



It's Not *That* You Said It, It's *How* You Said It: Exploring the Linguistic Mechanisms Underlying Values Affirmation Interventions at Scale

Nia M. M. Dowell 
University of California

Timothy A. McKay 
University of Michigan

George Perrett
New York University

Over the last decade, psychological interventions, such as the values affirmation intervention, have been shown to alleviate the male-female performance difference when delivered in the classroom, however, attempts to scale the intervention are less successful. This study provides unique evidence on this issue by reporting the observed differences between two randomized controlled implementations of the values affirmation intervention: (a) successful in-class and (b) unsuccessful online implementation at scale. Specifically, we use natural language processing to explore the discourse features that characterize successful female students' values affirmation essays to gain insight on the underlying mechanisms that contribute to the beneficial effects of the intervention. Our results revealed that linguistic dimensions related to aspects of cohesion, affective, cognitive, temporal, and social orientation, independently distinguished between males and females, as well as more and less effective essays. We discuss implications for the pipeline from theory to practice and for psychological interventions.

Keywords: values affirmation intervention, natural language processing, gender differences in STEM, educational data science

In the American higher education system, achievement gaps between male and female students persist despite gradual progress, and are particularly pronounced in STEM fields (i.e., science, technology, engineering, and mathematics; Miyake et al., 2010). For instance, across several STEM disciplines women consistently earn lower exam grades and lower scores on standardized tests of conceptual mastery (Brewer et al., 2010; Creech & Sweeder, 2012; Eddy et al., 2014; Matz et al., 2017; Pollock et al., 2007; Tai & Sadler, 2001). This is counterintuitive given that an important systematic bias exists in the population of males and females who attend college, wherein college-bound women systematically have higher high school GPAs than men who attend college (Conger & Long, 2010). This implies that, all else being equal, females ought to do better in their college classes than males (Eddy & Brownell, 2016). Despite this, research has consistently shown gendered performance differences (GPDs) that favor males, with male students outperforming their female counterparts in STEM courses. In particular, these observed GPDs in education endure even

when accounting for various measures of prior performance, including high school GPA, standardized tests, and prior college performance (Eddy & Brownell, 2016; Koester et al., 2016; Matz et al., 2017). While the causes and consequences of underachievement of female students in STEM are numerous and complex, the GPD has undoubtedly contributed to women remaining underrepresented in leadership roles across all STEM disciplines (National Research Council of the National Academies, 2011; National Science Foundation, 2019). Social identity threat has consistently been shown to be one factor which contributes to these GPDs and features a psychological basis (Dasgupta & Stout, 2014; Steele et al., 2002).

To address this issue, the current study focuses on evaluating the effectiveness of the values affirmation (VA) intervention for reducing stereotype threat and improving performance for female students in STEM. We explicitly focus on gender and not on other demographic variables (e.g., race, ethnicity, socioeconomic status) because the STEM courses under investigation have shown to have significant GPD, wherein



male students consistently outperformed their female counterparts. Toward this effort, we provide a novel assessment of student-generated VA essays using Educational Data Science and Learning Analytics techniques. In particular, we capture the language and discourse properties of students' VA essays using two established natural language processing (NLP) tools, Coh-Metrix (Graesser et al., 2004; McNamara et al., 2014) and linguistic inquiry word count (LIWC; Pennebaker et al., 2015; described in a later section). We explore the differences in the content of affirmation essays as a function of gender and successful and not successful VA intervention implementations. In doing so, we demonstrate how these two analytical techniques complement each other in the assessment of VA interventions. Although both are established analytical approaches within the learning analytics community, thus far, the unique combination of these approaches has not been utilized in the context of educationally focused psychological interventions. As such, this study provides unique evidence on this issue by reporting the observed differences between two randomized field implementations of the VA intervention at scale: (a) a successful traditional in-class intervention and (b) an unsuccessful online implementation.

The subsequent sections of the article are organized as follows. First, we provide a discussion of the psychological interventions situated within the context of relevant literature on the VA intervention and the underlying theoretical framework. Second, we move on to outline the promise of NLP for psychological interventions situated within the limited current efforts. We then provide an overview of the current research, before moving into the methodological features of the current investigation, including the principal component analyses (PCAs) that were used to identify specific writing profiles and mixed effects analyses to address the four research questions. Finally, we conclude the article with a detailed discussion of the results in the context of theory, as well as a general discussion of the theoretical, methodological, and practical implications for peer interaction research.

Psychological Interventions at Scale: A Way Forward?

Stereotype threat is a well-established social-psychological phenomenon. When an individual is placed in an evaluative environment in which they know others might expect them to confirm a negative stereotype (e.g., implicit stereotypes that engineering is a masculine field), they expend some cognitive resources on this concern, modestly reducing their ability to perform. Indeed, several studies suggest that students who feel at risk of upholding stereotypes or being judged based on stereotypes (i.e., stereotype threat) experience lower academic performance (Jordt et al., 2017; Nguyen & Ryan, 2008; Steele & Aronson, 1995).

Struggling with the issues induced by stereotype threat, either consciously or unconsciously, can prove detrimental for student performance by reducing working memory (Schmader

& Johns, 2003), which can activate hypervigilance (Forbes et al., 2008), and consequentially may distract students from tasks. The issues brought on by stereotype threat can be particularly detrimental for performance on challenging tasks (Beilock et al., 2007), such as high-stakes exams that require more of a student's mental faculties. The implications of students experiencing stereotype threat are not limited to short-term impacts, such as those on working memory. Indeed, stereotype threat can have nontrivial, long-term impacts, such as students distancing themselves from a discipline with which they once identified (Dasgupta, 2011; Dasgupta et al., 2015; Fogliati & Bussey, 2013; Thoman et al., 2013; van Veelen et al., 2019).

This disassociation, coupled with lower performance, could contribute to a student's decision to leave STEM and the resulting underrepresentation of minorities and women in STEM. For instance, Dasgupta et al. (2015) investigated the experience of female engineering students in teams of varying gender ratios: female-minority, sex-parity, and female-majority and found that female students participated more actively, and felt less threat/anxiety in female-majority groups than female-minority groups with sex-parity groups in-between. Moreover, when assigned to female-minority groups, women who harbored implicit masculine stereotypes about engineering reported less confidence and engineering career aspirations.

Implications of Stereotype Threat for Women in STEM

The long-term consequences highlighted by aforementioned studies and other research (e.g., Cheryan et al., 2009; Dasgupta, 2011; Dasgupta & Stout, 2014; London et al., 2011; Thoman & Sansone, 2016; van Veelen et al., 2019) raises several concerns. A growing proportion of employment in the United States requires expertise in STEM. Despite the demand for a STEM-educated workforce, insufficient numbers of U.S. college graduates have STEM expertise, producing a substantial and persistent gap between demand and supply (National Science Board, 2016; Skrentny & Lewis, 2013). This workforce shortage problem is intertwined with an equity problem because the undersupply of Americans who have STEM degrees is larger for women than for men (Dasgupta & Stout, 2014; National Science Board, 2015; National Science Foundation, 2019). Fewer women pursue academic majors and jobs in STEM relative to their proportions in the U.S. population, even though these jobs are growing rapidly, lucrative, and of high value. The relative scarcity of women entering and persisting in STEM majors in college limits their opportunities to access high-demand jobs in science, technology, and engineering after graduation, slowing down socioeconomic mobility.

Clearly, women are untapped human capital that, if leveraged, could increase the STEM workforce substantially. Accomplishing this goal involves identifying academic stages in the STEM pipeline, where women are less likely to

enter STEM fields and more likely to exit these fields than men, and developing interventions to address this “leaky pipeline.” There have been considerable research efforts devoted to using psychological interventions to address this problem. Indeed, since the discovery of stereotype threat in the 1990s, social psychologists have developed a variety of interventions, which reduce its effects during evaluations. These include interventions such as the growth mind-set, utility-value belonging, and VA, which is the focus of the current research. Walton (2014) referred to them as “wise interventions” because they are wise to specific underlying psychological processes that contribute to social problems or prevent people from flourishing. Wise interventions are brief, low-cost interventions that can be implemented in a variety of contexts and address a psychological need or process that is responsible for negative outcomes (Casad et al., 2018). In recent years, an important body of literature has emerged to explain why such brief interventions may create lasting impacts (see Harackiewicz & Priniski, 2018, for a review). However, far less research has explored the underlying mechanisms of these psychological interventions that result in more or less beneficial outcomes for women. To address this issue, the present study focuses on evaluating the effectiveness of the VA intervention for reducing stereotype threat and improving performance for female students in STEM.

Values Affirmation Intervention

The VA intervention is based on *self-affirmation theory* (Steele, 1988), which argues that individuals are motivated to maintain an overall sense of self-integrity. That is, how individuals maintain the integrity of the self, especially when it comes under threat, forms the heart of self-affirmation theory (Aronson et al., 1999; Sherman & Cohen, 2006; Steele, 1988). Under this perspective, self-affirmations bring about a more expansive view of the self and an individual’s available resources (see Cohen & Sherman, 2014, for a review). They can involve simple everyday activities. In this context, spending quality time with friends, participating in volunteer activities, or attending religious services all aid in securing a sense of adequacy in a higher purpose. According to the self-affirmation theory, these affirmations remind individuals of psychosocial resources beyond a specific threat and as such broaden their perspective beyond it (Sherman & Hartson, 2011). Under normal circumstances, people tend to narrow their attention on an immediate threat (e.g., the possibility of not meeting expectations), a response that promotes swift self-protection. But when self-affirmed, students are able to reorient and see the many anxieties of daily life in the context of the big picture (Schmeichel & Vohs, 2009). As such, the specific threat and the associated implications for the self become less potent and attract less attention.

The VA intervention has been a widely used strategy to improve educational outcomes (Casad et al., 2018; Harackiewicz & Priniski, 2018). Although several versions

of self-affirmation exist, the most examined experimental manipulation has students write about core personal values (McQueen & Klein, 2006; Napper et al., 2009), this implementation was also used in the current research. Personal values are the internalized standards used to evaluate the self (Cohen & Sherman, 2014). During the intervention, students first review a list of values and then choose a few values most important to them. The list typically excludes values relevant to a domain of threat (e.g., physics, biology, etc.) in order to broaden a student’s focus beyond that context. For example, if a student experiences threats to their identity in an important academic domain, such as a woman taking a physics test, then their self-integrity around this topic is called into question. To buffer against threatening negative gender stereotypes (e.g., women are bad at physics), physics-related information would be excluded from the list. Students then write a brief essay about why the selected values are important to them and a time when they were important. Thus, a key aspect of the affirmation intervention is that its content is self-generated text and tailored to tap into each student’s particular valued identity (Sherman, 2013). Often students write about their relationships with friends and family, but they also frequently write about religion, humor, and kindness. A central tenet of the VA intervention is that when they affirm their core values in a threatening environment, students reestablish a perception of personal integrity and worth, which in turn can provide them with the internal resources needed for coping effectively (Miyake et al., 2010).

The VA intervention has been shown to have a beneficial impact on closing achievement gaps in STEM for threatened groups, such as African American and Hispanic middle and high school students (Borman et al., 2020; Cohen et al., 2009; Sherman et al., 2013), undergraduate minority students (Brady et al., 2016; Jordt et al., 2017), first-generation (FG) college students (Harackiewicz et al., 2014), and women (Miyake et al., 2010; Walton et al., 2015). For instance, Brady et al. (2016) explored the long-term effects of a VA intervention. In the intervention condition, students picked one option from a preselected list of personal values and wrote about why that value was important on a personal level. Two years later, the same students were recruited for a follow-up. Interestingly, their findings showed that the racial achievement gap among Latinx students was reduced and their grades increased. Brady et al. attributed these results to more self-affirming and less self-threatening thoughts and feelings in response to adversity in school. The VA intervention has also been shown to be beneficial for women in STEM. Walton et al. (2015) explored the VA intervention with women in engineering found similarly positive effects. Specifically, self-affirmation helped women in engineering improve their academic attitudes as well as their GPAs. Interestingly, Walton et al. found that women who self-affirmed developed stronger gender identification, experienced less threat, and performed on par with their male peers on mathematics tests.

Most relevant for the current research is Miyake et al.'s (2010) study, which tested the effectiveness of a VA intervention with women in an undergraduate physics course. The study was a randomized, double-blind study in which students were assigned to write about their most important (intervention group) or least important (control group) values two times during the course. Miyake et al. found that female students in the intervention group improved their course grade by a full letter grade, on average, and improved their scores on a standardized physics test.

While encouraging, the positive results of the VA intervention have been largely limited to implementations within a single classroom or lab experiment. Efforts to move beyond a boutique remedy and close achievement gaps for large numbers of students have been inconsistent at best (Borman et al., 2018; Serra-Garcia et al., 2020), and unsuccessful at worst, highlighting the potential fragility of the VA in educational settings at scale, and the need for new quantifiable measures and evidence regarding the necessary conditions for effective VA interventions (Hanselman et al., 2017). Implementation fidelity and intervention processes have been used as a way to explain the inconsistencies of results (Bradley et al., 2015; Yeager & Walton, 2011). These investigations have primarily focused on external and static features such as implementation and delivery details (e.g., timing of the intervention, manner in which the intervention is framed) and contextual conditions (e.g., location of the writing—identity threats “in the air” in a particular setting). However, there is a comparatively limited body of research that has explored the actual content of student's essays (e.g., Tibbetts et al., 2016). This is surprising given that the dynamic, cognitive, and psychological mechanisms are externalized in the language and discourse features that characterize students' VA essays. The current research addresses this gap by leveraging automated NLP and computational modeling to characterize the linguistic features of students' VA essays which are related to more or less beneficial outcomes.

Text as Data: Linguistic Analysis in Psychological Interventions

Student-generated written responses are a critical component of many psychological interventions, including the values affirmation intervention (Akcaoglu et al., 2018; Riddle et al., 2015). Students' essays produced during such interventions can provide a valuable window into the processes that may contribute (more or less) to the beneficial effect of interventions. However, to date, there has been only a handful of studies that have investigated the language and discourse features underlying psychological interventions more broadly (Harackiewicz et al., 2014; Klebanov et al., 2017), and the VA intervention in particular (Hanselman et al., 2017; Shnabel et al., 2013; Tibbetts et al., 2016).

Some researchers have relied on more conventional approaches that require human examination (i.e., manual content analysis; Krippendorff, 2003) to characterize the content of student's intervention essays (e.g., Borman et al., 2018; Hanselman et al., 2017; Harackiewicz et al., 2016; Shnabel et al., 2013). Many of these studies coded content toward a goal of manipulation checks, wherein essays were coded to assess the degree to which they showed evidence of self-affirming reflection (e.g., Hanselman et al., 2017) or the level of utility value articulated in an essay (e.g., Harackiewicz et al., 2016). While useful, this approach is not focused on characterizing features of the texts, such as themes, sentiment, or cohesion. In contrast, Shnabel et al. (2013) used manual content analysis to qualitatively examine whether student's VA essays explicitly articulated their values as connected to some sense of “social belonging” (e.g., one values an activity because it is done with others). Qualitative text analysis approaches can provide useful information, but are also known to carry biases and other methodological limitations (Krippendorff, 2004). In particular, the laborious nature of manual coding essays make them a less viable option with the increasing scale of data (Crossley et al., 2019; Dowell, Graesser, et al., 2016; Joksimović et al., 2018; Li et al., 2018; McNamara et al., 2017).

As such, researchers have been incorporating automated linguistic analysis, including more shallow-level word counts and deeper level discourse analysis approaches. Both levels of linguistic analysis are informative. Content analysis using word-counting methods allows getting a fast overview of learners' participation levels, as well as assessing specific words and word categories. Advances in artificial intelligence methods, such as NLP (Kao & Poteet, 2007), have made it possible to automatically (a) harness vast amounts of educational discourse data being produced in technology-mediated learning environments, (b) quantify aspects of human cognition, affective, and social processes that (c) would otherwise not be possible or extremely time-consuming for human coders to capture, given the multifaceted characteristics of human discourse. Indeed, NLP and automated text analysis approaches have proven quite useful in quantifying and characterizing psychological, affective, cognitive, and social phenomena from a learner-generated discourse (Bell et al., 2012; Cade et al., 2014; D'Mello et al., 2009; D'Mello & Graesser, 2012; Dowell et al., 2017, 2019, 2020; Dowell & Graesser, 2015; Eichstaedt et al., 2018; Kern et al., 2020; Lin et al., 2020; McNamara et al., 2014; Schwartz et al., 2013; Tausczik & Pennebaker, 2010; Zedelius et al., 2019).

In the context of wise interventions, there has been growing efforts devoted toward exploring the content of student's psychological intervention essays using automated linguistic analysis tools, namely, LIWC (Pennebaker et al., 2015). While the majority of this research has been conducted in the contexts of the utility-value intervention paradigm (Akcaoglu

et al., 2018; Harackiewicz et al., 2016; Hecht et al., 2019; Klebanov et al., 2017, 2018; Priniski et al., 2019), there have been a notable few devoted toward understanding the linguistic mechanisms underlying student's values affirmation intervention essays (Riddle et al., 2015; Tibbetts et al., 2016). For instance, Tibbetts et al. (2016) conducted a follow-up study of the Harackiewicz et al. (2014) sample and found that the VA intervention was beneficial for FG students' overall postintervention GPAs over the course of a 3-year period. Particularly, relevant to the current research, they used LIWC to automatically quantify the degree to which student's essays exhibited independent and interdependent individual orientations. Words included in the independent dictionary included themes of individual interest and achievement, self-discovery, uniqueness, and leadership. Words in the interdependent dictionary reflected interpersonal themes of belonging, family, support, and empathy. They found that the effects of the VA intervention on course grades, academic belonging, and overall GPA 3 years later were all mediated by independent themes. In other words, for FG students, writing about independence in their VA essays led to higher grades in the biology course, higher levels of academic belonging, and higher GPAs over a 3-year period (Harackiewicz & Priniski, 2018).

As evident from this research, language can provide a powerful and measurable behavioral signal that can be used to capture the semantic processes and psychological constructs elicited during psychological interventions, and offer new insights into how different groups internalize intervention messages, and the linguistic mechanisms that incur the greatest benefits for students (Harackiewicz & Priniski, 2018; Hecht et al., 2019; Priniski et al., 2019). In the current research, we explored student's VA intervention essays using two well-established and complementary automated text analysis tools, namely, LIWC (Pennebaker et al., 2015) and Coh-Metrix (Graesser et al., 2004; McNamara et al., 2014; see Method section for more details).

This novel combination allowed us to quantify both psychologically meaningful word categories (i.e., LIWC) and discourse elements (i.e., cohesion; Coh-Metrix). In particular, LIWC allows us to quantify constructs directly relevant to the VA intervention, including references to family, independent, and interdependent individual orientations (e.g., pronouns). Moving past what has been explored previously, we additionally include constructs that situate these constructs within students' awareness, including temporal orientation, drives, cognition, and sentiment.

Additionally, Coh-Metrix, which employs more sophisticated NLP, allows us to dive deeper and explicitly focus on the cohesion within students' essays along two dimensions, namely, referential and deep cohesion. In line with Kintsch's (1998) construction-integration theory, Coh-Metrix distinguishes between multiple types of cohesion which fall under two main forms, namely, textbase (i.e., referential cohesion)

and situation model cohesion (i.e., deep cohesion). Referential or textbase cohesion is primarily maintained through the bridging devices, that is, the overlap in words, or semantic references, whereas deep cohesion related to the situation model dimension and reflects causation, intentionality, space, and time (McNamara et al., 2014). Together this NLP approach allows us to begin to address the need for data-driven insights (Paxton & Griffiths, 2017) and research efforts devoted toward "text analysis of students' essays may offer new insights into how different groups internalize intervention messages and what types of writing interventions have the greatest benefits for students" (Harackiewicz & Priniski, 2018).

Current Research

To address this need, the current research provides unique evidence on this issue by reporting the observed differences between two randomized field implementations of the VA intervention at scale: (a) a successful traditional in-class intervention and (b) an unsuccessful online implementation. The classroom intervention was delivered to 515 students in an introductory physics course at a Midwestern university that has experienced considerable GPDs over the years. As shown in Figure 1, female students in the in-class intervention experienced an increase in both exam grades and their overall course grade as reported by Koester and McKay (2021). During the same semester, we implemented an online VA intervention to 1,936 students across five STEM courses, which have also experienced considerable GPDs over the years, using ECoach, a well-established computer-tailored communication system (Huberth et al., 2015). As shown in Figure 2, there was no observed improvement in the GPD for female students in the affirmation condition of the online intervention.

The first three research questions are aimed at determining if there is a more successful engagement profile and identifying the language and discourse features that characterize it. The final research question not only achieves this aim but also allows us to determine if students' language and discourse might be an important factor in the effectiveness of the VA interventions. Overall, these research questions allowed us to explore new factors and gain a deeper understanding of the underlying mechanisms associated with effective VA interventions for alleviating the GPD in STEM courses.

Research Question 1 (RQ1): Are there unique linguistic features that differentiate affirmation essays from the control essays in the classroom intervention (i.e., affirmation vs. control)?

Research Question 2 (RQ2): What are the language and discourse features that differentiate between male and female VA essays (i.e., not control essays) in the classroom intervention?

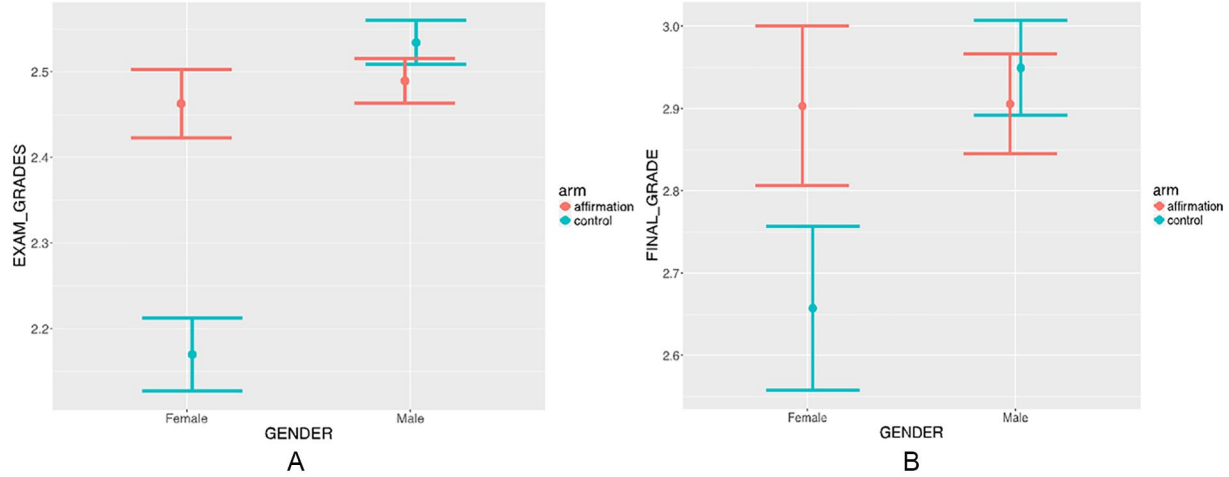


FIGURE 1. Exam grades (A) and final course grade (B) by gender and condition for the in-class VA intervention. Error bars represent 95% confidence intervals. Reprinted with permission from Koester and McKay (2021).

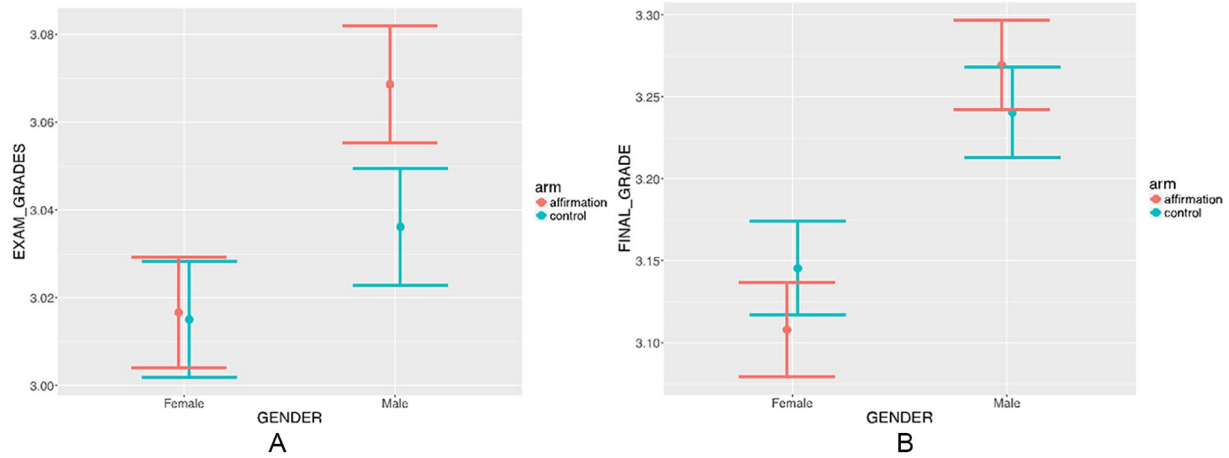


FIGURE 2. Exam grades (A) and final course grade (B) by gender and condition for the online VA intervention. Error bars represent 95% confidence intervals. Reprinted with permission from Koester and McKay (2021).

Research Question 3 (RQ3): For the classroom intervention, are students' linguistic profiles associated with their expected performance?

Research Question 4 (RQ4): What are the language and discourse features that differentiate between female students in-class affirmation essays, *which was successful*, and female students' affirmation essays in the online intervention, *which was not successful*?

Method

Participants

In-Class Intervention Participants. A total of 515 students enrolled in an electromagnetism-based physics course participated in the study. Of the 515 students, two were removed from the analysis due to participant

error. Of the remaining 513 students, 144 were female and 369 were male. Students were randomly assigned to receive either the VA intervention ($n = 255$) or control ($n = 258$).

Online Intervention Participants. A total of 1,936 students across five STEM courses participated in the study. In the current study, we focus only on the female students' affirmation essays ($n = 538$) to address RQ4.

Sample Size. The sample was sufficient to reliably detect effect sizes (ds) as small as 0.248; 95% confidence interval (CI) [0.074, 0.421] among VA and control essays (RQ1; $n = 513$, $\alpha = .05$, $1 - \beta = 0.80$), ds as small as 0.352; 95% CI [0.106, 0.599] among male and female essays (RQ2; $n = 255$, $\alpha = .05$, $1 - \beta = 0.80$), as small as 0.656; 95% CI = [0.106,

0.599] among high-performing male and female essays (RQ3; $n = 75$, $\alpha = .05$, $1 - \beta = 0.80$), and as small as 0.248; 95% CI [0.074, 0.248] among female students in-class and online affirmation essays (RQ4; $n = 609$, $\alpha = .05$, $1 - \beta = 0.80$).

Experimental Procedure: Psychological Interventions Design and Delivery

Experimental Procedure: In-Class. Students were blocked by race, gender, cumulative GPA and year in school and then randomly assigned to either the VA or the control condition. Following randomization, 255 students (71 female) received the VA exercise while 258 (73 female) students received the control exercise.

Participants completed the VA writing exercise or control writing exercise during the lab section of the course. Following from previously successful VA interventions in comparable college settings (Harackiewicz et al., 2014; Miyake et al., 2010) and in accordance with the suggested implementation standards, the writing exercises took place early in the semester (i.e., Week 3 of the 15-week semester) and preceded any course exams. Unlike previous VA research conducted in college settings (Harackiewicz et al., 2014; Miyake et al., 2010), this study does not include a second writing exercise midway through the term. While this is a departure from past implementations, a single dose of the writing exercise has shown to be sufficient. In the original test of the VA intervention in middle school classrooms, Cohen et al. (2006) only used a single dose of the writing exercise administered at the beginning of the year. Moreover, while Miyake and colleagues provided students the opportunity to complete the writing exercise a second time in the middle of the term, this second opportunity was optional and administered online, rather than in the classroom.

In accordance with the procedures used by Harackiewicz et al. (2014), the writing exercise was administered by TAs in the weekly lab section of the course. TAs were naive to the purpose of the study and were blinded to the students' condition. During Week 3 of the semester, a member of the research team reported to the lab prior to the start of class and handed the TAs a packet of manila envelopes labeled with students names. Within each packet was either the VA writing exercise or the control condition writing exercise predetermined for each student. TAs also received a standardized script to introduce the exercise to their students (see the online Supplemental Material for more details on the script).

After reading aloud the instructions, TAs proceeded to pass the labeled envelopes to corresponding students. While the envelope contained one of two different writing exercises, the exercises closely resembled one another in size and appearance. In both conditions, the envelope contained a two-page packet with a list of 14 values on the front of the first page. The list of values closely resembled those used in

previous college VA interventions (Harackiewicz et al., 2014; Miyake et al., 2010). After opening the envelope, the first page of the packet instructed students in the control condition to mark the two to three values that were the least important to them and write on the next page why they could be important to someone else. Conversely, students assigned to the affirmation condition were instructed to mark the two to three values that were the most important to them and then on the following page write why these values were important to themselves. Students were given 5 minutes to complete the writing exercise. After the exercise was completed, students placed their packet back into the original manila envelope with their name label. The envelopes were then collected by TAs and then returned to the member of the research team monitoring from the hallway.

Experimental Procedure: Online Intervention. Again, students were blocked by race, gender, cumulative GPA, and year in school and then randomly assigned to either the VA or the control condition. Following randomization, 538 female students received the VA exercise. To deliver our intervention, we used ECoach, a well-established computer-tailored communication system, already delivering personalized feedback, encouragement, and advice to thousands of students per term (Huberth et al., 2015). Students were invited to complete a writing exercise within the online platform around Week 3 of the semester and preceded any course exams. Some courses offered extra credit for the exercise, while others did not. Students who agreed to participate were randomized to receive either the intervention writing prompt or a control writing prompt. Following the same procedure as the classroom intervention, students in the control condition were asked to mark the two to three values that were the least important to them and write about why they could be important to someone else. Conversely, students assigned to the affirmation condition were instructed to mark the two to three values that were the most important to them and then write why these values were important to them. In both conditions, students were required to write for at least 5 minutes.

Performance Measurement

The performance measure used in the current analyses, for RQ3, was a relative performance measure that has been referred to as "better than expected" (BTE; Wright et al., 2014). Unlike more traditional performance measures (e.g., course grade), BTE is a relative estimate of student performance—whether a student performed better or worse than expected (Huberth et al., 2015; Matz et al., 2017). Expected performance, which has been shown to play a key role in motivation and achievement, is derived from student characteristics such as prior GPA and standardized test scores. In this approach, a student receiving a C in physics might be

TABLE 1

Linguistic Descriptive Statistics for Student Essays Across Intervention Conditions, Environment, and Gender

Measure	Affirmation				Control			
	Females		Males		Females		Males	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Online								
Mean word count	165.22	2.91	158.75	2.99	139.47	2.69	135.74	2.65
Mean words per sentence	19.66	0.24	19.29	0.26	20.79	0.24	21.23	0.48
In-class								
Mean word count	79.68	2.32	65.61	1.58	67.04	2.14	61.61	1.45
Mean words per sentence	18.89	0.60	17.77	0.41	20.63	0.82	19.33	0.44

considered BTE if peers with a similar background typically fail. Likewise, a student with a 4.0 GPA receiving her first B+ (which others might consider a good grade) would have a performance that is considered to be worse than expected.

Values Affirmation Intervention Essays

Participant essays were processed to quantify individual linguistic differences. These analyses yielded individual summary measures of the engagement and the nature of participation. Table 1 reports the descriptive statistics for average words written and average sentence length between conditions (intervention and control), genders, and environments (online and in-class).

A number of conclusions can be drawn from Table 1. First, the comparisons between student essays constructed within online versus in-class show that across both conditions, all students (i.e., females and males) wrote substantially more in the online environment, compared with the in-class. However, the average number of words across both environments does reflect appropriate task engagement. A comparison between the intervention and control conditions, across both environments, shows all students (i.e., females and males) wrote more in the intervention conditions; however, the average sentence length was longer in the control conditions. Finally, a gender comparison shows that females wrote more than males across both conditions and environments, but this difference is slightly more pronounced in the in-class intervention condition.

Computational Evaluation Tools

Prior to computational evaluation, the logs were cleaned and parsed to facilitate a student-level evaluation. Thus, text files were created that included each learner's essay, yielding a total of 513 text files for the in-class intervention and 538 (female essays only) for the online intervention, one for each student essay. All files were then analyzed using Coh-Metrix and LIWC.

The linguistic features explored in the current research were motivated by both related research and because of potential alignment with the VA intervention.

Coh-Metrix. Coh-Metrix (www.cohmetrix.com) is an automated linguistics facility that analyzes features of language and discourse (McNamara et al., 2014). Coh-Metrix incorporates automated computational methods of NLP, such as syntactic parsing and cohesion computation, to capture language characteristics at the word-level, sentence-level, and deeper levels of discourse. Coh-Metrix provides useful insights into learners' affective, social, and cognitive processes in a variety of digital learning environments (Choi et al., 2018; D'Mello & Graesser, 2012; Dowell et al., 2014; Dowell, Graesser, et al., 2016; Graesser et al., 2011; Graesser et al., 2018; McNamara & Graesser, 2012). Coh-Metrix has been extensively validated through more than 150 published studies, which have demonstrated that Coh-Metrix indices can be used to detect subtle differences in text and discourse (Graesser, 2011; Graesser et al., 2011; McNamara et al., 2006; McNamara et al., 2014). In the current research, we were particularly interested in utilizing Coh-Metrix to quantify properties of cohesion in students' VA essays. The two Coh-Metrix cohesion measures used in the current investigation are briefly described:

- *Deep Cohesion.* The extent to which the ideas in the text are cohesively connected at a deeper conceptual level that signifies causality or intentionality.
- *Referential Cohesion.* The extent to which explicit words and ideas in the text are connected with each other as the text unfolds.

Linguistic Inquiry Word Count. LIWC is an automated text analysis tool designed for studying the various emotional, cognitive, structural, and process components present in text (Pennebaker et al., 2015). As a word-count program, the key component of LIWC is its embedded dictionary. LIWC processes individual or multiple textual files by searching and

counting words that are listed in the predesignated dictionary. The dictionary itself has been revised and validated over the course of two decades, and the most recent version consists of 6,400 English words/word stems, covering a range of social and psychological constructs such as affect, cognition, and biological processes (see Pennebaker et al., 2015, for details). Currently, LIWC is one of the most popular and reliable programs for text analysis available; it has been utilized in hundreds of studies across the social sciences, including psychology, education, sociology, communication, political sciences, and economics (Borowiecki, 2017; Boyd et al., 2020; Cade et al., 2014; Dowell, Windsor, et al., 2016; Kacewicz et al., 2014; Lin et al., 2020; Newman et al., 2008; Pennebaker & Chung, 2014; Pennebaker et al., 2014). A total of 29 linguistic variables from six LIWC categories were included in the analysis. The LIWC categories used in the current investigation are briefly described below, and a full list of the associated 29 LIWC variables can be found in the online Supplemental Material:

- *Affective Processes*: Words expressing positive and negative affect, such as love, nice, sweet and hurt, ugly, nasty, respectively.
- *Cognitive Processes*: Words suggestive of individuals organizing and intellectually understanding the issues addressed in their writing (e.g., because, would, maybe, but).
- *Pronouns*: Words indicating attentional focus such as I, we, they.
- *Temporal Focus*: Words expressing temporal focus, including past focus (e.g., ago, did, talked), present focus (e.g., today, is, now), and future focus (e.g., may, will, soon).
- *Drives*: Words expressing an individual's motivations, including power, affiliation, and achievement.
- *Family*: Words indicating family relationships (e.g., mother, sister, aunt).

Statistical Analyses

Principal Component Analysis

In-class intervention. A PCA approach was adopted to discover language and discourse patterns associated with students' VA and control essays. PCA is a common data mining technique that involves reducing multidimensional data sets to lower dimensions for analysis (Tabachnick & Fidell, 2007). In the current research, it was used to reduce the 31 linguistic features (two Coh-Metrix indices and 29 LIWC indices) to create meaningful, broader variables with which to describe the students' VA intervention essays. PCA has been applied in previous studies of psychological interventions and has proven useful in building an understanding of language characteristics in student essays and discourse more broadly (Cade et al., 2014; Dowell &

Graesser, 2015; Pennebaker et al., 2014). Prior to analysis, the data were normalized, centered, and checked for factorability (for more details, see online Supplemental Material). The loadings, which quantify the strength of the relationship between the component and each linguistic variable, were used to describe and name each component. Table 2 provides a description of the 10 principal components. Due to word limit constraints, we are not able to provide illustrative examples here; however, example student essays from the current data are provided in the online Supplemental Material as an illustrative example of the linguistic features that comprise a few of the component scores.

Women in-class and online intervention. A separate PCA was conducted to create meaningful, broader variables with which to describe the female students' VA intervention essays written in the in-class and online intervention. The same procedure was followed as in the previous analysis, including data normalization, centering, and factorability evaluation (for more details, see online Supplemental Material). The loadings, which quantify the strength of the relationship between the component and each linguistic variable, were used to describe and name each component. Table 3 provides a description of the 10 principal components.

Generalized Logistic Mixed-Effects Regressions. A generalized logistic mixed-effects modeling approach was adopted for all analyses due to the structure of the data (e.g., interindividual word count variability; Baayen et al., 2008). Mixed-effects models include a combination of fixed and random effects that assess the influence of the fixed effects on dependent variables after accounting for any extraneous random effects. Mixed-effect modeling provides a robust and flexible approach that allows for a wide set of correlation patterns to be modeled.

The analyses for the in-class intervention consisted of testing for linguistic differences in VA intervention essays (affirmation vs. control; RQ1) and between males' and females' affirmation essays (RQ2 and RQ3). There were two sets of dependent measures for the in-class intervention analyses: (a) essay type (affirmation vs. control) and (b) gender. For RQ2, gender was the dependent variable, and we focused on the language and discourse features that differentiate between male and female VA essays only (i.e., not control essays) in the classroom intervention. This analysis was motivated to investigate any potential influence of gender on essay construction, wherein it could be possible that observed differences on RQ1 were due to males and females constructing essays in a similar fashion. For RQ3, gender was also the dependent variable; however, learners were grouped into performance bins (i.e., based on a quartile split of their BTE scores) to investigate any potential influence of performance on essay construction. Across all models, the independent

TABLE 2
Description of Principal Components (PCs) for In-Class Intervention

Measure	Language Characteristics
Intrapersonal family (PC1)	Very intrapersonal focus (I), family references, with less future focused and tentative language (maybe, perhaps, cognitive processes)
Positive emotion and affiliation (PC2)	Positive emotion, affiliation (e.g., social connected references “care,” “help,” “intimate,” “kind,” “neighbor,” and “volunteer”), and reward drives
Sad anxious (PC3)	Very negative, sad, and anxious language
Achievement (PC4)	Reward and achievement orientation (take, prize, benefit), work and deep cohesion
Negative past focus (PC5)	Less positive emotion, more complex (longer) past-oriented language, and much less present focused, and certainty
Cohesive future self (PC6)	Referential cohesion, more “I” references, and less “we” and more future focused
Tentative future focus (PC7)	Very tentative, relative, and future-focused language
Differentiation (PC8)	High differentiation (hasn’t, but, else) language
Uncertain perceptions (PC9)	Visual perception and uncertain language
Present family cohesion (PC10)	Deep and referential cohesion coupled with present focus and family references

TABLE 3
Description of Principal Components (PCs) for Women in the Intervention (In-Class and Online)

Measure	Language Characteristics
Positive emotion and affiliation focus (PC1)	Positive emotion, affiliation (e.g., social connected references “care,” “help,” “intimate,” “kind,” “neighbor,” and “volunteer”), and reward drives
Confident intrapersonal family focus (PC2)	Very intrapersonal focus (I), family references, with less future focused and less tentative language (maybe, perhaps, cognitive processes)
Sad and anxious affiliation focus (PC3)	Very negative, sad, and anxious language, coupled with affiliation references (e.g., social connected references)
Achievement (PC4)	Reward and achieve orientation (take, prize, benefit), work and deep cohesion
Interpersonal and future focus (PC5)	Interpersonal references (“we”, affiliation), more future focused, and contemplative orientation (discrepancy—should, would)
Present and future personal achievement (PC6)	Achievement and work references, cohesive language, more “I” references, and more future/present focused
Perceptions (PC7)	Perceptual oriented language
Uncertain differentiation (PC8)	High differentiation (hasn’t, but, else), and low certainty (always, never) language, coupled with slightly intrapersonal references (I)
Positive and certain (PC9)	Positive emotion, certain, relative (then, finally, after) and a time references (never, always, end, until)
Deep cohesion (PC10)	High levels of deep cohesion, more complex words, adverbs, and a power reward orientation

fixed-effect variables consisted of the 10 linguistic dimensions (i.e., principal components, Table 2).

The analysis for RQ4 consisted of testing for linguistic differences in VA intervention essays between women who performed the intervention in-class (i.e., successful), compared with those who performed it online (i.e., unsuccessful). The dependent variable for this analysis was environment (in-class vs. online), independent fixed-effect variables consisted of the 10 linguistic dimensions (Table 3).

In addition to constructing the models with the 10 discourse features as fixed effects, *null models* with the random effects (*learner*) but no fixed effects were also constructed.

A comparison of the null, random effects only model with the fixed-effect models allows us to determine whether discourse predicts essay type, gender, and environment above and beyond the random effects (i.e., individual differences in learners). Akaike information criterion (AIC), log likelihood (LL), and a likelihood ratio test were used to determine the best fitting and most parsimonious model. Additionally, the effect sizes (R^2) for each model were estimated according to Nagelkerke (1991) and Cragg and Uhler’s (1970) pseudo- R^2 statistic. The generalized logistic mixed-effects regression models were conducted using R Version 3.0.1 software for statistical analysis.

Results

The likelihood ratio test indicated that the full model was the most parsimonious and best fit for *Intervention Condition* (RQ1), and *Affirmation Gender* (RQ2) models with $\chi^2(1) = 478.03, p < .001, R^2 = .81$, and $\chi^2(1) = 53.28, p < .001, R^2 = .27$, respectively. A number of conclusions can be drawn from this initial model fit evaluation and inspection of R^2 variance. First, the model comparisons suggest that the discourse features were able to add a significant improvement in differentiating between the learners' VA and control essays and between males' and females' construction of VA essays. Second, linguistic characteristics explained about 81% and 27% of the predictable variance in essay type and gender, respectively. The linguistic characteristics that were predictive of Intervention Condition type and Gender are presented in Figures 3 and 4. The reference group was the Control Essays and Males—meaning that higher odds ratio indicates higher probability of being a VA Essay or Female student for each model, respectively.

Table 4 shows the coefficients for the discourse features that successfully differentiated VA essays from controls, and female students' affirmation essays from male students' affirmation essays. As shown in Table 4 and Figure 3, for the *Intervention Condition Model*, affirmation essays were characterized by Intrapersonal Family Focus, Positive Emotion and Affiliation Focus, Achievement, Sad and Anxious Discourse. However, results also indicate the opposite association between Uncertain Perceptions, Tentative Future Focus, Negative Past Focus, and Differentiation Language with the predicted probability of being a VA Essay.

The *Affirmation Gender Model*, explored linguistic differences in male and female students' VA essays for the in-class intervention (RQ2). The *Affirmation Gender Model* (Figure 4) shows that female students' affirmation essays, compared with males', were characterized more by Intrapersonal Family Focus, and Present Family Cohesion. Compared with males, female student's affirmation essays also used significantly less Sad and Anxious Discourse, Tentative Future Focus, Negative Past Focus language.

It is possible that the observed linguistic differences in male and female students VA essays are simply a product of performing similarly within the class (RQ3). Thus, the third set of analyses involved a more fine-grained investigation of how higher and lower performing males and females constructed their essays (both VA and control). In order to explore higher and lower performing students, we created three bins of learners based on a quartile split of their BTE scores. This resulted in roughly 75 learners per bin. The lower and higher bins were used for analysis while the middle was excluded to reduce noise. Four models were constructed, where two models were VA essays for higher and lower performing students, and two models were control essays for higher and lower performing students. Particularly,

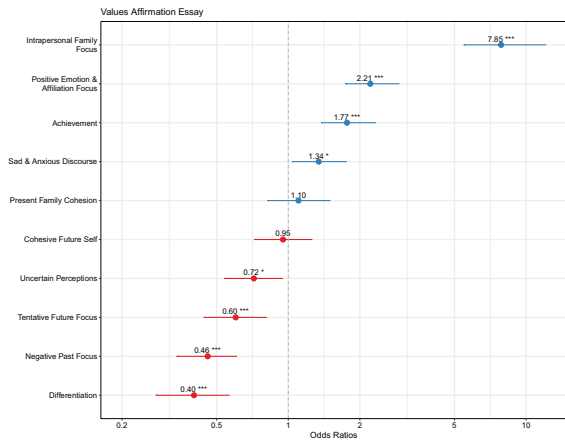


FIGURE 3. Odds ratios for intervention condition model. Error bars represent 95% confidence intervals. The reference group was the control essays, meaning that higher odds ratio indicates higher probability of being a values affirmation essay. * $p < .05$. ** $p < .01$. *** $p < .001$.

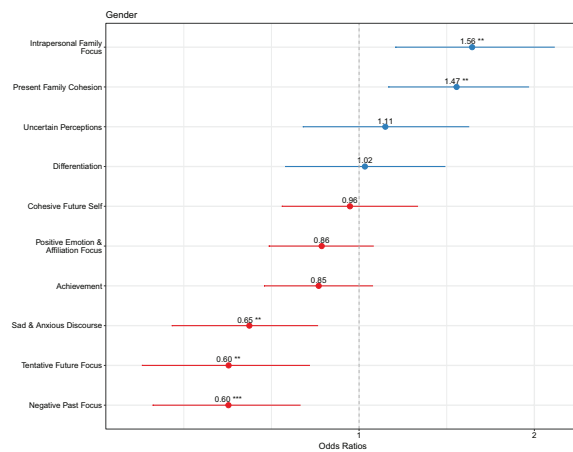


FIGURE 4. Odds ratios for affirmation gender model. Error bars represent 95% confidence intervals. The reference group was males, meaning that higher odds ratio indicates higher probability of being a female student essay. * $p < .05$. ** $p < .01$. *** $p < .001$.

for both conditions (i.e., VA and control), we constructed a *higher BTE model*, and a *lower BTE model*. For all models, the linguistic characteristics were the independent variables, and gender (i.e., male or female) was the dependent variable.

For the VA analyses, the likelihood ratio test indicated that the full model was the most parsimonious and best fit for *VA higher BTE model*, with $\chi^2(10) = 22.76, p < .01, R^2 = .46$, but not the *VA lower BTE model* with $\chi^2(10) = 15.87, p = .10$. Interestingly, when these relationships were further explored in the control essay analyses, the likelihood ratio tests indicated that the full models for *control higher BTE*

TABLE 4

Mixed-Effects Model Coefficients for Predicting Intervention Condition Type, and Gender With Language Characteristics

Measure	Intervention Condition		Affirmation Gender	
	β	<i>SE</i>	β	<i>SE</i>
Intrapersonal family	2.06***	0.20	0.45*	0.18
Positive emotion and affiliation	0.79***	0.13	-0.15	0.11
Sad anxious	0.29*	0.13	-0.44**	0.16
Achievement	0.57***	0.13	-0.16	0.11
Negative past focus	-0.78***	0.15	-0.52**	0.17
Cohesive future self	-0.05	0.14	-0.04	0.14
Tentative future focus	-0.51***	0.15	-0.52**	0.19
Differentiation	-0.91***	0.18	0.02	0.16
Uncertain perceptions	-0.33*	0.14	0.10	0.17
Present family cohesion	0.10	0.15	0.39*	0.16

Note. Intervention Condition Model $N = 513$, Affirmation Gender Model $N = 255$. The reference group was the Control Essays and Males, meaning that higher odds ratio indicates higher probability of being a VA Essay or Female student for each model, respectively. β = Fixed-effect coefficient; *SE* = standard error; VA = values affirmation.

* $p < .05$. ** $p < .01$. *** $p < .001$.

and control lower BTE model did not yield a significantly better fit than the null model with $\chi^2(10) = 4.71$, $p = .90$, and $\chi^2(10) = 9.66$, $p = .47$, respectively.

The model comparisons suggest that the discourse features were able to add a significant improvement in differentiating between the higher performing male and female learners' VA essays, however, no discerning difference was observed between lower performing male and female learners' VA essays, or higher and lower performing learners' control essays. This suggests that good essays may be an important mechanism for effective VA interventions. More specifically, this indicates that perhaps *how* students construct their essays in terms of linguistic characteristics may be an important construct in the underlying mechanisms driving the beneficial effect of the intervention for women. Second, the linguistic features of high-performing learners' essays explained about 46% of the predictable variance in gender differences. The linguistic characteristics that significantly discriminate between high-performing male and female essays are presented in Figure 5 with the odds ratios and confidence intervals. The reference group was Males, meaning that higher odds ratio indicates higher probability of being a female student's essay. As highlighted in Figure 5, we observed a significant effect for Intrapersonal Family Focus ($\beta = 1.19$, $SE = 0.58$, $p < .05$), and a marginally significant effect for Sad and Anxious Discourse ($\beta = 1.01$, $SE = 0.51$, $p = .05$).

The final analysis focused on investigating the language and discourse features that characterize female learners' VA essays in-class (i.e., successful) and in the online version (i.e., unsuccessful) of the intervention. Here, we constructed one model, *Women Environment Model* (RQ4), with environment (in-class vs. online) as the dependent variable, the 10

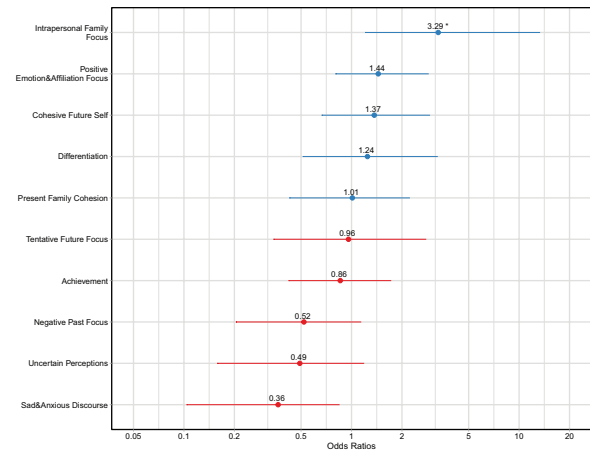


FIGURE 5. Odds ratios and 95% confidence intervals for values affirmation higher "better than expected" model. The reference group was males, meaning that higher odds ratio indicates higher probability of being a female student's essay.

* $p < .05$. ** $p < .01$. *** $p < .001$.

linguistic dimensions as the independent variables, and participant as the random effect. The likelihood ratio test indicated that the full model was the best fit for the data with $\chi^2(10) = 84.95$, $p < .001$, $R^2 = .25$. The linguistic characteristics that significantly discriminated between females in the classroom intervention and online intervention are presented in Figure 6 with the odds ratio and confidence intervals. The reference group was online female students' essays, meaning that higher odds ratio indicates higher probability of being a female student's essay in the classroom intervention. As highlighted in Figure 6, we observed a significant difference

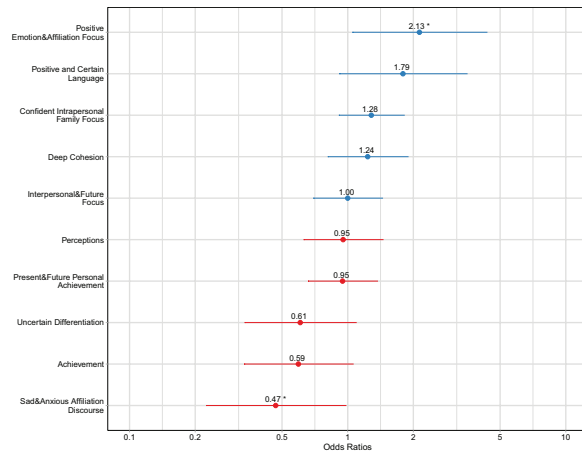


FIGURE 6. Odds ratios and 95% confidence intervals for the women environment model. The reference group for this analysis is women in the online intervention, with higher odds indicating an increased likelihood of it being a female in-class. * $p < .05$. ** $p < .01$. *** $p < .001$.

for Positive Emotion and Affiliation Focus ($\beta = 0.76$, $SE = 0.35$, $p < .05$), and Sad & Anxious Affiliation Focus ($\beta = -0.76$, $SE = 0.37$, $p < .05$). Additionally, there were several linguistic dimensions that were marginally significant, namely Achievement ($\beta = -0.52$, $SE = 0.29$, $p = .07$), Uncertain Differentiation ($\beta = -0.50$, $SE = 0.30$, $p = .09$) and Positive and Certain Language ($\beta = 0.58$, $SE = 0.34$, $p = .09$). As noted in Koester and McKay (2021), female learners in the online version of the intervention did not experience the same performance change as the female students in the classroom intervention. Had we observed a similar linguistic pattern for both female populations, this might lend evidence toward the context hypothesis. However, as it stands, the results suggest that what might be more important is *how* individuals construct the VA essays, than *where* they construct the essays.

Discussion

There have been attempts to use VA interventions to address gender-based achievement gaps (e.g., Miyake et al., 2010), identify the linguistic features associated with its beneficial effects (Tibbetts et al., 2016), and most recently scale the intervention (Borman et al., 2018). However, to our knowledge, this is the first study designed to alleviate gender-based achievement gaps that has also attempted to scale the intervention and disentangle the beneficial psychological constructs elicited, from a linguistic point of view, between classroom and online implementations. Specifically, we explored the extent to which characteristics of discourse diagnostically reveal the unique linguistic profile associated with students' VA essays, and in particular, more and less

successful VA intervention essays. The findings present some methodological and theoretical implications for both intervention scientists and teachers. First, as a methodological contribution, we have highlighted the rich contextual information that can be garnered from using NLP techniques to reveal more proximal underlying intervention mechanisms. Indeed, NLP has been advocated for in the literature (Harackiewicz & Priniski, 2018) as a means to gain additional insights into how different groups internalize intervention messages and what types of writing interventions have the greatest benefits for students. Particularly, in the current study, students' discourse features added significant improvement in predicting the essential characteristics of the intervention including the essay type, gender, and intervention context.

We first established that there was a distinct linguistic profile that distinguished VA essays from control essays, and then explored the linguistic differences in how female and male students constructed their VA essays (i.e., *Affirmation Gender Model*). Our results here were somewhat contradictory to previous research (Tibbetts et al., 2016). Tibbetts et al. (2016) used LIWC to analyze FG students' VA essays and found that students who employed more linguistic features associated with independence in their VA essays, rather than interdependence, led to higher grades in their biology course. Shnabel et al. (2013) identified social belonging (i.e., writing that reminds students of their interdependence) as the mechanism facilitating the positive effects of VA for Black middle school students. In contrast, we identified, among others, that a mix of independent and interdependent language (e.g., intrapersonal family and positive affiliation focus), was a potential mechanism driving the positive effects for female students. This discrepancy in findings highlights the importance of giving careful consideration to the *target* population before transferring VA strategies, and clearly demonstrates the need for additional research to understand how these linguistic constructs operate across different groups.

We investigated whether the observed difference between male and female students' essay construction was a product of similar performance (i.e., the *high- and low-performing student models*). A noteworthy finding from this concerns the fact that, when we grouped students by performance, we only observed a difference between high-performing male and female students' VA essay construction, however, no difference was detected for the other groups (low VA, low and high control). This provides confidence that the observed linguistic differences are not simply a product of students being high performers, but instead offers evidence suggesting that *how* students construct their essays in terms of linguistic characteristics may be an important construct in the underlying mechanisms driving the beneficial effect of the intervention for women.

Our final analysis sought to gain insight into whether language and discourse might help explain why female learners

who completed the intervention online did not experience the same performance change as the female students in the classroom intervention. If female students wrote similarly in both environments, this would provide evidence in favor of the social context hypothesis (Steele, 1997), which states that the effectiveness of self-affirmation approaches depends on the identity threats “in the air” in a particular setting (i.e., the classroom). However, as a theoretical contribution, our results suggest that what might be more important is *how* individuals construct the VA essays, than *where* they construct the essays. It is important to note that the findings are not a product of females in class simply being more prolific, because students actually wrote more in the online intervention. We cannot entirely rule out the social context hypothesis, however, but our results do suggest that there are at least other important elements at play. In our future research, we will be designing a scaled online intervention specifically geared toward eliciting the identified linguistic features from students writing. Additionally, we plan to use a causal modeling approach to identify actual causal relationships between specific linguistic features in VA essays and beneficial educational outcomes.



While the automated text analysis approaches utilized in the current research provide several opportunities, applying these methodological approaches to real-world data brings new risks (Iliev et al., 2015; Schwartz & Ungar, 2015). For instance, word-count-based methods, such as LIWC, lack the contextual information that is available with human judgment. An interesting illustrative example of this was highlighted in the Back et al. (2011) work that explored the emotional content of text messages sent in the aftermath of September 11, 2001. A notable finding of this work was that the timeline of anger-related words showed an intense trend that kept constantly increasing for several hours after the attack. However, revisiting this research showed that many of the text messages were automatically generated by phone servers (“critical” server problem), and although unrelated to the theoretical question, they were identified as anger-related words by the system (Pury, 2011). This cautionary tale highlights the need for careful consideration when utilizing automated text analytic approaches with real-world data.

Despite these limitations, the present research does help advance our understanding of the VA intervention by highlighting critical language and discourse features that qualify VA effectiveness in buffering against identity threat with the potential to alleviate GPDs in STEM courses. In doing so, it furthers research beyond knowing that it *can* work in one context, to understand *how to potentially make it work* in educational settings at scale and close achievement gaps for larger numbers of students. Overall, this work helps inform affirmation theory by suggesting that the processes set in motion through self-affirmation interventions, for women in STEM, may be facilitated when these interventions involve specific language and discourse features.

Acknowledgments

This work was partially supported by the National Science Foundation 15-585, National Science Foundation No. 1535300, and by the office of Academic Innovation, and the Holistic Modeling of Education project funded by the Michigan Institute of Data Science.

ORCID iDs

Nia M. M. Dowell  <https://orcid.org/0000-0002-9839-8947>
 Timothy A. McKay  <https://orcid.org/0000-0001-9036-6150>

References

- Akcaoglu, M., Rosenberg, J. M., Ranellucci, J., & Schwarz, C. V. (2018). Outcomes from a self-generated utility value intervention on fifth and sixth-grade students' value and interest in science. *International Journal of Educational Research*, 87, 67–77. <https://doi.org/10.1016/j.ijer.2017.12.001>
- Aronson, J., Cohen, G., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones, & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (Vol. 2, pp. 159–174). American Psychological Association. <https://doi.org/10.1037/0000135-008>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Back, M. D., Küfner, A. C., & Egloff, B. (2011). “Automatic or the people?” Anger on September 11, 2001, and lessons learned for the analysis of large digital data sets. *Psychological Science*, 22(6), 837–838. <https://doi.org/10.1177/0956797611409592>
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill-over. *Journal of Experimental Psychology: General*, 136(2), 256–276. <https://doi.org/10.1037/0096-3445.136.2.256>
- Bell, C. M., McCarthy, P. M., & McNamara, D. S. (2012). Using LIWC and Coh-Metrix to investigate gender differences in linguistic styles. In P. M. McCarthy, & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation and resolution* (pp. 545–556). IGI Global. <https://doi.org/10.4018/978-1-60960-741-8.ch032>
- Borman, G. D., Choi, Y., & Hall, G. J. (2020). The impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. *Journal of Educational Psychology*. Advance online publication. <https://doi.org/10.1037/edu0000570>
- Borman, G. D., Grigg, J., Rozek, C. S., Hanselman, P., & Dewey, N. A. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science*, 29(11), 1773–1784. <https://doi.org/10.1177/0956797618784016>
- Borowiecki, K. J. (2017). How are you, my dearest Mozart? Well-being and creativity of three famous composers based on their letters. *Review of Economics and Statistics*, 99(4), 591–605. https://doi.org/10.1162/REST_a_00616
- Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 6(32), Article eaba2196. <https://doi.org/10.1126/sciadv.aba2196>

- Bradley, D., Crawford, E., & Dahill-Brown, S. E. (2015). Fidelity of implementation in a large-scale, randomized, field trial: Identifying the critical components of values affirmation. *Proceedings of the Society for Research on Educational Effectiveness* (ED562183). Society for Research on Educational Effectiveness. <https://eric.ed.gov/?id=ED562183>
- Brady, S. T., Reeves, S. L., Garcia, J., Purdie-Vaughns, V., Cook, J. E., Taborsky-Barba, S., Tomasetti, S., Davis, E. M., & Cohen, G. L. (2016). The psychology of the affirmed learner: Spontaneous self-affirmation in the face of stress. *Journal of Educational Psychology*, 108(3), 353–373. <https://doi.org/10.1037/edu0000091>
- Brewer, E., Sawtelle, V., Kramer, L. H., O'Brien, G. E., Rodriguez, I., & Pamelá, P. (2010). Toward equity through participation in Modeling Instruction in introductory university physics. *Physical Review Special Topics—Physics Education Research*, 6(1), Article 010106. <https://doi.org/10.1103/PhysRevSTPER.6.010106>
- Cade, W. L., Dowell, N. M., Graesser, A. C., Tausczik, Y. R., & Pennebaker, J. W. (2014). Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the Seventh International Conference on Educational Data Mining* (pp. 399–400). Springer.
- Casad, B. J., Oyler, D. L., Sullivan, E. T., McClellan, E. M., Tierney, D. N., Anderson, D. A., Greeley, P. A., Fague, M. A., & Flammang, B. J. (2018). Wise psychological interventions to improve gender and racial equality in STEM. *Group Processes & Intergroup Relations*, 21(5), 767–787. <https://doi.org/10.1177/1368430218767034>
- Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97(6), 1045–1060. <https://doi.org/10.1037/a0016239>
- Choi, H., Dowell, N. M., & Brooks, C. (2018). Social comparison theory as applied to MOOC student writing: Constructs for opinion and ability. In J. Kay, & R. Luckin (Eds.), *Proceedings of the 13th International Conference for the Learning Sciences* (pp. 1421–1422). International Society of the Learning Sciences.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, 313, 1307–1310.
- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, 324(5925), 400–403. <https://doi.org/10.1126/science.1170769>
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371. <https://doi.org/10.1146/annurev-psych-010213-115137>
- Conger, D., & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *Annals of the American Academy of Political and Social Science*, 627(1), 184–214. <https://doi.org/10.1177/0002716209348751>
- Cragg, S. G., & Uhler, R. (1970). The demand for automobiles. *Canadian Journal of Economics*, 3, 386–406. <https://doi.org/10.2307/133656>
- Creech, L. R., & Sweeder, R. D. (2012). Analysis of student performance in large-enrollment life science courses. *CBE Life Sciences Education*, 11(4), 386–391. <https://doi.org/10.1187/cbe.12-02-0019>
- Crossley, S. A., Kim, M., Allen, L., & McNamara, D. (2019). Automated summarization evaluation (ASE) using natural language processing tools. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial intelligence in education: AIED 2019* (Lecture Notes in Computer Science, Vol. 11625, pp. 84–95). Springer. https://doi.org/10.1007/978-3-030-23204-7_8
- Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry*, 22(4), 231–246. <https://doi.org/10.1080/1047840X.2011.607313>
- Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 112(16), 4988–4993. <https://doi.org/10.1073/pnas.1422822112>
- Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 21–29. <https://doi.org/10.1177/2372732214549471>
- D'Mello, S., Dowell, N. M., & Graesser, A. C. (2009). Cohesion relationships in tutorial dialogue as predictors of affective states. In V. Dimitrova, R. Mizoguchi, B. DuBoulay, & A. Graesser (Eds.), *Artificial intelligence in education* (Vol. 200, pp. 9–16). IOS Press.
- D'Mello, S., & Graesser, A. C. (2012). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5(4), 304–317. <https://doi.org/10.1109/TLT.2012.10>
- Dowell, N. M., Brooks, C., Kovanović, V., Joksimović, S., & Gašević, D. (2017). The changing patterns of MOOC discourse. In C. Urrea, J. Reich, & C. Thille (Eds.), *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (pp. 283–286). Association for Computing Machinery. <https://doi.org/10.1145/3051457.3054005>
- Dowell, N. M., Cade, W. L., Tausczik, Y. R., Pennebaker, J. W., & Graesser, A. C. (2014). What works: Creating adaptive and intelligent systems for collaborative learning support. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Twelfth International Conference on Intelligent Tutoring Systems* (pp. 124–133). Springer.
- Dowell, N. M., & Graesser, A. C. (2015). Modeling learners' cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, 1(3), 183–186. <https://doi.org/10.18608/jla.2014.13.18>
- Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh-Metrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, 3(3), 72–95. <https://doi.org/10.18608/jla.2016.33.5>
- Dowell, N. M., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills and outcomes: A group communication analysis. *Journal of Learning Analytics*, 7(1), 38–57. <https://doi.org/10.18608/jla.2020.71.4>
- Dowell, N. M., Nixon, T., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach

- for detecting sociocognitive roles in multi-party interactions. *Behavior Research Methods*, 51(3), 1007–1041. <https://doi.org/10.3758/s13428-018-1102-z>
- Dowell, N. M., Windsor, L. C., & Graesser, A. C. (2016). Computational linguistics analysis of leaders during crises in authoritarian regimes. *Dynamics of Asymmetric Conflict*, 9(1-3), 1–12. <https://doi.org/10.1080/17467586.2015.1038286>
- Eddy, S. L., & Brownell, S. E. (2016). Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Physical Review Physics Education Research*, 12(2), Article 020106. <https://doi.org/10.1103/PhysRevPhysEducRes.12.020106>
- Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE Life Sciences Education*, 13(3), 478–492. <https://doi.org/10.1187/cbe.13-10-0204>
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), 11203–11208. <https://doi.org/10.1073/pnas.1802331115>
- Fogliati, V. J., & Bussey, K. (2013). Stereotype threat reduces motivation to improve: Effects of stereotype threat and feedback on women's intentions to improve mathematical ability. *Psychology of Women Quarterly*, 37(3), 310–324. <https://doi.org/10.1177/0361684313480045>
- Forbes, C. E., Schmader, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience*, 3(3), 253–261. <https://doi.org/10.1093/scan/nsn012>
- Graesser, A. C. (2011). Learning, thinking, and emoting with discourse technologies. *American Psychologist*, 66(8), 746–757. <https://psycnet.apa.org/journals/amp/66/8/746/>
- Graesser, A. C., Dowell, N., Hampton, A. J., Lippert, A. M., Li, H., & Williamson, S. D. (2018). Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 Program for International Student Assessment. In *Building intelligent tutoring systems for teams* (Vol. 19, pp. 173–211). Emerald. <https://doi.org/10.1108/S1534-08562018000019012>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. <https://doi.org/10.3102/0013189X11413260>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202. <https://www.ncbi.nlm.nih.gov/pubmed/15354684>
- Hanselman, P., Rozek, C. S., Grigg, J., & Borman, G. D. (2017). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of Educational Psychology*, 109(3), 405–424. <https://doi.org/10.1037/edu0000141>
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, 106(2), 375–389. <https://doi.org/10.1037/a0034679>
- 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>
- Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, 69(1), 409–435. <https://doi.org/10.1146/annurev-psych-122216-011725>
- Hecht, C. A., Harackiewicz, J. M., Priniski, S. J., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Promoting persistence in the biological and medical sciences: An expectancy-value approach to intervention. *Journal of Educational Psychology*, 111(8), 1462–1477. <https://doi.org/10.1037/edu0000356>
- Huberth, M., Chen, P., Tritz, J., & McKay, T. A. (2015). Computer-tailored student support in introductory physics. *PLOS ONE*, 10(9), Article e0137001. <https://doi.org/10.1371/journal.pone.0137001>
- Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, 7(2), 265–290. <https://psycnet.apa.org/doi/10.1017/langcog.2014.30>
- Joksimović, S., Dowell, N., Gašević, D., Mirriahi, N., Dawson, S., & Graesser, A. C. (2018). Linguistic characteristics of reflective states in video annotations under different instructional conditions. *Computers in Human Behavior*, 96, 211–222. <https://doi.org/10.1016/j.chb.2018.03.003>
- Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E., King, K., & Freeman, S. (2017). Values Affirmation Intervention reduces achievement gap between underrepresented minority and White students in introductory biology classes. *CBE Life Sciences Education*, 16(3). <https://doi.org/10.1187/cbe.16-12-0351>
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143. <https://doi.org/10.1177/0261927X13502654>
- Kao, A., & Poteet, S. R. (2007). *Natural language processing and text mining*. Springer. <https://www.amazon.com/Natural-Language-Processing-Text-Mining/dp/184628175X>
- Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences of the United States of America*, 117(19), 10165–10171. <https://www.pnas.org/content/117/19/10165.short>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Klebanov, B. B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective writing about the utility value of science as a tool for increasing STEM motivation and retention—Can AI help scale up? *International Journal of Artificial Intelligence in Education*, 27(4), 791–818. <https://doi.org/10.1007/s40593-017-0141-4>
- Klebanov, B. B., Priniski, S., Burstein, J., Gyawali, B., Harackiewicz, J., & Thoman, D. (2018). Utility-value score: A case study in system generalization for writing analytics. *Journal of Writing Analytics*, 2, 314–328. <https://www.ncbi.nlm.nih.gov/pubmed/31565684>

- Koester, B. P., Grom, G., & McKay, T. A. (2016). *Patterns of gendered performance difference in introductory STEM courses*. arXiv. <https://arxiv.org/abs/1608.07565>
- Koester, B. P., & McKay, T. A. (2021). *Gendered performance in introductory STEM courses* [Manuscript submitted for publication]. Department of Physics, University of Michigan.
- Krippendorff, K. (2003). *Content analysis: An introduction to its methodology*. Sage.
- Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Li, H., Cai, Z., & Graesser, A. C. (2018). Computerized summary scoring: Crowdsourcing-based latent semantic analysis. *Behavior Research Methods*, 50(5), 2144–2161. <https://doi.org/10.3758/s13428-017-0982-7>
- Lin, Y., Yu, R., & Dowell, N. (2020). LIWCs the same, not the same: Gendered linguistic signals of performance and experience in online STEM courses. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Proceedings of the 21st International Conference: AIED 2020* (Artificial Intelligence in Education: Part I; Vol. 12163, pp. 333–345). Springer International. https://doi.org/10.1007/978-3-030-52237-7_27
- London, B., Rosenthal, L., & Gonzalez, A. (2011). Assessing the role of gender rejection sensitivity, identity, and support on the academic engagement of women in nontraditional fields using experience sampling methods. *Journal of Social Issues*, 67(3), 510–530. <https://doi.org/10.1111/j.1540-4560.2011.01712.x>
- Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open*, 3(4). <https://doi.org/10.1177/2332858417743754>
- McNamara, D. S., Allen, L. K., Crossley, S. A., Dascalu, M., & Perret, C. A. (2017). Natural language processing and learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gaevic (Eds.), *Handbook of learning analytics* (1st ed., pp. 93–104). Society for Learning Analytics Research. <https://www.solaresearch.org/hla-17/hla17-chapter8/>
- McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In *Applied natural language processing: Identification, investigation and resolution* (pp. 188–205). IGI Global. <https://doi.org/10.4018/978-1-60960-741-8.ch011>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- McNamara, D. S., Ozuru, Y., Graesser, A. C., & Louwerse, M. (2006). Validating Coh-Metrix. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 573–578). <https://www.academia.edu/download/30813803/fpo444-mcnamara.pdf>
- McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity: The Journal of the International Society for Self and Identity*, 5(4), 289–354. <https://doi.org/10.1080/15298860600805325>
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008), 1234–1237. <https://doi.org/10.1126/science.1195996>
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692. <https://doi.org/10.1093/biomet/78.3.691>
- Napper, L., Harris, P. R., & Epton, T. (2009). Developing and testing a self-affirmation manipulation. *Self and Identity: The Journal of the International Society for Self and Identity*, 8(1), 45–62. <https://doi.org/10.1080/15298860802079786>
- National Research Council of the National Academies. (2011). A review of gender differences at critical transitions in the careers of science, engineering, and mathematics faculty [S. Bell, Reviewer]. *International Journal of Gender, Science and Technology*, 3(1). <https://genderandset.open.ac.uk/index.php/genderandset/article/download/147/249>
- National Science Board. (2015, February 4). *Revisiting the STEM Workforce, A comparison to science and engineering indicators 2014* (NSB-2015-10). National Science Foundation. <https://www.nsf.gov/pubs/2015/nsb201510/nsb201510>
- National Science Board. (2016). *Developing a National STEM Workforce strategy: A workshop summary*. National Academies Press. <https://www.nap.edu>. <https://doi.org/10.17226/21900>
- National Science Foundation. (2019). *Women, minorities, and persons with disabilities in science and engineering*. <https://ncses.nsf.gov/pubs/nsf19304/>
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211–236. <https://doi.org/10.1080/01638530802073712>
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334. <https://doi.org/10.1037/a0012702>
- Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods*, 49(5), 1630–1638. <https://doi.org/10.3758/s13428-017-0874-x>
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. <https://repositories.lib.utexas.edu/handle/2152/31333>
- Pennebaker, J. W., & Chung, C. K. (2014). Counting little words in big data: The psychology of individuals, communities, culture, and history. In J. P. Forgas, O. Vincze, & J. László (Eds.), *Sydney symposium of social psychology. Social cognition and communication* (pp. 25–42). Psychology Press. <https://psycnet.apa.org/fulltext/2013-28261-002.pdf>
- Pennebaker, J. W., Chung, C. K., Frazee, J., Laverne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, 9(12), Article e115844. <https://doi.org/10.1371/journal.pone.0115844>
- Pollock, S. J., Finkelstein, N. D., & Kost, L. E. (2007). Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics—Physics Education Research*, 3(1), Article 010107. <https://doi.org/10.1103/PhysRevSTPER.3.010107>
- Priniski, S. J., Rosenzweig, E. Q., Canning, E. A., Hecht, C. A., Tibbetts, Y., Hyde, J. S., & Harackiewicz, J. M. (2019). The benefits of combining value for the self and others in utility-value interventions. *Journal of Educational Psychology*, 111(8), 1478–1497. <https://doi.org/10.1037/edu0000343>

- Pury, C. L. (2011). Automation can lead to confounds in text analysis: Back, Küfner, and Egloff (2010) and the not-so-angry Americans. *Psychological Science*, 22(6), 835–836. <https://doi.org/10.1177/0956797611408735>
- Riddle, T., Bhagavatula, S. S., Guo, W., Muresan, S., Cohen, G., Cook, J. E., & Purdie-Vaughns, V. (2015, June 26–29). Mining a written values affirmation intervention to identify the unique linguistic features of stigmatized groups. *Proceedings of the Eighth International Conference on Educational Data Mining* (pp. 274–281). International Educational Data Mining Society. <https://eric.ed.gov/?id=ED560575>
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–452. <https://doi.org/10.1037/0022-3514.85.3.440>
- Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology*, 96(4), 770–782. <https://doi.org/10.1037/a0014635>
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9), Article e73791. <https://doi.org/10.1371/journal.pone.0073791>
- Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: A systematic overview of automated methods. *Annals of the American Academy of Political and Social Science*, 659(1), 78–94. <https://doi.org/10.1177/0002716215569197>
- Serra-Garcia, M., Hansen, K. T., & Gneezy, U. (2020). Can short psychological interventions affect educational performance? Revisiting the effect of self-affirmation interventions. *Psychological Science*, 31(7), 865–872. <https://doi.org/10.1177/0956797620923587>
- Sherman, D. K. (2013). Self-affirmation: Understanding the effects. *Social and Personality Psychology Compass*, 7(11), 834–845. <https://doi.org/10.1111/spc3.12072>
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 183–242). Elsevier. [https://doi.org/10.1016/s0065-2601\(06\)38004-5](https://doi.org/10.1016/s0065-2601(06)38004-5)
- Sherman, D. K., & Hartson, K. A. (2011). Reconciling self-protection with self-improvement: Self-affirmation theory. In M. D. Alicke (Ed.), *Handbook of self-enhancement and self-protection* (Vol. 524, pp. 128–151). Guilford Press. <https://psycnet.apa.org/fulltext/2011-04015-006.pdf>
- Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A. D., & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, 104(4), 591–618. <https://doi.org/10.1037/a0031495>
- 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 & Social Psychology Bulletin*, 39(5), 663–676. <https://doi.org/10.1177/0146167213480816>
- Skrentny, J., & Lewis, K. (2013). *Building the innovation economy? The challenges of defining, building and maintaining the STEM Workforce* (No. 1). Center for Comparative and Immigration Studies.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60229-4](https://doi.org/10.1016/S0065-2601(08)60229-4)
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379–440. [https://doi.org/10.1016/S0065-2601\(02\)80009-0](https://doi.org/10.1016/S0065-2601(02)80009-0)
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.
- Tai, R. H., & Sadler, P. M. (2001). Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA. *International Journal of Science Education*, 23(10), 1017–1037. <https://doi.org/10.1080/09500690010025067>
- 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>
- Thoman, D. B., & Sansone, C. (2016). Gender bias triggers diverging science interests between women and men: The role of activity interest appraisals. *Motivation and Emotion*, 40(3), 464–477. <https://doi.org/10.1007/s11031-016-9550-1>
- Thoman, D. B., Smith, J. L., Brown, E. R., Chase, J., & Lee, J. Y. K. (2013). Beyond performance: A motivational experiences model of stereotype threat. *Educational Psychology Review*, 25(2), 211–243. <https://doi.org/10.1007/s10648-013-9219-1>
- Tibbetts, Y., Harackiewicz, J. M., Canning, E. A., Boston, J. S., Priniski, S. J., & Hyde, J. S. (2016). Affirming independence: Exploring mechanisms underlying a values affirmation intervention for first-generation students. *Journal of Personality and Social Psychology*, 110(5), 635–659. <https://doi.org/10.1037/pspa0000049>
- van Veelen, R., Derks, B., & Endeldijk, M. D. (2019). Double trouble: How being outnumbered and negatively stereotyped threatens career outcomes of women in STEM. *Frontiers in Psychology*, 10, Article 150. <https://doi.org/10.3389/fpsyg.2019.00150>
- 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>
- Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468–485. <https://doi.org/10.1037/a0037461>
- Wright, M. C., McKay, T., Hershock, C., Miller, K., & Tritz, J. (2014). Better than expected: Using learning analytics to promote student success in gateway science. *Change: The*

Magazine of Higher Learning, 46(1), 28–34. <https://doi.org/10.1080/00091383.2014.867209>

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>

Zedelius, C. M., Mills, C., & Schooler, J. W. (2019). Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods*, 51(2), 879–894. <https://doi.org/10.3758/s13428-018-1137-1>

Authors

NIA M. M. DOWELL is an assistant professor in the School of Education at the University of California, Irvine. Dowell's team conducts basic research on sociocognitive and affective processes

across a range of educational technology interaction contexts and develops computational models of these processes and their relationship to learner outcomes.

TIMOTHY A. MCKAY is an Arthur F. Thurnau Professor of Physics, Astronomy, Education, and associate dean for Undergraduate Education at the University of Michigan. His research focuses on exploring grading patterns and performance disparities both at Michigan and across the CIC, developing a variety of data-driven student support tools like E2Coach through the Digital Innovation Greenhouse, an innovation space for exploring the personalization of education, and launching the National Science Foundation-funded REBUILD project.

GEORGE PERRETT is the director of Research and Data Analysis at New York University. His work sits at the intersection of machine learning and causal inference.