

Who sticks to the instructions—and does it matter? Antecedents and effects of students' responsiveness to a classroom-based motivation intervention

**Brigitte Maria Brisson · Chris S. Hulleman · Isabelle Häfner ·
Hanna Gaspard · Barbara Flunger · Anna-Lena Dicke ·
Ulrich Trautwein · Benjamin Nagengast**

Published online: 10 March 2020
© The Author(s) 2020

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11618-019-00922-z>) contains supplementary material, which is available to authorized users.

Dr. B. M. Brisson (✉)

DIPF | Leibniz Institute for Research and Information in Education, Rostocker
Straße 6, 60323 Frankfurt am Main, Germany
E-Mail: brigitte.brisson@dipf.de

Prof. Dr. C. S. Hulleman

Curry School of Education and Human Development, CASTL Piedmont Center, University of
Virginia, 2405 Ivy Rd, Charlottesville, VA 22903, USA
E-Mail: chris.hulleman@virginia.edu

Dr. I. Häfner · Dr. H. Gaspard

Hector Research Institute of Education Sciences and Psychology, Universität Tübingen,
Walter-Simon-Straße 12, 72072 Tübingen, Germany
E-Mail: hanna.gaspard@uni-tuebingen.de

Dr. B. Flunger

Universität Utrecht, Martinus J. Langeveldgebouw, Heidelberglaan 1, Kamer E 329, 3584 CS Utrecht,
The Netherlands
E-Mail: b.flunger@uu.nl

Dr. A.-L. Dicke

Irvine School of Education, University of California, 401 E. Peltason Drive,
Suite 3200, Irvine, CA 92617, USA
E-Mail: adicke@uci.edu

Prof. Dr. U. Trautwein · Prof. Dr. B. Nagengast

Hector Research Institute of Education Sciences and Psychology, Universität Tübingen,
Europastraße 6, 72072 Tübingen, Germany

Prof. Dr. U. Trautwein

E-Mail: ulrich.trautwein@uni-tuebingen.de

Prof. Dr. B. Nagengast

E-Mail: benjamin.nagengast@uni-tuebingen.de

Abstract Why do some students benefit from interventions and others do not? By investigating the antecedents and effects of students' responsiveness to a classroom-based motivation intervention, the current study aims to shed light on the intervention processes that make educational interventions in real-life settings work. Using data from a cluster-randomized controlled experiment with 1916 ninth-grade students, students' responsiveness to two written intervention activities about the personal relevance of mathematics (evaluating quotations or writing a text) was assessed. Based on the hypothesized theory of change, 1280 student essays were coded on three indicators of responsiveness (positive arguments, personal connections, in-depth reflections) which were combined into a continuous index. Linear regression analyses showed that students' conscientiousness, gender, math-related motivation, and achievement predicted the responsiveness index. This research highlights the importance of investigating intervention processes in order to optimize the theories and designs of classroom interventions

Keywords Classroom intervention · Motivation · Mathematics · Relevance intervention · Student responsiveness · Utility-value intervention

Wer hält sich an die Anweisungen – und macht es einen Unterschied? Determinanten und Wirkungen der Reaktion von Schüler*innen auf eine Motivationsintervention im Klassenzimmer

Zusammenfassung Warum profitieren einige Schüler*innen von Interventionen und andere nicht? Die vorliegende Studie untersucht die Determinanten und Wirkungen der Reaktionen von Schüler*innen auf eine Motivationsintervention im Klassenzimmer. Ziel ist es, die Prozesse zu beleuchten, mit denen psychologische Interventionen in der Praxis funktionieren. Anhand von Daten aus einem randomisiert-kontrollierten Experiment zur persönlichen Relevanz der Mathematik mit 1916 Neuntklässler*innen wurde erfasst, wie gut sich die Schüler*innen an die Anweisungen von zwei schriftlichen Interventionsaufgaben (Bewertung von Zitaten oder Verfassen eines Textes) gehalten hatten. Basierend auf der angenommenen Wirkungsweise der Intervention wurden 1280 Schüleraufsätze auf die Bearbeitungsqualität hin kodiert. Die drei Qualitätsindikatoren (positive Argumente, persönlicher Bezug, Reflexionstiefe) wurden zu einem kontinuierlichen Index zusammengefasst. Lineare Regressionsanalysen zeigten, dass die Gewissenhaftigkeit, das Geschlecht, die Motivation in Mathematik und die Schülerleistung den Qualitätsindex vorhersagten. Diese Studie zeigt, wie wichtig es ist, Interventionsprozesse zu untersuchen, um Theorien und Designs von Interventionen im Klassenzimmer zu optimieren.

Schlüsselwörter Intervention · Klassenzimmer · Mathematik · Motivation · Nützlichkeitsintervention · Relevanzintervention · Schülerreaktion

1 Introduction

Experimental studies in real-life classroom settings are valuable to advance psychological theories in education and to find ways to improve instructional practice (e.g., Shadish et al. 2002). However, researchers or teachers may deliver classroom-based interventions as intended, but the students may not do what they are expected to do. For example, in order to trigger changes in students’ personal beliefs and thereby affect students’ academic behavior or achievement, individual writing assignments for students (e.g., Yeager and Walton 2011) are a common intervention tool. Yet, if students do not stick to the instructions of such intervention tasks, the processes initiating the targeted change in students’ beliefs, also called intervention processes (Murrah et al. 2017), may not unfold. Assessing the extent to which students complete the intervention activities as intended (*student responsiveness*, Dane and Schneider 1998), and analyzing its effect on the target psychological process is therefore essential to unravel how educational interventions work (e.g., Nelson et al. 2012)—and why they sometimes do not work (e.g., Husman et al. 2017; Karabenick et al. 2017). In other words, studying student responsiveness helps to draw the desired theoretical and practical implications from educational experiments.

Having students write about the personal relevance of a task or subject is an effective intervention strategy to improve students’ academic motivation and behavior (for reviews, see Durik et al. 2015; Rosenzweig and Wigfield 2016). Such so-called relevance interventions were found to foster students’ utility value beliefs as the most proximal outcome, which in turn also boosted students’ interest, effort, and test scores (Hulleman et al. 2010). The intervention processes that instigate the target psychological processes of relevance interventions yet remain unclear: How do students have to complete the writing activity so that changes in their utility value beliefs will be triggered? To tackle this question, students’ responsiveness to the intervention activities needs to be systematically defined, assessed, and related to the target psychological process, utility value beliefs. In addition, identifying the features of highly and lowly responsive students or classrooms, respectively, helps to unravel if certain students stick less to the instructions than others and to optimize the intervention materials accordingly. The current study addresses these gaps in research by using data from the intervention study “Motivation in Mathematics” (MoMa), which was shown to foster students’ value and competence beliefs, effort, and achievement in mathematics—particularly for females (Gaspard et al. 2015a; Brisson et al. 2017). By analyzing students’ essays on the relevance of mathematics written during the classroom intervention, the current study aims at: (a) providing a theory of change to assess and combine the core elements of students’ responsiveness to the intervention activities, and (b) exploring individual student characteristics and classroom perceptions as predictors of student responsiveness.

2 Theoretical Background

2.1 Theoretical Background and Effectiveness of Relevance Interventions

Based in expectancy-value theory (Eccles et al. 1983; Eccles and Wigfield 2002), relevance interventions aim at improving academic outcomes by raising students' beliefs about the utility value of an academic task or subject, that is its perceived usefulness to a student's current or future goals. Numerous correlational studies have shown that students who perceive the learning contents as highly useful feature high levels of interest and attainment values, self-efficacy beliefs and ability perceptions, as well as effort (for overviews, see e.g., Roeser et al. 2000; Wigfield et al. 2017; for correlations in the current sample, see Brisson et al. 2017; Gaspard et al. 2015a). Accordingly, fostering students' utility value beliefs in targeted interventions may lead to positive effects on other important academic outcomes.

Like other social-psychological interventions in education, relevance interventions typically rely on the assumption that a change in students' personal beliefs can be caused through individual writing exercises (cf., Yeager and Walton 2011). Such writing tasks are supposed to trigger students' reflection about relevance and to enable the personalization of relevance-related messages provided during the intervention (Lazowski and Hulleman 2016). They may require students, for example, to write an essay about the personal relevance of a self-chosen topic from their science class (e.g., Hulleman and Harackiewicz 2009).

The effectiveness of relevance interventions is typically estimated in comparison to a control group of students who either performed an unrelated writing task (e.g., Hulleman and Harackiewicz 2009) or who did not do any assignment. Overall, relevance interventions using writing exercises have been shown to promote students' utility value beliefs as its most proximal outcome as well as more distal outcomes including students' competence beliefs, interest, effort, and achievement in subjects like mathematics, physics, biology, and psychology (Hulleman and Harackiewicz 2009; Hulleman et al. 2010, study 2, 2017, study 2; Gaspard et al. 2015a; Harackiewicz et al. 2016; Brisson et al. 2017). The MoMa relevance interventions, which are the focus of the present study, fostered students' utility value beliefs as well as students' intrinsic and attainment value beliefs, academic self-concept, homework-related self-efficacy, teacher-rated effort, and achievement in mathematics until up to five months after the intervention (Gaspard et al. 2015a; Brisson et al. 2017). The quotations condition, in which students answered questions on the personal relevance of interview statements about mathematical skills, had stronger effects on all outcomes than the conventional text condition. Furthermore, the text condition promoted girls' value beliefs more than boys' (Gaspard et al. 2015a).

Prior studies about the processes leading to intervention effects have shown that utility value beliefs mediate intervention effects on more distal outcomes, for example interest and achievement (Hulleman et al. 2010). Yet, not much is known about the processes through which a change in students' beliefs about the utility of a task or subject is triggered. Experimental studies have revealed links between the length and contents of students' essays written during the intervention and intervention effects on students' utility value beliefs (Hulleman and Cordray 2009;

Harackiewicz et al. 2016; Hulleman et al. 2017). However, to identify the processes that contribute to the effects of classroom-based relevance interventions, the role of students' intervention responsiveness in manipulating students' utility value beliefs must be studied more systematically (cf., Nelson et al. 2012).

2.2 Student Responsiveness in Relevance Interventions

2.2.1 Intervention and psychological processes

Sometimes classified as one aspect of intervention fidelity (i.e., the extent to which an intervention is implemented as designed), student responsiveness describes the extent to which students are engaged in the intervention activities (Dane and Schneider 1998). Such responses from the participants can be classified as *intervention processes*. Intervention processes are assumed to induce *psychological processes*, the most proximal outcomes of an intervention, which in turn instigate more distal outcomes (Murrah et al. 2017). This chain of processes leading to intervention effects is also called an intervention *theory of change* (or *change model*; Nelson et al. 2012). An increase in utility value beliefs represents the target psychological process in relevance interventions, which initiates further intervention effects: These value beliefs lead students to become more interested and engaged in learning, and attain better learning outcomes (cf., study by Hulleman et al. 2010).

2.2.2 Core elements of student responsiveness

Asking students to write about the relevance of a learning matter is typically the most basic instruction of individual tasks used in relevance interventions (cf., Durik et al. 2015). *Using relevance arguments* is thus assumed to be the first and most important core element students have to respond to so that the writing task will have a positive effect on their value beliefs. In addition, as utility value beliefs are high if students consider academic tasks or subjects useful for personal, not impersonal, goals (Eccles et al. 1983), students are instructed to connect the relevance arguments to their personal lives. *Making personal connections* thus constitutes the second core element of students' responsiveness to the writing activity (see also Hulleman et al. 2017). Furthermore, relevance interventions are assumed to be effective when the students are strongly cognitively engaged in the writing activity (Harackiewicz et al. 2016). Accordingly, students who reformulate relevance information provided during an intervention in their own words or who transfer it to personally important contexts in their writings may learn about relevance in a more sustained way than students who merely reproduce previously encountered relevance arguments without much reflection. *Using in-depth reflections* may thus represent the third core element of student responsiveness.

2.2.3 Empirical studies on student responsiveness

Hulleman and Cordray (2009) provided some initial evidence of the importance of students' responsiveness to classroom-based relevance intervention tasks. They

investigated why a relevance intervention during which students wrote about the personal relevance of science topics was less effective in high-school classrooms than a similar intervention in the laboratory. Analyses on the content of students' essays produced during the intervention showed that students in the classroom failed to make high quality, personal connections to the learning material in their essays, resulting in decreased relative intervention strength compared to the laboratory experiment.

In a second study, Harackiewicz et al. (2016) examined students' responsiveness within an online relevance intervention at university. For first-generation underrepresented minority (FG-URM) students, course achievement improved after writing about the personal relevance of concepts from their biology class. As FG-URM students wrote longer essays and used more words indicative of social processes and of cognitive involvement than students not belonging to this group, the authors concluded that these aspects contributed to the success of the intervention for FG-URM students.

Going beyond descriptive analyses, Nagengast et al. (2018) used the measure of student responsiveness developed in the current paper to compare the effects of the MoMa interventions obtained from complier-average causal effects (CACE) models, which take into account student responsiveness (e.g., Sagarin et al. 2014), with those obtained from intent-to-treat (ITT) analyses, which are only based on students' group assignment (i.e., experimental vs. control group; e.g., Boruch 1997). Using various outcomes like students' math-related motivational beliefs and achievement, the authors not only found the estimates obtained from CACE models to be greater than those calculated with ITT analyses but also detected further effects when looking at responsive and nonresponsive students separately. Interestingly, the CACE estimates differed more from the ITT estimates in the text condition than in the quotations condition, hinting at a higher importance of the responsiveness measure in the text condition than in the quotations condition. For example, the text condition fostered students' math-related utility value beliefs only when students were highly responsive to the intervention whereas in the quotations condition positive effects on utility value were observed for both responsive and nonresponsive students. Students' responsiveness was also found to partially explain differential effects favoring girls over boys (cf., Gaspard et al. 2015a).

These initial studies notwithstanding, further research based on a theory of change (see Nelson et al. 2012) that guides the assessment of core elements of students' responsiveness to relevance interventions is needed. In addition, it is unclear which student and classroom characteristics drive responsiveness to instructions in relevance interventions. We will address these questions using data from the MoMa intervention study (Gaspard et al. 2015a; Brisson et al. 2017).

2.3 Potential Predictors of Student Responsiveness

2.3.1 Stable student characteristics and domain-specific motivation

Several research studies found secondary school students' cognitive abilities, conscientiousness, and domain-specific motivation (e.g., self-concept, homework-related

self-efficacy, and value beliefs related to mathematics) to be positively associated with homework compliance in diverse school subjects (e.g., Trautwein et al. 2006; Trautwein and Lüdtke 2009). Because the writing tasks of the MoMa interventions were completed in a similar manner as typical homework tasks—which are guided but not necessarily controlled by the teacher (e.g., Cooper 1989)—these variables might also predict students' responsiveness to written intervention activities. In addition, as girls seem to be particularly compliant with homework in language-related subjects (Trautwein et al. 2009), which often include coherent text writing based on reasoning, girls might also respond better to intervention activities that resemble such tasks (like the text condition in MoMa) than boys.

2.3.2 *Academic achievement*

Numerous studies have found high positive correlations between students' academic achievement and their task engagement (for reviews, see e.g., Reschly and Christenson 2012; Fredricks et al. 2004). Although achievement is sometimes considered an outcome of high task engagement, the relationship is probably reciprocal: Students tend to engage in subjects they are good at (e.g., Eccles and Wigfield, 2002; Finn and Zimmer 2012). This is why particularly high achievers might get involved in writing activities completed during relevance interventions and thus feature high levels of responsiveness. Accordingly, Harackiewicz et al. (2016) found positive links between students' achievement and the number of words and of personal connections in their relevance essays.

2.3.3 *Classroom perceptions*

One important influence on students' academic engagement in the classroom are peers—especially in teenage years (for reviews, see e.g., Fredricks et al. 2004; Juononen et al. 2012). More precisely, perceived classmates' math-related value beliefs have been found to positively correlate with students' interest, value beliefs, and positive emotions in math class (e.g., Frenzel et al. 2007, 2010; Schreier et al. 2014). Furthermore, a good classroom structure as indicated by a high disciplinary climate is known to positively correlate with time on task and task involvement (for reviews, see e.g., Fredricks et al. 2004; Reschly and Christenson 2012). If students perceive their classmates to highly value a subject and the disciplinary climate to be high—in short, the classroom atmosphere to be good—they might also be ready to work on in-class intervention tasks in a concentrated and thorough way.

2.4 **Aims of the Current Study**

The current study presents a theoretical framework to assess students' responsiveness to written tasks completed during an intervention about the relevance of mathematics (evaluating quotations or writing a text; MoMa study). By investigating the antecedents of student responsiveness and taking into account results from causal analyses based on the current responsiveness measure (Nagengast et al. 2018) further insights into the mechanisms contributing to differences in the effectiveness of

the two MoMa intervention approaches are to be gained. This research can serve as a blueprint for in-depth investigations on the role of students' intervention responsiveness for the effectiveness of classroom experiments (e.g., Shnabel et al. 2013).

Based on the assumed theory of change (cf., Nelson et al. 2012; see Methods), students' essays written during the 90-minute relevance intervention in the classroom were coded on three core elements of students' responsiveness: relevance arguments, personal connections, and in-depth reflections, which were combined into an index. In contrast to prior research on relevance interventions (Hulleman and Cordray 2009; Harackiewicz et al. 2016; Hulleman et al. 2017), the assessment of students' responsiveness in the current study was guided by a previously determined theory of change.

Two research questions were addressed. First, how did students comply with the core elements of the writing tasks completed during the MoMa relevance interventions? Second, which individual student characteristics and classroom perceptions predicted students' responsiveness to the writing tasks? We expected students' gender, cognitive ability, conscientiousness, math-related achievement and motivation (self-concept, homework-related self-efficacy, intrinsic value, utility value), and classroom perceptions (classmates' math-related value beliefs, disruptions in math class) to be associated with students' intervention responsiveness in both conditions. However, as the writing tasks of the MoMa interventions consisted of either writing short comments on quotations or producing an essay-like coherent text, the two conditions might appeal to different types of students. For example, girls might respond better to the text condition than boys but not to the quotations condition.

3 Method

3.1 Sample and Procedure

Data were collected as part of the cluster-randomized field experiment "Motivation in Mathematics" (MoMa) with 82 ninth-grade classes in 25 German academic track schools ("Gymnasium"). A total of 1978 students with active parental consent participated in the study (participation rate: 96.0%). Sixty-two students absent during the intervention were excluded from the current study, yielding a total sample of 1916 students ($M_{age} = 14.62$, $SD = 0.47$; 53.5% female).

The design of the MoMa intervention study is presented in Fig. 1. In the beginning of the study, teachers and their classes were randomly assigned within each school to either one of two intervention conditions, "quotations" (25 classes, 561 students; 52.8% female) or "text" (30 classes, 720 students; 52.4% female), or to the waiting

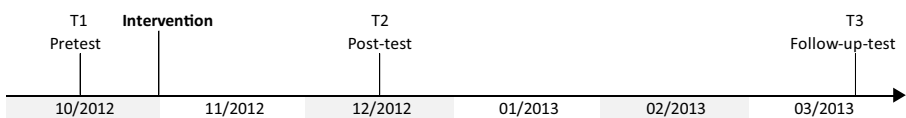


Fig. 1 Design of the MoMa intervention study

control group (27 classes, 635 students; 55.6% female). Afterwards, students took part in three data collections from autumn 2012 to spring 2013: Students in the experimental conditions completed questionnaires before the intervention as well as six weeks and five months after the intervention (see Fig. 1). Students in the waiting control group completed the same questionnaires at the same time points but did not receive any intervention before the last data collection. All 82 classes fully completed all waves of data collections.

Data on students' responsiveness were obtained by coding a total of 1280 essays produced during the interventions in class. Data on students' gender and math achievement at the beginning of the school year were provided by the teachers. Students' cognitive abilities, conscientiousness, math-related motivation, and classroom perceptions were measured at the pretest.

3.2 The MoMa Relevance Interventions

After the first data collection, students in both experimental groups took part in a 90-minute standardized intervention in the classroom about the relevance of mathematics led by trained researchers. The interventions started with a psychoeducational presentation, of which the first part served to reinforce students' competence beliefs. In the second part, students were provided with various examples of the utility of mathematics for future education, career opportunities, and leisure time activities. Right after the presentation, students completed an individual writing assignment differing by condition. Based on theories postulating that students can learn from persons they identify with (e.g., Bandura 1977; Markus and Nurius 1986; Oyserman and Destin 2010), students in the quotations condition were asked to provide short answers to several questions about the personal relevance of six interview statements from young adults describing everyday situations in which they needed mathematics. Students in the text condition were asked to collect arguments for the personal relevance of mathematics to their current and future lives and to then write a coherent text on their notes (see Online Supplement, Part A).

Researchers collected students' handwritten essays and recorded the actual procedure in the minutes at the end of the intervention. Overall, the implementation of the interventions seemed highly standardized: All students in the experimental groups had the occasion to complete the writing assignment during the in-class intervention session. Deviations from the standard procedure occurred in only two out of 55 classes (both in the text condition): In one class, the initial presentation had to be held without any projector due to technical problems; in the other class, students did not work quietly on their individual writing tasks.

Students in classes in the waiting control condition did not watch any presentation or complete any individual writing tasks. However, they received the on average more powerful intervention, namely evaluating quotations (e.g., Brisson et al. 2017) after the last measurement point.

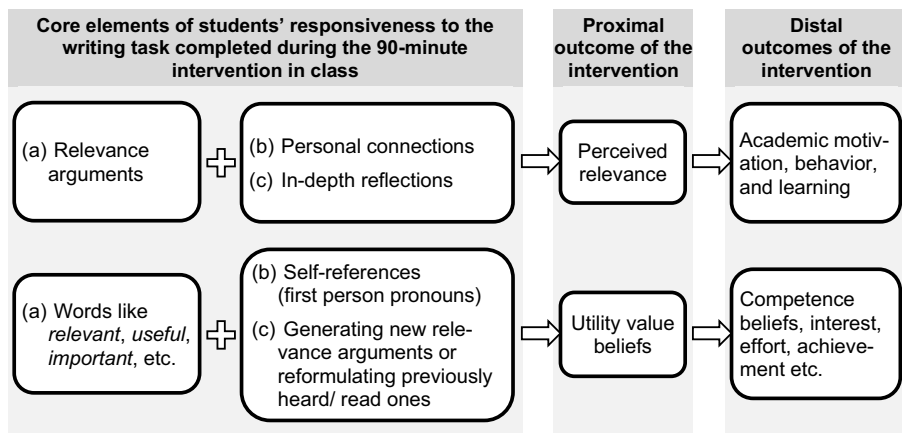


Fig. 2 Theory of change (*above*) and operational model (*below*) depicting the hypothesized processes underlying the effectiveness of the MoMa relevance interventions

3.3 Assessing Student Responsiveness

To produce the responsiveness data, we followed a five-step procedure suggested by Nelson et al. (2012). We first defined the core elements of the *theory of change* and then set up an *operational model* which specifies how these elements are operationalized in the intervention activities—in the current study, in the contents of students' essays (step 1). Responsiveness data were then produced by coding the extent to which the elements of the operational model had been executed in the essays as intended (step 2). After determining the reliability and validity of the responsiveness data (step 3), various indicators were combined into an index which took into account their assumed theoretical importance in affecting the intervention outcome (step 4). Finally, the index was analyzed by relating it to the target psychological process, students' utility values beliefs (Nagengast et al. 2018), as well as to predictors (step 5). The first four steps are described in more detail in the following.

3.3.1 Step 1: intervention models

The theory of change and the operational model of the MoMa interventions are presented in Fig. 2. As shown in the theory of change, we specified the intervention processes as (a) describing arguments about the usefulness of mathematics (*relevance arguments*), which (b) relate to the individual (*personal connections*) and which (c) are not reproductions of previously presented arguments (*in-depth reflections*). Writing about the usefulness of the learning matter—the essential element of relevance interventions (cf., Durik et al. 2015)—was considered a prerequisite for initiating the desired change and was thus considered more important than the other two components in the theory of change.

The operational model (see Fig. 2) was developed assuming that the intervention components were realized in the use of certain key words or types of words in students' essays (see e.g., Pennebaker et al. 2003; Harackiewicz et al. 2016). Relevance

Table 1 Coding Values and Reliabilities of the Indicators of Students’ Responsiveness to the Intervention Tasks

Indicator	Value			κ_{mean}	
	1	2	3	Quot	Text
Relevance arguments	Only negative arguments (against the utility of math)	$\leq 50\%$ of all arguments are positive (for the utility of math)	$> 50\%$ of all arguments are positive (for the utility of math)	.66	.81
Personal connections	Only other-references and/or impersonal references, e.g., <i>they, one, he, his, it, anyone</i>	$< 50\%$ of all personal references are self-references, e.g., <i>I, me, my</i>	$\geq 50\%$ of all personal references are self-references, e.g., <i>I, me, my</i>	.53	.81
In-depth reflections	Only arguments from the presentation and/or the quotes ^a	$< 50\%$ of all arguments were new, reformulated or transferred (i.e., not directly copied from the presentation/quotes ^a)	$\geq 50\%$ of all arguments were new, reformulated or transferred (i.e., not directly copied from the presentation/quotes ^a)	.64	.71

^ain the quotations condition only

κ_{mean} weighted Cohen’s kappa coefficient (Cohen 1968), *Quot* quotations condition, *Text* text condition

arguments are reflected in words such as “useful”, “relevant”, “important”, and the like. Personal connections materialize in the use of self-references (i.e., first person pronouns, Tausczik and Pennebaker 2010). A high degree of reflection is represented by describing relevance arguments that go beyond the ones presented in the initial part of the intervention. In line with the theory of change, we assumed that only if students use words indicative of the usefulness of mathematics during the intervention activity, the intervention can trigger an increase in the target psychological process, students’ math-related utility value beliefs.

3.3.2 Steps 2 and 3: coding values, coding procedure, and reliability measures

As indicated in the intervention models, students participating in the MoMa study were supposed to adhere to the three indicators of responsiveness, which are presented in Table 1 along with their coding values and reliabilities. All indicators were coded with the values 1 (*low responsiveness*), 2 (*medium responsiveness*), and 3 (*high responsiveness*). The coding values reflect proportions of (a) positive vs. negative relevance arguments for *relevance arguments*, (b) self-references vs. other-references for *personal connections*, and (c) relevance arguments which have been newly generated or reformulated from the intervention material vs. arguments which have been reproduced from the intervention material for *in-depth reflections* (see Table 1).

Six trained students coded the essays on the indicators of responsiveness using a coding manual (for examples of coded essays, see Online Supplement, Part A). At first, each coder independently coded a randomly chosen set of 10 essays per condition. Inter-coder agreement was determined by calculating weighted Cohen’s

kappa, which is applicable to ratings using ordinal categories and measures the proportion of weighted agreement corrected by chance (Cohen 1968). Mean weighted Cohen's κ was moderate to almost perfect depending on the coding category and condition (Landis and Koch 1977). The coders discussed remaining inconsistencies and agreed on one common value for each essay and coding category. Subsequently, the rest of the essays except a random set of 20 essays per condition was distributed randomly within condition among the coders and coded only once. Four of the coders each coded 244 essays individually and two of the coders each coded 122 essays individually. After half of the individual codings, the randomly chosen 40 essays were coded by all of the coders independently. Intercoder agreements calculated from the second set of multiple-coded essays were substantial to almost perfect for all categories and conditions (see Table 1)—excepting “personal connections” in the quotations condition, for which agreement remained moderate (Landis and Koch 1977). Finally, the second half of the randomly distributed essays was coded by the coders individually.

3.3.3 Step 4: combining the indicators of responsiveness into one index

The three indicators were combined into an index with a scale ranging from 1 (lowest responsiveness) to 11 (highest responsiveness). Guided by the theory of change which considers writing about the usefulness of mathematics a prerequisite for the intervention to have positive effects (see Fig. 2), students' score on the index was most dependent on their scores on “relevance arguments”. Students with the lowest score on “relevance arguments”, who had written nonsense or about the uselessness of mathematics, received the lowest value on the responsiveness index (the value of 1); their scores on the other two indicators, “personal connections” and “in-depth reflections”, were neglected. Students with a medium score on “relevance arguments”, who had partly argued for the utility of mathematics, received a medium value on the index between 2 and 6, depending on their scores on “personal connections” and “in-depth reflections”: the higher their sum, the higher their score on the index. Similarly, students with the highest score on “relevance arguments”, who had mainly argued for the utility of mathematics, received a high value on the index between 7 and 11, depending on their scores on “personal connections” and “in-depth reflections” (see also Fig. 4 in the Results).

3.4 Assessing Potential Predictors of Student Responsiveness

3.4.1 Stable student characteristics

Information on students' gender (0 = female, 1 = male) was provided by the teachers. Students' cognitive ability scores were obtained from a figural cognitive ability test (Heller and Perleth 2000) with 25 items (Cronbach's $\alpha = 0.79$). Students' conscientiousness was assessed with a German version of the NEO-FFI (Borkenau and Ostendorf 1991) in a questionnaire with a 4-point-Likert type scale ranging from 1 (*totally disagree*) to 4 (*totally agree*). The scale consisted of eleven items (e.g., “I am a productive person who always gets the job done.”, $\alpha = 0.80$).

3.4.2 Math achievement

Teachers provided students’ results from a curriculum-based standardized math test in the state of Baden-Württemberg taken at the beginning of Grade 9.

3.4.3 Initial math-related motivation

Students’ math-related competence beliefs were assessed with two scales that were adapted from previous studies (Trautwein and Köller 2003; Schwanzer et al. 2005). Math-related self-concept was measured with five items (e.g., “I am good at math.”, $\alpha=0.93$). Homework-related self-efficacy in mathematics was measured with four items (e.g., “When I try hard, I can solve my math homework correctly.”, $\alpha=0.76$). Students’ math-related intrinsic value beliefs (four items, e.g., “I like doing math.”, $\alpha=0.93$) and math-related utility value beliefs (twelve items, e.g., “I will often need math in my life.”, $\alpha=0.84$) were measured using a newly developed value instrument by Gaspard et al. (2015b). All items were answered on a four-point Likert type scale ranging from 1 (*totally disagree*) to 4 (*totally agree*).

3.4.4 Classroom perceptions

The scale measuring students’ perception of classmates’ math-related value beliefs consisted of five items (e.g., “Most students in my class consider math an important subject.”, $\alpha=0.75$). Students’ perceived disruptions in math class scale contained three items (e.g., “Our math lessons are often disrupted.”, $\alpha=0.88$). The scales were taken or adapted from previous studies (e.g., Baumert et al. 2009). All items were answered on a four-point Likert type scale ranging from 1 (*totally disagree*) to 4 (*totally agree*).

Table 2 Numbers, Means, Standard Deviations, and Intraclass Correlation Coefficients of All Variables under Investigation

	Quotations				Text			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>ICC</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>ICC</i>
<i>Intervention responsiveness</i>								
Responsiveness index	544	8.25	2.27	.06	712	8.58	2.43	.07
<i>Predictors</i>								
Cognitive ability	519	19.96	4.00	.04	681	19.99	4.22	.05
Conscientiousness	518	2.90	0.44	.05	682	2.91	0.43	.02
Math test score	517	48.67	16.49	.08	676	49.85	18.19	.21
Math self-concept	515	2.76	0.79	.03	678	2.74	0.81	.04
Math homework self-efficacy	427	2.80	0.62	.03	599	2.72	0.62	.05
Math intrinsic value	515	2.31	0.84	.04	675	2.29	0.86	.09
Math utility value	517	2.56	0.49	.05	680	2.52	0.47	.07
Classmates’ math valuing	505	1.98	0.57	.24	669	1.96	0.57	.25
Disruptions in math class	486	2.33	0.73	.29	652	2.35	0.77	.40

N number, *M* mean, *SD* standard deviation, *ICC* intraclass correlation coefficient relating to the amount of variance explained by differences between classes

3.5 Statistical Analyses

3.5.1 Descriptive statistics

The means, standard deviations, and intraclass correlation coefficients for the responsiveness index, students' individual characteristics and classroom perceptions are presented per condition in Table 2. The intercorrelations between these variables are accessible in the Online Supplement (Part B).

3.5.2 Regression analyses

The association of students' individual characteristics and classroom perceptions with students' intervention responsiveness was analyzed by running regression models in *Mplus 7* (Muthén and Muthén 1998-2012). For each intervention condition, the responsiveness index was regressed on individual student characteristics and classroom perceptions as predictors simultaneously to compare their predictive strength. Standard errors were corrected to account for the nesting of students within classes by using design-based correction of standard errors and test statistics (see McNeish et al. 2017, for a justification of this approach). Before running the analyses, all continuous (but not dichotomous) variables were standardized.

3.5.3 Missing data

In the quotations condition, missing data (see also Table 2) amounted to 3.0% for the responsiveness index and ranged from 7.5 to 23.9% for the predictors (i.e., individual student characteristics and classroom perceptions). In the text condition, missing values amounted to 1.1% for the responsiveness index and ranged from 5.3 to 16.8% for the predictors. The full information maximum likelihood (FIML) method was used in all analyses (e.g., Graham 2009).

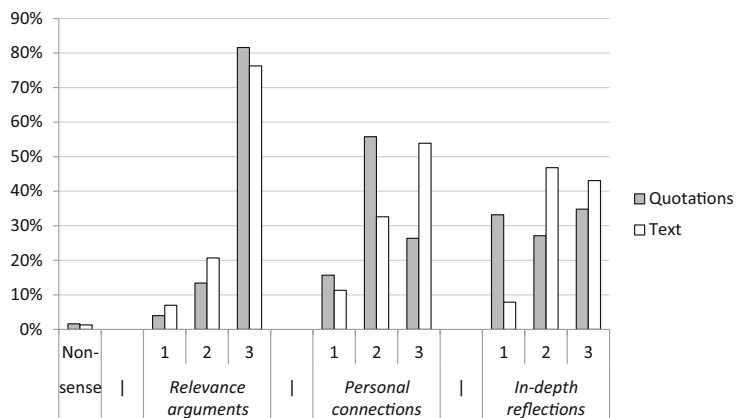


Fig. 3 Frequencies of students' values on the indicators of responsiveness per condition

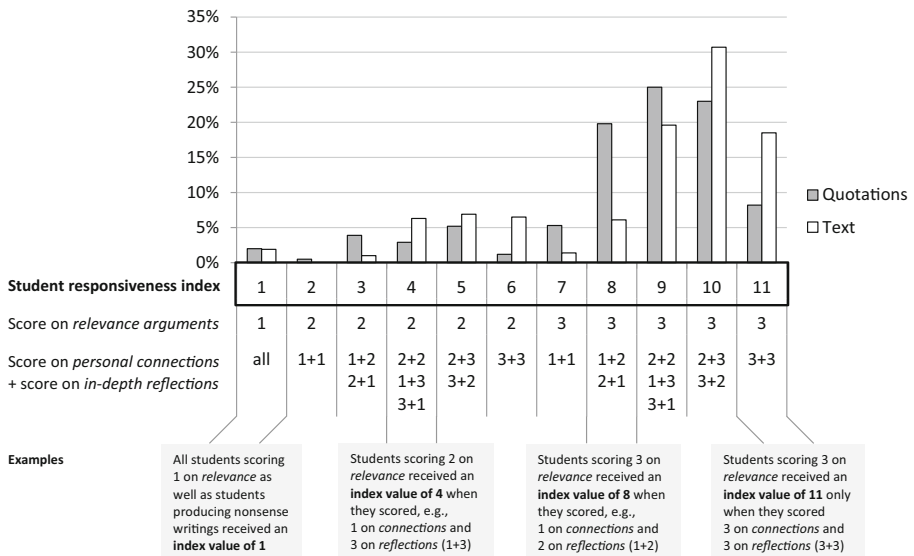


Fig. 4 Frequencies of students’ values on the responsiveness index (and respective scores on the indicators of responsiveness) per condition

4 Results

4.1 Students’ Responsiveness to the Writing Tasks about Relevance

The frequency distributions of the three indicators of responsiveness are presented in Fig. 3. A very small amount of students produced nonsense writings. Results on “relevance arguments” show that in both conditions the majority of students wrote mostly or only about the relevance of mathematics, rather than about its uselessness. Concerning “personal connections”, most students in the quotations condition used more other-references than self-references. In the text condition, most students used at least the same number of self-references as other-references. As for “in-depth reflections”, about one third of the students in the quotations condition did not use any new relevance arguments in their writings. In the text condition, the majority of students used at least one new relevance argument in their essays.

Combing all indicators, the frequency distributions of the responsiveness index are presented in Fig. 4. In both the quotations and the text condition, the frequency distributions were skewed to the left indicating overall high levels of responsiveness to the writing task: Only a few students in both conditions received values between 1 and 7 on the responsiveness index, whereas most students received values between 8 and 11. The medians were the value of 9 in the quotations condition and the value of 10 in the text condition.

Table 3 Predicting Intervention Responsiveness from Students' Individual Characteristics and Classroom Perceptions

	Quotations			Text		
	β	(SE)	<i>p</i>	β	(SE)	<i>p</i>
<i>Basic student characteristics</i>						
Gender (1 = male)	-0.07	(0.08)	.401	-0.29	(0.10)	.003
Cognitive ability	0.02	(0.03)	.593	0.04	(0.04)	.381
Conscientiousness	0.09	(0.05)	.050	0.09	(0.04)	.014
<i>Math achievement</i>						
Test score	0.18	(0.06)	.001	0.06	(0.04)	.143
<i>Math motivation</i>						
Self-concept	-0.15	(0.10)	.136	-0.03	(0.05)	.642
Homework self-efficacy	0.05	(0.06)	.440	0.07	(0.05)	.179
Intrinsic value	0.14	(0.07)	.050	0.09	(0.06)	.116
Utility value	0.07	(0.05)	.216	0.14	(0.04)	.000
<i>Classroom perceptions</i>						
Classmates' math valuing	0.06	(0.04)	.206	0.06	(0.04)	.161
Disruptions in math class	-0.03	(0.06)	.648	-0.05	(0.04)	.154

4.2 Individual Characteristics and Classroom Perceptions Predicting Responsiveness

Results concerning the prediction of students' responsiveness to the writing tasks through students' individual characteristics and classroom perceptions are shown in Table 3. Comparing the relative predictive strength of all predictors, students' conscientiousness was a statistically significant predictor of the responsiveness index in both conditions (quotations: $\beta=0.09$, $p=.050$; text: $\beta=0.09$, $p=.014$). In the quotations condition, students' math achievement ($\beta=0.18$, $p=.001$) and math-related intrinsic value ($\beta=0.14$, $p=.050$) predicted the responsiveness index positively, indicating that high-achievers and students who were highly intrinsically motivated for math responded to the quotations assignments significantly better than low-achievers and students with low intrinsic value beliefs of math. In the text condition, students' gender ($\beta=-0.29$, $p=.003$) emerged as the strongest predictor of students' responsiveness to the relevance essays when controlling for all other predictors, indicating that females were more responsive than males. Furthermore, students with high initial utility value beliefs of math had significantly higher values on the responsiveness index ($\beta=0.14$, $p<.001$) than students with low initial math-related utility value. Students' cognitive ability and classroom perceptions were not associated with students' responsiveness in either of the two conditions, controlling for all other predictors.

5 Discussion

Although writing tasks are a common tool used in psychological interventions to change students' personal beliefs (Yeager and Walton 2011), comprehensive studies assessing if intervention processes are related to psychological processes in ways that support the theory of change are missing. In this study, we sought to fill that gap in the literature by investigating whether students did what they were asked to do during the intervention and what student characteristics predicted responsiveness. We found highly conscientious students, girls, high achievers, and students with high math-related motivation to be most responsive to written intervention tasks about the relevance of mathematics.

5.1 The Logic of Relevance Interventions: Different Tasks, Different Intervention Processes

Prior analyses of the MoMa dataset have shown that evaluating quotations about the relevance of mathematics led to stronger effects on students' math-related motivation, effort, and achievement than writing a text about the personal relevance of mathematics, and that girls benefited more than boys from the text condition (Gaspard et al. 2015a; Brisson et al. 2017). The current study introduced a systematically derived theoretical framework for assessing and analyzing responsiveness that can be used to investigate the intervention processes leading to changes in students' utility value beliefs and to overall differences in the effectiveness of the two conditions (cf., complier-average causal effects analyses by Nagengast et al. 2018). Based on the assumed theory of change, students' responsiveness to the MoMa intervention tasks was assessed by coding the degree of positive argumentation, personal connections, and in-depth reflections about relevance in students' essays (e.g., Eccles et al. 1983). For theoretical reasons, the codings were combined into an overall responsiveness index by giving a stronger weight to the degree of positive argumentation than to the other two indicators (cf., recommendations by Nelson et al. 2012).

Overall responsiveness was similarly high in both conditions, indicating that student responsiveness per se cannot explain the differences in the strength of the two intervention approaches. Interestingly, linear regression analyses investigating the relation of students' individual characteristics and classroom perceptions with responsiveness revealed that the different conditions appealed to different kinds of students. The current results thus imply that different intervention processes are at work in the two conditions. This assumption is reinforced when considering that Nagengast et al. (2018) found that interventions effects on math-related utility value compared to controls differed more strongly between responsive and nonresponsive students in the text condition than in the quotations condition.

5.1.1 *Writing a text about relevance: new insights through analyzing responsiveness*

In the text condition, students with high initial utility value beliefs, girls, and highly conscientious students had the highest levels of responsiveness to the relevance task, holding other individual and classroom characteristics constant. Nagengast

et al. (2018) found positive effects of the text condition on students' motivational beliefs for responsive but not for nonresponsive students, whose perceived utility of mathematics could not be fostered. Which insights do these findings provide into the processes that make the text condition trigger a change in students' motivation—or not?

First, freely writing an essay about the relevance of mathematics without getting ideas from situations described by young adults seemed to be a very difficult task for students with low initial utility value beliefs. As a consequence, the intervention effects might pertain to positive and negative self-reinforcing processes (Yeager and Walton 2011): Students who initially had high value beliefs possibly had several ideas about the usefulness of mathematics and writing them down might have reinforced their positive beliefs. In contrast, students with low initial math utility value might have found it hard to come up with a lot of utility arguments and therefore have (partially) argued against the relevance of mathematics, resulting in no or even negative intervention effects.

Second, boys might have liked the text writing task less than girls and thus responded less well to the task. Indeed, boys comply less with activities done in language subjects (e.g., Trautwein et al. 2009), and writing a coherent text based on reasoning resembles such typical tasks more than reading quotations and then reflecting on their personal relevance by answering questions. The current results in fact indicate that girls' high degrees of intervention responsiveness might have contributed to the gender effects found in the text condition (for more detailed analyses, see Nagengast et al. 2018).

Finally, the text condition promoted students' utility value beliefs to a lesser degree than the quotations condition (Gaspard et al. 2015a). The current findings might also be interpreted in a way that the effectiveness of the MoMa interventions might have resulted from an interaction of students' responsiveness to the writing task and their reaction to the initial presentation of the utility of mathematics. In the text condition, the positive effect of the presentation might have been undermined by the rather difficult subsequent task of having to write the essays. Only students who responded well to the writing assignment might have benefited from the positive effect of the presentation. In other words, it might be necessary for the writing task to be easy enough for all students so that the input on relevance given in the presentation can unfold its full potential on the students.

5.1.2 Evaluating quotations about relevance: extending the intervention models

Conscientious students, high achievers, and students with high math-related intrinsic value beliefs responded best to the quotations-based writing assignment, controlling for all other individual and classroom characteristics. The quotations condition might have appealed to other types of students than the text condition for several reasons. First, reading and evaluating the personal importance of relevance quotations about mathematics required students to make judgments based on their own standards and those defined by the task, which represent deep-level cognitive processes (Krathwohl 2002). Fundamental math-related knowledge might have been helpful to respond well to this task, especially when interviewees mentioned quite specific math top-

ics in their descriptions (see sample quotations in the Online Supplement, Part A). Second, interest in the subject matter has been found to support increased attention, cognitive processing, and persistence on reading tasks (e.g., Schiefele 2009). Students who perceived mathematics to be intrinsically enjoyable might have been more susceptible to learning about particular situations in which they could apply math skills, leading to more engagement, and more success in performing the deep-level cognitive processes required in the task.

Considering that math-related utility value beliefs of both responsive and nonresponsive students were fostered through the quotations task (Nagengast et al. 2018), more in-depth analyses of responsiveness are needed to fully explain the processes leading to a change in students’ motivational beliefs through the quotations task. In fact, students who read the quotations were provided more relevance information than students in the text condition. The criterion “reading the quotations” might thus also have contributed to the strength of this approach. Furthermore, as deep-level cognitive processes were needed to perform the quotations-based task (cf., Krathwohl 2002), a more complex measure going beyond the novelty of the relevance arguments might be needed to fully capture the degree of “in-depth reflections”. Last but not least, the degree of “personal connections” might need to be measured in more detail by taking into account students’ the degree of identification and emotional closeness with the interviewees.

5.2 Paving the Way for Relevance Interventions to Enter Educational Practice

Throughout secondary school, students’ motivational beliefs are declining—in particular their utility value beliefs (e.g., Jacobs et al. 2002; Gaspard et al. 2017). Relevance interventions have yet been shown to be a powerful tool to halt this decrease in motivation and thereby support students’ academic interests, behavior, and achievement in real-life classroom settings (for an overview, see e.g., Rosenzweig and Wigfield 2016). The current results show which students are most likely to be nonresponsive to the writing activities. Future research should use these findings to investigate ways to enhance students’ intervention responsiveness, to optimize the designs of relevance interventions, and thereby pave their way for entering educational practice. Interviewing students who are lower in responsiveness might be useful in exploring reasons of students’ negative reactions to intervention activities. If the assignment was not intelligible enough for them, providing more scaffolding in the instructions might be helpful. If the intervention activity was not attractive to students, changes to the instructions that reduce reactance need to be developed and tested.

5.3 Advancing Research on Psychological Interventions in Education

Many field-based psychological interventions in education, even if brief in nature, can be effective in raising important academic outcomes (Lazowski and Hulleman 2016). The demand for expertise to successfully adapt such interventions to diverse educational contexts is thus growing—otherwise, it is difficult to replicate their effects (e.g., Yeager and Walton 2011). Indeed, in educational settings, the source of

the effectiveness or non-effectiveness of interventions may be blurred by the number of factors which cannot be kept constant across classrooms (Weiss et al. 2014). Using theoretically sound research designs is therefore crucial to deal with unobservable variations across classrooms (e.g., Rubin 1974). Variation across classrooms in program implementation, in contrast, can be made at least partially observable, for example by measuring participants' reaction to the intervention.

The current study showed that assessing indicators of student responsiveness and analyzing their antecedents and effects is helpful to provide an empirical account on the role of core intervention elements in contributing to the effectiveness of field-based interventions. Such an empirical understanding of the intervention processes underlying a change in target psychological processes is essential to advance psychological theorizing and to enhance the "psychological precision" (Walton 2014, p. 74) of classroom interventions. Intervention responsiveness needs to be researched within diverse learning contexts in order to enable an evidence-based adaptation of specific intervention components to specific educational settings. In addition, studies on intervention responsiveness can help to inform researchers and educators about any potential risks associated with participants' nonresponsiveness to a certain intervention program. Responsiveness studies should thus not be an exception, but the rule to go along with any experimental research in the field.

5.4 Limitations and Suggestions for Future Research

As to all research studies, several limitations apply to the current investigation pertaining to, for example, the specific sample investigated in the current study. To ensure generalizability, the current results need to be replicated with other samples including students in other education systems as well as German students in non-academic track schools (e.g., vocational track schools). Depending on the focus of the education system or school track, the contents of the intervention material would probably have to be adapted.

Second, to enable a comparison between the two relevance intervention conditions, the same intervention models and, therefore, the same indicators were used to assess students' responsiveness to the different writing tasks. However, the coding categories did not seem to be equally straightforward in both conditions, as reliability measures were lower in the quotations condition than in the text condition. Future studies could investigate the importance of other indicators of responsiveness such as students' identification with the interviewees or cognitive engagement in relevance interventions, for example, by combining different media such as reading versus hearing quotations with computerized tasks or think-aloud methods (Ericsson and Simon 1980).

Third, while a comprehensive amount of variables was investigated as predictors of responsiveness, no information was available on students' literacy in reading and writing—skills required by both intervention tasks. Future researchers should investigate the impact of students' reading and writing competencies on students' responsiveness to intervention tasks requiring these skills. As girls typically outperform boys in reading and writing skills (e.g., Reilly et al. 2019), such research might also help to shed light on the question whether the gender effects in students' respon-

siveness to and the effectiveness of the text condition is ruled out when controlling for students' literacy in reading and writing.

Finally, the individual writing tasks constituted the core element of the MoMa relevance interventions and thus were in the focus of the current investigation on student responsiveness. Nevertheless, students' experiences during the introductory psychoeducational presentation such as their cognitive engagement while learning about diverse examples for the utility of math might also have affected the intervention effects. Students' experiences during such a pre-writing part of the intervention could be taken into account in future research, for example, by using experience sampling methods (Rosenzweig and Wigfield 2016) or observational methods (cf., Fredricks et al. 2004).

6 Conclusion

Relevance interventions show a huge potential to raise important learning outcomes (Durik et al. 2015). The results of the present study on student responsiveness imply that when designing written relevance intervention tasks aimed for implementation in real-life educational settings, is it important to consider that individual student characteristics such as conscientiousness, gender, and domain-specific motivation may determine how well the students follow the instructions of the assignments. Responsiveness in turn affects the strength of the intervention effects (Nagengast et al. 2018)—in the current interventions especially in the text condition, an approach used in numerous other relevance intervention programs (e.g., Hulleman and Harackiewicz 2009). Adapting the intervention material to the needs of students who are most likely to be nonresponsive should thus be the goal of future research studies. Using this knowledge to further improve the theories and designs of relevance interventions might help to eventually pave the way for relevance interventions to enter educational practice at a larger scale.

Acknowledgements We thank Katharina Allgaier and Evelin Herbein for their help conducting this research.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funding This research was funded in part by German Research Foundation Grant TR 553/7-1 awarded to Ulrich Trautwein, Oliver Lüdtke, and Benjamin Nagengast. Brigitte Brisson, Isabelle Häfner, and Hanna Gaspard were members of the Cooperative Research Training Group of the University of Education, Ludwigsburg, and the University of Tübingen, which was supported by the Ministry of Science, Research and the Arts in Baden-Württemberg. This research was supported as part of the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments, and by the Tübingen Postdoctoral Academy for Research on Education (PACE) at the Hector Research Institute of Education Sciences and Psychology, Tübingen; PACE is funded by the Baden-

Württemberg Ministry of Science, Research and the Arts. We are also indebted to the Baden-Württemberg Stiftung for the financial support of this research project by the Eliteprogramme for Postdocs.

Funding Open Access funding provided by Projekt DEAL.

References

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., & Krauss, S. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV). Dokumentation der Erhebungsinstrumente. Materialien aus der Bildungsforschung Nr. 83 [Documentation of the data collection instruments of the COACTIV study. Materials of Research in Education Nr. 83]*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Borkenau, P., & Ostendorf, F. (1991). Ein Fragebogen zur Erfassung fünf robuster Persönlichkeitsfaktoren [A questionnaire for assessing five robust personality factors]. *Diagnostica*, 37(1), 29–41.
- Boruch, R. F. (1997). *Randomized experiments for planning and evaluation*. Thousand Oaks: SAGE.
- Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal*, 54(6), 1048–1078. <https://doi.org/10.3102/0002831217716084>.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>.
- Cooper, H. (1989). *Homework*. New York: Longman.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23–45. [https://doi.org/10.1016/S0272-7358\(97\)00043-3](https://doi.org/10.1016/S0272-7358(97)00043-3).
- Durik, A. M., Hulleman, C. S., & Harackiewicz, J. M. (2015). One size fits some: Instructional enhancement to promote interest. In K. A. Renninger, M. Nieswandt & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 49–62). Washington, DC: American Educational Research Association.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>.
- 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: Psychological and sociological approaches* (pp. 75–146). New York: Freeman.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. <https://doi.org/10.1037/0033-295x.87.3.215>.
- Finn, J. D., & Zimmer, K. S. (2012). Student engagement: what is it? Why does it matter? In S. L. Christenson, A. L. Reschly & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97–132). New York: Springer.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507–537. <https://doi.org/10.1111/j.1532-7795.2010.00645.x>.
- Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: a multilevel analysis of mathematics classrooms. *Learning and Instruction*, 17(5), 478–493. <https://doi.org/10.1016/j.learninstruc.2007.09.001>.
- Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Trautwein, U., & Nagengast, B. (2015a). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, 51(9), 1226–1240. <https://doi.org/10.1037/dev0000028>.
- Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015b). More value through greater differentiation: gender differences in value beliefs about math. *Journal of Educational Psychology*, 107, 663–677. <https://doi.org/10.1037/edu0000003>.
- Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade

- level, gender, and academic subject. *Contemporary Educational Psychology*, 48, 67–84. <https://doi.org/10.1016/j.cedpsych.2016.09.003>.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>.
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: disentangling race and social class. *Journal of Personality and Social Psychology*, 111(5), 745–765. <https://doi.org/10.1037/pspp0000075>.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klasse [Cognitive Ability Test for Grades 4 through 12]*. Göttingen: Hogrefe.
- 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. <https://doi.org/10.1080/1934574080253925>.
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958), 1410–1412. <https://doi.org/10.1126/science.1177067>.
- 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. <https://doi.org/10.1037/A0019506>.
- Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology*, 109(3), 387–404. <https://doi.org/10.1037/edu0000146>.
- Husman, J., Nelson, K., & Cheng, K. (Aug., 2017). *Intervening to promote engineering student success in technical, nonmajor, required barrier courses*. Poster presented at the 17th Biennial EARLI Conference. Tampere: Finland.
- 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. <https://doi.org/10.1111/1467-8624.00421>.
- Juvonen, J., Espinoza, G., & Knifsend, C. (2012). The role of peer relationships in student academic and extracurricular engagement. In I. S. L. Christenson, A. L. Reschly & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 387–492). New York: Springer.
- Karabenick, S. A., Albrecht, J., & Rausch, N. (Aug., 2017). *Examining the self-generation of subjective task value: A mixed-methods investigation*. Paper presented at the 17th Biennial EARLI Conference. Tampere: Finland.
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2.
- Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: a meta-analytic review. *Review of Educational Research*, 86(2), 602–640. <https://doi.org/10.3102/0034654315617832>.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954–969. <https://doi.org/10.1037/0003-066X.41.9.954>.
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140. <https://doi.org/10.1037/met0000078>.
- Murrah, W. M., Kosovich, J. J., & Hulleman, C. S. (2017). A framework for incorporating intervention fidelity in educational evaluation studies. In I. G. Roberts, S. Vaughn, T. Beretvas & V. Wong (Eds.), *Treatment fidelity in studies of educational intervention* (pp. 39–60). New York: Routledge.
- Nagengast, B., Brisson, B. M., Hulleman, C. S., Gaspard, H., Häfner, I., & Trautwein, U. (2018). Learning more from educational intervention studies: Estimating complier average causal effects in a relevance intervention. *The Journal of Experimental Education*, 86(1), 105–123. <https://doi.org/10.1080/00220973.2017.1289359>.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, 39(4), 374–396. <https://doi.org/10.1007/s11414-012-9295-x>.
- Oyserman, D., & Destin, M. (2010). Identity-based motivation: Implications for intervention. *The Counseling Psychologist*, 38(7), 1001–1043. <https://doi.org/10.1177/0011000010374775>.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>.

- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, 74(4), 445–458. <https://doi.org/10.1037/amp0000356>.
- Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–20). New York: Springer.
- Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: a summary of research findings. *Elementary School Journal*, 100(5), 443–471. <https://doi.org/10.1086/499650>.
- Rosenzweig, E. Q., & Wigfield, A. (2016). STEM Motivation interventions for adolescents: a promising start, but further to go. *Educational Psychologist*, 51(2), 146–163. <https://doi.org/10.1080/00461520.2016.1154792>.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. <https://doi.org/10.1037/H0037350>.
- Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods*, 19(3), 317–333. <https://doi.org/10.1037/met0000013>.
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197–222). New York: Routledge.
- Schreier, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Lüdtkke, O., & Trautwein, U. (2014). Der Wert der Mathematik im Klassenzimmer – Die Bedeutung relevanzbezogener Unterrichtsmerkmale für die Wertüberzeugungen der Schülerinnen und Schüler [The value of mathematics in the classroom: The importance of a relevance-oriented learning environment for students' value beliefs]. *Zeitschrift für Erziehungswissenschaft*, 17(2), 225–255. <https://doi.org/10.1007/s11618-014-0537-y>.
- Schwanzer, A. D., Trautwein, U., Lüdtkke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener [Development of an instrument to assess young adults' self-concept]. *Diagnostica*, 51(4), 183–194. <https://doi.org/10.1026/0012-1924.51.4.183>.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth Cengage Learning.
- Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying values-affirmation interventions: writing about social belonging is a key to buffering against identity threat. *Personality and Social Psychology Bulletin*, 39(5), 663–676. <https://doi.org/10.1177/0146167213480816>.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>.
- Trautwein, U., & Köller, O. (2003). Was lange währt, wird nicht immer gut: Zur Rolle selbstregulativer Strategien bei der Hausaufgabenerledigung [Time investment does not always pay off: The role of self-regulatory strategies in homework execution]. *Zeitschrift für Pädagogische Psychologie*, 17(3), 199–209. <https://doi.org/10.1024/1010-0652.17.34.199>.
- Trautwein, U., & Lüdtkke, O. (2009). Predicting homework motivation and homework effort in six school subjects: the role of person and family characteristics, classroom factors, and school track. *Learning and Instruction*, 19(3), 243–258. <https://doi.org/10.1016/j.learninstruc.2008.05.001>.
- Trautwein, U., Lüdtkke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, 97(6), 1115–1128. <https://doi.org/10.1037/a0017048>.
- Trautwein, U., Lüdtkke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, 98(2), 438–456. <https://doi.org/10.1037/0022-0663.98.2.438>.
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73–82. <https://doi.org/10.1177/0963721413512856>.
- Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778–808. <https://doi.org/10.1002/pam.21760>.
- Wigfield, A., Rosenzweig, E. Q., & Eccles, J. S. (2017). Achievement values: Interactions, interventions, and future directions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 116–134). New York, NY: The Guilford Press.
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, 81(2), 267–301. <https://doi.org/10.3102/0034654311405999>.