Please cite as:

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology, 51,* 1226–1240. doi:10.1037/dev0000028

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

Hanna Gaspard, Anna-Lena Dicke, Barbara Flunger, Brigitte Maria Brisson, Isabelle Häfner, Benjamin Nagengast, and Ulrich Trautwein

Hector Research Institute for Education Sciences and Psychology, University of Tübingen, Germany

Author Note

Anna-Lena Dicke is now affiliated with the University of California, Irvine.

This research was funded in part by German Research Foundation Grant TR 553/7-1 awarded to Ulrich Trautwein, Oliver Lüdtke, and Benjamin Nagengast. Hanna Gaspard, Brigitte M. Brisson, and Isabelle Häfner are members of the Cooperative Research Training Group of the University of Education, Ludwigsburg, and the University of Tübingen, which is supported by the Ministry of Science, Research and the Arts in Baden-Württemberg. Hanna Gaspard and Isabelle Häfner are also doctoral students of the LEAD Graduate School [GSC 1028], funded by the Excellence Initiative of the German federal and state governments. We thank Katharina Allgaier and Evelin Herbein for their help conducting this research.

Correspondence concerning this article should be addressed to Hanna Gaspard, University of Tübingen, Europastraße 6, 72072 Tübingen, Germany. Email: hanna.gaspard@uni-tuebingen.de

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 (β = .73) to detect intervention effects of δ = 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 ≥ .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, 2015). 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: χ^2 = 2098.03, df = 574, CFI = .957, TLI = .950, RMSEA = .038; T2: χ^2 = 1818.56, df = 574, CFI = .964, TLI = .958, RMSEA = .035; T3: χ^2 = 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: χ^2 = 2789.13, df = 614, CFI = .939, TLI = .933, RMSEA = .044 for T1; T2: χ^2 = 2645.43, df = 614, CFI = .942, TLI = .937, RMSEA = .043; T3: χ^2 = 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).

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 post-test 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

---

[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.

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 (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).

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.

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.

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.

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*(4), 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*(1), 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*(1), 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*(2), 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*(3), 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*(6), 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*(4), 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*(1), 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*(3), 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*(3), 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*(2), 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*(3), 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*(1), 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*(2), 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*(4), 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*(4), 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*(6), 641-651. doi: 10.1177/1745691614551642

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Jornal of Counseling Psychology, 28*(6), 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*(1), 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*(8), 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)*. 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*(2), 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
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*(1), 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*(4), 880-895. doi: 10.1037/a0019506
Hulleman, C. S. & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(5958), 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*(2), 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*(4), 1003-1013. doi: 10.3758/s13428-011-0094-8
Markus, H. & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9), 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*(6), 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*(2), 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*(9), 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*(1), 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*(15), 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*(8), 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*(3), 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*(4), 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*(7), 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*(1), 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
Plewis, I. & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation, 4*(1), 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*(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*(3), 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*(1), 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*(2), 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*(2), 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*(3), 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*. 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*(5), 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*. 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), 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*(3), 283-292. doi: 10.1177/0361684313482109

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.

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

		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).

Table 5

Intervention Effects on Value Components at Posttest and Follow-up

Variable	Utility value			Attainment value			Intrinsic value			Cost		
	Est.		(SE)	Est.		(SE)	Est.		(SE)	Est.		(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 Est.		Utility for school (SE)	Utility for daily life Est.		Utility for daily life (SE)	Social utility Est.		Social utility (SE)	Utility for job Est.		Utility for job (SE)	General utility Est.		General utility (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.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

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

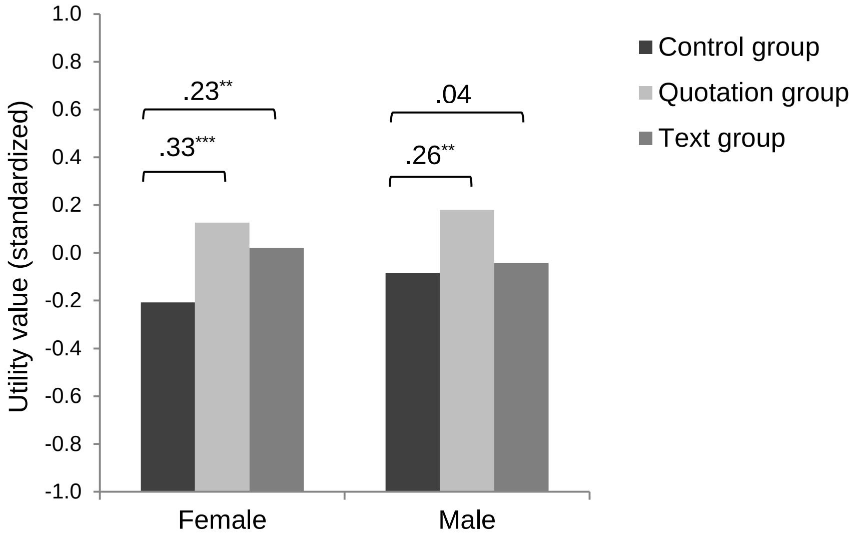

Figure 1: Adjusted means for utility value at posttest by gender and intervention group. Effect sizes for the quotations condition and the text condition as compared to the control condition are displayed separately for females and males.

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

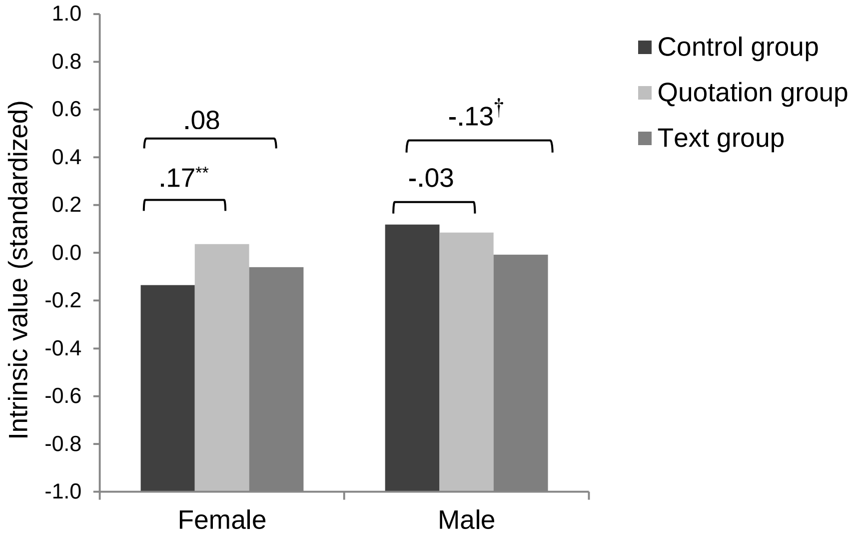

Figure 2. Adjusted means for intrinsic value at posttest by gender and intervention group. Effect sizes for the quotations condition and the text condition as compared to the control condition are displayed separately for females and males.

** $p < 0.01$; † $p < 0.10$

Supplement with Results for Additional Analyses

Table 1

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 2

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

	Imp. of achievement			Personal importance			Effort required			Emotional cost			Opportunity cost		
Variable	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 3

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

	Utility for school		Utility for daily life		Social utility		Utility for job		General utility	
Variable	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. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.

Table 4

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