation Coefficients for Value Components and Facet Scales at All Measurement Waves

Variable	Sample item	Items	α_{T1}	α_{T2}	α_{T3}	ICC_{T1}	ICC_{T2}	ICC_{T3}
Intrinsic value	Math is fun to me.	4	0.94	0.93	0.92	0.07	0.08	0.08
Attainment value		10	0.91	0.92	0.92	0.05	0.06	0.07
Importance of achievement	Good grades in math are very important to me.	4	0.88	0.89	0.90	0.04	0.05	0.07
Personal importance	Math is very important to me personally.	6	0.85	0.86	0.86	0.06	0.06	0.07
Utility value		12	0.84	0.86	0.87	0.06	0.09	0.08
General utility for future life	I will often need math in my life.	2	0.79	0.82	0.81	0.05	0.07	0.06
Utility for school	Being good at math pays off, because it is simply needed at school.	2	0.51	0.64	0.64	0.03	0.07	0.06
Utility for job	Learning math is worthwhile, because it improves my job and career chances.	2	0.68	0.76	0.77	0.04	0.07	0.07
Utility for daily life	Understanding math has many benefits in my daily life.	3	0.84	0.85	0.86	0.06	0.06	0.06
Social utility	I can impress others with intimate knowledge in math.	3	0.75	0.80	0.82	0.05	0.04	0.07
Cost		11	0.93	0.94	0.94	0.04	0.05	0.06
Effort required	Doing math is exhausting to me.	3	0.83	0.86	0.88	0.04	0.05	0.07
Emotional cost	Doing math makes me really nervous.	4	0.90	0.91	0.92	0.04	0.05	0.06
Opportunity cost	I have to give up a lot to do well in math.	4	0.87	0.88	0.89	0.02	0.04	0.04

Note. *ICC= Intraclass Correlation Coefficient. Sample items are translated from the original version of the survey, which was given in German. The complete set of items can be found in Gaspard at el. (2014).*

Table 2

Intercorrelations among Value Component Scales Across All Measurement Waves

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Intrinsic value T1	-											
(2)	Intrinsic value T2	0.79	-										
(3)	Intrinsic value T3	0.74	0.79	-									
(4)	Attainment value T1	0.64	0.54	0.49	-								
(5)	Attainment value T2	0.52	0.60	0.52	0.76	-							
(6)	Attainment value T3	0.47	0.50	0.59	0.68	0.76	-						
(7)	Utility value T1	0.54	0.46	0.42	0.68	0.56	0.50	-					
(8)	Utility value T2	0.41	0.51	0.42	0.54	0.67	0.54	0.69	-				
(9)	Utility value T3	0.38	0.42	0.51	0.46	0.53	0.66	0.62	0.69	-			
(10)	Cost T1	-0.68	-0.56	-0.55	-0.41	-0.34	-0.33	-0.31	-0.24	-0.24	-		
(11)	Cost T2	-0.61	-0.61	-0.58	-0.38	-0.39	-0.36	-0.30	-0.26	-0.27	0.78	-	
(12)	Cost T3	-0.56	-0.54	-0.59	-0.35	-0.34	-0.36	-0.27	-0.24	-0.24	0.73	0.82	-

Note. All correlations are significant at $p < .001$.

Table 3

Intercorrelations among Value Facet Scales at T1

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Intrinsic value	-										
(2)	Importance of achievement	0.46^{***}	-									
(3)	Personal importance	0.68^{***}	0.74^{***}	-								
(4)	Utility for school	0.45^{***}	0.41^{***}	0.57^{***}	-							
(5)	Utility for daily life	0.27^{***}	0.46^{***}	0.47^{***}	0.34^{***}	-						
(6)	Social utility	0.32^{***}	0.43^{***}	0.49^{***}	0.55^{***}	0.39^{***}	-					
(7)	Utility for job	0.44^{***}	0.38^{***}	0.52^{***}	0.70^{***}	0.31^{***}	0.39^{***}	-				
(8)	General utility for future life	0.38^{***}	0.34^{***}	0.38^{***}	0.27^{***}	0.20^{***}	0.21^{***}	0.30^{***}	-			
(9)	Effort required	-0.39^{***}	-0.13^{***}	-0.23^{***}	-0.18^{***}	-0.06^{**}	-0.09^{**}	-0.12^{***}	-0.02	-		
(10)	Emotional cost	-0.65^{***}	-0.24^{***}	-0.42^{***}	-0.28^{***}	-0.09^{***}	-0.17^{***}	-0.23^{***}	-0.18^{***}	0.55^{***}	-	
(11)	Opportunity cost	-0.68^{***}	-0.31^{***}	-0.51^{***}	-0.36^{***}	-0.22^{***}	-0.25^{***}	-0.29^{***}	-0.18^{***}	0.55^{***}	0.79^{***}	-

Note. N = 1809. Correlations pattern at T2 and T3 are comparable.
*** p < 0.001; ** p < 01.

Statistical Analyses

Multilevel regression analyses. Given the multilevel structure of the data, we conducted two-level regression analyses[1] with Mplus (Version 7; Muthén & Muthén, 1998-2012) to examine the effects of the interventions on students' value beliefs. Multilevel regression analyses provide corrected estimates of the standard errors of regression coefficients that take the nesting of students in classrooms into account (Raudenbush & Bryk, 2002). Multilevel regression analyses were carried out separately for all value components and facets at posttest and follow-up, respectively. To estimate the effects of the intervention more precisely (Raudenbush, 1997), all models included the respective value indicator at the pretest as a covariate at the student level as well as at the class level. The effects at both levels were freely estimated to account for contextual effects (Korendijk, Hox, Moerbeek, & Maas, 2011; Marsh et al., 2009). The pretest indicator at the student level was group-mean centered (Enders & Tofighi, 2007), and manifest aggregation was used for the class level predictor (Marsh et al., 2009). To assess the main effects of the intervention, we regressed value beliefs at the posttest/follow-up on two class-level dummy variables that indicated the intervention conditions (quotations, text) with the control condition as a reference group. To assess if intervention effects varied depending on gender, we specified additional two-level regression models with non-randomly varying slopes of student gender (Raudenbush & Bryk, 2002) and included two cross-level interaction effects (Quotations × Gender, Text × Gender). Significant interactions were probed assessing intervention effects separately for males and females. To facilitate the interpretation of the results, all continuous variables were standardized before running the analyses. Thereby, the regression coefficients of the dummy variables indicating the effects of the intervention conditions compared to the control condition can directly be interpreted as effect sizes (for effect sizes in multilevel models, see Marsh et al., 2009; Tymms, 2004).

Missing data. Due to the absence of students at single measurement waves and non-response to single items, missing data ranged from 6 to 13 % for the relevant variables. All analyses were conducted using full information maximum likelihood estimation implemented in Mplus (Graham, 2009). All analyses used the total sample

[1] As there was no significant variance between schools (0.3 - 2.8 %) for any of the outcome variables, the school level was neglected in the analyses.

(N=1916) if not stated otherwise. To make the assumption of missing-at-random more plausible, a nonverbal cognitive ability score, gender, previous math grade and achievement data for math at Time 1 were used as auxiliary variables by including correlations between these variables and the predictor variables as well as the residuals of the dependent variables at both levels (see Collins, Schafer, & Kam, 2001; Enders, 2010).

Results

After testing if the randomization was successful in establishing comparable groups, we report our findings regarding our three main research questions: effects of the two intervention conditions on value beliefs in terms of the four value components (research question 1), intervention effects on value beliefs depending on the facet under consideration (research question 2), and intervention effects depending on students' gender (research question 3).

Descriptive Statistics and Randomization Check

Descriptive statistics for all value scales are reported in Table 4. To test if the randomization of classes to conditions was successful, we conducted multilevel multi-group models (with each experimental condition as a group) for value beliefs and a standardized achievement test at pretest. Differences regarding means were tested for statistical significance by Wald-χ^2-tests (using the "Model Test" command in Mplus), which is asymptotically equivalent to the likelihood ratio test (cf. Bollen, 1989). We found no significant differences between the groups prior to the intervention, neither for any of the four value components (utility value: χ^2 (2) = 0.79, p = .675; attainment value: χ^2 (2) = 3.34, p = .188; intrinsic value: χ^2 (2) = 2.52, p = .284; cost: χ^2 (2) = 1.38, p = .501), nor for achievement (χ^2 (2) = 2.42, p = .298).

Table 4

Descriptive Statistics for Value Components and Facet Scales in the Three Conditions at All Measurement Waves

		Quotations condition (*N*=561)		Text condition (*N*=720)		Control condition (*N*=635)	
		M	SD	M	SD	M	SD
Intrinsic value	T1	2.31	0.84	2.29	0.86	2.18	0.84
	T2	2.26	0.83	2.17	0.82	2.10	0.82
	T3	2.34	0.80	2.25	0.82	2.14	0.79
Attainment value	T1	2.83	0.61	2.78	0.59	2.74	0.57
	T2	2.91	0.62	2.78	0.62	2.75	0.62
	T3	2.91	0.61	2.83	0.60	2.76	0.61
Importance of achievement	T1	3.00	0.65	2.94	0.65	2.92	0.62
	T2	3.06	0.66	2.94	0.67	2.91	0.68
	T3	3.05	0.65	2.95	0.68	2.91	0.66
Personal importance	T1	2.72	0.64	2.67	0.61	2.62	0.62
	T2	2.81	0.65	2.68	0.65	2.65	0.65
	T3	2.82	0.64	2.75	0.61	2.67	0.63
Utility value	T1	2.56	0.49	2.52	0.47	2.52	0.49
	T2	2.64	0.50	2.53	0.51	2.45	0.50
	T3	2.60	0.49	2.53	0.49	2.44	0.51
General utility for future life	T1	2.74	0.74	2.69	0.72	2.70	0.76
	T2	2.85	0.74	2.69	0.76	2.53	0.75
	T3	2.74	0.73	2.68	0.70	2.57	0.77
Utility for school	T1	3.13	0.58	3.10	0.60	3.13	0.58
	T2	3.20	0.60	3.08	0.63	3.07	0.62
	T3	3.15	0.59	3.10	0.62	3.05	0.61

(continued)

		Quotations condition (*N*=561)		Text condition (*N*=720)		Control condition (*N*=635)	
		M	SD	M	SD	M	SD
Utility for job	T1	3.10	0.71	3.08	0.69	3.10	0.71
	T2	3.25	0.66	3.13	0.74	2.98	0.72
	T3	3.17	0.67	3.09	0.71	2.97	0.74
Utility for daily life	T1	2.47	0.74	2.37	0.70	2.40	0.74
	T2	2.48	0.76	2.35	0.75	2.23	0.72
	T3	2.41	0.73	2.31	0.72	2.22	0.72
Social utility	T1	1.77	0.64	1.77	0.65	1.69	0.61
	T2	1.87	0.71	1.83	0.70	1.80	0.67
	T3	1.92	0.66	1.88	0.73	1.82	0.65
Cost	T1	2.08	0.69	2.08	0.71	2.14	0.68
	T2	2.08	0.76	2.12	0.73	2.18	0.71
	T3	2.04	0.71	2.07	0.76	2.15	0.71
Effort required	T1	1.69	0.71	1.70	0.75	1.76	0.73
	T2	1.84	0.82	1.84	0.80	1.90	0.78
	T3	1.79	0.78	1.87	0.84	1.91	0.77
Emotional cost	T1	2.46	0.84	2.43	0.84	2.52	0.82
	T2	2.33	0.86	2.38	0.86	2.45	0.83
	T3	2.29	0.83	2.30	0.85	2.39	0.84
Opportunity cost	T1	1.98	0.77	2.02	0.80	2.05	0.77
	T2	2.00	0.80	2.06	0.80	2.11	0.78
	T3	1.96	0.76	2.00	0.79	2.08	0.77

Note. Sample size varied for individual scales (quotations condition: *N* = 497 - 530; text condition: *N* = 606 - 680; control condition: *N* = 546 – 607).

Intervention Effects on Value Components

All results for two-level regression models assessing intervention effects on value components at posttest and follow-up are reported in Table 5. For utility value, we found positive effects—as compared to the waiting control condition—for both intervention conditions at the posttest with the quotations condition having stronger effects than the text condition. An additional Wald-χ^2-test comparing the two parameters indicated that the effects of the two intervention conditions differed significantly, χ^2 (1) = 9.10, p = .003. At the follow-up, we still found effects of both conditions on utility value, and again, the effect tended to be larger for the quotations condition, χ^2 (1) = 2.77, p = .096. For attainment value, we found positive effects of the quotations condition at the posttest as well as at the follow-up. The text condition did not show a statistically significant effect on attainment value. For intrinsic value, no significant intervention effects were found at the posttest, whereas at the follow-up, students in classes in the quotations condition reported higher intrinsic value compared to students in classes in the control condition. No effect of the text condition on intrinsic value was found. For cost, no effects of the intervention conditions were found.

Intervention Effects on Specific Value Facets

Additional analyses assessed intervention effects depending on the value facets. Utility value was assessed in terms of different life domains, and as the intervention focused on the utility of mathematics for some of these life domains such as future career and job opportunities and daily life, we wanted to assess if intervention effects on subfacets were in line with the focus of the intervention. The results of the intervention effects on the five subfacets of utility at the posttest and the follow-up are displayed in Table 6. At the posttest, both interventions showed positive effects on utility for daily life, utility for job, and general utility for future life. The quotations condition also had positive effects on utility for school. No effects on social utility were found. At the follow-up, the effects were somewhat smaller, but we still found significant effects of both intervention conditions on utility for daily life, utility for job, and general utility for future life and a significant effect of the quotations condition on utility for school. Altogether, the intervention effects of both conditions and at both time points were, thus, stronger for some utility facets than for others with stronger effects for those facets directly targeted in the intervention, namely utility for job and utility for daily life.

Table 5

Intervention Effects on Value Components at Posttest and Follow-up

Variable	Utility value Est.		Utility value (SE)	Attainment value Est.		Attainment value (SE)	Intrinsic value Est.		Intrinsic value (SE)	Cost Est.		Cost (SE)
Posttest												
Student level												
Value T1	0.67	***	(0.02)	0.76	***	(0.02)	0.78	***	(0.02)	0.77	***	(0.02)
Class level												
Value T1	0.79	***	(0.06)	0.76	***	(0.07)	0.82	***	(0.06)	0.85	***	(0.07)
Quotations	0.30	***	(0.06)	0.12	*	(0.05)	0.08		(0.06)	-0.08		(0.06)
Text	0.14	*	(0.06)	-0.01		(0.05)	-0.02		(0.05)	-0.01		(0.05)
Residual variance												
Student level	0.49			0.40			0.35			0.38		
Class level	0.02			0.02			0.02			0.01		
Follow-Up												
Student level												
Value T1	0.60	***	(0.02)	0.67	***	(0.02)	0.73	***	(0.02)	0.72	***	(0.02)
Class level												
Value T1	0.78	***	(0.07)	0.77	***	(0.08)	0.77	***	(0.07)	0.88	***	(0.08)
Quotations	0.26	***	(0.06)	0.15	**	(0.05)	0.14	*	(0.06)	-0.06		(0.06)
Text	0.16	**	(0.06)	0.06		(0.06)	0.04		(0.05)	0.00		(0.06)
Residual variance												
Student level	0.59			0.51			0.43			0.44		
Class level	0.02			0.02			0.02			0.02		

Note. Est. = Estimated parameters. Students' gender, pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 6

Intervention Effects on Utility Facets at Posttest and Follow-up

Variable	Utility for school		Utility for daily life		Social utility		Utility for job		General utility	
	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)
Posttest										
Student level										
Value T1	0.44 ***	(0.02)	0.64 ***	(0.02)	0.56 ***	(0.02)	0.54 ***	(0.02)	0.58 ***	(0.02)
Class level										
Value T1	0.79 ***	(0.11)	0.63 ***	(0.07)	0.71 ***	(0.08)	0.66 ***	(0.08)	0.65 ***	(0.08)
Quotations	0.20 **	(0.07)	0.27 ***	(0.05)	0.00	(0.07)	0.38 ***	(0.06)	0.38 ***	(0.05)
Text	0.05	(0.07)	0.19 ***	(0.05)	-0.07	(0.06)	0.22 ***	(0.06)	0.19 **	(0.06)
Residual variance										
Student level	0.74		0.56		0.66		0.66		0.62	
Class level	0.02		0.01		0.02		0.02		0.01	
Follow-Up										
Student level										
Value T1	0.39 ***	(0.03)	0.58 ***	(0.02)	0.43 ***	(0.03)	0.49 ***	(0.03)	0.52 ***	(0.03)
Class level										
Value T1	0.72 ***	(0.10)	0.70 ***	(0.06)	0.63 ***	(0.09)	0.64 ***	(0.10)	0.70 ***	(0.09)
Quotations	0.17 **	(0.06)	0.21 ***	(0.06)	0.10	(0.06)	0.30 ***	(0.07)	0.21 ***	(0.06)
Text	0.12 †	(0.07)	0.15 *	(0.06)	0.00	(0.07)	0.17 *	(0.07)	0.18 **	(0.06)
Residual variance										
Student level	0.80		0.63		0.76		0.71		0.69	
Class level	0.03		0.01		0.03		0.03		0.01	

Note. Est. = Estimated parameters. Students' gender, pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

We also investigated intervention effects on different facets of attainment value and cost to see if we could find a differentiated pattern of intervention effects for these value components. For attainment value, similar to the results of the total scale, we found positive effects of the quotations condition on importance of achievement and personal importance at both time points. No effects of the text condition on facets of attainment value were found. For cost, classes in the quotations condition perceived less effort required and tended to perceive less emotional cost at the posttest compared to classes in the waiting control condition. Whereas we found no effects on the total cost scale, there was, thus, some support for effects on subscales. At the follow-up, we did not find these effects any more. No effects on opportunity cost were found. The text condition did not show any significant effect on perceived cost. All results for the intervention effects on attainment and cost facets can be found in the supplement.

Differential Intervention Effects Depending on Gender

Female students in this sample reported lower value beliefs before the intervention (Gaspard et al., 2014). As we wanted to see if these gender differences could be reduced by relevance interventions, we tested whether the intervention effects differed between female and male students. We, therefore, added gender and two cross-level interactions (Quotation × Gender and Text × Gender) into our models. Again, we first present the results for the four major value components, before reporting the results for the value facets. Results for two-level regressions on value components depending on gender at the posttest and the follow-up are reported in Table 7.

For utility value, the effects of the text condition at the posttest tended to differ between males and females, but the interaction term missed significance (p = .051). Whereas the text condition had positive effects on utility value for females (β = .23, p = .001), it did not show a significant effect for males (β = .04, p = .618). The intervention effects on utility value for females and males at the posttest are displayed in Figure 1. At the follow-up, no significant interaction between the intervention conditions and student gender was found for utility value.

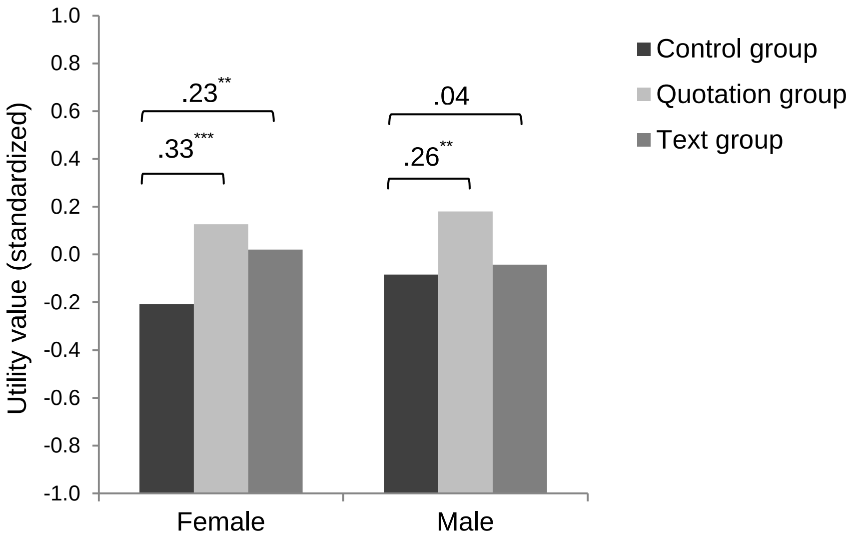

Figure 1. Adjusted means for utility value at posttest by gender and intervention group. Effect sizes for the quotations condition and the text condition as compared to the control condition are displayed separately for females and males.
*** $p < 0.001$; ** $p < 0.01$.

For intrinsic value, we found a significant interaction with gender for the quotations condition at the posttest. Whereas there was a positive effect of the quotations condition on girls' intrinsic value (β = .17, *p* = .009), there was no significant effect for boys (β = -.03, *p* = .644). Effects of the text condition also depended on gender: There was no significant effect on intrinsic value for females (β = .08, *p* = .180), but a marginally significant negative effect for males (β = -.13, *p* = .085). These differential intervention effects on intrinsic value for females and males are displayed in Figure 2. At the follow-up, these interactions were no longer found to be significant. No differential effects for boys and girls were found for attainment value and cost.

Interactions of the intervention effects with gender were also tested for all facets (see supplement). For facets of utility value, we found an interaction between the quotations condition and gender at the follow-up for utility for daily life (β = -.19, *p* = .025), indicating that intervention effects were limited to females (β = .29, *p* < .001 compared to β = .11, *p* < .157 for males). No differential effects were found for the other facets of utility value, attainment value, or cost. Altogether, all significant interactions pointed to both intervention conditions having stronger positive effects on value beliefs for females than for males.

Table 7

Intervention Effects Depending on Gender on Value Components at Posttest and Follow-up

Variable	Utility value Est.	Utility value (SE)	Attainment value Est.	Attainment value (SE)	Intrinsic value Est.	Intrinsic value (SE)	Cost Est.	Cost (SE)
Posttest								
Student level								
Intercept gender	0.12 †	(0.07)	-0.02	(0.07)	0.25 ***	(0.06)	-0.03	(0.06)
Value T1	0.67 ***	(0.02)	0.76 ***	(0.02)	0.78 ***	(0.02)	0.77 ***	(0.02)
Class level								
Value T1	0.78 ***	(0.06)	0.76 ***	(0.07)	0.79 ***	(0.06)	0.84 ***	(0.07)
Quotations	0.33 ***	(0.07)	0.11	(0.07)	0.17 *	(0.07)	-0.07	(0.07)
Text	0.23 ***	(0.07)	0.02	(0.06)	0.08	(0.06)	0.01	(0.06)
Quotations × Gender	-0.07	(0.08)	0.03	(0.09)	-0.21 **	(0.08)	-0.01	(0.08)
Text × Gender	-0.19 †	(0.10)	-0.07	(0.09)	-0.20 **	(0.08)	-0.03	(0.09)
Residual variance								
Student level	0.49		0.40		0.34		0.38	
Class level	0.02		0.02		0.02		0.01	
Follow-Up								
Student level								
Intercept gender	0.14 *	(0.06)	-0.04	(0.06)	0.21 ***	(0.06)	-0.10	(0.06)
Value T1	0.60 ***	(0.02)	0.68 ***	(0.02)	0.72 ***	(0.02)	0.72 ***	(0.02)
Class level								
Value T1	0.77 ***	(0.07)	0.78 ***	(0.08)	0.75 ***	(0.07)	0.86 ***	(0.08)
Quotations	0.32 ***	(0.07)	0.12	(0.08)	0.13 †	(0.07)	-0.03	(0.08)
Text	0.19 **	(0.07)	0.11	(0.07)	0.10	(0.06)	-0.02	(0.07)
Quotations × Gender	-0.14 †	(0.08)	0.05	(0.09)	0.01	(0.08)	-0.05	(0.10)
Text × Gender	-0.09	(0.09)	-0.11	(0.08)	-0.13 †	(0.08)	0.05	(0.08)
Residual variance								
Student level	0.59		0.51		0.42		0.44	
Class level	0.02		0.02		0.02		0.02	

Note. Est. = Estimated parameters. Gender was coded 0=female, 1=male. Pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$

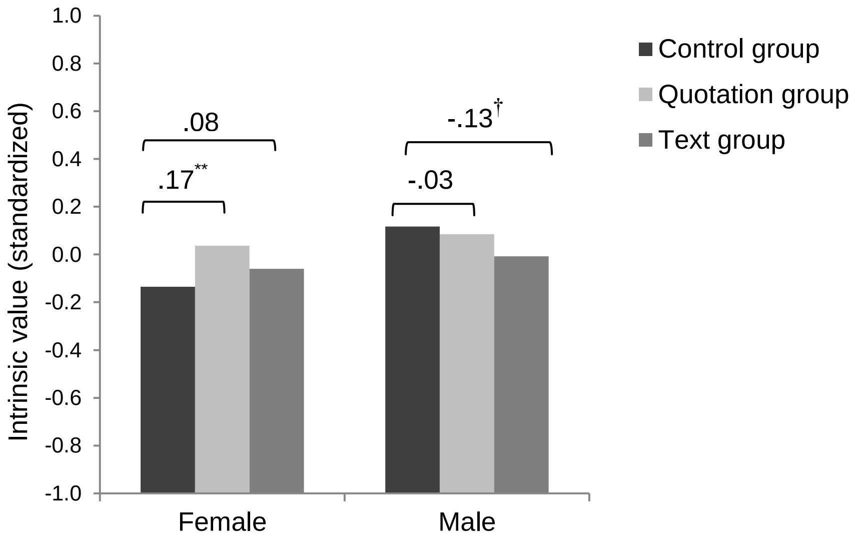

Figure 2. Adjusted means for intrinsic value at posttest by gender and intervention group. Effect sizes for the quotations condition and the text condition as compared to the control condition are displayed separately for females and males.
** $p < 0.01$; † $p < 0.10$

Discussion

This study aimed at testing whether adolescents' value beliefs for mathematics could be promoted by a relevance intervention in the classroom. We conducted a cluster randomized controlled study with 82 ninth-grade classes comparing two different relevance interventions with a non-treated control group. Our findings show that a 90-minute intervention in the classroom and two short reinforcements had long-term effects on students' value beliefs for mathematics. Reflecting on quotations about the usefulness of mathematics was shown to be more beneficial than writing texts about the personal relevance of mathematics. Whereas the quotations condition had stronger effects on utility value and also affected attainment and intrinsic value, the text condition only had significant main effects on utility value. Regarding students' beliefs about the usefulness of mathematics, we found stronger effects for those life domains that were targeted by the intervention than for other life domains. There was some evidence that both intervention conditions were more effective for girls, who are one target group of motivational interventions within mathematics.

Intervening on the Development of Students' Value Beliefs

Our study could show that it is possible to affect adolescents' value beliefs longitudinally with the help of relevance interventions in the classroom. We compared two different tasks to induce perceived relevance. Whereas one of these tasks (i.e., self-generating arguments for the utility of the subject material) has already been applied successfully in previous studies (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009), the other task (i.e., reflecting on given arguments) is a rather new approach in utility value interventions with students (see also Harackiewicz et al., 2012). Evaluating quotations, as implemented in our intervention, is a combination of providing information (Durik & Harackiewicz, 2007; Shechter et al., 2011) with more active elements of elaboration as students were guided to create links to their own lives. Why did this relevance-inducing task have more beneficial effects than the text condition? One possible explanation is that structured reflection on the personal relevance of mathematics is potentially easier and more enjoyable for adolescents compared to writing an essay which is a typical task done at school that might even cause aversive reactions. Students potentially would not have been able to produce the breadth of arguments presented in the quotations from six different individuals. Also, the people that were interviewed for these quotations mostly were young adults (for example college students) that could have served as possible role models for students in our study (Markus & Nurius, 1986). Quotations from these interviews seem to be more authentic and persuasive than just providing this information without giving any specific source. This task might, thus, have fitted better to the developmental needs and preferences of the ninth-grade students participating in our study.

There were some differences between the relevance-inducing tasks in our study and those used before. In previous studies (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009), essays were written at home, which can lead to lower completion rates, whereas in our study all students worked on their tasks in school. However, compared to graded assignments as used before, the level of engagement when writing these essays might still have been lower in our study. It is, thus, possible, that students simply were more engaged in the quotations task and, in consequence, put more effort into it. Another difference was that students were asked to generate arguments for the usefulness of mathematics as a general domain, whereas in studies by Hulleman and colleagues, students were asked to write essays about the topic currently covered in class. This less topic-specific intervention was used to ensure comparability across the classrooms participating in our study and to facilitate long-lasting effects of a short intervention. However, coming up with arguments for the usefulness of a subject might have been harder for the students under these conditions.

When interpreting the results of our study, it is important to keep in mind that the intervention implemented in the classroom combined different elements, all drawing on EVT: Not only did students work on individual relevance-inducing tasks in class, but they were also prepared for these tasks by a psychoeducational presentation that provided them with some information on the potential relevance of mathematics for their future lives as well as with research results on the importance of attitudes for future performance, and got two additional reinforcement tasks. These different elements were combined to create maximum impact, while buffering potential detrimental effects for subgroups of students. When implementing interventions in the classroom at large scale, ethical reasons call for the consideration of potential negative effects. These might occur if the importance of achievement within a subject is highlighted, but students believe that they cannot improve their achievement. We, therefore, applied the aforementioned buffering strategy in our intervention. Whereas our relatively short intervention was effective in fostering students' value beliefs until five months after the beginning of the intervention, it is not possible to tease apart effects of the individual components of our intervention. However, as our two intervention conditions had different effects on students' value beliefs, the relevance-inducing tasks that students worked on seem to be one crucial part of our intervention. To gain a better understanding on how utility value interventions work, laboratory studies that test combinations of different intervention elements (with and without confidence boost, see Durik et al., 2015) and forms (communication vs. self-generation of utility value) are promising means, even if generalizability remains an issue in laboratory experiments. Hence, it is important to also test interventions based on motivational theories in the field, as researchers can encounter new challenges compared to a controlled laboratory setting (Hulleman & Cordray, 2009).

The intervention effects in our study were all rather small applying conventional standards (e.g., Cohen, 1988). However, when evaluating the size of these effects, several aspects need to be considered. First, the intervention consisted only of a 90-minute session in the classroom and two short reinforcement tasks and can therefore be seen as a minimal intervention. Second, interventions in the field often show smaller effects than interventions in the laboratory due to variations in the implementation and the context (Hulleman & Cordray, 2009). Third, effect sizes in empirical studies are only estimates of true effects (see Gelman & Carlin, 2014). The reliability of these effect estimates depends on the sample size. Given that the sample size was carefully chosen to achieve an acceptable power, the estimated effects will on average be close to the true effects of the intervention. In comparison, effect estimates from studies with smaller sample sizes are more variable and thus often overestimate the true effect size when statistically significant (Gelman & Carlin, 2014).

As the intervention effects were stable at the follow-up measurement, the intervention actually seemed to affect the development of students' value beliefs. In terms of the trajectory over time, we observed a decline in value beliefs in the control condition as has been consistently reported in the literature (e.g., Watt, 2004). The intervention, thus, buffered against this negative development. There was evidence for an additional increase in value beliefs in the quotations condition—at least for utility value. The follow-up measurement took place approximately five months after the initial intervention. However, students in the intervention conditions additionally received two reinforcements that were embedded into a homework diary. These might have been important for sustaining intervention effects.

From a developmental perspective, an important question is at what age such interventions can be applied successfully. As declines in student motivation have been observed from elementary school on (Fredricks & Eccles, 2002; Jacobs et al., 2002), one might call for earlier interventions. However, interventions should be applied within the "motivational zone of proximal development" to effectively foster student motivation (Brophy, 1999) and younger students may have difficulties to reflect on the relevance of engagement in a subject to their future careers (cf. Gottfredson, 1981; Wigfield, 1994). To date, relevance interventions have been successfully applied from ninth grade on (Hulleman & Harackiewicz, 2009). This might be a critical developmental period when adolescents start to think about their future in a more elaborate way and, therefore, students' age might be a decisive point for the intervention effects we found.

Gender Differences in Reacting to the Intervention

Females are one target group of motivational interventions within the domain of mathematics as they seem to have lower value beliefs than males and these gender differences in beliefs can also explain differences in choices (e.g., Chow et al., 2012; Nagy et al., 2006; Watt et al., 2012). Female students tended to benefit more from the intervention than male students. Differential effects were found for those value components (i.e., utility value and intrinsic value) where females showed a more negative development in the control group than males. It seems thus that the relevance interventions prevented widening of a gender gap in mathematics motivation rather than reducing such a gap.

There are several possible reasons for why these differential effects for boys and girls occurred. First, the interventions were conducted by female researchers only, and therefore, role model effects may be one factor. Competent same-gender role models have been found to buffer effects of stereotype threat on women within STEM and enhance women's STEM-related self-concept as well as their identification with and aspirations towards STEM (Marx & Roman, 2002; Stout, Dasgupta, Hunsinger, & McManus, 2011; Young, Rudman, Buettner, & McLean, 2013). Second, the intervention examples on the usefulness of mathematics for future career opportunities included both more male- and more female-typed domains. As applying mathematics is more in line with what students typically think about more male-typed domains such as engineering, the information that mathematics is also relevant for more female-typed domains such as psychology might have been new for students. Third, the writing assignments that were implemented within the intervention to induce relevance are a rather prototypical female activity. Girls might have enjoyed those activities more than boys and might even have worked more conscientiously on these assignments (cf., Meece, Glienke, & Burg, 2006). Fourth, girls at that age group might just be more mature compared to boys at the same age (Eisenberg, 2006), and therefore, might have benefited more from an intervention referring to their future. These different reasons are, however, not mutually exclusive and all of them might have contributed to differential intervention effects for boys and girls. Future research is, thus, needed to further explore the processes at play.

Theoretical Implications for EVT

The results of our study have several implications for the future development of EVT. With regards to the conceptualization of subjective task value, they provide strong support for the usefulness of differentiating components and also subfacets of these components. Although thinking of task value as one general factor might be preferable in terms of parsimony, a set of different beliefs are subsumed under the heading of value. These aspects of value appear to be malleable through interventions to different extents. In line with the focus of our intervention, we found stronger effects on some facets than on others. More experimental or longitudinal research is, however, needed to further understand the processes through which these facets are affected. Generally, our study can be understood as testing EVT under real life conditions. The intervention we developed was based on theoretical assumptions as well as empirical results – many of them stemming from correlational research. Intervention research, thus, can be seen as the next step that is useful for practice, but also provides a strong test of the theory applying it in the educational context (Pintrich, 2003). Our results show that such applications are possible and that the elaborate view of motivational factors in EVT actually helps in developing such interventions.

Limitations and Future Research

Although our study could show beneficial effects of a relevance intervention in the classroom, there are several noteworthy limitations to our research. First, whereas we found effects on students' value beliefs until several months after the intervention, in this paper we only used self-report to assess the effects of our intervention. We were interested in how relevance interventions in the classroom affect value beliefs as one important outcome of motivational interventions (Pintrich, 2003; Wigfield et al., 2009), and self-report is an adequate means to assess changes in students' subjective beliefs (Wigfield & Cambria, 2010). Changes in value beliefs should, however, also translate into changes in students' behavior and choices in the long term (Harackiewicz et al., 2012; Wigfield et al., 2009). Future studies should, therefore, follow students' development for a longer time period and also take other outcomes into account. To measure value beliefs in a comprehensive way, we used a newly developed scale including different aspects that were important within the context of our study (Gaspard et al., 2014). However, some of the scales for sub-facets of utility value consisted of only two items and had low reliabilities (especially utility for school), thereby undermining the potential to find substantial intervention effects on these facets.

Second, whereas we used a large sample to thoroughly test effects of our intervention in the classroom, the sample of our study was limited to ninth-grade students within the highest track in Germany. Future research should test if our findings are replicable with younger students and other school types. Implementing this intervention in other samples might, however, require some changes in terms of specific content. As the intervention strategy applied in our study required students to reflect on their future career plans, implementing a similar intervention with younger students might be difficult.

Third, whereas we compared two relevance-inducing tasks and found different effects of these tasks, more research is needed to examine the processes through which relevance interventions work. Qualitative analyses of students' answers to such relevance-inducing tasks might be a way to clarify why some tasks work better than others and also why some students benefit more than others. Moreover, the intervention was only implemented by female researchers, which represents a limitation in terms of interpreting differential intervention effects according to students' gender. Last, whereas our intervention is relatively short and easy to implement in the classroom, it should be tested if teacher-implemented interventions have the same effect as interventions implemented by researchers.

Practical Applications

Interventions fostering students' value beliefs are highly relevant for practice as value beliefs influence students' academic development in terms of effort and persistence in various school subjects as well as academic choices (Durik et al., 2006; Nagengast et al., 2011; Nagy et al., 2006; Simpkins et al., 2006; Trautwein & Lüdtke, 2007). Our study extended previous studies by Hulleman and colleagues (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009) by implementing a relatively short relevance intervention in the classroom at a larger scale. This intervention was designed to meet the practical needs and challenges of classroom-based intervention studies. Given its duration and the use of standardized intervention material, the intervention is cost-efficient and could easily be implemented as part of a regular math curriculum. Our results show positive effects on students' value beliefs that sustained for several months. Our sample size was adequate to find small, yet realistic effects of this intervention. Scaling up this intervention could be realized effectively by training researchers that can be deployed to classrooms. In a next step, it would be important to investigate whether teacher-implemented interventions have the same effect. Extending the findings from previous studies, reflection on arguments from quotations was more beneficial than self-generating arguments. This new strategy is, thus, an effective tool to promote students' value beliefs. Instead of directly providing information on the usefulness of a subject, it draws on the experiences from young adults who can function as role models for students.

References

Acee, T. W. & Weinstein, C. E. (2010). Effects of a value-reappraisal intervention on statistics students' motivation and performance. *The Journal of Experimental Education, 78,* 487-512. doi: 10.1080/00220970903352753

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Battle, A. & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior*, *62,* 56-75. doi: 10.1016/s0001-8791(02)00037-4

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task-value, and achievement goals. *Journal of Educational Psychology*, *93*, 23-34. doi: 10.I037//0022-0663.93.1.23

Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist, 34*, 75-85. doi: 10.1207/s15326985ep3402_1

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504. doi: 10.1080/10705510701301834

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233-255. doi: 10.1207/S15328007SEM0902_5

Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology, 48*, 1612-1628. doi: 10.1037/a0030194

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330-351. doi: 10.1037/1082-989X.6.4.330

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

Craven, R. G., Marsh, H. W., Debus, R. L., & Jayasinghe, U. (2001). Diffusion effects: control group contamination threats to the validity of teacher-administered interventions. *Journal of Educational Psychology, 93,* 639-645. doi: 10.I037//0022-0663.93.3.639

Durik, A. M. & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology, 99*, 597-610. doi: 10.1037/0022-0663.99.3.597

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,* 104-118. doi: 10.1007/s11031-014-9419-0

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology, 98*, 382-393. doi: 10.1037/0022-0663.98.2.382

Eccles, J. S. (2005). Subjective task values and the Eccles et al. model of achievement related choices. In A. J. Elliott & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York, NY: Guilford.

Eccles, J. S. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development, 35,* 195-201. doi: 10.1177/0165025411398185

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74-146). San Francisco, CA: W. H. Freeman.

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

Eisenberg, N. (Ed.). (2006). *Handbook of child psychology. Volume 3: Social, emotional, and personality development* (6 ed.). Hoboken, NJ: Wiley.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 103-127. doi: 10.1037/a0018053

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

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138. doi: 10.1037/1082-989X.12.2.121

Fredricks, J. A. & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38,* 519-533. doi: 10.1037//0012-1649.38.4.519

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics - A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 12*, 497-514. doi: 10.1007/BF03173468

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2014). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology,* Advance online publication. doi: 10.1037/edu0000003

Gelman, A. & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science, 9*, 641-651. doi: 10.1177/1745691614551642

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Jornal of Counseling Psychology, 28,* 545-579. doi: 10.1037/0022-0167.28.6.545

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science, 23,* 899-906. doi: 10.1177/0956797611435530

Harackiewicz, J. M., Tibbetts, Y., Canning, E., & Hyde, J. S. (2014). Harnessing values to promote motivation in education. In S. A. Karabenick & T. Urdan (Eds.), *Motivational interventions (Advances in motivation and achievement, Volume 18)* (pp. 71-105). Bingley, UK: Emerald Group Publishing Limited.

Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4.-12. Klassen, Revision (KFT 4–12+ R) [Cognitive abilities test for 4th to 12th grades, revision (KFT 4-12+ R)]*. Göttingen, Germany: Hogrefe.

Hidi, S. & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*, 151-179. doi: 10.3102/00346543070002151

Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111-127. doi: 10.1207/s15326985ep4102_4

Hulleman, C. S. & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88-110. doi: 10.1080/19345740802539325

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*, 880-895. doi: 10.1037/a0019506

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

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509-527. doi: 10.1111/1467-8624.00421

Karabenick, S. A. & Urdan, T. (2014). *Motivational interventions (Advances in motivation and achievement, Volume 18)*. Bingley, UK: Emerald Group Publishing Limited.

Korendijk, E. J., Hox, J. J., Moerbeek, M., & Maas, C. J. (2011). Robustness of parameter and standard error estimates against ignoring a contextual effect of a subject-level covariate in cluster-randomized trials. *Behavior Research Methods, 43*, 1003-1013. doi: 10.3758/s13428-011-0094-8

Markus, H. & Nurius, P. (1986). Possible selves. *American Psychologist, 41*, 954-969. doi: 10.1037/0003-066X.41.9.954

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*, 764-802. doi: 10.1080/00273170903333665

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects of causal ordering. *Child Development, 76*, 397-416. doi: 10.1111/j.1467-8624.2005.00853.x

Marx, D. M. & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*, 1183-1193. doi: 10.1177/01461672022812004

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. Journal of School Psychology, 44(5), 351-373. doi: 10.1016/j.jsp.2006.04.004

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*, 60-70. doi: 10.1037/0022-0663.82.1.60

Moerbeek, M. (2006). Power and money in cluster randomized trials: when is it worth measuring a covariate? *Statistics in Medicine, 25*, 2607-2617. doi: 10.1002/sim.2297

Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition.* Los Angeles, CA: Muthén & Muthén.

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*, 1058-1066. doi: 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. doi: 10.1080/00273171.2013.775060

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12,* 323-345. doi: 10.1080/13803610600765687

Nuttin, J. R. & Lens, W. (1985). *Future time perspective and motivation: Theory and research method.* Leuven, Belgium: Leuven University Press.

OECD. (2004). *Education at a Glance 2004: OECD Indicators.* Paris, France: OECD Publishing.

Oyserman, D. & Destin, M. (2010). Identity-based motivation: Implications for intervention. *The Counseling Psychologist, 38,* 1001-1043. doi: 10.1177/0011000010374775

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106*, 315-329. doi: 10.1037/a0034027

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667-686. doi: 10.1037/0022-0663.95.4.667

Plewis, I. & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation, 4,* 13-26. doi: 10.1076/edre.4.1.13.13014

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173-185. doi: 10.1037/1082-989X.2.2.173

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X.-f., Martinez, A., & Bloom, H. (2011). *Optimal design software for multi-level and longitudinal research (Version 3.01)* [Software]. Available from www.wtgrantfoundation.org.

Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. *Journal of Educational Psychology, 107,* 195-206. doi: 10.1037/a0036981

Shechter, O. G., Durik, A. M., Miyamoto, Y., & Harackiewicz, J. M. (2011). The role of utility value in achievement behavior: the importance of culture. *Personality and Social Psychology Bulletin, 37*, 303-317. doi: 10.1177/0146167210396380

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42,* 70-83. doi: 10.1037/0012-1649.42.1.70

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 100,* 255-270. doi: 10.1037/a0021385

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*, 432-444. doi: 10.1037/0022-0663.99.2.432

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,* 763-777. doi: 10.1037/a0027470

Trautwein, U., Nagengast, B., Marsh, H. W., Gaspard, H., Dicke, A.-L., Lüdtke, O., & Jonkmann, K. (2013). Expectancy-value theory revisited. From expectancy-value theory to expectancy-valueS theory? In D. M. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self-processes and human development* (pp. 233-249). Charlotte, NC: Information Age Publishing.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55-66). London, UK: National Foundation for Educational Research.

Wang, M. T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology, 2,* 1643-1657. doi: 10.1037/a0027247

Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade

Australian students. *Child Development, 75*, 1556-1574. doi: 0.1111/j.1467-8624.2004.00757.x

Watt, H. M. (2008). What motivates females and males to pursue sex-stereotyped careers? In H. M. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 87-113). Washington, DC: American Psychological Association.

Watt, H. M. & Eccles, J. S. (Eds.). (2008). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences.* Washington, DC: American Psychological Association.

Watt, H. M., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Developmental Psychology, 48,* 1594-1611. doi: 10.1037/a0027838

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6,* 49-78. doi: 10.1007/BF02209024

Wigfield, A. & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30,* 1-35. doi: 10.1016/j.dr.2009.12.001

Wigfield, A. & Eccles, J. S. (1992). The development of achievement task values. *Developmental Review, 12,* 265-310. doi: 10.1016/0273-2297(92)90011-P

Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. W. Wentzel, A. (Ed.), *Handbook of motivation at school* (pp. 55-75). New York, NY: Routledge.

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly, 37,* 283-292. doi: 10.1177/0361684313482109

Supplemental Material

Table S1

Tests of Measurement Invariance Across Time and Intervention Condition

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Across Time						
M1: Configural invariance	9382.22	5353	.967	.963	.020	.038
M2: Weak measurement invariance	9501.05	5405	.967	.962	.020	.039
M3: Strong measurement invariance	9799.78	5457	.965	.961	.020	.039
M4: Strict measurement invariance	10095.91	5531	.963	.959	.021	.040
Across Intervention Condition						
T1						
M1: Configural invariance	3379.34	1721	.955	.948	.040	.054
M2: Weak measurement invariance	3398.55	1771	.956	.951	.039	.053
M3: Strong measurement invariance	3464.86	1823	.956	.952	.039	.053
M4: Strict measurement invariance	3522.52	1897	.956	.954	.038	.055
T2						
M1: Configural invariance	3173.24	1721	.961	.954	.037	.046
M2: Weak measurement invariance	3198.78	1771	.961	.956	.036	.045
M3: Strong measurement invariance	3256.11	1823	.961	.957	.036	.046
M4: Strict measurement invariance	3301.90	1897	.962	.960	.035	.047
T3						
M1: Configural invariance	2835.46	1721	.967	.962	.034	.043
M2: Weak measurement invariance	2884.11	1771	.967	.963	.033	.043
M3: Strong measurement invariance	2940.74	1823	.967	.964	.033	.043
M4: Strict measurement invariance	2982.10	1897	.968	.966	.032	.044

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Tests across time: *N* = 1912; Tests across intervention condition: T1: *N* = 1810; T2: *N* = 1816; T3: *N*= 1709. All models fit statistics reported are robust fit statistics. Correlated residuals were allowed between identical items for analyses across time and for two negatively worded attainment items for all analyses.

Table S2

Intervention Effects on Attainment and Cost Facets at Posttest and Follow-up

Variable	Imp. of achievement			Personal importance			Effort required			Emotional cost			Opportunity cost		
	Est.		(SE)	Est.		(SE)	Est.		(SE)	Est.		(SE)	Est.		(SE)
Posttest															
Student level															
Value T1	0.70	***	(0.02)	0.72	***	(0.02)	0.73	***	(0.02)	0.71	***	(0.02)	0.64	***	(0.02)
Class level															
Value T1	0.72	***	(0.08)	0.77	***	(0.06)	0.85	***	(0.07)	0.76	***	(0.07)	0.71	***	(0.11)
Quotations	0.12	*	(0.05)	0.12	*	(0.05)	-0.10	*	(0.05)	-0.09	†	(0.06)	0.00		(0.07)
Text	0.02		(0.05)	-0.04		(0.05)	-0.01		(0.04)	-0.03		(0.05)	0.01		(0.06)
Residual variance															
Student level	0.48			0.44			0.44			0.48			0.57		
Class level	0.02			0.01			0.01			0.01			0.02		
Follow-Up															
Student level															
Value T1	0.61	***	(0.02)	0.65	***	(0.02)	0.68	***	(0.02)	0.65	***	(0.02)	0.61	***	(0.02)
Class level															
Value T1	0.83	***	(0.10)	0.73	***	(0.07)	0.88	***	(0.08)	0.82	***	(0.08)	0.56	***	(0.12)
Quotations	0.12	*	(0.05)	0.15	**	(0.05)	-0.04		(0.07)	-0.07		(0.06)	-0.06		(0.07)
Text	0.04		(0.06)	0.07		(0.06)	0.00		(0.06)	-0.05		(0.06)	0.04		(0.07)
Residual variance															
Student level	0.59			0.54			0.50			0.55			0.60		
Class level	0.02			0.02			0.02			0.02			0.03		

Note. Imp. = Importance, Est. = Estimated parameters. Students' gender, pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables.

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

Table S3

Intervention Effects Depending on Gender on Utility Facets at Posttest and Follow-up

Variable	Utility for school		Utility for daily life		Social utility		Utility for job		General utility	
	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)
Posttest										
Student level										
Intercept Gender	-0.20**	(0.08)	0.16*	(0.07)	0.26**	(0.08)	0.04	(0.07)	0.02	(0.07)
Value T1	0.44***	(0.02)	0.64***	(0.02)	0.55***	(0.02)	0.54***	(0.02)	0.58***	(0.02)
Class level										
Value T1	0.77***	(0.11)	0.62***	(0.07)	0.71***	(0.08)	0.66***	(0.08)	0.65***	(0.08)
Quotations	0.15*	(0.08)	0.33***	(0.07)	0.02	(0.07)	0.37***	(0.08)	0.40***	(0.07)
Text	0.05	(0.07)	0.27***	(0.07)	0.00	(0.06)	0.27***	(0.08)	0.23**	(0.07)
Quotations × Gender	0.12	(0.10)	-0.12	(0.09)	-0.05	(0.11)	0.02	(0.10)	-0.05	(0.09)
Text × Gender	-0.01	(0.10)	-0.18†	(0.11)	-0.14	(0.11)	-0.13	(0.11)	-0.09	(0.10)
Residual variance										
Student level	0.74		0.56		0.65		0.66		0.62	
Class level	0.02		0.01		0.01		0.02		0.01	
Follow-Up										
Student level										
Intercept Gender	-0.20**	(0.08)	0.16*	(0.06)	0.29***	(0.07)	0.03	(0.08)	0.18*	(0.08)
Value T1	0.38***	(0.03)	0.58***	(0.02)	0.42***	(0.03)	0.49***	(0.03)	0.52***	(0.03)
Class level										
Value T1	0.71***	(0.10)	0.69***	(0.06)	0.63***	(0.10)	0.65***	(0.10)	0.69***	(0.09)
Quotations	0.12	(0.08)	0.29***	(0.07)	0.12	(0.08)	0.37***	(0.09)	0.30***	(0.07)
Text	0.14†	(0.08)	0.19*	(0.08)	-0.01	(0.08)	0.23**	(0.09)	0.23**	(0.08)
Quotations × Gender	0.11	(0.10)	-0.19*	(0.09)	-0.04	(0.11)	-0.14	(0.12)	-0.19†	(0.10)
Text × Gender	-0.04	(0.09)	-0.10	(0.08)	0.02	(0.11)	-0.13	(0.11)	-0.11	(0.11)
Residual variance										
Student level	0.79		0.62		0.75		0.71		0.69	
Class level	0.02		0.02		0.03		0.03		0.01	

Note. Est. = Estimated parameters. Gender was coded 0=female, 1=male. Pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.10$.

Table S4

Intervention Effects Depending on Gender on Attainment and Cost Facets at Posttest and Follow-up

	Imp. of achievement		Personal imp.		Effort required		Emotional cost		Opportunity cost	
Variable	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)
Posttest										
Student level										
Intercept Gender	-0.02	(0.07)	0.01	(0.07)	-0.09	(0.07)	-0.06	(0.08)	-0.02	(0.07)
Value T1	0.70***	(0.02)	0.72***	(0.02)	0.72***	(0.02)	0.71***	(0.02)	0.65***	(0.02)
Class level										
Value T1	0.71***	(0.08)	0.77***	(0.06)	0.82***	(0.06)	0.75***	(0.07)	0.71***	(0.11)
Quotations	0.11	(0.07)	0.10	(0.07)	-0.11†	(0.06)	-0.09	(0.07)	0.00	(0.09)
Text	0.05	(0.06)	-0.01	(0.07)	0.03	(0.06)	-0.02	(0.07)	0.03	(0.08)
Quotations × Gender	0.03	(0.10)	0.04	(0.09)	0.02	(0.08)	-0.01	(0.10)	-0.01	(0.11)
Text × Gender	-0.07	(0.09)	-0.07	(0.09)	-0.07	(0.09)	-0.01	(0.10)	-0.05	(0.10)
Residual variance										
Student level	0.48		0.44		0.44		0.48		0.56	
Class level	0.02		0.01		0.01		0.01		0.02	
Follow-Up										
Student level										
Intercept Gender	-0.06	(0.07)	-0.01	(0.07)	-0.13*	(0.07)	-0.13**	(0.06)	-0.01	(0.07)
Value T1	0.61***	(0.02)	0.65***	(0.02)	0.67***	(0.02)	0.64***	(0.02)	0.61***	(0.02)
Class level										
Value T1	0.83***	(0.10)	0.74***	(0.07)	0.85***	(0.08)	0.79***	(0.09)	0.56***	(0.12)
Quotations	0.10	(0.07)	0.13	(0.08)	0.01	(0.08)	-0.09	(0.08)	-0.06	(0.10)
Text	0.09	(0.06)	0.11	(0.08)	-0.01	(0.07)	-0.05	(0.08)	0.03	(0.08)
Quotations × Gender	0.05	(0.09)	0.05	(0.09)	-0.11	(0.10)	0.03	(0.09)	0.00	(0.11)
Text × Gender	-0.12	(0.08)	-0.09	(0.09)	0.04	(0.09)	0.02	(0.08)	0.01	(0.10)
Residual variance										
Student level	0.58		0.54		0.49		0.54		0.60	
Class level	0.02		0.02		0.02		0.02		0.03	

Note. Imp. = Importance, Est. = Estimated parameters. Gender was coded 0=female, 1=male. Pretest cognitive ability score, math achievement test scores and previous math grades were included in the models as auxiliary variables. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

4

Adverse or Desired Side Effects of STEM Interventions? Effects of a Motivational Math Intervention on Motivation in Verbal Domains

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Trautwein, U., & Nagengast, B. (revise and resubmit). Adverse or Desired Side Effects of STEM Interventions? Effects of a Motivational Math Intervention on Motivation in Verbal Domains. *AERA Open.*

Abstract

One way to address the leaking pipeline towards science, technology, engineering, and mathematics (STEM) related careers is intervening on students' STEM motivation in school. However, a neglected question in intervention research is how such interventions affect motivation in subjects not targeted by the intervention. This question was addressed using data from a cluster randomized study, in which a value intervention was successfully implemented in 82 ninth-grade math classrooms. Side effects on value, self-concept, and effort in German as students' native language and English as a foreign language were assessed six weeks and five months after the intervention. Negative effects on value for German, but not for English, were found five months after the intervention. The discussion focuses on intraindividual hierarchies in motivation and the question if negative side effects on non-targeted subjects are to be seen as a desired outcome.

Keywords: dimensional comparisons; motivational intervention; self-concept; value

Adverse or Desired Side Effects of STEM Interventions? Effects of a Motivational Math Intervention on Motivation in Verbal Domains

In many Western countries, concerns have been raised about a lack of young people choosing careers in science, technology, engineering, and mathematics (STEM; e.g., National Science Board, 2007). Important precursors of career choices are students' motivational beliefs in high school about their expectancies and values for different subjects (Eccles et al., 1983; for a review, see Wang & Degol, 2013). One possible way to address the leaking pipeline towards STEM careers at an early stage is thus to foster motivation for related subjects such as math in high school. Recently, researchers have developed a number of successful motivational interventions in STEM (for an overview, see Karabenick & Urdan, 2014). Some of these interventions draw on expectancy-value theory (Eccles et al., 1983), aiming at helping students understand the value of STEM courses. Previous studies have shown that value interventions can be effective in promoting motivation and performance in STEM courses as well as STEM course choices (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009).

However, previous intervention studies in STEM neglected potential effects on motivation in non-STEM areas. Students' expectancies and values are highly domain-specific (Bong, 2001). Students tend to see themselves as either mathematically or verbally oriented, irrespective of whether their achievement in these domains differs substantially (Marsh & Hau, 2004). Academic choices, in turn, are influenced by intraindividual hierarchies in motivational beliefs: The probability that a student intends to pursue a career in STEM increases not only with his/her motivation in STEM becoming higher, but also with his/her motivation in other domains becoming lower (Chow, Eccles, & Salmela-Aro, 2012; Eccles, 2009; Parker et al., 2012). What happens to these motivational patterns when motivation in one domain is fostered through interventions?

In this study, we argue that intervention research should not only consider effects on differences between persons, but also address effects on intraindividual differences. This paper uses an intraindividual difference perspective for evaluating effects of motivational interventions in STEM. We propose that the "true" effect of such interventions can be expressed as the effect on STEM motivation minus the effect on non-STEM motivation. This broader conceptualization is clearly relevant for domain-specific motivational interventions, but could be important for other educational interventions as well.

Student Motivation and Dimensional Comparisons

According to expectancy-value-theory (Eccles et al., 1983), academic choices such as choosing a university major are made based on two beliefs: (a) the expectancy that one can succeed in a task and (b) the value that one attaches to a task. Expectancies are closely related to academic self-concepts, referring to students' evaluation of their abilities in a given domain (Bong & Skaalvik, 2003; Eccles & Wigfield, 2002). Task value comprises several components: attainment value or the personal importance to do well, intrinsic value or enjoyment, utility value or the usefulness for personal goals, and cost or the perceived negative aspects of engaging in a task. Previous research has found high correlations between these components and many studies collapsed the positive value aspects into a single scale (Trautwein et al., 2013). An extensive line of research demonstrates that expectancies and values are indeed important predictors for achievement-related behaviors such as effort and for academic choices (for reviews, see Wang & Degol, 2013; Wigfield, Tonks, & Klauda, 2009).

Expectancies and values are developed through experiences with different domains in the school context. These experiences provide students with a set of possible comparisons including other students' achievement, but also comparisons between domains. Dimensional comparison theory (Möller & Marsh, 2013) assumes that individuals compare their ability in one domain with their ability in another domain (e.g., "How good am I in English compared with math?"). In educational psychology, research has mainly investigated dimensional comparisons in the context of self-concept. Using the internal/external frame of reference model (Marsh, 1986), path-analytic studies have found that achievement in one domain (e.g., math) can have negative effects on self-concept in another domain (e.g., English; Marsh & Hau, 2004; Möller, Pohlmann, Köller, & Marsh, 2009). Such contrast effects have mainly been supported for comparisons between math and the native language, but also for other comparisons between numerical (e.g., physics) and verbal domains (e.g., foreign language; Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2015). Beyond self-concept, effects of dimensional comparisons were found on interest (Schurtz, Pfost, Nagengast, & Artelt, 2014) and enjoyment (Goetz, Frenzel, Hall, & Pekrun, 2008). Evidence for the use of dimensional comparisons largely stems from correlational research, whereas there is less evidence from experimental research (for an exception, see Möller & Köller, 2001). Dimensional comparisons also play a crucial role in expectancy-value theory. Academic choices are supposed to be informed by intraindividual hierarchies of expectancies and values (Eccles, 2009). Recent research has addressed this assumption showing that choices (e.g., beginning a math- vs. verbal-intensive major) are not only affected by expectancies and values in the target domain, but also by expectancies and values in other domains (Chow et al., 2012; Nagy et al., 2008; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Parker et al., 2012).

Given the findings on dimensional comparisons, we propose that motivational interventions in STEM can have adverse effects on motivation in verbal domains. Such effects could ultimately reinforce intervention effects on STEM choices because of increased effects on intraindividual differences between STEM and non-STEM subjects. The intraindividual comparison between math on the one hand and the verbal domain (especially students' native language) on the other hand seems to be particularly relevant for the pursuit of a STEM career (see also Wang, Eccles, & Kenny, 2013).

The Present Study

For the present investigation, we use data from a large cluster randomized trial conducted in Germany to test whether an intervention in ninth-grade math classrooms had negative side effects on motivation for verbal subjects. Motivation in math was chosen as the target of the intervention as high school math courses are one important prerequisite for future careers in STEM fields. In this intervention study, 82 classrooms were randomly assigned to one of two intervention conditions or a waiting control group. Drawing on expectancy-value theory, the intervention consisted of a 90-minute session in a math classroom focusing on the value of math for students' lives. Students in the two intervention conditions either evaluated interview quotations describing the usefulness of math or wrote a text on the relevance of math. Both intervention conditions were shown to positively affect students' value and to some extent also expectancy beliefs for math with more comprehensive effects for the quotations condition (Brisson et al., 2014; Gaspard et al., in press). Here, we test effects of this intervention on the patterns of motivation across several domains. Next to math, we explore intervention effects on motivation in: (a) German as students' native language and therefore the main target of dimensional comparisons and (b) English as the first foreign language and thus a potential alternative target of dimensional comparisons. To examine the breadth of intervention effects, we consider effects on value as the most proximal intervention outcome and self-concept and effort as more distal outcomes. Taking an intraindividual difference perspective, we test effects on the difference between motivation for math and German as well as math and English.

Methods

Sample

Data for the study "Motivation in Mathematics" (MoMa) were collected in 82 ninth-grade classes in 25 academic track schools in the German state of Baden-Württemberg. The sample size was based on a power analysis for a multi-site cluster randomized trial aiming at an acceptable power ($\beta > .70$) to detect intervention effects of $\delta = 0.20$ when comparing a single intervention condition to the control condition (for more details, see Gaspard et al., in press). A total of 1978 students with active parental consent participated in the study, corresponding to a 96% participation rate. For the current study, 62 students in the two intervention conditions were excluded as they were absent during the intervention. Data analyses were thus based on a sample of 1916 students (mean age at the beginning of the study = 14.62, SD = 0.47, 53.5% female). The study consisted of three waves of data collection from September 2012 to March 2013. Students were administered questionnaires by trained research assistants before the intervention (pretest = T1), six weeks after the intervention (posttest = T2), and five months after the intervention (follow-up = T3).

Value intervention in math

Before the first data collection, within each school, the participating teachers and their classes were randomly assigned to one of two intervention conditions or a waiting control condition. Unequal class sample sizes for different conditions (quotations condition: 25 classes; text condition: 30 classes; waiting control condition: 27 classes) resulted from the fact that for teachers participating with two classes (N=9), both classes were included in the same experimental condition.

Students in the intervention conditions received a 90-minute standardized value intervention led by five trained researchers. The intervention consisted of a psychoeducational presentation on the relevance of math for the whole class and tasks for individual students. The psychoeducational presentation had two main components. First, research results on the importance of effort and self-concept for math achievement were presented. Students were also told about frame of reference effects (i.e., effects of social comparisons in the classroom) and the benefits of relying on temporal instead of social comparisons. This first part aimed at inoculating students against potential negative effects of highlighting the importance of a subject. These might occur if students judge their own achievement in this subject as low and are therefore threatened by information on its importance (cf., Durik, Shechter, Noh, Rozek, & Harackiewicz, 2015). Second, students were provided with various examples on the relevance of math for future education, career opportunities, and leisure time activities. This presentation was identical for both intervention conditions. After this presentation, students worked on individual tasks which differed between the two conditions. In the quotations condition, students were asked to read quotations of young adults describing situations in which math was useful to them and to evaluate these quotations based on their personal relevance. In the text condition, students were asked to make a list of arguments for the personal relevance of math to their current and future lives and to write an essay explaining these arguments.

Additionally, each intervention group received two reinforcements that were embedded into a homework diary, which was filled out by all classes for four weeks after the intervention. The first reinforcement, in which students were asked to reproduce what they remembered from their individual tasks, was filled out one week after the intervention. The second reinforcement was filled out two weeks after the intervention and resembled the individual tasks assigned to the students during the intervention lesson (for more details on the intervention, see Gaspard et al., in press). Classes in the waiting control condition also filled out homework diaries, but these did not include any intervention reinforcements. Students in the waiting control condition received the intervention that was shown to be more successful after the last wave of data collection.

The intervention focused only on the subject of math. No dimensional comparisons (i.e., highlighting the importance of math as compared to the verbal domain) were made at any time.

Measures

We assessed value beliefs, self-concept, and effort for math, German, and English with parallel scales (i.e., the wording was identical except for the subject name). All items used a four-point Likert scale ranging from *completely disagree* to *completely agree* as response format.

Value beliefs. Value beliefs were assessed with four items for each subject. The items tapped different value aspects: attainment value ("It is important to me to be good at [subject]."; "It is important to me to know a lot of [subject]."), intrinsic value ("I like doing [subject].") and utility value ("[Subject] is very useful to me."). The scales for German and English were constructed using a subset of items out of a larger questionnaire assessing value beliefs in math (Gaspard et al., 2014). Based on preliminary factor analyses, we excluded one item assessing cost ("[Subject] is a real burden to me."). All resulting scales exhibited good internal consistency (math: α = .77/.78/.77; German: .85/.85/.86; English: .83/.84/.84 at T1/T2/T3, respectively).

Self-concept. Self-concept was assessed with four items (e.g., "I just have no talent for Math/German/English" [reverse scored]). All items were well validated stemming from previous German large-scale studies (e.g., Trautwein, Lüdtke, Köller, & Baumert, 2006). The internal consistency of this scale was good for all subjects at all measurement waves (math: α = .93/.92/.92; German: .89/.89/.90; English: .90/.90/.91 at T1/T2/T3, respectively).

Effort. Effort in the subjects math, German, and English was assessed with four items for each subject (e.g., "I really try hard in [subject]."; adapted from Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009). The scale showed good internal consistency for all subjects (math: α = .80/.84/.86; German: .88/.89/.89; English: .85/.87/.88 at T1/T2/T3, respectively).

As a prerequisite for our analyses, we conducted tests of measurement invariance across time, subjects, and intervention conditions in several steps (see tables S1-S3 in the supplemental material for more details). Specifically, we tested invariance of factor loadings (strict measurement invariance) and invariance of item intercepts (strong measurement invariance) to be able to compare differences in latent means (Widaman & Reise, 1997). In the first step, we conducted tests of measurement invariance for value, self-concept, and effort across the three time points. As the analyses suggested that strong measurement invariance across time was acceptable for all three constructs, we used these models constraining factor loadings and intercepts to be equal across time for further tests of measurement invariance across subjects and intervention conditions. For the tests across subjects, a model with equal intercepts (in addition to factor loadings) was not tenable for value. We therefore assessed partial strong measurement invariance (Steenkamp & Baumgartner, 2009) by freely estimating the intercept for one item (assessing utility value). As this model yielded an acceptable fit, partial strong measurement invariance was defensible for value. For effort and self-concept, the tests of measurement invariance across subjects did not suggest any problem. The tests across intervention conditions indicated that strong measurement invariance could be accepted for all three constructs. Comparability of latent means across time, subjects, and intervention conditions was, therefore, established.

Statistical Analyses

Multilevel structural equation modeling. Given the multilevel structure of the data, we used multilevel structural equation modeling with Mplus (Version 7; Muthén & Muthén, 1998-2012) to examine the effects of the intervention on students' value beliefs, self-concept, and effort. Multilevel structural equation modeling (Mehta & Neale, 2005) combines the advantages of multilevel analyses, which take the nesting of students in classrooms into account (Raudenbush & Bryk, 2002), and latent variable modeling, which controls for measurement error (Bollen, 1989). An additional advantage of structural equation modeling is its flexibility; for instance, it allows explicit modeling of the measurement properties that were established based on prior confirmatory factor analyses.

Multilevel structural equation analyses were carried out separately for the latent constructs value, self-concept, and effort at the post-test and follow-up, respectively (for a graphical representation of the estimated model, see Figure 1). To be able to assess intervention effects on intraindividual differences between subjects, we combined all three subjects into one model for each construct and time point. In line with the recommendations for the evaluation of cluster randomized trials (Raudenbush, 1997), the respective pretest constructs in all three subjects were included as control variables at the student as well as at the class level. The effects at both levels were freely estimated to account for contextual effects (Marsh et al., 2009). The indicators of the latent constructs at the student level were group-mean centered, and manifest aggregation was used for the class level indicators (Marsh et al., 2009). Factor loadings were set to be equal across levels to ensure a common metric at student and class level (Marsh et al., 2009). Additionally, the factor loadings and item intercepts were constrained to be the same across time and subjects (with the exception of one value item, for which the intercept was freely estimated across subjects, see above). To assess effects of the intervention, we regressed the latent constructs at the posttest/follow-up on two class-level dummy variables that indicated the intervention conditions (quotations, text) with the control condition as a reference group.

To investigate effects on intraindividual differences in motivation for different subjects (i.e., math vs. German/English), we computed the difference between the effects on math and the other two subjects (i.e., b_{math} - b_{German}; b_{math} - $b_{English}$). Standard errors for these comparisons were obtained using the multivariate delta method implemented in Mplus (Muthén & Muthén, 1998-2012). These comparisons were possible as we used parallel scales across subjects. To obtain standardized effect sizes (for effect sizes in multilevel models, see Marsh et al., 2009; Tymms, 2004), the raw coefficients of intervention effects were divided by the total variance of the outcome variables out of null models (i.e., allowing all variables to correlate instead of estimating path coefficients). These effect sizes thus represent the adjusted difference between the interventions condition and the control condition in the outcome variable in total standard deviations.

Missing data. Due to the absence of students at single measurement waves and non-response to single items, missing data ranged from 5.4 to 12.6 % for the indicators of the focal motivational constructs. All analyses were conducted using full information maximum likelihood estimation (Graham, 2009) implemented in Mplus. To make the assumption of missing-at-random more plausible (see Enders, 2010), a nonverbal cognitive ability score, gender, previous math grades and achievement data for math at Time 1 were used as auxiliary variables by including correlations between these variables and the predictor variables as well as residuals of the dependent variables at both levels.

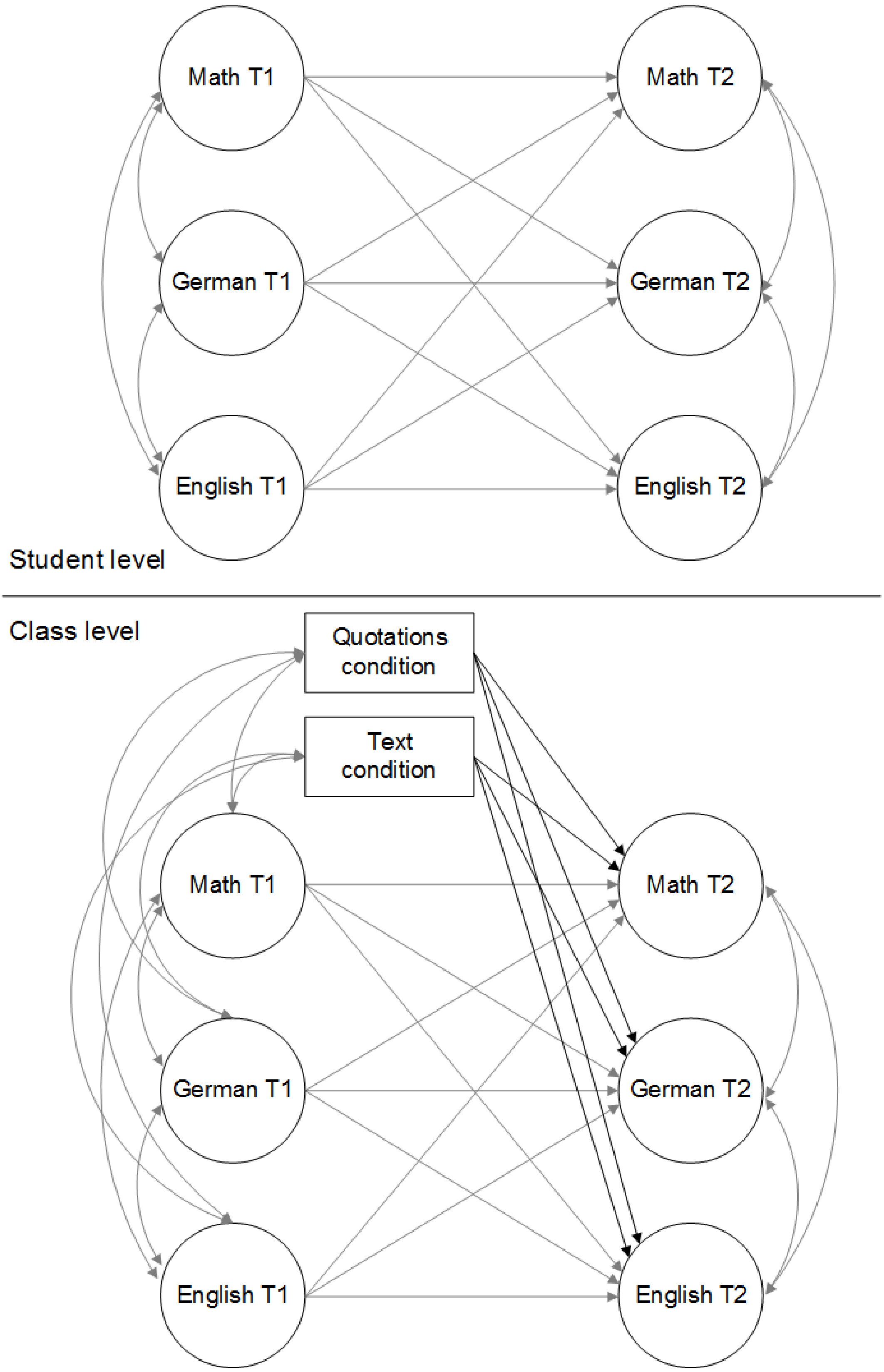

Figure 1. Multilevel structural equation modeling to estimate intervention effects on value, self-concept, and effort in math, German, and English. Measurement models and auxiliary variables are not depicted.

Results

Descriptive Statistics and Randomization Check

Descriptive statistics for all scales are displayed in Table 1. Correlations between value, self-concept, and effort in all subjects from a confirmatory factor analysis with the pretest data are presented in Table 2 . The confirmatory factor analysis supported the separability of the three constructs across the three subjects ($\chi2$ = 2663.12, *df* = 513, CFI = .940, TLI = .926, RMSEA = .048, SRMR = .051). Several aspects can be noted with regards to the correlation pattern. First, value, self-concept, and effort within one subject showed moderate to high intercorrelations. Second, value, self-concept and effort for German and English showed low to moderate intercorrelations. Third, value and self-concept for math and the verbal domains tended to be negatively correlated. Fourth, the correlation pattern for effort indicated a lower degree of domain-specificity with moderate intercorrelations between all three subjects. Correlation pattern at T2 and T3 were comparable.

To test if there were any differences between the three experimental conditions before the intervention, we specified multilevel multi-group models (with each experimental condition as a group) for initial value, self-concept, and effort in math, German, and English as well as for the auxiliary variables (i.e., cognitive abilities, gender composition, math grades, math achievement test, see Table S4 in the Supplemental Material for details). We conducted omnibus tests comparing the means of the three groups by Wald-$\chi2$-tests (using the "model test" command in Mplus), which are asymptotically equivalent to likelihood ratio tests (cf., Bollen, 1989). We found no significant differences between the groups prior to the intervention, neither for the focal constructs (all p's $\geq .620$), nor for the auxiliary variables (all p's $\geq .125$).

Table 1

Descriptive Statistics for All Study Variables by Intervention Condition

	Quotations condition			Text condition			Control condition		
Variable	M	SD	ICC	M	SD	ICC	M	SD	ICC
Math									
Value T1	2.68	0.64	.02	2.65	0.62	.07	2.61	0.60	.07
Value T2	2.74	0.63	.04	2.61	0.64	.07	2.55	0.64	.05
Value T3	2.75	0.61	.03	2.67	0.61	.08	2.58	0.61	.07
Self-concept T1	2.71	0.84	.02	2.68	0.86	.04	2.61	0.86	.04
Self-concept T2	2.75	0.84	.02	2.68	0.84	.05	2.59	0.84	.06
Self-concept T3	2.80	0.79	.01	2.75	0.79	.05	2.67	0.79	.05
Effort T1	2.78	0.59	.05	2.77	0.62	.01	2.78	0.61	.03
Effort T2	2.79	0.65	.07	2.73	0.67	.03	2.75	0.69	.02
Effort T3	2.71	0.68	.05	2.70	0.71	.04	2.68	0.69	.04
German									
Value T1	2.76	0.73	.08	2.73	0.71	.06	2.78	0.69	.04
Value T2	2.81	0.75	.10	2.81	0.75	.06	2.87	0.72	.05
Value T3	2.79	0.73	.07	2.87	0.74	.07	2.91	0.72	.07
Self-concept T1	2.96	0.75	.08	3.01	0.72	.07	2.99	0.73	.03
Self-concept T2	2.95	0.77	.09	3.02	0.73	.06	3.02	0.73	.02
Self-concept T3	2.91	0.78	.07	3.03	0.73	.08	3.00	0.73	.02
Effort T1	2.78	0.67	.06	2.81	0.64	.04	2.82	0.65	.06
Effort T2	2.83	0.71	.06	2.83	0.72	.04	2.85	0.70	.03
Effort T3	2.80	0.69	.06	2.84	0.73	.07	2.90	0.70	.04
English									
Value T1	3.35	0.57	.08	3.34	0.61	.06	3.33	0.58	.05
Value T2	3.37	0.58	.06	3.30	0.66	.04	3.33	0.61	.04
Value T3	3.38	0.60	.07	3.35	0.61	.06	3.31	0.62	.05
Self-concept T1	3.16	0.73	.04	3.14	0.70	.03	3.13	0.74	.07
Self-concept T2	3.17	0.70	.03	3.10	0.76	.02	3.12	0.72	.08
Self-concept T3	3.20	0.72	.04	3.13	0.73	.04	3.12	0.73	.06
Effort T1	3.06	0.61	.03	3.09	0.63	.06	3.13	0.60	.06
Effort T2	3.13	0.63	.03	3.09	0.69	.04	3.09	0.66	.06
Effort T3	3.06	0.65	.01	3.09	0.69	.06	3.08	0.69	.10

Note. ICC = intraclass correlation coefficient. Due to the absence of students at single measurement waves and non-response to single items, the sample sizes for the scales range from 509 to 530 in the quotations condition, from 619 to 680 in the text condition, and from 550 to 609 in the control condition.

Table 2

Correlations Between Study Variables at T1 (Corrected for Measurement Error)

	Variable	1	2	3	4	5	6	7	8	9
1	Math value	-								
2	Math self-concept	0.79 ***	-							
3	Math effort	0.54 ***	0.28 ***	-						
4	German value	-0.06	-0.27 ***	0.17 ***	-					
5	German self-concept	-0.25 ***	-0.27 ***	0.02	0.73 ***	-				
6	German effort	0.08 *	-0.13 ***	0.46 ***	0.77 ***	0.49 ***	-			
7	English value	-0.10 **	-0.22 ***	0.09 ***	0.31 ***	0.30 ***	0.23 ***	-		
8	English self-concept	-0.13 ***	-0.18 ***	0.07 *	0.21 ***	0.31 ***	0.17 ***	0.86 ***	-	
9	English effort	0.05	-0.13 ***	0.44 ***	0.30 ***	0.19 ***	0.56 ***	0.66 ***	0.53 ***	-

Note. Bivariate correlations from confirmatory factor analyses using pretest data are presented. Correlation pattern at T2 and T3 are comparable.

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

Intervention Effects on Value, Self-Concept, and Effort in Math, German, and English

Effects of the two intervention conditions as compared to the control condition were assessed on value, self-concept and effort in math, German, and English at posttest and follow-up (see Table 3 for effect sizes and Table S5 in the Supplemental Material for the complete models including the effects of pretest variables). Effects on value and self-concept in math as the target subject of the intervention have already been reported by Gaspard et al. (in press) and Brisson et al. (2014). For value, previous analyses used a more differentiated measure with subscales for different value components. Here, the results on math as the target subject of the intervention are reported as a comparison to effects on German and English using parallel scales. Additionally, we computed model-implied effects on the intraindividual difference between math and German as well as between math and English. All effects of the two intervention conditions reported in the text are standardized effect sizes with respect to the total variance of the outcome variable.

For value at the posttest, we found a positive effect of the quotations condition on math value ($\beta = .27$, $p < .001$, 95% CI [0.12, 0.41]). The quotations condition did not show a significant effect on German ($p = .123$) or English value ($p > .250$) at the posttest. Concerning between-subject differences, we found effects of the quotations condition on the difference between math and German ($\beta = .35$, $p < .001$, 95% CI [0.17, 0.52]) as well as on the difference between math and English ($\beta = .27$, $p = .001$, 95% CI [0.12, 0.42]), which can be explained by the effects on math value. At the follow-up, we still observed a positive effect of the quotations condition on math value ($\beta = .24$, $p = .002$, 95% CI [0.09, 0.39]). In addition, we found a negative effect of the quotations condition on German value ($\beta = -.18$, $p = .004$, 95% CI [-0.30, -0.06]). In consequence, the effect of the quotations condition on the difference between math and German was significant ($\beta = .42$, $p < .001$, 95% CI [0.21, 0.63]). An additional Wald test indicated that this effect on the difference was larger than the net effect on math value (χ^2 (1) = 8.15, $p = .004$). The quotations condition did not show an effect on English value ($p = .101$) or the difference between math and English ($p = .222$) at the follow-up. The text condition did not have any significant effect on students' value beliefs at the posttest (all p's $\geq .065$) and the follow-up (all p's $\geq .135$). The effects of the two intervention conditions on students' value beliefs are further displayed in Figure 2.

For self-concept, the quotations condition did not show a significant effect on math (p = .065) or German (p = .300) at the posttest. However, the difference between math and German self-concept in the quotations condition increased as compared to the control condition (β = .15, p = .037, 95% CI [0.01, 0.30]). No effects of the quotations condition for English (p = .233) or the difference between math and English (p > .250) were found. The observed effect pattern at the follow-up was similar. The effect on the difference between math and German self-concept, however, missed significance (p = .077). No other effects of the quotations condition on self-concept were found at the follow-up (p ≥ .074). In line with the results for value, the text condition did not show any significant effect on students' domain-specific self-concepts at the posttest or the follow-up (all p's > .250).

For effort, students in the quotations condition did not report higher effort in math than students in the control condition at the posttest (p = .100). In line with this, no effects of the quotations condition were observed for German effort (p > .250) and the difference between math and German (p = .193) at the posttest. However, at the posttest, students in the quotations condition reported higher effort in English compared to students in the control condition (β = .14, p = .008, 95% CI [0.04, 0.24]). The effect of the quotations condition on the difference between math and English was not significant (p > .250). At the follow-up, no effects of the quotations condition on students' effort were observed (all p's ≥ .068). Again, we found no effect of the text condition on effort in any subject, neither at the posttest (all p's > .250) nor at the follow-up (all p's ≥ .242).

Table 3

Effects of the Intervention Conditions as Compared to the Control Condition

Outcome	Quotations condition						Text condition					
	Posttest			Follow-Up			Posttest			Follow-Up		
	β	*p*	95% CI	β	*p*	95% CI	β	*p*	95% CI	β	*p*	95% CI
Value												
Math	0.27	<.001	[0.12, 0.41]	0.24	.002	[0.09, 0.39]	0.04	.633	[-0.11, 0.18]	0.13	.088	[-0.02, 0.28]
German	-0.08	.123	[-0.18, 0.02]	-0.18	.004	[-0.30, -0.06]	-0.07	.231	[-0.17, 0.04]	-0.02	.788	[-0.15, 0.11]
English	-0.01	.925	[-0.11, 0.10]	0.12	.101	[-0.02, 0.25]	-0.09	.065	[-0.19, 0.01]	0.10	.135	[-0.03, 0.22]
Derived effects												
Math - German	0.35	<.001	[0.17, 0.52]	0.42	<.001	[0.21, 0.63]	0.10	.300	[-0.09, 0.29]	0.15	.190	[-0.07, 0.37]
Math - English	0.27	.001	[0.12, 0.42]	0.13	.222	[-0.07, 0.32]	0.13	.100	[-0.02, 0.28]	0.03	.723	[-0.14, 0.21]
Self-concept												
Math	0.10	.065	[-0.01, 0.19]	0.07	.216	[-0.04, 0.17]	0.03	.508	[-0.06, 0.12]	0.03	.572	[-0.07, 0.12]
German	-0.06	.300	[-0.16, 0.05]	-0.09	.168	[-0.21, 0.04]	-0.03	.481	[-0.12, 0.06]	0.02	.776	[-0.11, 0.14]
English	0.06	.233	[-0.04, 0.16]	0.10	.074	[-0.01, 0.22]	-0.03	.539	[-0.12, 0.06]	0.02	.789	[-0.10, 0.13]
Derived effects												
Math - German	0.15	.037	[0.01, 0.30]	0.15	.077	[-0.02, 0.32]	0.06	.327	[-0.06, 0.19]	0.01	.926	[-0.16, 0.18]
Math - English	0.04	.621	[-0.11, 0.18]	-0.03	.690	[-0.20, 0.13]	0.06	.398	[-0.08, 0.20]	0.01	.900	[-0.15, 0.18]
Effort												
Math	0.12	.100	[-0.02, 0.26]	0.06	.426	[-0.08, 0.20]	0.01	.912	[-0.12, 0.13]	0.03	.651	[-0.10, 0.15]
German	0.00	.992	[-0.09, 0.09]	-0.10	.123	[-0.23, 0.03]	-0.03	.542	[-0.13, 0.07]	-0.06	.378	[-0.21, 0.08]
English	0.14	.008	[0.04, 0.24]	0.07	.291	[-0.06, 0.21]	0.03	.568	[-0.07, 0.13]	0.08	.242	[-0.05, 0.21]
Derived effects												
Math - German	0.12	.193	[-0.06, 0.29]	0.16	.068	[-0.01, 0.33]	0.04	.643	[-0.12, 0.20]	0.09	.193	[-0.05, 0.23]
Math - English	-0.02	.827	[-0.19, 0.15]	-0.01	.877	[-0.19, 0.17]	-0.02	.756	[-0.17, 0.12]	-0.05	.481	[-0.18, 0.09]

Note. CI = confidence interval. Regression coefficients represent standardized effect sizes.

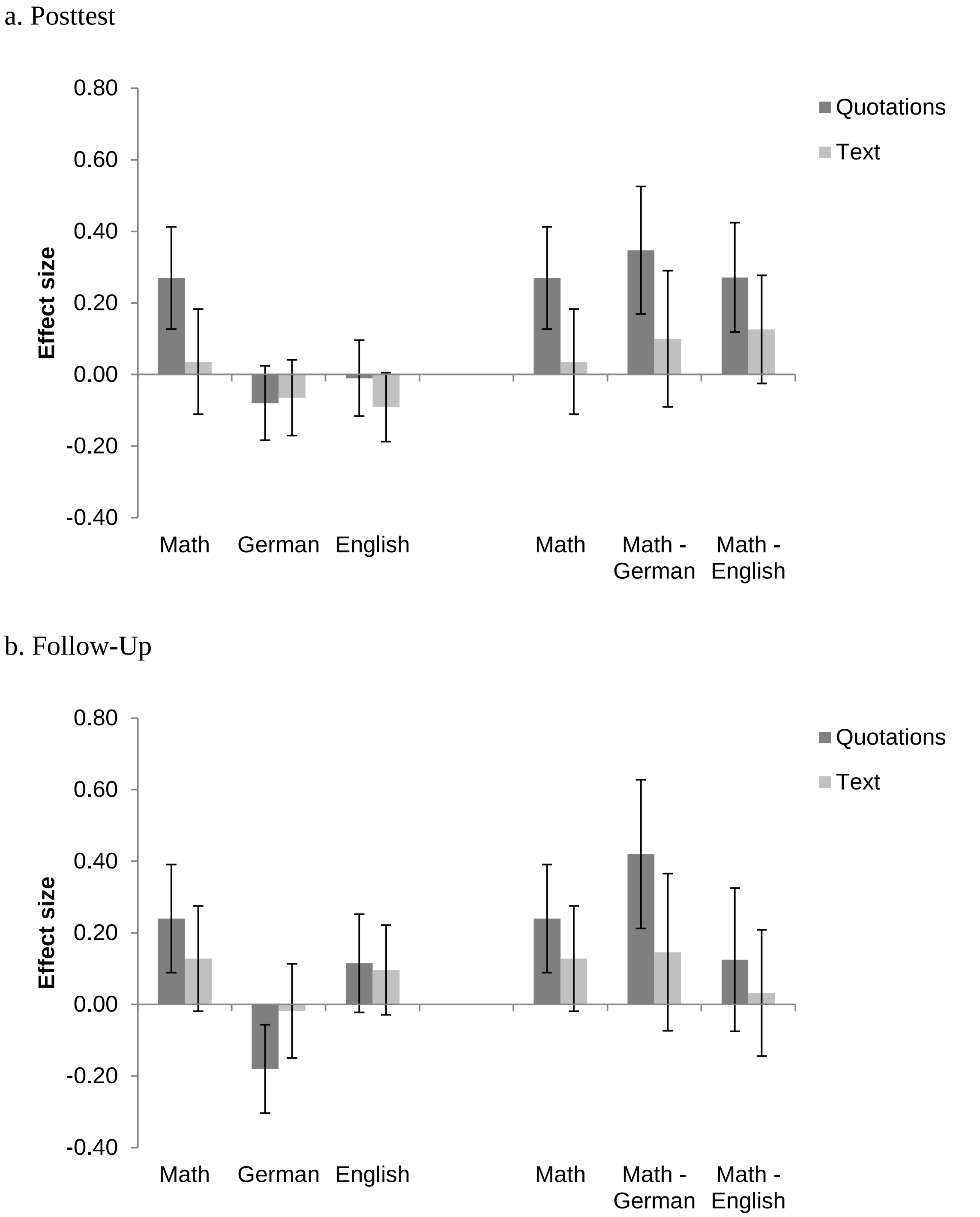

Figure 2. Effects of the intervention conditions (as compared to the control condition) on value at the posttest (a) and at the follow-up (b). On the left, effects on value in different subjects are presented. On the right, unadjusted effects on math value are presented as compared to the effects adjusted for German and English value, respectively. Error bars represent 95% confidence intervals.

Discussion

Within a cluster randomized study aiming at fostering motivation in math, we addressed the important question how motivational interventions in one subject affect motivational patterns across subjects. Based on the literature on dimensional comparisons, side effects were assessed in the verbal domain. The intervention condition that was successful in fostering math value showed adverse effects on value for students' native language (i.e., German) five months after the intervention. Regarding the effect size, we could show that the effect on the difference between math and German as a measure of intraindividual comparisons was larger than the effect on math. No effects were found on value in English. Whereas this effect pattern was observed for value as the focal construct of the intervention, the effects did not generalize to students' self-concepts and effort.

This pattern of effects can be explained by dimensional comparisons. Most support has previously been found for dimensional comparisons between math and the native language (Möller & Marsh, 2013). Students in our study generally reported high value for English. The intervention aimed at triggering reflections on the relevance of the current learning material for future careers. In Germany, English is generally perceived as highly relevant for almost every career, including math-intensive fields, and students' knowledge about this might have buffered negative effects on English value. By comparison, the curriculum in German in secondary school is highly focused on literature and students might therefore perceive its relevance as more limited (for students' perceptions of domain characteristics, see Goetz et al., 2014). Regarding the development over time, negative effects on German were only found at the follow-up, whereas effects on math were already found at the posttest. Dimensional comparisons processes seem to rely on changes in motivation in the targeted domain incurring prior to spillover effects on other domains. These effects did not generalize beyond the focal construct (i.e., value); they were not found for self-concept and effort. We even found a positive effect of the quotations condition on effort in English at the posttest, which could, however, no longer be found at the follow-up. One noticeable finding, which might contribute to this effect, is that students' reported effort seemed to be less domain-specific as it showed positive correlations across the three subjects.

Our study provides strong evidence for the importance of dimensional comparisons in real life settings as it experimentally manipulated motivation in one domain in the classroom context and assessed effects on non-targeted domains. In addition to self-concept, where dimensional comparison effects have been extensively supported (Marsh & Hau, 2004; Möller et al., 2009), our study suggests that dimensional comparisons also affect task value (see also Nagy et al., 2008, 2006). This is in line with the assumptions of expectancy-value-theory discussing the role of individual hierarchies for the value attached to different choice options (Eccles, 2009). Students have limited time and energy. Highlighting the value of one subject can lead to changes in students' hierarchies of importance and therefore increase the subjective costs of engaging in another subject.

Our results have important implications for interventions in STEM fields. Adverse side effects as found in our study should no longer be neglected in intervention research. We included two verbal domains that we thought of as the most important candidates for such effects and found negative effects for students' native language. Future research including a broader range of domains is needed to see which domains are affected through spillover effects and how. Depending on the content of an intervention as well as the similarity between domains, positive side effects via assimilation processes are also possible (e.g., between math and sciences; cf. Jansen et al., 2015; Marsh et al., 2015).

At first sight, adverse effects of motivational interventions are troubling. However, if the ultimate goal of STEM interventions is to recruit more people for STEM, such effects might not only be a risk that needs to be accepted, but can actually help in having more people choosing careers in STEM. By taking an intraindividual difference perspective, we demonstrated that adverse side effects are good news in terms of the effectiveness of interventions. As positive effects on math motivation as well as negative effects on verbal motivation positively affect intraindividual comparisons between math and verbal domains, these side effects could increase the likelihood for students to pursue math-related careers. Future research using longer follow-up designs should assess such potential effects on later career choices. Ethical considerations, on the other hand, seem to speak against diminishing motivation in one subject as a means to foster motivation for another subject. Researchers developing motivational interventions should therefore carefully consider side effects and seek for possibilities to foster motivation and engagement across subjects.

References

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task-value, and achievement goals. *Journal of Educational Psychology, 93*, 23–34. doi:10.I037//0022-0663.93.1.23

Bong, M., & Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*, 1–40. doi:10.1023/A:1021302408382

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2014). Short intervention, lasting effects: Promoting students' competence beliefs, effort, and achievement in mathematics. *Manuscript Submitted for Publication.*

Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology, 48*, 1612–1628. doi:10.1037/a0030194

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*, 104–118. doi:10.1007/s11031-014-9419-0

Eccles, J. S. (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*, 78–89. doi:10.1080/00461520902832368

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74–146). San Francisco, CA: W. H. Freeman.

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

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

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (in press). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2014). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, Advance online publication.* doi:10.1037/edu0000003

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology, 33*, 9–33. doi:10.1016/j.cedpsych.2006.12.002

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C., & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology, 5,* 1153. doi:10.3389/fpsyg.2014.01153

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: *An experimental test of a utility-value intervention. Psychological Science, 23,* 899–906. doi:10.1177/0956797611435530

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*, 880–895. doi:10.1037/a0019506

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

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *Journal of Educational Psychology, Advance online publication.* doi:10.1037/edu0000021

Karabenick, S. A., & Urdan, T. (Eds.). (2014). *Motivational interventions (Advances in motivation and achievement, Volume 18)*. Bingley, UK: Emerald Group Publishing Limited.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research, 23*, 129–149. doi:10.3102/00028312023001129

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67. doi:10.1037/0022-0663.96.1.56

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional Comparison Theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16–32. doi:10.1016/j.learninstruc.2014.08.005

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*, 764–802. doi:10.1080/00273170903333665

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*, 259–284. doi:10.1037/1082-989X.10.3.259

Möller, J., & Köller, O. (2001). Dimensional comparisons : An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93*, 826–835. doi:10.1037//O022-O663.93.4.826

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*, 544–560. doi:10.1037/a0032459

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129–1167. doi:10.3102/0034654309337522

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus user’s guide. Seventh edition. Los Angeles, CA: Muthén & Muthén.

Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115–143). Washington, DC: American Psychological Association.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12,* 323–345. doi:10.1080/13803610600765687

National Science Board. (2007). *A national action plan for addressing the critical needs of the U.S. science, technology, engineering, and mathematics education system*. Arlington, VA: National Science Foundation.

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology, 48,* 1629–1642. doi:10.1037/a0029167

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2,* 173–185. doi:10.1037/1082-989X.2.2.173

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student’s mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction, 34,* 32–41. doi:10.1016/j.learninstruc.2014.08.001

Steenkamp, J.-B. E. M., & Baumgartner, H. (2009). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78–90. doi:10.1086/209528

Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*, 334–349. doi:10.1037/0022-3514.90.2.334

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*, 1115–1128. doi:10.1037/a0017048

Trautwein, U., Nagengast, B., Marsh, H. W., Gaspard, H., Dicke, A.-L., Lüdtke, O., & Jonkmann, K. (2013). Expectancy-value theory revisited. From expectancy-value theory to expectancy-valueS theory? In D. M. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self-processes and human development* (pp. 233–249). Charlotte, NC: Information Age Publishing.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.

Wang, M.-T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*, 304–340. doi:10.1016/j.dr.2013.08.001

Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*, 770–775. doi:10.1177/0956797612458937

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). New York, NY: Routledge.

Supplemental Material

Table S1

Results for Tests of Measurement Invariance Across Time

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Value						
M1: configural invariance	1136.41	414	.976	.963	.030	.056
M2: weak measurement invariance	1162.77	432	.976	.964	.030	.057
M3: strong measurement invariance	1270.83	450	.973	.962	.031	.057
Self-concept						
M1: configural invariance	501.12	333	.996	.993	.016	.020
M2: weak measurement invariance	524.28	351	.996	.993	.016	.021
M3: strong measurement invariance	571.36	369	.995	.992	.017	.021
Effort						
M1: configural invariance	1364.23	414	.973	.960	.035	.047
M2: weak measurement invariance	1404.43	432	.973	.960	.034	.047
M3: strong measurement invariance	1550.76	450	.969	.957	.036	.048

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All models fit statistics reported are robust fit statistics. Correlated residuals were allowed between identical items across time, between parallel items across subjects and between negatively worded items (for self-concept).

Table S2

Results for Tests of Measurement Invariance Across Subjects

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Value						
M1: configural invariance	1270.83	450	.973	.962	.031	.057
M2: weak measurement invariance	1383.17	456	.969	.957	.033	.062
M3: strong measurement invariance	3420.03	464	.901	.866	.058	.143
M3a: partial strong measurement invariance	2522.04	456	.931	.904	.049	.107
Self-concept						
M1: configural invariance	571.36	369	.995	.992	.017	.021
M2: weak measurement invariance	654.46	375	.993	.989	.020	.027
M3: strong measurement invariance	987.64	383	.986	.977	.029	.061
Effort						
M1: configural invariance	1550.76	450	.969	.957	.036	.048
M2: weak measurement invariance	1720.96	456	.965	.951	.038	.053
M3: strong measurement invariance	2520.27	464	.943	.922	.048	.066

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All models fit statistics reported are robust fit statistics. Correlated residuals were allowed between identical items across time, between parallel items across subjects and between negatively worded items (for self-concept). In all models, factor loadings and intercepts were set to be equal across time.

Table S3

Results for Tests of Measurement Invariance Across Intervention Conditions

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Value						
M1: configural invariance	2430.21	1350	.966	.953	.035	.063
M2: weak measurement invariance	2442.11	1368	.967	.954	.035	.063
M3: strong measurement invariance	2449.48	1386	.967	.955	.035	.064
Self-concept						
M1: configural invariance	1486.60	1107	.992	.986	.023	.029
M2: weak measurement invariance	1515.28	1125	.992	.986	.023	.031
M3: strong measurement invariance	1533.43	1143	.992	.986	.023	.031
Effort						
M1: configural invariance	2640.92	1350	.966	.953	.039	.053
M2: weak measurement invariance	2659.14	1368	.966	.953	.038	.054
M3: strong measurement invariance	2672.70	1386	.966	.954	.038	.054

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All models fit statistics reported are robust fit statistics. Correlated residuals were allowed between identical items across time, between parallel items across subjects and between negatively worded items (for self-concept). In all models, factor loadings and intercepts were set to be equal across time.

Table S4

Randomization Check for Study and Auxiliary Variables at the Beginning of the Study

	Quotations condition		*Text condition*		*Control condition*		*Randomization check*	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wald statistic (*df*=2)	*p*
Math value	0.01	0.52	0.00[a]	0.50	-0.04	0.49	0.84	.657
Math self-concept	0.02	0.77	0.00[a]	0.79	-0.04	0.80	0.96	.620
Math effort	0.01	0.54	0.00[a]	0.58	-0.01	0.56	0.20	.905
German value	0.02	0.61	0.00[a]	0.60	0.03	0.59	0.29	.867
German self-concept	-0.04	0.68	0.00[a]	0.64	-0.03	0.65	0.43	.808
German effort	-0.01	0.64	0.00[a]	0.61	0.00	0.62	0.08	.962
English value	0.02	0.47	0.00[a]	0.52	-0.01	0.49	0.30	.860
English self-concept	0.02	0.67	0.00[a]	0.64	-0.01	0.67	0.25	.882
English effort	-0.04	0.57	0.00[a]	0.59	0.01	0.56	0.87	.648
Math grade	2.81	0.97	2.74	0.96	2.88	0.90	4.16	.125
Math achievement test	48.63	16.44	49.84	18.19	47.15	17.03	1.51	.470
Cognitive abilities	19.99	4.00	19.97	4.23	19.69	4.23	1.34	.512
Gender (1=male)	0.48	0.50	0.48	0.50	0.46	0.50	0.18	.913

Note. Standard deviations are computed out of the total variance across levels. Means and standard deviations for value, self-concept, and effort are means and standard deviations of latent constructs. Loadings and intercepts are constrained to be equal across intervention conditions. Math grades were coded 1 (very good) to 6 (unsufficient); math achievement test scores are percent correct; cognitive abilities test scores are number correct (max. 25).

[a] fixed parameter.

Table S5

Structural Coefficients from Multilevel Structural Equation Models Predicting Value, Self-Concept, and Effort in Math, German, and English

	Posttest						Follow-Up					
	Math		German		English		Math		German		English	
Outcome and Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Value												
Student level												
Math	0.87 ***	0.03	0.00	0.03	0.01	0.03	0.73 ***	0.03	-0.08 *	0.04	-0.04 †	0.03
German	0.01	0.02	0.86 ***	0.03	0.03	0.03	-0.02	0.02	0.71 ***	0.03	0.04	0.02
English	0.01	0.03	0.03	0.03	0.81 ***	0.03	-0.01	0.03	0.12 ***	0.03	0.73 ***	0.04
Class level												
Math	0.93 ***	0.14	-0.05	0.11	0.14 †	0.08	0.87 ***	0.14	0.15	0.14	-0.01	0.09
German	0.14 †	0.07	1.00 ***	0.07	0.01	0.07	0.15 *	0.07	0.91 ***	0.08	0.10	0.09
English	0.12	0.09	0.07	0.10	0.88 ***	0.09	0.15	0.12	0.24 *	0.11	0.69 ***	0.12
Quotations	0.14 ***	0.04	-0.05	0.03	0.00	0.03	0.12 **	0.04	-0.11 **	0.04	0.06	0.04
Text	0.02	0.04	-0.04	0.04	-0.05 †	0.03	0.06 †	0.04	-0.01	0.04	0.05	0.03
Self-concept												
Student level												
Math	0.86 ***	0.02	-0.03 *	0.01	-0.02	0.02	0.77 ***	0.02	-0.03 †	0.02	-0.02	0.02
German	-0.06 **	0.02	0.87 ***	0.02	0.07 **	0.02	-0.08 ***	0.02	0.79 ***	0.03	0.06 **	0.02
English	0.05 **	0.02	0.02	0.02	0.86 ***	0.02	0.04 †	0.02	0.05 *	0.02	0.80 ***	0.02
Class level												
Math	0.84 ***	0.07	-0.05	0.06	-0.04	0.06	0.75 ***	0.07	-0.03	0.10	-0.06	0.08
German	-0.10	0.07	0.76 ***	0.08	0.01	0.07	-0.15 †	0.09	0.65 ***	0.10	0.09	0.07
English	0.10	0.07	0.10	0.09	0.88 ***	0.09	0.15 †	0.08	0.14	0.13	0.77 ***	0.07
Quotations	0.07 †	0.04	-0.04	0.04	0.04	0.03	0.05	0.04	-0.06	0.04	0.07 †	0.04
Text	0.03	0.04	-0.02	0.03	-0.02	0.03	0.02	0.03	0.01	0.04	0.01	0.04

(continued)

Outcome and Predictor	Posttest						Follow-Up					
	Math		German		English		Math		German		English	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Effort												
Student level												
Math	0.80 ***	0.04	0.02	0.03	0.03	0.04	0.71 ***	0.04	0.10 **	0.03	0.10 **	0.03
German	0.05	0.03	0.74 ***	0.03	0.01	0.03	0.03	0.04	0.61 ***	0.04	0.06	0.04
English	0.02	0.03	0.08 *	0.04	0.77 ***	0.03	0.08 †	0.04	0.08 *	0.04	0.65 ***	0.04
Class level												
Math	1.00 ***	0.15	0.07	0.10	0.07	0.15	1.01 ***	0.23	-0.04	0.17	0.37 *	0.17
German	-0.14	0.10	0.91 ***	0.06	0.07	0.09	0.10	0.14	0.79 ***	0.11	0.00	0.11
English	0.12	0.14	-0.10	0.09	0.71 ***	0.11	-0.01	0.16	0.07	0.15	0.68 ***	0.12
Quotations	0.08	0.05	0.00	0.03	0.09	0.03	0.04	0.05	-0.07	0.05	0.05	0.04
Text	0.00	0.04	-0.02	0.03	0.02	0.03	0.02	0.04	-0.04	0.05	0.05	0.04

Note. Parameter estimates are non-standardized. The model included student gender, previous math grades, a math achievement test and cognitive abilities as auxiliary variables.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

5

General Discussion

5 General Discussion

Students' value beliefs for subjects such as mathematics are important predictors for achievement-related choices at school and beyond (Wang & Degol, 2013; Wigfield, Tonks, & Klauda, 2009). Many students, however, have a hard time to come up with specific reasons for why they should learn mathematics and students' value beliefs for mathematics decline over the school years. The present dissertation therefore investigated how students' value beliefs for mathematics can be promoted in the classroom context. Within a cluster randomized trial, a utility value intervention was implemented in 82 ninth grade math classrooms. Three empirical studies using data from this intervention project were conducted within this dissertation. In the following, the findings of the three empirical studies will be summarized and situated in their broader research context. The discussion of these findings is arranged around three major topics: (a) the importance of a multidimensional perspective on value, (b) the role of gender for value beliefs in mathematics, and (c) the effectiveness of value interventions. Next, some strengths and limitations of the dissertation will be explored. In the final section, implications of the dissertation for future research and educational practice will be elaborated.

5.1. Discussion of General Findings

5.1.1. The case for a multidimensional perspective on values

All three empirical studies in this dissertation took a multidimensional perspective on value beliefs. Whereas Study 1 and 2 assessed value beliefs in mathematics with a newly developed multifaceted instrument, Study 3 examined value beliefs in several subjects.

A multifaceted approach to values

Although the four value components are separable on a theoretical basis, previous research mostly did not include separate measures for all of the components (Trautwein et al., 2013). Moreover, substantial differences can be noted in the items that have been used to indicate these components with some items tapping qualitatively different aspects than others (see section 1.1.3). Although these different aspects are all implied within the broad conceptualization of value beliefs in expectancy-value theory, they have not been systematically differentiated thus far. To overcome such inconsistencies in the operationalization of value beliefs, this dissertation proposed a multifaceted approach assuming that most of the value components include multiple facets. Supporting this assumption, Study 1 found that a total of eleven subfacets of value beliefs could be empirically distinguished. Attainment value was divided into the facets importance of achievement and personal importance, utility value was differentiated into five facets referring to short- and long-term goals in different life domains (i.e., school, daily life, social life, job, future life in general), and cost was divided into the facets opportunity cost, effort required, and emotional cost. In line with a higher order model, facets within components were highly correlated. Still, the results of both Study 1 and 2 supported the usefulness of considering subfacets in the measurement of value beliefs. Pronounced gender differences favoring boys were found for some, but not all of these value facets. These gender differences depended not only on the higher-order value component, but on the specific subfacet under consideration. The facet approach can thus help in achieving a better understanding of gender differences in value beliefs in mathematics (see also section 5.1.2). Building on the findings of Study 1, Study 2 examined effects of the utility value intervention on separate value facets. Here, the facet approach was especially helpful in the case of utility value, where a differential pattern of intervention effects on subfacets was found that was in line with the focus of the intervention.

An intraindividual-difference approach to values

Expectancy-value theory assumes that choices are affected by intraindividual hierarchies in expectancies and values (Eccles, 2009). From this point, it seems important to investigate value beliefs in several subjects and the processes through which intraindividual hierarchies develop. However, research based on expectancy-value theory has only recently begun to explicitly model the hierarchies in task values (e.g., Chow, Eccles, & Salmela-Aro, 2012). In the tradition of the internal/external frame of reference model (Marsh, 1986), on the other hand, there is extensive evidence that students' self-concepts are affected not only by external (or social) comparisons, but also by internal (or dimensional) comparisons (Marsh & Hau, 2004; Möller, Pohlmann, Köller, & Marsh, 2009). In addition to self-concept, effects of dimensional comparisons have also been found for interest and intrinsic value (Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Schurtz, Pfost, Nagengast, & Artelt, 2014). Based on these findings, Study 3 explored whether an intervention that targets students' motivation in math had negative side effects on motivation in two verbal domains (i.e., German and English). Negative effects of the intervention were found on value in German, but not in English. In line with the effects in math, these effects did not generalize to self-concept and effort in German and English. As the study used an experimental design in the classroom context, it provides strong empirical support for the role of dimensional comparisons in the development of students' value beliefs. However, the study only assessed value beliefs in two subjects other than math. Based on previous research, German and English were chosen as the two subjects that were most susceptible to contrast effects (i.e., negative effects resulting from comparisons between different subjects). Still, more research is needed assessing a broader range of subjects. Different school subjects can be ordered on a continuum from more math like to more verbal like domains (Möller & Marsh, 2013). Depending on the closeness of subjects on this continuum, assimilation effects between subjects (i.e., positive effects resulting from comparisons between similar subjects such as math and physics) are also possible. Contrast and assimilation effects between a broader range of subjects have only been examined for self-concept thus far (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2015). However, such effects are potentially also relevant for the development of students' value beliefs. When students realize the usefulness of physics, for instance, this might also lead to an increased perceived utility for mathematics as knowledge in mathematics is relevant for physics.

5.1.2. Gender differences in value beliefs for mathematics

Females do not choose math-related courses and careers as frequently as do males (Watt & Eccles, 2008). Although value beliefs are assumed to be the central factor for explaining such gender differences in choices (Eccles, 2009), previous research did not consistently find gender differences in math values. Some of these inconsistencies in previous findings might be due to differences in the operationalization of value beliefs. Therefore, Study 1 examined gender differences in math values on the facet level. Whereas the factor structure was found to be invariant across gender, considerable differences in mean levels favoring boys were found for some of the value facets: Compared to girls, boys reported higher intrinsic value, personal importance and utility for long-term goals (i.e., future life in general as well as job). Girls, on the other hand, perceived higher emotional cost and effort required than boys. No significant gender differences were found for the importance of achievement, opportunity cost, and the usefulness for short-term goals (i.e., school, daily life, social life). These results point to differences and similarities in girls' and boys' value beliefs about math. Summarizing this complex pattern of results, it seems that girls perceive math as a subject that is important in the school context, while it is less important to them personally and rather unrelated to their future plans. Girls, thus, tend to perceive math as a necessary evil that they do not intrinsically value. This constellation of value beliefs in math could potentially lead to an impaired affective experience of math for females (Frenzel, Pekrun, & Goetz, 2007). Such gender differences in value beliefs likely also contribute to later gender differences in choices in the domain of math.

Drawing on these findings, Study 2 explored whether gender differences in value beliefs can be reduced by interventions that highlight the usefulness of math for various career opportunities, including both male- and female-typed domains. Indeed, the intervention tended to be more beneficial in fostering intrinsic and utility value for girls than for boys. Several factors might generally contribute to differential effects of utility value interventions for boys and girls. First, one important factor might be students' age. Students in this study were at an age where boys and girls are differentially mature (Eisenberg, 2006). Second, the effects of utility value interventions for boys and girls might depend on the implementation of the intervention and the specific intervention strategy. In this study, the intervention was implemented by female researchers only which might have led to role model effects. The information on the usefulness of math that students were provided with referred to a variety of careers. However, the novelty of this information for students potentially differed between STEM and non-STEM careers. The individual tasks that were used to foster students' perceptions of usefulness both involved writing, which is an activity that females typically enjoy more than males (Meece, Glienke, & Burg, 2006). Therefore, the answer to the question whether utility value interventions help in reducing gender differences is probably complex and depends on several factors. Nevertheless, the results of Study 2 suggest that both boys' and girls' value beliefs for math can be promoted by such interventions. It is important to note that even interventions that specifically target girls' motivation in STEM subjects have been shown to foster both boys' and girls' motivation (Häussler & Hoffmann, 2002), and such interventions can still be seen as very successful.

5.1.3. Effectiveness of value interventions

Previous studies have demonstrated that utility value interventions can be an effective tool for promoting motivation and performance in STEM courses as well as STEM course choices (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009). The empirical studies that were conducted within this dissertation further the findings on the effectiveness of utility value interventions in several ways. Study 2 added to the positive results of previous studies in showing that students' value beliefs in math were promoted until five months after the intervention. As mixed effects for different intervention strategies were previously found, the effects of two intervention strategies on students' math value beliefs were systematically compared. One of these strategies (i.e., self-generating arguments for the usefulness in an essay format) was adapted from previous studies (Hulleman et al., 2010; Hulleman & Harackiewicz, 2009), whereas the other strategy (i.e., reflecting on given arguments from interview quotations) was newly tested. It was found that evaluating quotations had more beneficial effects on students' value beliefs than writing an essay. The task to evaluate such quotations seems to be an effective way to trigger reflection processes and personal connections in students. Interview quotations in which older students describe the usefulness of subject-specific knowledge to their lives can be seen as an easily implementable approach to provide students with role models. To maximize the effectiveness, other elements were embedded in the intervention in addition to these individual tasks: A psychoeducational presentation that included information on the importance of one's attitudes for achievement as well as on the relevance of math for students' future prepared students for their individual tasks. These intervention components were, however, not separately tested. The cost of testing different intervention elements and their combinations is considerably higher in the classroom context than in laboratory studies because of the required sample size as well as because of ethical considerations. The information on the importance of one's attitudes was included to buffer against potential negative effects of highlighting the importance of a subject for low-performing students (for a laboratory study, see Durik, Shechter, Noh, Rozek, & Harackiewicz, 2015). Such potential negative effects should be avoided when implementing an intervention in a real-life setting, especially at a larger scale.

For the effectiveness of value interventions, side effects on subjects that are not targeted by the intervention are also highly relevant to consider. If the ultimate goal of interventions within STEM subjects is to have more people in STEM-related careers, such side effects can actually increase the effectiveness of STEM interventions. STEM career choices are affected both by value beliefs for STEM subjects as well as by value beliefs for non-STEM subjects (Chow et al., 2012). Therefore, effects on intraindividual differences in value beliefs are especially relevant for career choices. The results of Study 3 could demonstrate that intervention effects on intraindividual differences in value beliefs—in this case between math and German—were larger than the effects on math value alone. Whereas this is good news for the effectiveness of value interventions, research should also carefully consider such side effects from an ethical perspective. It is worrisome that increasing motivation in one subject comes with lowering motivation in another subject. To avoid such side effects, research should seek to develop interventions that foster motivation across subjects.

5.1.4. Strengths and limitations of the present dissertation

When interpreting the findings of the studies conducted within this dissertation, it is important to keep some general strengths and limitations in mind. The present dissertation generally benefited from the use of a strong research design. Within a cluster randomized trial, a specifically developed intervention was implemented in an adequately large sample of classes, and an experimental design including a control group and pretest, posttest, and follow-up measurements was used to thoroughly test the effects of the intervention. Also, the data were analyzed using appropriate state-of-the-art statistical methodology that takes the multilevel nature of the data into account (Raudenbush & Bryk, 2002; Raudenbush, 1997).

Although the intervention was intentionally implemented in ninth grade math classrooms (see section 1.4), this specific study focus represents a limit to the generalizability of the findings. With regards to the characteristics of the sample, further research is needed to examine how the structure of value beliefs develops across age, whether gender differences in these beliefs depend on students' age and at what age utility value interventions can be successfully applied. Students' beliefs about the usefulness of a subject are assumed to develop at a later time point than more intrinsic types of value beliefs (Wigfield, 1994). Students' value beliefs might therefore be less differentiated at an earlier age and younger students might also benefit less from interventions referring to their future. The value facets that were assessed were chosen in line with the context of this study. Future research applying a similar facet approach to value beliefs might, therefore, need to adapt the specific facets and items in order to increase the fit to the characteristics of the specific sample (e.g., university students). As regards gender differences, previous longitudinal studies found gender differences in value beliefs to be rather stable over time (Frenzel, Goetz, Pekrun, & Watt, 2010; Watt, 2004). However, gender differences might still show different trajectories depending on the specific value aspect under consideration. The sample of the intervention study was also selective in terms of the school type: Only classes from academic track schools participated in the study. The intervention was specifically designed to fit to the future plans of students in this sample; for instance, examples for the usefulness of mathematics referred to specific university majors. Implementing such an intervention in other samples would, therefore, require changes in the specific utility information that is provided. More research is also needed to see how the effects of value interventions, the structure of value beliefs, and gender differences in these beliefs varies according to the school type.

The intervention focused intentionally on math because of its relevance for careers within and beyond STEM. A domain-specific approach to motivational beliefs in general and to motivational interventions in particular is in line with the state of the art. However, the focus on math comes with several limitations. First of all, it has been shown that an intervention targeting values in math had negative side effects on values in German. One approach to avoid such side effects would be to develop broader interventions that highlight the relevance of various school subjects for students' future lives (for a similar intervention strategy, see Woolley, Rose, Orthner, Akos, & Jones-Sanpei, 2013). It would, however, still be important to evaluate the effects of such an intervention on value beliefs in different subjects. With regards to the assessment of value beliefs in subjects other than math, short scales assessing general value were included. Whereas this consideration of different subjects is an improvement compared to previous intervention studies, assessing a broader range of subjects and more differentiated measures would be necessary to learn more about intraindividual hierarchies in value beliefs and how these are affected through interventions. The questionnaire that was developed to assess value beliefs in math should theoretically be applicable for other subjects as well, but this needs to be established first.

Lastly, this dissertation focused solely on self-reports as outcome of the intervention. Although self-reports are an adequate means to assess students' subjective beliefs in general and value beliefs in particular (Wigfield & Cambria, 2010), it should also be investigated how changes in students' beliefs translate into changes in more behavioral outcomes (see also section 5.2.1 for a further discussion of this point). Whereas intervention effects were still found until five months after the intervention, it would be necessary to follow students' development over a longer period of time to assess potential effects on later academic choices.

5.2. General Implications and Future Directions

The findings of the present dissertation have implications for future research as well as for practice. On the one hand, some questions arise from these findings that can lead to potential avenues for future research. On the other hand, several implications can be derived for educational practice and policy. Both types of implications will be discussed in the following.

5.2.1. Implications for future research

Implications for future research will be discussed referring to three different lines of research. First, the findings of the present dissertation provide support for the usefulness of taking a multidimensional perspective on value beliefs. However, the validity of multiple value aspects for predicting students' choices should be explored further in future research. Second, the present dissertation found that students' value beliefs can be promoted by interventions in the classroom. Future research is needed to explore how these interventions affect more distal outcomes, including behavioral measures and long-term choices. Third, the dissertation provided evidence for the effectiveness of value interventions in the classroom. However, more needs to be known about the mechanisms through which these interventions work.

Exploring the validity of value beliefs for choices

Students' value beliefs have been shown to be important predictors for achievement-related choices such as course and career choices, but also engagement in different activities (Wigfield et al., 2009). However, as previous research mostly did not include separate scales for different value aspects, there is only limited evidence on the differential validity of these value aspects for predicting students' choices (see also section 1.1.5). Given the results of this dissertation, which have shown a complex pattern of gender differences in value facets as well as differential intervention effects on these facets, this seems to be one important direction of future research. With respect to this, the high intercorrelations between value facets represent a methodological challenge. When including all of these facets simultaneously into one regression, multicollinearity becomes an issue (Cohen, Cohen, West, & Aiken, 2003). To avoid the problem of multicollinearity, it is possible to model the hierarchical structure of the underlying constructs in terms of mutually unrelated general and specific factors (see Brunner, Lüdtke, & Trautwein, 2008; Brunner, Nagy, & Wilhelm, 2012). These general and specific value constructs can then be associated with students' choices. However, it is more reasonable from a theoretical point of view that different value facets combine to influence students' choices instead of affecting them independently (see Wigfield & Cambria, 2010). The analytical techniques that are used to examine the validity of value beliefs need to adequately model these dynamics. One possible approach is using person-centered techniques that take the intraindividual combinations of different motivational beliefs into account (Bergman & El-Khouri, 2003). Such techniques (e.g., latent class analysis; Vermunt & Magidson, 2002) classify students into homogenous groups with similar profiles across the considered motivational dimensions. After establishing groups, differences in students' outcomes between these groups can be investigated. If groups show distinct patterns of motivational beliefs and differ in the outcomes considered, this analysis yields valuable information about the validity of these motivational beliefs.

More research is also warranted that examines the role of intraindividual hierarchies in value beliefs for students' choices. A broader range of subjects should be considered, which brings similar challenges for the statistical modeling. Previous studies (Chow et al., 2012; Chow & Salmela-Aro, 2011) applied latent class analysis to study the effect of intraindividual hierarchies in values on educational and career aspirations. These analyses are rather exploratory in that they pick up the naturally occurring patterns of motivational beliefs. Confirmatory factor analyses, on the other hand, allow modeling theoretically assumed structures of value beliefs in different subjects. To tease apart general and subject-specific elements of value beliefs, nested models that separately model these aspects might be valuable (for a similar approach to the structure of academic self-concepts, see Brunner et al., 2010). The statistical modeling can also represent complex structures such as a circumplex models, which have been used to represent vocational interests (Nagy, Trautwein, & Lüdtke, 2010). Similar analytical techniques could be used to represent the structure of value beliefs across several subjects in a parsimonious way.

Future research should also continue to examine how gender differences in choices can be explained by an adequate modeling of value beliefs. Considering multiple value facets as well as value beliefs in multiple subjects should help to better understand why females do not pursue STEM related careers as often as do males. Studies showing that intraindividual hierarchies in value beliefs across multiple subjects can explain gender differences in choices are a first step in this direction (Chow et al., 2012; Chow & Salmela-Aro, 2014).

Exploring the effects of value interventions on more distal outcomes

The empirical studies that were conducted within this dissertation further supported the effectiveness of value interventions in the classroom. As value beliefs have been shown to predict students' choices (Wigfield et al., 2009), changes in students' value beliefs should translate into changes in students' behavior and choices. Further research should, therefore, examine the effects of value interventions on more distal outcomes. To avoid effects of social desirability, these outcomes should not only rely on self-reports. To date, effects of value interventions in the field have been shown on course grades (Hulleman & Harackiewicz, 2009) as well as course choices in high school (Harackiewicz et al., 2012). In addition to achievement and course choices, important outcomes that should be considered include long-term career choices as well as short-term behavioral measures. A laboratory study provided first evidence that value interventions can foster behavioral effort (Shechter, Durik, Miyamoto, & Harackiewicz, 2011). Examining the effects of value interventions on students' behavior would also allow gaining a better understanding of the processes at play. An experience sampling approach (see Csikszentmihalyi & Larson, 1987) that assesses students' motivation and engagement in different situations could be used to examine intervention effects on students' behavior and the trajectories of these effects over time. The perspective of external observers (e.g., teachers or parents) could also provide valuable insight in how students' behavior is affected through value interventions and should be considered in addition to students' self-reports.

Exploring the mechanisms of value interventions

Further research is needed to understand the mechanisms through which value interventions work. Different strategies were shown to promote value beliefs to different extents, but it is not clear why these differential effects occurred. Investigating the implementation fidelity, in this case analyzing the quality of students' responses to different tasks, could be a way to understand why some tasks work better than others.

Possible indicators that can be taken into account are the quantity and the quality of connections made in these tasks as well as the number of tasks that were completed (for similar coding of students' essays, see Hulleman & Cordray, 2009; Hulleman et al., 2010). With regards to the quality of connections, the degree of personal identification and specificity seem to be especially important (Harackiewicz, Tibbetts, & Canning, 2014; Hulleman & Kosovich, 2014). Two sets of questions can be answered using data on students' responses to relevance-inducing tasks. First, how do different students respond to value interventions? To answer this question, students' characteristics can be associated with the quality of their responses. Students with high initial motivation will likely provide more elaborate responses on the usefulness of the learning material than students with low initial motivation. Analyses of students' responses might also explain differential intervention effects according to student gender, as it is possible that females tend to work more conscientiously on such writing assignments than males. Second, does the quality of students' responses have an effect over and above initial differences in motivation? To be able to draw causal inferences about the mechanisms at play, initial differences between students need to be considered in analyzing the effects of different responses on subsequent outcomes.

In addition to students' responsiveness to different tasks, more needs to be known about the specific psychological processes that are initiated by value interventions and the conditions under which these interventions work best. The findings of this dissertation suggest that interventions in the classroom can be an effective means for promoting students' value beliefs, but provide only limited insight into the processes involved. To achieve this, potential mediating variables such as competence valuation, task involvement, and perceived competence (cf., Shechter et al., 2011) need to be measured in adequate ways. Considering such process measures across tasks and situations could help in better understanding the effects of value interventions. With regards to the context in which value interventions are implemented, it is not yet clear which conditions need to be met for value interventions to be effective. Research based on self-determination theory suggests that utility information needs to be framed in terms of intrinsic goals and presented in an autonomy-supportive manner to be beneficial (Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004). Although the intervention that was implemented in the classroom context was relatively short, it combined several elements. Future research should therefore examine which intervention elements and strength are necessary for value interventions to be effective. The moderators of intervention effects also warrant further examination. Potential moderators that should be investigated include both student characteristics (e.g., expectancy) as well as characteristics of the context (e.g., teaching quality).

5.2.2. Implications for educational policy and practice

Although future research is needed to extend the findings of the present dissertation, a number of implications for educational policy and practice can be derived. Interventions that promote students' value beliefs are highly relevant for educational practice as value beliefs are important predictors for students' effort and choices in- and outside the school context (e.g., Durik, Vida, & Eccles, 2006; Trautwein & Lüdtke, 2007). The design of the intervention targeting students' value beliefs for mathematics specifically catered to the practical needs and challenges of classroom-based intervention studies. The intervention was relatively short (i.e., 90 minutes in school plus two additional short reinforcements) and standardized material was used to facilitate the implementation at a larger scale through various persons. It could, therefore, be easily implemented as part of a regular math curriculum. The empirical studies conducted within this dissertation have shown that the effects of this intervention, which was implemented in 82 classrooms by a total of five researchers, sustained for several months. Scaling up this intervention would be possible by training researchers that can be deployed to classrooms. The intervention could also be implemented by teachers. However, it would be important to test if teacher-implemented interventions have the same effect—especially given the fact that the intervention contained some psychoeducational elements.

Before scaling up the developed intervention, one should, however, carefully consider the general aims of motivational interventions given that such interventions can have negative side effects on non-targeted domains. If the aim is to recruit more young people for STEM-related careers, subject-specific motivational interventions in related domains can be an effective tool. If the aim is, however, to increase general achievement motivation, the effects of such subject-specific interventions can be considered as a zero-sum game. Given that students have limited time and energy, it might be better to "steal" from students' time for leisure activities instead of other school subjects. One way to achieve this could then be developing interventions that foster the perceived relevance of various school subjects or school in general. Using similar intervention techniques as developed to promote the relevance of mathematics would still be possible.

In the end, whereas there is still a great deal to be learned about value beliefs in general and value interventions in particular, the present dissertation has demonstrated that theoretically driven intervention studies can yield practically important results. The intervention that was implemented in the school setting was developed and evaluated based on theoretical principles and empirical research results. Only like this, the findings of interventions studies can further scientific understanding and educational practice at the same time (Pintrich, 2003).

References

Bergman, L. R., & El-Khouri, B. M. (2003). A person-oriented approach: Methods for today and methods for tomorrow. *New Directions for Child and Adolescent Development, 101*, 25–38. doi:10.1002/cd.80

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*, 964–981. doi:10.1037/a0019644

Brunner, M., Lüdtke, O., & Trautwein, U. (2008). The internal/external frame of reference model revisited: Incorporating general cognitive ability and general academic self-concept. *Multivariate Behavioral Research*, *43*, 137–172. doi:10.1080/00273170701836737

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846. doi:10.1111/j.1467-6494.2011.00749.x

Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology*, *48*, 1612–1628. doi:10.1037/a0030194

Chow, A., & Salmela-Aro, K. (2011). Task-values across subject domains: A gender comparison using a person-centered approach. *International Journal of Behavioral Development*, *35*, 202–209. doi:10.1177/0165025411398184

Chow, A., & Salmela-Aro, K. (2014). Does priority matter? Gendered patterns of subjective task values across school subject domains. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment* (pp. 247–265). Cambridge, UK: Cambridge University Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, *175*, 526–536. doi:10.1097/00005053-198709000-00004

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*, 104–118. doi:10.1007/s11031-014-9419-0

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98,* 382–393. doi:10.1037/0022-0663.98.2.382

Eccles, J. S. (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*, 78–89. doi:10.1080/00461520902832368

Eisenberg, N. (2006). Handbook of child psychology. Volume 3: Social, emotional, and personality development. Hoboken, NJ: Wiley.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*, 507–537. doi:10.1111/j.1532-7795.2010.00645.x

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics - A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *12*, 497–514. doi:10.1007/BF03173468

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, *23*, 899–906. doi:10.1177/0956797611435530

Harackiewicz, J. M., Tibbetts, Y., & Canning, E. (2014). Measuring perceived utility value and personal relevance in college biology courses. *Paper presented at the annual conference of the American Educational Research Association, Philadelphia, PA.*

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest , self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, *39*, 870–888. doi:10.1002/tea.10048

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110. doi:10.1080/19345740802539325

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*, 880–895. doi:10.1037/a0019506

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

Hulleman, C. S., & Kosovich, J. J. (2014). Fostering perceptions of relevance for schoolwork: A mixed-methods study. *Paper presented at the annual conference of the American Educational Research Association, Philadelphia, PA.*

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *Journal of Educational Psychology*, Advance online publication. doi:10.1037/edu0000021

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research, 23*, 129–149. doi:10.3102/00028312023001129

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67. doi:10.1037/0022-0663.96.1.56

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional Comparison Theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16–32. doi:10.1016/j.learninstruc.2014.08.005

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology, 44*, 351–373. doi:10.1016/j.jsp.2006.04.004

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*, 544–560. doi:10.1037/a0032459

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*, 1129–1167. doi:10.3102/0034654309337522

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, *12*, 323–345. doi:10.1080/13803610600765687

Nagy, G., Trautwein, U., & Lüdtke, O. (2010). The structure of vocational interests in Germany: Different methodologies, different conclusions. *Journal of Vocational Behavior*, *76*, 153–169. doi:10.1016/j.jvb.2007.07.002

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*, 315–341. doi:10.1007/s10648-006-9029-9

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*, 667–686. doi:10.1037/0022-0663.95.4.667

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*, 173–185. doi:10.1037/1082-989X.2.2.173

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student’s mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. doi:10.1016/j.learninstruc.2014.08.001

Shechter, O. G., Durik, A. M., Miyamoto, Y., & Harackiewicz, J. M. (2011). The role of utility value in achievement behavior: The importance of culture. *Personality and Social Psychology Bulletin*, *37*, 303–17. doi:10.1177/0146167210396380

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*, 432–444. doi:10.1037/0022-0663.99.2.432

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*, 246–260. doi:10.1037/0022-3514.87.2.246

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.

Wang, M.-T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*, 304–340. doi:10.1016/j.dr.2013.08.001

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*, 1556–1574. doi:10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., & Eccles, J. S. (Eds.). (2008). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences*. Washington, DC: American Psychological Association.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49–78. doi:10.1007/BF02209024

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*, 1–35. doi:10.1016/j.dr.2009.12.001

Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). New York, NY: Routledge.

Woolley, M. E., Rose, R. A., Orthner, D. K., Akos, P. T., & Jones-Sanpei, H. (2013). Advancing academic achievement through career relevance in the middle grades: A longitudinal evaluation of CareerStart. *American Educational Research Journal*, *50*, 1309–1335. doi:10.3102/0002831213488818