The Journal of Experimental Education

Routledge
Taylor & Francis Group

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

Promises and Pitfalls of Adapting Utility Value Interventions for Online Math Courses

Emily Q. Rosenzweig, Chris S. Hulleman, Kenneth E. Barron, Jeffery J. Kosovich, Stacy J. Priniski & Allan Wigfield

To cite this article: Emily Q. Rosenzweig, Chris S. Hulleman, Kenneth E. Barron, Jeffery J. Kosovich, Stacy J. Priniski & Allan Wigfield (2019): Promises and Pitfalls of Adapting Utility Value Interventions for Online Math Courses, The Journal of Experimental Education, DOI: 10.1080/00220973.2018.1496059

To link to this article: https://doi.org/10.1080/00220973.2018.1496059

Published online: 09 Feb 2019.

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2018.1496059

Promises and Pitfalls of Adapting Utility Value Interventions for Online Math Courses

Emily Q. Rosenzweig[a], Chris S. Hulleman[b], Kenneth E. Barron[c], Jeffery J. Kosovich[d], Stacy J. Priniski[e], and Allan Wigfield[f]

[a]University of Maryland, College Park, Maryland, USA; [b]University of Virginia, Charlottesville, VA, USA; [c]James Madison University, Harrisonburg, VA, USA; [d]University of Virginia, Greensboro, VA, USA; [e]University of Wisconsin-Madison, Wisconsin, USA; [f]University of Maryland, College Park, Maryland, USA

ABSTRACT

A growing body of research suggests that interventions promoting students' utility value for a subject can improve their academic outcomes. However, numerous questions remain regarding how much to adapt prior intervention materials to promote utility value in new educational contexts and how implementation constraints of an educational context may impact the success of these interventions. In this study, using a design-based process we developed and tested three utility value interventions in a new educational context (online high school math). We found that one of the interventions increased utility value compared to control conditions, but we also encountered constraints on intervention implementation that limited the effectiveness of our intervention and the conclusions we could draw from this work. We use our experience as a case study to illustrate the costs and benefits of making certain implementation choices when partnering with practitioners to administer utility value interventions in new contexts.

KEYWORDS

Expectancy value theory; motivation; intervention; mathematics education; adolescence

"Algebra is a waste of time … Schools need to focus on what will help us in life not useless stuff."

—High school student enrolled in an online Algebra I course

ONE CRITICAL FACTOR that affects students' choices to pursue courses and careers in science, technology, engineering, and math (STEM) is whether the students perceive that what they are learning in STEM courses is useful (i.e., utility value; Eccles-Parsons et al., 1983). Students' perceptions of utility value tend to decline during middle and high school, and these declines are strongest in math and science (Gaspard, Häfner, Parrisius, Trautwein, & Nagengast, 2017; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Watt, 2004). These trends are important, because students who perceive more utility value are more likely to take STEM courses and pursue STEM majors in the future (e.g., Musu-Gillette, Wigfield, Harring, & Eccles, 2015; Simpkins, Davis-Kean, & Eccles, 2006).

Fortunately, a growing body of research demonstrates that students' perceptions of utility value can improve after students participate in a brief educational intervention in which they relate what they are learning in their courses to their lives. Hulleman and colleagues (e.g., Hulleman, 2007; Hulleman & Harackiewicz, 2009; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman, Kosovich, Barron, & Daniel, 2017) call this type of intervention a *utility value*

CONTACT Emily Q. Rosenzweig emily.rosenzweig@wisc.edu Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, United States.
The first author is now at University of Wisconsin-Madison.

intervention. Utility value interventions have been shown to increase students' academic performance in STEM fields (e.g., Harackiewicz, Canning, Tibbetts, Priniski, & Hyde, 2016; Tibbetts, Harackiewicz, Priniski, & Canning, 2016, for reviews). However, researchers have not fully explored how effective utility value interventions are in different types of STEM educational contexts and how to transfer utility value interventions across contexts. In this research, we tried to adapt utility value interventions and improve students' perceived utility value in online high school math courses, while accommodating constraints on intervention implementation from our partner school.

Utility value interventions

Utility value interventions are based in Eccles and colleagues' expectancy-value model of motivation and choice (Eccles-Parsons et al., 1983; see Wigfield, Tonks, & Klauda, 2016, for a recent review). Expectancy-value researchers posit that students' perceptions of the value of a task combine with the students' expectations of success (i.e., their competence-related beliefs) to determine later educational behavior, such as choices of which activities to pursue and performance. Utility value is one of three influences on task value that are most often studied by researchers—the others being intrinsic (enjoyment) and attainment (identity-based) value. The higher students' perceptions of task value and competence-related beliefs, the better students perform in a subject and the more likely they are to choose activities related to that subject.

The present research builds on previous expectancy-value–theory-based interventions that aim to improve students' utility value by asking them to write about how what they are learning connects to their lives (e.g., Gaspard et al., 2015; Harackiewicz et al., 2016; Hulleman & Harackiewicz, 2009; Hulleman et al., 2010, 2017). For these utility value interventions to be effective, several processes need to occur. Researchers argue that it is critical that students make specific connections between themselves and their course material during an intervention (Hulleman & Cordray, 2009; Hulleman et al., 2017). Students also need to relate personally to the connections being made (Harackiewicz & Priniski, 2018).

Considering this, it is unclear whether researchers can successfully transfer utility value interventions for use across all learning contexts. Yeager and Walton (2011) argue that motivational interventions are not magic bullets and the same intervention materials may not produce the same effects across all educational settings (e.g., Dee, 2015). All past utility value interventions have been conducted in face-to-face courses and the intervention was part of class activities or students received some kind of incentive (e.g., course points) to participate. It is unclear whether a utility value intervention would be effective if implemented in contexts with less oversight or that are not taught face to face. Would students engage sufficiently with intervention materials to make specific, personal connections between course materials and their lives?

In this research, we focus on testing utility value interventions in one such setting: online high school math courses. These courses are rapidly growing in popularity within the K–12 educational market (Watson, Pape, Murin, Gemin, & Vashaw, 2014). Several states and many school districts require that students pass at least one online course in order to graduate high school (Sheehy, 2012). Even when online learning is not required, nearly all school districts in the United States offer online course options to students (Watson et al., 2014; Picciano & Seaman, 2009). Although high school students taking online math courses have performed equally well as students taking the same classes face to face, students in the online courses have reported less satisfaction with the courses and less confidence about their ability to master course material (Bernard et al., 2004; O'Dwyer, Carey, & Kleiman, 2007). Further, students in online courses have shown 10% to 20% higher dropout rates than have students in similar face-to-face classes (Cavanaugh, Gillan, Kromrey, Hess, & Blomeyer, 2004; Rice, 2006). High dropout rates in online high school math courses have negative long-term consequences. These courses are critical students need to complete these courses successfully early in high school in order to continue on in STEM (Stone, 1998). Students who fail to complete these courses are at a distinct disadvantage for enrolling in competitive colleges and for pursuing STEM-related majors (Speilhagen, 2006). Through a unique partnership with an online school provider who was interested in fostering students' value, we had the opportunity to assess the effectiveness of utility value interventions to increase students' utility value in online algebra and geometry courses.

Transferring utility value interventions to new educational contexts

In exploring whether we could adapt previous utility value interventions so that they would be effective in online math courses, we made certain assumptions about whether and how the same motivational intervention materials can be effective across educational contexts. Some researchers take a very situated approach to motivational interventions, assuming that the same intervention materials or instructional practices will target motivational processes very differently across different educational contexts (Kaplan, Katz, & Flum, 2012). Conversely, other educational policy-makers and researchers assume that it is possible to use the same or very similar intervention protocols across contexts and derive "optimal" rules of intervention administration for those programs (Al-Ubaydli & List, 2013). Many utility value intervention researchers fall somewhere in between these perspectives: They assume that the same types of intervention activities can target similar psychological processes across different educational contexts, but they acknowledge that the context might interact with the effectiveness of intervention activities (Harackiewicz & Priniski, 2018; Hulleman & Barron, 2016; Rosenzweig & Wigfield, 2016). In the present study, we assumed that utility value intervention activities did have the potential to improve students' utility value across different educational contexts, but the specifics of the intervention prompts and materials would need to change in order to target utility value effectively within the educational context of our study (i.e., online high school math courses).

One aspect of adapting utility value materials so the intervention can be transferred to a new educational context is ensuring that the materials are meaningful to the target population of students. For example, in the present study we would be working with a group of online students who were at risk for high dropout and low engagement. A second aspect of adapting materials is ensuring that the intervention materials meet the concerns and requirements of the partner school. For example, our partner school requested that we implement the intervention within a very brief time frame. In addressing these concerns, researchers need to balance consideration for the educational context—wanting to implement an intervention that teachers would actually be willing to use and that students would enjoy—with scientific considerations.

There is little guidance in the published literature for utility value intervention researchers on how to negotiate these tensions. Many utility value intervention researchers are aware that they should adapt their materials in a new setting, but the little published work that has addressed this topic has focused on college students (e.g., Harackiewicz & Priniski, 2018). As a result, new utility value intervention researchers or practitioners interested in implementing these interventions may not be aware of how much they should adapt a prior study's materials for a new context, how to adapt these materials, or what implementation constraints (e.g., a restricted number of sessions, limited ability to measure outcomes) can be accommodated while maintaining an intervention's scientific integrity. To address this gap, our study aimed to provide a specific case study by which to illustrate how we adapted the intervention materials to meet contextual constraints and fit our target population and to discuss the costs and benefits of each of these adaptation choices.

We partnered with the Project for Education Research That Scales (PERTS) and a large virtual high school to develop and implement our intervention materials. The high school administrators were interested in implementing interventions with their students as a means of addressing high rates of dropout and low engagement in their courses. However, working with this school brought with it significant practical constraints. Beyond the concerns already noted regarding the potential for limited engagement among students learning math online, the school wanted to minimize the time and resources needed to implement the interventions. In general, researchers recommend addressing such constraints by conducting careful pilot work that could account for unique attributes of a particular educational context (Yeager & Walton, 2011; Yeager et al., 2016). Yeager et al. (2016) recommended that researchers develop motivational intervention materials via an iterative, design-thinking process, in which they solicit feedback from both students and educators who are familiar with the educational context (Kelley & Kelley, 2013). This procedure would make it more likely that intervention materials would evoke the intended psychological responses from students.

We adopted this process for the present study. We gathered information about the design constraints associated with our target educational context, and we made methodological choices about how to develop and implement the intervention in response to the constraints. In this paper, we evaluate the costs and benefits of each choice in order to inform future utility value intervention research (Collins, Joseph, & Bielaczyc, 2004).

The current research

The present study had two goals: (1) develop materials for and test the efficacy of three interventions designed to boost students' perceptions of utility value in online algebra and geometry and (2) report our experience as a case study from which to provide recommendations regarding what implementation choices seem to have positive and negative consequences when adapting utility value interventions for use in novel educational contexts.

We developed and tested three types of utility value interventions in this study, because prior work suggests that not all utility value intervention materials promote students' engagement and participation equally well (Gaspard et al., 2015). The first of the three intervention conditions was similar to Hulleman and colleagues' interventions (i.e., Hulleman & Harackiewicz, 2009) and prompted students to write a personal essay about how math related to their lives (a condition that we called Utility Essay). In the second intervention condition, students were prompted to read quotations about utility value and evaluate them (i.e., Utility Quotes + Evaluation). This was adapted from an intervention conducted by Gaspard et al. (2015), who reported that German high school students showed larger gains in utility value when they read and responded to quotations from other students regarding how math was related to those students' lives compared to when the students wrote essays. Gaspard et al. (2015) suggested that a quotation evaluation task was more effective than an essay task because it provided students examples of ways that math could be useful and asked them to do a more pleasant activity than essay writing. Third, we created an intervention condition that was a hybrid of these two tasks, asking students to read quotations about utility value and then write a personal essay (i.e., Utility Quotes + Essay). Prior research did not lend itself to clear hypotheses about which intervention condition might impact utility value most positively, so we were primarily concerned with whether any utility value condition was more effective than either of two control conditions: one in which students did nothing and one in which they summarized course material.

Our research was carried out in three phases. In the *information gathering* phase, we learned more about the students in our sample and how best to target their perceived utility value. In the *intervention prototyping* phase, we incorporated this information into a set of preliminary intervention materials, which we then revised in response to feedback from students. In the *randomized field experiment* phase, we tested whether each of the three interventions affected online algebra and geometry students' perceived utility value compared with either of the two control groups. To further evaluate intervention implementation, we coded two indicators of whether students made personal and specific connections to their course material during each intervention.

These were (a) students' use of personal pronouns in their writing and (b) the extent to which students articulated specific and personal connections to their course material while writing essays. We present Method, Results, and Discussion sections for each of our three phases of research. In each phase, in addition to reporting the data we collected, we provide a description of the decisions we made regarding intervention implementation and design. Then, in the General Discussion, we evaluate what promising findings we were able to observe and what "pitfalls" we encountered as a result of the implementation decisions we made. We conclude by using our experience to make recommendations for future researchers who want to implement utility value interventions in novel educational contexts.

Phase 1: Information gathering

Method

To begin developing intervention materials, we gathered feedback to learn more about our target population of online students and the constraints we would need to accommodate from our partner school. We asked the instructional programs manager at the partner school about the backgrounds of students who enrolled in these courses, the specifics of the timing and incentives that we would be able to use during intervention implementation, and the content of the algebra and geometry courses students would complete.

Results

Findings regarding the target population

Most students in the algebra and geometry courses tended to be enrolled full-time at other schools, with a smaller proportion being home-schooled. At the time of the intervention (June), most enrolled students were likely to be taking online math courses as summer supplemental courses for their typical face-to-face public schools. This was because many students needed to satisfy requirements that they take at least one online course prior to graduating high school. Overall, our sample was likely to perform at an average level (i.e., students were not all high or low achieving), with some students taking this course for enrichment and others for remediation. Our sample also was likely to be socioeconomically diverse, with approximately one-third of students qualifying for free and reduced-price lunch. In terms of race, students were expected to be mostly White (70%), with about 20% Black or African American and 10% Asian or Asian American. Approximately one-third of students were likely to identify as Latinx.

Both the algebra and geometry courses were administered in 10 self-paced modules, representing different topics (e.g., simple linear equations in algebra; angles in geometry). The modules included text and interactive examples, quizzes, tests, homework, and collaborative assignments. Students had a designated course instructor whom they could contact with questions. Both courses had fairly high dropout rates. Five months after this study was completed, only 48.2% of the algebra students who had been enrolled at the time of the study had completed the course (over 99% of them with passing grades) whereas 40% had dropped out; the other students had special circumstances or were still enrolled. In geometry, 60% of students had completed the course (over 99% with passing grades), while 27% had dropped out.

We asked if we could obtain preliminary feedback from students and/or teachers at the school to learn more about how students would respond to prototypes of our intervention materials. The administrator said that we could connect with a group of students enrolled in online science courses at the high school for this purpose but due to privacy concerns we would not be able to collect consent forms from them.

Findings regarding constraints on intervention implementation

Although our partner school was eager to participate in the study, there were regulations that would limit how we could implement the utility value interventions. First, as outside personnel we would not be allowed to interact with students directly and would send all intervention and recruitment materials to students via their course instructors. Second, the data-management system at the school was such that we would not be able to follow up with students after initial intervention implementation. Therefore, we were asked to implement the intervention at a single time point. Specifically, we were asked to implement pretests, posttests, and the intervention materials all within a single 15-minute window. Third, students were not permitted to receive incentives to participate in the interventions. Fourth, we were not able to embed intervention materials within regularly assigned course activities, which has been done in other utility value intervention research to ensure high participation rates in the absence of material incentives (e.g., Harackiewicz et al., 2016; Hulleman et al., 2010, 2017). Finally, we asked, but were not given approval, to offer an alternative incentive for study participation besides giving something to every student, such as entry in a raffle.

Discussion

As a result of our information gathering, we made several major decisions regarding the next phase of our research. First, we concluded that our intervention sample likely would be diverse in terms of student backgrounds and prior experiences with math. Therefore, we planned to create materials that would appeal to a broad group of students by including a wide variety of quotations for students to read and/or by allowing students to make connections to a variety of aspects of their lives in their essays.

Second, the very brief time frame for implementing the study meant we would have to use a more limited dosage than previous interventions had done. This could reduce the impacts that our interventions might have, so we would try to make the materials as engaging as possible. Third, the short implementation time frame would prevent us from thoroughly measuring students' responses to intervention activities and the mechanisms by which interventions might impact students. We decided that we would focus on measuring whether the interventions improved students' perceived utility value and would not collect detailed data to assess the mechanisms of that effect (e.g., engagement, perceived competence). We determined that this goal would be worthwhile for our target population and that it would be interesting to test whether very brief utility value interventions could be effective at increasing utility value.

Finally, students would have no incentive to participate in our study. This could result in low participation rates and/or high attrition, which would limit the generalizability of any results we obtained. We concluded that even with a high amount of attrition we would obtain a fairly large number of students, so the information that could be gained from this study would still be informative if the sample were not fully representative. However, we would try to make intervention materials very engaging, so that students would be less likely to drop out.

Phase 2: Intervention prototyping

Method

In this phase of the research, we obtained feedback on drafts of our intervention materials from 184 students enrolled in science courses at the online high school in which the intervention was conducted. We also obtained feedback from eight students not enrolled in online courses: four undergraduate college students who were lab research assistants and four K–12 students who knew one of the investigators (ages 12, 14, 15, and 16). The research team first adapted materials

Table 1. Changes made to quotations during the intervention prototyping phase.

Changes Made	Rationale	Excerpt from Original Quotations	Excerpt from Final Quotations
"Math Normally Not Useful" Statement *Added phrases stating that students were surprised to learn that math related to their lives.*	Students had negative responses to quotations that expressed enthusiasm about making math connections. They expressed that the quotations felt insincere and unrealistic.	"I can see how Geometry is connected to my life because it can help me improve my sports playing."	"I never thought math skills were important for me to learn, but they can come in handy."
Interesting Topics *Made quotations more aligned to students' interests based on their feedback.*	Students expressed that they were not interested in many topics referenced in the quotations.	"I can use the MPG = miles/gallons equation to figure out about how many miles I can drive in my car until I run out of gas."	"At the movie theater they have a deal to pay a little more and get bottomless popcorn and soda. I never can figure out whether it's a rip-off or not. So I set up an Algebra equation to figure out how much money I was spending on each cup of popcorn with the regular and bottomless sizes."
Novel Connections *Provided examples that students were unlikely to have heard from their teachers.*	Students expressed frustration over connections they had heard before that seemed cliché. They expressed that these quotations were insincere and boring.	"I always … wanted to be an Architect. Now that I'm in this geometry class I can see how the angles of beams all fit together to make a sturdy bridge that also looks cool."	"Geometry is actually really important for art or graphic design because I will need to understand how to scale different shapes to avoid making things that are too big."
More Natural Phrasing *Include language that felt more naturally a part of students' vernacular.*	Students expressed that some aspects of the quotations did not sound realistic.	"Geometry is also really important to understand how things fit together in space to figure out how to create my designs, and knowing about concepts like volume and area can help me do this."	"Let me tell you, when I work for a long time on a painting and realize that the whole thing won't fit on the canvas, it really messes up the whole day."

from prior utility value intervention studies to create an initial draft of materials for each condition. We made all materials briefer and more engaging than what had been used in prior studies. We asked students to provide feedback on the materials and then revised them. As noted above, we could not report on students' specific responses to this pilot work. Therefore we report in detail how we changed our intervention materials during this phase but we discuss only in general terms the actual responses we obtained from students.

Results

Our piloting and findings differed by condition. First, for the *Utility Essay* condition, we began by shortening an essay prompt used by Hulleman and Harackiewicz (2009) in high school science. The original prompt asked students to summarize a topic they were studying in the course, then write an essay with at least five sentences describing the relevance of that topic to their lives. We omitted the summary portion of this task in our materials to reduce length. Even with this change, the non–online-course student participants reported that the task was fairly difficult, so we added additional instructions listing a few specific ways that prior students had written about utility value in math before giving the essay prompt to help students make their own connections. We also shortened the essay requirement to three to five sentences. Students who provided feedback on the revised materials informed us that the revisions made the task less challenging.

Second, for the *Utility Quotes + Evaluation* condition, we needed to develop quotations that would appeal to the diverse group of students enrolled in the online math courses. We developed an initial set of quotations about how math related to students' lives by finding high-quality examples of student essays from previous utility value interventions in college biology (e.g., Harackiewicz et al., 2016) and developmental math (e.g., Kosovich, Hulleman, Phelps, & Lee, 2016). We used this material to develop five potential quotations for each of three different types of utility value (utility for everyday activities/hobbies, utility for career, utility for helping the community) in each math subject (algebra, geometry). We provided the pool of quotations to the 184 high school students enrolled in online science courses. Each student provided feedback on five quotations, rating how interesting each was and if it represented their own or their classmates' experiences. They also rated how much each quotation helped them connect math to their lives and provided suggestions for improving each quotation.

Online students did not respond positively to this activity; their responses and the changes we made to address them are outlined in Table 1. Some students reported strong, overt, negative reactions to many of the quotations, noting that (a) they would never discuss math in the way described in the quotations, (b) the quotations were not believable, or (c) they were not interested in the content of the quotations or in being told about the utility of math. Moving forward, we eliminated quotations that had provoked strong negative reactions from any students. We chose three quotations for each subject that had overall high positive average ratings, and that, when possible, provided the same situations in which students could use math across both algebra and geometry (i.e., in sports, art, and volunteer work). We also wrote one additional quotation for each subject based on students' suggestions (about social media and going to the movies). To reduce reactance, we added statements to each quotation emphasizing that students typically disliked math but had been surprised to learn that math related to their lives in some ways. Our eight additional students who were not part of the online sample made a few minor suggestions to improve these quotations further. See Appendix for a final draft of the quotations. Another issue was that there was very high attrition: We asked students only to read and evaluate briefly five quotations, but 60% of students dropped out before completing the task.

To develop the quotation evaluation task, we adapted materials used by Gaspard et al. (2015). These materials prompted students to read quotations, rank them in terms of personal relevance, and write about what they found to be convincing and/or relevant in each quotation. Students then re-wrote their top-rated quotation in their own words. We shortened this approach by having students write only about their top-rated quotation instead of all quotations. The non–online-course students reported that this task was reasonably engaging and easy to complete.

Third, for the *Utility Quotes + Essay* condition, students read the four quotations we developed for the Utility Quotes + Evaluation condition but instead of evaluating the quotations they completed the revised writing task we developed for the Utility Essay condition. Task instructions and framing were taken verbatim from the other conditions. We confirmed with the non–online-course students that the task was not too long.

Fourth, we created the two control conditions. The *Summary Control* condition was consistent with what prior utility value intervention researchers have used (Harackiewicz et al., 2016; Hulleman & Harackiewicz, 2009; Hulleman et al., 2010, 2017). This condition controlled for writing about math and had the potential to boost student learning by helping students to process the material deeply through elaboration (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). To create materials, we adapted prompts used in work by Hulleman and Harackiewicz (2009), Kosovich et al. (2016), and Hulleman et al. (2010). Students chose a topic from their course, wrote about calculations needed to solve problems related to that topic, and wrote a sample problem and solution related to that topic. We confirmed that this task was not too long with the non–online-course students. In the *Survey Control* condition, students completed the same surveys that students in all of the other conditions completed without doing an intervention or summary activity. This condition was shorter than the other conditions and contained only questionnaires, so we did not pilot test it.

Discussion

The prototyping process provided information regarding which aspects of the intervention materials were too long or difficult and how to frame examples so they were representative of students' own experiences with math. The major ways in which we revised materials compared to prior studies were making the materials shorter and removing summarizing activities from the essay conditions. It was possible that these changes could weaken the impact of the utility-value essay prompts, but we wanted to maximize student engagement and needed to fit our materials within the time limitations provided by our partner school. The other major change was that, in the conditions with quotations (Utility Quotes + Evaluation and Utility Quotes + Essay), we wrote the quotations more informally compared with prior studies (i.e., Gaspard et al., 2015). The quotations also included more language indicating that students typically did not expect to use math in their lives. Specific examples of these changes are reported in Table 1; see Appendix for the final text of the quotations. Our intention was that the modifications we made would make it more likely that students would connect with the quotations and would be engaging, which could reduce attrition.

We knew that even with these changes, there was still potential for high attrition in the next phase of our research. We therefore decided that we would shorten our intervention measurement approach, compared to what previous researchers have done, by using individual items from previously published scales rather than using entire scales. This approach is consistent with a growing movement in the field to develop valid and reliable measures of classroom motivation and engagement that also minimize the amount of time that teachers and students devote to measurement (e.g., Kosovich, Hulleman, & Barron, 2017; Krumm et al., 2016). This approach might reduce the sensitivity of our measure of utility value to detect effects. However, the measure we chose was face valid, allowed us to maximize the time spent giving intervention activities, and reduced the possibility that students would drop out of the study before completing the post-intervention questionnaires.

Phase 3: Randomized field experiment

In the final phase of this research, we tested the refined intervention materials with high school students in online algebra and geometry courses using an experimental procedure in which students were randomly assigned to one of five conditions: Utility Essay, Utility Quotes + Evaluation, Utility Quotes + Essay, Summary Control, or Survey Control.

Method

Participants

Nine hundred thirty-one high school students taking algebra 1 and geometry courses in a large virtual high school participated on a volunteer basis. These students represented 7% of a larger group of students enrolled in the two courses whom we invited to participate in the study (total students invited = 13,756: algebra, $n = 6{,}329$; geometry, $n = 7{,}427$). We were able to match students' self-reported school IDs to demographic data from school records for 797 of the 931 study participants. The participants were 67.7% female, 46.4% White, non-Latinx; 19.5% Latinx; 18.0% African American; 5.6% Asian; and 10.5% other ethnicities. Students' mean age was 15.39, $SD = 1.38$, and 75.4% of participants attended a public or private face-to-face school as their predominant form of instruction (as opposed to being home-schooled).

Procedure

All enrolled participants in the algebra and geometry courses received an email from their course instructor asking them to complete an online activity about how math related to their lives. Volunteer participants were directed to an online survey and completed a brief measure of baseline interest and self-efficacy. Next, within each course, participants were randomly assigned to one of the five conditions and completed the activities associated with that condition for approximately 5 to 15 minutes. Materials for these conditions were those we had developed in the intervention prototyping phase. At the end of the study, participants completed a brief measure to assess perceived utility value and distraction during study activities.

Measures of baseline interest and self-efficacy

We assessed participants' baseline interest in algebra or geometry using the average of two items adapted from Hulleman and Harackiewicz (2009; e.g., "How interesting is what you're learning in algebra/geometry?"; $\alpha = .85$). We assessed participants' baseline self-efficacy using the average of two items adapted from Kosovich, Hulleman, Barron, and Getty (2015; e.g., "How confident are you that you can learn the material in algebra/geometry?"; $\alpha = .89$). Participants responded to all items using a five-point response scale from 1 (*not at all*) to 5 (*extremely*).

Measure of post-intervention utility value

We assessed participants' perceptions of utility value using one item from Hulleman and Harackiewicz (2009; "How useful is what you will learn in algebra/geometry class?"). We chose this item because it is included in many psychometrically sound utility value scales (e.g., Eccles, Wigfield, Harold, & Blumenfeld, 1993; Kosovich et al., 2015; Musu-Gillette et al., 2015) and because this item has been found to be the most sensitive to intervention effects (Kosovich, 2017). Participants responded using a five-point response scale from 1 (*not at all*) to 5 (*extremely*).

Measure of post-intervention distraction

Students reported the extent to which they felt distracted while they completed intervention activities by answering one question, "How distracted were you as you completed this activity today?" Responses were made on a five-point scale from 1 (*not distracted at all)* to 5 *(extremely distracted).*

Coding of written responses

In the four conditions in which students wrote responses (Utility Essay, Utility Quotes + Evaluation, Utility Quotes + Essay, and Summary Control), we used the Linguistic Inquiry and Word Count application (Pennebaker, Boyd, Jordan, & Blackburn, 2015) to measure students' use of personal pronouns in their writing. This variable was intended to be an indicator of students' personal connections to course material while writing. It is measured as the percentage of total words in students' written responses that are personal pronouns. Students who dropped out before completing the essays, did not write in answers, wrote in answers unrelated to the question, or wrote in answers that expressed the opposite of the prompt (i.e., I do not think math is useful) earned a score of 0 (no answers = 49.9% of students; nonsensical answers = 0.9%; oppositional answers = 4.2%).

Second, we coded the *strength of personal connections* in students' essays, which measured the extent to which students made specific and personal connections to their course material while writing essays. We coded this for the two conditions in which students wrote essays (i.e., Utility

Table 2. Descriptive statistics and correlations among motivation variables in the randomized field experiment phase.

	n	*M*	*SD*	*1.*	*2.*
1. Baseline Interest	931	2.44	1.09		
2. Baseline Self-Efficacy	931	3.27	1.08	.63*	
3. Post-Intervention Utility Value	682	2.99	1.24	.65*	.49*

*$p < .001$

Quotes + Essay and Utility Essay). A trained coder rated these essays on the following scale: 0 = no utility value (e.g., "I am learning about translations."); 1 = Nonpersonal utility value statement (e.g., "Geometry is helpful to make sure [a building] doesn't fall."); 2 = Personal, general utility-value statement (e.g., "If I was ever going to build something [geometry] would be helpful."); 3 = Personal, specific utility-value statement (e.g., "When I was moving my furniture around in my room, I actually did use [geometry] and a scale ratio to see if things would fit in there a certain way"); 4 = Personal, specific utility-value statement that elaborated on the connection being made (e.g.," "I am an artist so … it is very useful to know as much about [angles and] shapes as possible. This way I can make them properly."). This scale was adapted from one used in prior studies with longer written responses (e.g., Harackiewicz et al., 2016; Hulleman et al., 2010). Students who did not write in answers, wrote nonsensical answers that were unrelated to the topics of interest, or wrote in answers that expressed the opposite of the prompt (i.e., I do not think math is useful) earned a score of 0.

Results

Descriptive statistics and correlations

Descriptive statistics and correlations for self-report measures are presented in Table 2, descriptive statistics by condition on these measures are presented in Table 3, and descriptive statistics for the coding of written responses are presented in Table 4. Correlations among variables were in the expected directions. As a randomization check, we confirmed using linear regression and ANOVA that students' gender, age, school type, pretest interest, and pretest self-efficacy did not differ by condition.

Intervention participation and attrition

We received 931 surveys that represented unique, participating students from our target population. Participants reported low mean levels of distraction during the study (on a scale of 1 to 5, with 1 being the least distracted, for algebra, $M = 1.88$, $SD = 1.11$; for geometry, $M = 1.84$, $SD = 1.01$). However, this finding may underestimate the level of distraction because some participants dropped out of the intervention and thus did not provide data on the indicator of distraction (29.5%) or on the post-intervention outcome measure of value (22.5%; see Table 2). The condition with the largest amount of attrition was the Summary Control condition (47.1%), and the condition with the smallest attrition was the Survey Control condition (9.1%). Whether or not students dropped out was not significantly correlated with baseline interest, $r = .05$, or baseline self-efficacy, $r = .02$. The three intervention conditions fell in between the two control conditions and had similar attrition rates to one another (24.7% Utility Essay, 36.8% Utility Quotes + Evaluation, 31.7% Utility Quotes + Essay). Although the two control conditions had the highest and lowest attrition rates, the mean estimated utility value scores of both control conditions were similar. Thus it does not seem that the differential attrition by condition affected students' ratings of utility value.

Table 3. Unadjusted mean motivation scores by condition in the randomized field experiment phase.

	Survey Control			*Summary Control*			*Utility Essay*			*Utility Quotes + Evaluation*			*Utility Quotes + Essay*		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Baseline Interest	186	2.41	1.06	189	2.47	1.15	182	2.37	1.03	185	2.51	1.13	189	2.43	1.09
Baseline Self-Efficacy	186	3.39	1.08	189	3.27	1.06	182	3.10	1.08	185	3.21	1.09	189	3.36	1.06
Post-Intervention Utility Value	180	2.91	1.30	103	2.85	1.30	141	2.96	1.22	119	3.24	1.16	139	3.04	1.18

Note. The items used to assess baseline versus post-intervention self-efficacy and interest were different; score differences between these variables do not reflect changes in the overall level of the construct.

Table 4. Descriptive statistics for linguistic indicators by condition in the randomized field experiment phase.

	Summary Control			*Utility Essay*			*Utility Quotes + Evaluation*			*Utility Quotes + Essay*		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Percentage of Personal Pronouns	189	2.91	5.43	182	6.34	6.61	185	6.74	6.70	189	5.50	6.86
Coded Strength of Personal Connections	–	–	–	182	1.20	1.48	–	–	–	189	1.16	1.59

Note. The Survey Control condition is omitted because students did not produce written responses. For coded strength of personal connections, some conditions are omitted because students were not asked to write an example connecting their lives to course material. The percentage of personal pronouns variable was automatically coded by the Linguistic Inquiry and Word Count (LIWC) software.

Analysis strategy

We conducted an intent-to-treat analysis (Shadish & Cook, 2009) that evaluated the effects of the conditions for all students who received the intervention or control conditions regardless of whether or not they dropped out of the study. We evaluated data using linear multiple regression analyses in MPLUS, accounting for missing data using full-information maximum likelihood estimation. To make the estimation of missing data more accurate, we modeled missingness using the variables of ethnicity, age, and gender. We regressed utility value on a set of four dummy codes, which represented each intervention condition relative to a reference group (either the Survey Control or the Summary Control condition). We included course type (algebra versus geometry, dummy coded: algebra = 0, geometry = 1), pretest self-efficacy beliefs (standardized), and pretest interest (standardized) as covariates. Previous research has suggested that students' baseline competence-related beliefs or interest sometimes moderate the effects of motivational interventions (e.g., Hulleman & Harackiewicz, 2009; Hulleman et al., 2010; Rosenzweig & Wigfield, 2016). However, we found no significant interactions between any of the conditions and any of the three covariates on students' post-intervention utility value, so we trimmed these terms from the final model.

We also evaluated the effects of the different utility value intervention conditions compared to one another. To test this, we ran the same regression model just noted two more times. One additional model compared all conditions to the Utility Quotes + Evaluation condition, and a second compared all conditions to the Utility Quotes + Essay condition. Because these two models were exploratory, we employed a Bonferroni adjustment for them and reduced our threshold for significance to $\alpha = .025$.

Effects on post-intervention utility value

Figure 1 depicts estimated adjusted mean scores by condition and Table 5 presents regression results. The Utility Quotes + Evaluation condition showed the highest estimated posttest utility value, which was significantly higher than the mean score for the Summary Control condition ($\beta = 0.12$, $SE = 0.04$, $p = .003$) and the Survey Control condition ($\beta = 0.10$, $SE = 0.04$, $p = .004$). The Utility Essay condition showed a difference from the Summary Control condition, at significance, $\beta = 0.07$,

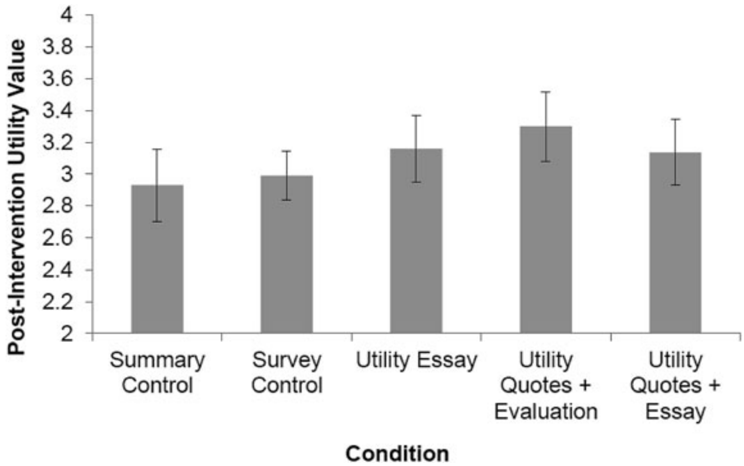

Figure 1. Mean adjusted utility value scores by condition in the Randomized Field Experiment phase. Each error bar represents +/- 2 standard errors of the adjusted mean. Adjusted score estimates are based on students' mean levels of self-efficacy and interest and the reference group of algebra class membership.

Table 5. Regression results for motivation with control conditions as reference groups in the randomized field experiment phase.

	Post-intervention Utility Value (versus Summary Control)		*Post-intervention Utility Value (versus Survey Control)*		*Percentage of Personal Pronouns (versus Summary Control)*	
	β	*SE*	*β*	*SE*	*β*	*SE*
Other Control Condition	0.02	0.04	−0.02	0.04	–	–
Utility Essay	0.07*	0.04	0.06	0.03	0.20**	0.04
Utility Quotes + Evaluation	0.12**	0.04	0.10**	0.04	0.27**	0.04
Utility Quotes + Essay	0.07	0.04	0.05	0.04	0.15**	0.04
Pre-Interest	0.57**	0.03	0.57**	0.03	−0.001	0.05
Pre-Self-Efficacy	0.14**	0.04	0.14**	0.04	0.04	0.05
Geometry Course Enrollment	−0.06*	0.03	−0.06*	0.03	−0.03	0.04

Note. Values shown are standardized regression coefficients. All study conditions were dummy coded as a 1 with either the Summary Control condition or Survey Control condition serving as the reference group and coded as a 0. Geometry course enrollment is dummy coded as a 1 with algebra enrollment serving as the reference group and coded as a 0. For the utility-value analysis, $n = 931$; for the personal pronouns analysis, $n = 707$.
*$p \leq .05$
**$p \leq .01$

$SE = 0.04$, $p = .05$, but the Utility Quotes + Essay condition did not, and neither of the two essay-based conditions showed significantly higher utility value than the Survey Control condition.

Evaluating the content of student essays

We evaluated the percentage of personal pronouns students used in their written responses using the Linguistic Inquiry and Word Count application (Pennebaker et al., 2015). As Table 5 reports, students in each of the three intervention conditions used a significantly larger proportion of personal pronouns than students in the Summary Control condition (for Utility Essay, $\beta = 0.20$; for Utility Quotes + Evaluation, $\beta = 0.27$; for Utility Quotes + Essay, $\beta = 0.15$).

Mean scores by condition on the coded strength of the personal connections variable can be found in Table 4. For this analysis, we evaluated data only from those students who also reported on utility value ($n = 141$ for the Utility Essay condition; $n = 139$ for the Utility Quotes + Essay condition). We found that students who had higher coded strength of personal connections reported higher utility value post-intervention in both the Utility Essay condition “r (139) = .27, $p = .01$” and the Utility Quotes + Essay condition “r (137) = .31, $p < .001$”.

Discussion

The goal of the randomized field experiment phase was to explore whether any of three utility-value interventions would increase students’ utility value for online math courses. To some extent the results were promising, because the utility value intervention in which students read and evaluated quotations (i.e., Utility Quotes + Evaluation) caused students to report higher utility value compared to both control conditions. Students who read and evaluated quotations also used more personal pronouns compared to the Summary Control condition, suggesting that the intervention helped students make personal connections to course material. However, results for the Essay-Based Intervention conditions did not show the same pattern, findings were based on a selective sample of students, and there was somewhat high attrition.

General discussion

In the sections that follow, we provide a candid discussion of the results we obtained and how they are a function of the design constraints of our partner school and the implementation choices that we made to address those constraints. Playing off the title of our paper, we first discuss the “promise” of the quotation and evaluation intervention and how our choice to conduct pilot work helped us observe this effect. Then we turn to the “pitfalls” we encountered, discussing why our implementation choices impaired our ability to understand fully the impacts of the utility value interventions we tested. We conclude with recommendations for designing and testing future utility value interventions.

Promises: The benefits of quotation evaluation activities in utility value interventions

The major conclusion we can draw from our results is that when students read and evaluated quotations from other students (i.e., the Utility Quotes + Evaluation condition), they showed significantly higher post-intervention utility value. This finding held in comparison to both control conditions, suggesting that it was not a function of simply asking students to write about their course material. It is important for several reasons. Online courses represent a growing segment of K–12 education that has been understudied in motivational intervention research, and educational administrators are eager to improve students’ retention and motivation for these courses. Second, our results replicate Gaspard et al.’s (2015) findings that a quotation evaluation intervention was more effective than an essay-based intervention for high school math students. Third, we obtained this finding despite students having no teacher oversight and little time to complete the intervention activities. Specifically, our intervention took 15 minutes or less; whereas Gaspard et al.’s (2015) similar intervention was 90 minutes in length for the initial session, followed by several, smaller booster activities that students completed with their homework. Our results are promising in showing an effective way to improve high school students’ valuing of STEM courses when time is limited. Results are also promising in suggesting that future researchers can develop utility value interventions that will be effective while meeting the needs of partner schools and requiring few resources to implement.

The Utility Quotes + Evaluation condition likely was successful for several reasons. First, the intervention provided students with opportunities to think about and evaluate how their course material related to their own lives (Gaspard et al., 2015). This conclusion is supported by the finding that students used a larger proportion of personal pronouns in the quotation-evaluation condition compared to control. Second, students may have been more engaged with the quotation-evaluation task than with a writing-based prompt or a summary task. Gaspard et al. (2015) argued that high school students generally dislike writing essays and they benefit from having questions that provide scaffolding about how to reflect on the utility value of what they are learning. Third, students had to think about and respond to each of the quotations they read. Therefore, they were prompted to consider carefully each example rather than skimming through all of the examples together (or not reading any examples) and then writing an essay.

A critical step in this process was our implementation choice to engage in careful pilot work to develop the wording of the quotation and evaluation intervention condition. We took great care to adapt the quotations so that they targeted utility value in a way that would reduce potential reactance from students and maximize engagement. In particular, through an iterative design-based prototyping process, we developed examples in the quotation evaluation intervention, which represented real ways that students would use math and sounded like real students' writing (see Table 1). This also maximized the likelihood that students would be engaged with the study materials and would experience neutral or positive reactions while completing them. Indeed, we were able to reduce student attrition from 60% in the prototyping phase to an average of 30% in the field experiment phase. We recommend that future utility value intervention researchers engage in pilot work to edit and refine their intervention prompts so that they are appropriate for a given educational context.

Pitfalls: Negative implications of some implementation choices

Two aspects of the results we obtained are less promising. The first is that we failed to replicate past research showing benefits of essay-based utility value interventions (i.e., Utility Quotes + Essay and Utility Essay). One condition showed an at-significance difference in post-intervention utility value compared to the Summary condition, but both conditions did not differ significantly from the Survey Control condition. We did observe that students in the two essay-based intervention conditions used significantly more personal pronouns than did students in the Summary condition, and students' amount of articulated specific and personal connections to course material was associated with their post-intervention utility value. That pattern of results is consistent with prior research and theory regarding how utility value interventions are expected to benefit students. However, an alternative interpretation is that students who perceived more utility value simply were more likely to articulate this in their essays.

What does this failure to clearly replicate past results mean? It might be that the two essay-based intervention conditions were not strong enough to help students make specific personal connections to their course material as compared to the quotation evaluation condition. Perhaps our revised essay prompt was too short for students to engage sufficiently with their assigned tasks. In a similar vein, our essay prompts omitted a subtask from prior studies asking students to summarize their course material before making connections to it; perhaps, as a result, students did not think about their material sufficiently to connect to it in their essays. Alternatively, perhaps the two essay-based intervention conditions did improve students' perceived utility value, but our implementation choice to use a one-item measure of utility value was not sensitive enough to capture these differences consistently.

We cannot say which interpretation is more plausible, because one of the major pitfalls of this study is that in trying to keep our intervention short, we did not collect adequate data to understand precisely how the different intervention conditions impacted students. We did not collect any measures exploring whether our changes to the essay-based prompts during the prototyping phase (e.g., shortening intervention materials from prior studies; omitting the summary task) impacted students' perceptions of utility value. Additionally, we used only a one-item measure of utility value. We chose this measure for its practical benefits and face validity. However, Kosovich

Table 6. Major intervention implementation choices and consequences.

Implementation Choice Made	Rationale	Consequence	Recommendations for Future Research
Pilot work	• Working with diverse population in terms of math experience • Make materials engaging and natural sounding so as to reduce reactance and attrition	• Reduced negative reactions to materials and reduced attrition between prototyping and field experiment	• Conduct pilot work to ensure materials are appropriately worded • Can use pilot work to collect additional data on potential differences between intervention materials
Very brief materials	• Make materials engaging and easy to complete • School required that intervention be brief and delivered in one session	• Shorter essay prompts showed mixed results; unclear what role the shorter materials played in these effects.	• Adapt material length and dosage from prior studies as closely as possible if implementing in new context.
Limited data collection	• Maximize time spent giving intervention in the limited time frame • School required that intervention be brief and delivered in one session	• Hard to evaluate psychometric properties of utility value measure • Unclear why essay-based interventions showed mixed results on utility value • Unclear how shorter essay prompts impacted students	• Use caution when utilizing a oneitem measure to assess central outcomes • Collect data on potential differences that could shed light on mixed effects between conditions • Collect data about how changes to intervention materials (e.g., shortening them) impacts students
Voluntary participation (no incentive)	• School did not allow incentives or for intervention to be part of course	• Obtained low response rate that likely is not representative of population • Limits generalizability of results • Low power to examine moderator effects	• Offer incentive (e.g., extra credit) for participating in intervention or embed intervention into course • Speak with institution about importance of obtaining representative sample

et al. (2017) have reported that brief measures of motivation cannot account for all the nuances of a particular motivational construct. Furthermore, it is difficult to evaluate one-item measures' reliability, distinctness from other items, and sensitivity to group differences using traditional methods (Kosovich et al., 2017). It is therefore possible that the measure we used did not capture utility value fully or reliably.

A second major pitfall is that we cannot determine whether *all* online high school students are likely to benefit from utility value interventions that include quotation evaluation or whether results are specific to the students who participated in this study. Students who clicked on the link to complete our study represented a small proportion of the total students in our target algebra and geometry courses (7%). It is possible that more engaged, self-efficacious, and interested students chose to click on the link compared to the target population of all online algebra and geometry students at the virtual school. This self-selected student sample might be more likely to benefit from interventions than other students because they would have been more likely to engage deeply with intervention materials. However, the opposite interpretation also is possible. Students who are less high-achieving and less highly motivated to learn often benefit more from utility value interventions (e.g., Hulleman & Harackiewicz, 2009; Hulleman et al., 2010). Therefore, with a larger proportion of participating students, we also might observe larger benefits of a quotation-based utility value intervention. During our information-gathering and prototyping phases, we acknowledged that the constraint against incentivizing student participation might reduce the possibility of obtaining a representative sample. However, we valued the unique opportunity to work in online math classes and anticipated that our results could be insightful for this population even if our sample was not fully representative.

Recommendations for adapting utility value interventions when facing implementation constraints

It is critical that researchers consider the practical feasibility of an intervention in order to conduct research that is informative for schools (Kaplan et al., 2012). However, it can be challenging to balance practical and scientific concerns in an intervention setting. In the present study, we tried to adapt intervention materials to see if they would target students' value when transferred to a novel and logistically challenging implementation context. In some ways our choices were beneficial, but in other ways they constrained the conclusions that we were able to draw. In Table 6, we summarize each of the major implementation choices we made, our rationale for the choice, and the consequence of it. In this section, we discuss our recommendations regarding each of these choices for future utility value intervention research. The process that we used to adapt intervention materials and our recommendations for the future might seem less or more novel based on readers' epistemological stances toward intervention design and implementation. For those who assume that the same intervention materials can be implemented across educational settings without adaptation, our findings will shed light on the importance of adapting materials for use in new contexts. In contrast, our discussion will resonate with the perspective of researchers who take a highly contextualized approach to motivational intervention research. Regardless, we hope our recommendations are informative for all readers (no matter their epistemological stance) in providing a specific illustration of negotiating practical and scientific concerns in the intervention setting.

First, we recommended above that researchers conduct pilot work to revise materials and prompts with respect to students' needs in the target educational context. As a corresponding recommendation, we advise researchers to think carefully about how any changes made to materials might impact students' engagement and perceptions of utility value and collect data regarding these possibilities. Similarly, if researchers want to include multiple intervention conditions, they could collect data to explain potential differences in how those conditions might impact students. This could include asking students to provide open-ended responses regarding the positive and negative aspects of different materials, asking students to report on their perceptions of materials' difficulty and interestingness or assessing students' utility value, interest, or competence-related beliefs using questionnaires. In the present study, we did not collect sufficient data to know if shortening the intervention prompts made them too brief to engage students deeply with the essay tasks or if students engaged differently with the essay versus the quotation-evaluation tasks. We were not able to collect this information due to time constraints in our randomized field experiment, and ultimately, this limited our ability to explain the null findings we obtained regarding the essay-based utility value intervention conditions.

A related recommendation is to consider carefully the costs and benefits of compromising on aspects of intervention timing and dosage. We chose to implement a brief, single-session intervention due to time constraints imposed by our partner school. Prior work suggested that utility-value interventions were effective for college students after only a single session (Canning et al., 2018), but that effect had been found with a different group of students and using a longer intervention task. Ultimately, we do not know whether the two essay-based interventions were ineffective because they did not last long enough. Our decision to compromise on dosage and timing resulted in a mixed picture of effectiveness. When possible, researchers may want to try to replicate the timing and dosage of prior intervention materials as closely as possible. By doing so, they can rule out the possibility that any observed mixed or null effects were due to the change in intervention timing and dosage. Alternatively, researchers could try to collect ancillary data to understand how shorter versus longer intervention implementation might affect students, either during the full intervention administration or through pilot work.

Third, one-item measures of dependent variables may achieve practical goals, but researchers are wise to be cautious if they plan to utilize these measures to assess central outcomes of interest (i.e., utility value). Our one-item outcome measure was sufficient to demonstrate that one of the intervention conditions promoted utility value, but it is unclear why the other two conditions did not show effects on the measure. Researchers need to be aware of the limits of single-item measures, such as not being able to estimate reliability and validity using traditional methods, and the potential that the measure lacks complete construct validity. When possible, researchers could try to measure major outcomes with at least a two- or three-item measure. This makes it easier to assess the quality of the measure used with the students in their sample, and it still would maintain practical utility. It also increases the likelihood that any observed mixed or null effects would be due to factors beyond measurement.

Finally, we recommend that researchers discuss with practitioners their plans for recruitment of students during interventions and stress the importance of obtaining a representative sample. We do not know if our results are generalizable to all online students because it is likely that our sample of students was not representative of the target population. Researchers can work toward obtaining high participation by offering students a traditional incentive or small amount of course extra credit to participate in a study or by embedding materials into a course itself. Researchers also need to obtain as much data as possible regarding students who were *not* in the sample in order to evaluate the representativeness of the students who were included in the study. These steps are critical to help ensure that researchers do not waste time and resources collecting data that cannot clearly be generalized to other populations.

Conclusion

Despite the pitfalls of compromising on aspects of intervention implementation, this study provided useful information about implementing utility value intervention in an online setting. In particular, we discovered that we could boost perceptions of math value for students randomized to one of our utility-value conditions. We also learned the boundary conditions of timing and dosage, with a one-time, 10-minute intervention being perhaps at the lower limit of the dosage needed to improve student outcomes. Finally, we were able to make visible our lessons learned, particularly the experience of collaborating with an educational institution, so that future researchers can learn from our example.

Our goal is not to recommend that researchers avoid working in logistically challenging educational contexts. These are settings in which students can benefit much from interventions, researchers can learn a lot, and educational psychology research can be fruitfully applied to address real-world problems. Rather, we hope that we can provide a specific example to researchers regarding *how* to navigate implementation challenges in educational settings, because doing so can be quite rewarding. We recommend that researchers take precautions in advance to avoid making implementation choices that have practical benefits but undermine the scientific validity of their studies. We did some, but not enough, work to address such pitfalls pre-emptively in the present study. We recommend that future intervention researchers learn from both the positive and negative aspects of our implementation choices, in order to create utility value intervention that are both practically useful and scientifically meaningful in future work.

Funding

The writing of this article was supported by a National Science Foundation (NSF) Postdoctoral Research Fellowship 1714481 and an NSF Graduate Research Fellowship DGE 1322106 to the first author; by NSF awards

1534835 and 1252463 to the second author; by Grant R305B090002 to the fourth author from the Institute of Education Sciences, U. S. Department of Education through the University of Virginia; by Grant R305B150003 to the fifth author from the Institute of Education Science, U.S. Department of Education through the University of Wisconsin-Madison; and by support from the PERTS Research Center at Stanford University. Any findings expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or Institute for Education Sciences.

References

Al-Ubaydli, O., & List, J. A. (2013). *On the generalizability of experimental results in economics* (No. *w19666*). Cambridge, MA, USA: National Bureau of Economic Research.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., … & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, *74*(3), 379–439. doi: 10.3102/00346543074003379

Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., & Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. *Journal of Educational Psychology 110*(6), 834–849. doi:10.1037/edu0000244.

Cavanaugh, C. S., Gillan, K. J., Kromrey, J., Hess, M., & Blomeyer, R. (2004). *The effects of distance education on K-12 student outcomes: A meta-analysis*. Naperville, IL: Learning Point Associates. Retrieved from http://eric.ed.gov/?id=ED489533

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, *13*(1), 15–42. doi: 10.1207/s15327809jls1301_2

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, *8*(2), 149–168. doi: 10.1080/19345747.2014.906009

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. doi: 10.1177/1529100612453266

Eccles-Parsons, 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 motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*(3), 830–847.

Gaspard, H., Dicke, A., Flunger, B., Brisson, B., Hafner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. doi: 10.1037/dev0000028

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.

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*, 745–765. doi: 10.1037/pspp0000075

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, *69*.

Hulleman, C. S. (2007). *The role of utility value in the development of interest and achievement*. Madison: University of Wisconsin-Madison. Retrieved from ERIC database (ED498264)

Hulleman, C. S., & Barron, K. E. (2016). Motivation interventions in education: Bridging theory, research, and practice. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., pp. 160–171). New York, NY: Routledge.

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*, 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

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. doi: 10.1037/edu0000146

Jacobs, J., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Ontogeny of children's self-beliefs: Gender and domain differences across grades one through 12. *Child Development*, *73*, 509–527. doi: 10.1111/1467-8624.00421

Kaplan, A., Katz, I., & Flum, H. (2012). Motivation theory in educational practice: Knowledge claims, challenges, and future directions. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook: Vol. 2. Individual differences in cultural and contextual factors* (pp. 165–194). Washington, DC: American Psychological Association.

Kelley, T., & Kelley, D. (2013). *Creative confidence: Unleashing the creative potential within us all*. New York, NY: Crown Business.

Kosovich, J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity for the expectancy-value-cost scale in middle school. *Journal of Early Adolescence*, *35*(5–6), 790–816. doi: 10.1177/0272431614556890

Kosovich, J. J. (2017). *Pragmatic measurement for education science: A method-substance synergy of validation and motivation* (Unpublished doctoral dissertation). Charlottesville, VA: University of Virginia.

Kosovich, J. J., Hulleman, C. S., & Barron, K. E. (2017). *Measuring motivation in educational settings: A case for pragmatic measurement. To appear in K. A. Renninger & S. E.* Hidi (Eds.), *The Cambridge handbook on motivation and learning* (pp. 39–60). New York, NY: Routledge.

Kosovich, J. J., Hulleman, C. S., Phelps, J., & Lee, M. (2016, April 9). *Improving student success in developmental math: A utility value intervention. Poster presented at the meeting of the American Educational Research Association Annual Conference*, Washington, DC

Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. *Journal of Learning Analytics*, *3*(2), 116–138. doi: 10.18608/jla.2016.32.6

Musu-Gillette, L. E., Wigfield, A., Harring, J., & Eccles, J. S. (2015). Trajectories of change in student's self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, *21*(4), 343. doi: 10.1080/13803611.2015.1057161

O'Dwyer, L. M., Carey, R., & Kleiman, G. (2007). A study of the effectiveness of the Louisiana Algebra I online course. *Journal of Research on Technology in Education*, *39*(3), 289–306. doi: 10.1080/15391523.2007.10782484

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin: University of Texas at Austin. Retrieved from https://utexas-ir.tdl.org

Picciano, A. G., & Seaman, J. (2009). *K-12 online learning: A 2008 follow-up of the survey of US school district administrators*. Newburyport, MA: The Sloan Consortium.

Rice, K. L. (2006). A comprehensive look at distance education in the K–12 context. *Journal of Research on Technology in Education*, *38*(4), 425–448. doi: 10.1080/15391523.2006.10782468

Rosenzweig, E. Q., & Wigfield, A. (2016). STEM motivation interventions for adolescents: A promising start, but further to go. *Educational Psychologist*, *51*(2), 146–163. doi: 10.1080/00461520.2016.1154792

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607–629. doi: 10.1146/annurev.psych.60.110707.163544

Sheehy, K. (2012). *States, districts require online ed for high school graduation*. US News and World Report.

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

Spielhagen, F. R. (2006). Closing the achievement gap in math: The long-term effects of eighth-grade algebra. *Journal of Advanced Academics*, *18*(1), 34–59. doi: 10.4219/jaa-2006-344

Stone, C. (1998). Leveling *the playing field: An urban school system examines equity in access to mathematics curriculum. Urban Review*, *30*(4), 295–307. doi: 10.1023/A:1023246618407

Tibbetts, Y., Harackiewicz, J. M., Priniski, S. J., & Canning, E. A. (2016). Broadening participation in the life sciences with social-psychological interventions. *CBE Life Sciences Education*, *15*(3), 1–10. doi: 10.1187/cbe.16-01-0001

Watson, J., Pape, L., Murin, A. L., Gemin, B., & Vashaw, L. (2014). *Keeping pace with K-12 digital learning: An annual review of policy and practice*. Evergreen, CO: Evergreen Education Group. Retrieved from www.kpk12.com

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: 10.1111/j.1467-8624.2004.00757.x

Wigfield, A., Tonks, S. M., & Klauda, S. L. (2016). Expectancy-value theory. *In K. R. Wentzel & D. B. Miele (Eds.), Handbook of motivation of school (2nd ed., pp. 55–74). New York, NY: Routledge.*

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

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., … Trott, J. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. doi: 10.1037/edu0000098

Appendix: Complete text of final quotations used in the randomized field experiment phase

Algebra

"I never thought math skills were important for me to learn, but they do come in handy sometimes. At the movie theater they have a deal to pay a little more and get bottomless popcorn and soda. I never can figure out whether it's a rip-off or not. So I set up an algebra equation to figure out how much money I was spending on each cup of popcorn with the regular and bottomless sizes. I actually found out that it wasn't worth the money for the bigger size, plus I usually eat more than I need to, so I save money and I don't overeat!" —Derek, 9th grade

"I didn't realize how useful math could be until I took algebra. Whenever I'm watching sports, the announcers always give statistics about different players on my team and on the opposing team. I know from algebra that an average is total points divided by the number of observations. I'm a huge basketball fan, so I can calculate the average points and fouls for players on both teams and figure out how much better my team's stats need to be to beat the other team, without having to wait for the announcers to tell me." —Stacey, 9th grade

"I used to think algebra was sort of useless outside math class, because I want to be an artist and definitely not go into any math job. But I'm taking an art class and we actually used an algebra equation to make sure we could fit our painting on the canvas. My teacher also showed us how to calculate how much all the paint and clay and stuff costs for all of her classes and how much we would need for our projects. It was cool to know she uses math because I never thought an artist would need to know anything about math, but I guess I do!" —Omar, 9th grade

"I volunteer with Habitat for Humanity, which builds homes for families who can't afford them. In order to build a house, we need materials like wood and nails, and sometimes I go with our supervisors to buy that stuff. I was surprised when I realized that I could use algebra to help them do this job better. I looked up the cost of housing materials from different stores online. Then I made an algebra equation to figure out how expensive the materials would be at different stores based on how much Habitat for Humanity needed, so we could go to the place with the best deal. That way we were able to build as many houses as possible." —Rachel, 9th grade

Geometry

"I have always liked playing basketball, but I never thought much about why different shots wouldn't go in, or how you can get trapped in different parts of the court. Taking geometry actually helped me play better because I could understand that stuff better. Once I learned about angles I could kind of explain why I might make a mistake shooting during a game. I learned I have to be standing at a good angle relative to other players to throw the ball in the net, or to make space to get off the right pass. I also have to shoot the ball a certain way with just the right arc." —Derek, 9th grade

"I never thought math skills were important for me to learn, but they can come in handy. I post a lot on Instagram with my friends, but sometimes it's hard to get everyone into a picture or I take it at an unflattering angle for some friends. Now that I understand angles from geometry class, it's been a lot easier for me to get good shots (pictures straight on aren't usually as flattering as pictures from above). Plus, I almost never cut anyone out of group shots anymore because I knew that you can increase the area of the photo by using angles to create depth, which provides more surface area by increasing the length of the picture frame!" —Stacey, 9th grade

"I used to think geometry was sort of useless outside of math class, because I want to be an artist or graphic designer and definitely not go into any job where you use a lot of math. But let me tell you, when I work for a long time on a painting and realize that the whole thing won't fit on the canvas, it really messes up the whole day. Geometry is actually really important for art or graphic design because I will need to understand how to scale different shapes to avoid making things that are too big. This requires knowing about shapes, area, and volume." —Omar, 10th grade

"I didn't realize how useful math could be until I took geometry. For instance, I volunteer with Habitat for Humanity, which builds homes for families who can't afford them. When I'm volunteering, certain pieces of wood for the frame of the house have to be at an exact right angle with each other, but others need to be parallel with one another. And some pieces of wood need to be a certain size to match up with the other beams of the house. Even though I am not in charge of the building process, it really helps me to know about angles, shapes, and area because I understand how to build stronger houses. Otherwise people might move into a new home that would fall apart right away!" —Rachel, 11th grade