Reducing Racial and Gender Achievement Gaps in STEM: Use of Natural Language Processing to Understand Why Affirmation Interventions Improve Performance

Overview

Academic achievement gaps, especially in science, technology, engineering and mathematics (STEM), have long concerned American policymakers.1,2 The causes of achievement gaps are multifaceted and arise from a reciprocal interaction between societal, social, and environmental factors that unfold from infancy into adulthood. Poor performance in challenging academic STEM domains can eventually lead students to withdraw and disengage,3 further perpetuating a lifelong process by which many members of intellectually stereotyped groups are unable to reach their potential as scholars and innovators. The long-term goal of this interdisciplinary ROADS proposal is to address this issue – with the anticipation that the initial steps described here will substantially increase our competitiveness for large-scale funding from agencies who have already shown interest in this work, such as the Spencer Foundation’s initiative in Education and Social Opportunity.

Over the last several years, our team has successfully developed and implemented a novel “values-affirmation” writing exercise based on self-affirmation theory4,5 that mitigates the effect of the identity threat experienced by members of intellectually stereotyped groups. Values-affirmation interventions are typically centered on a brief writing exercise that encourages students to focus on important values (e.g., religion, relationships with friends and family) and then write about how these are important to them. Because the intervention benefits intellectually stereotyped students the most, it has closed achievement gaps, as documented in peer-reviewed scientific research using randomized controlled trials.6-9 What explains this benefit? Can its content be distilled and refined to be made more effective? These two questions lie at the heart of this proposal. The proposed research will allow us for the first time to synthesize and systematically analyze affirmation writing exercises --- drawing from data of students who vary by race, ethnicity, age, gender, and social class, with data in many cases tied to longitudinal outcomes (e.g., grades) that extends for years. We will use recent advances in data science and natural language processing (NLP) to compile and analyze affirmation essays and determine the key ingredients of the intervention’s effectiveness.

The proposed work will be done by a team of researchers from the Graduate School of Arts and Sciences (Department of Psychology) and The Fu Foundation School of Engineering and Applied Science (Center for Computational Learning Systems). It leverages our expertise in intergroup relations and diversity, and natural language processing and semantics in particular. If successful, this project will advance our understanding of how to develop and test a more robust intervention for reducing the effects of social identity threat. This ROADS proposal allows us to digitize a large-scale collection of essays and to develop and test NLP technologies to analyze these essays. We expect that collecting and analyzing our data in the way described will greatly strengthen our applications for funding from agencies such as the NSF’s directorate of Education and Human Resources.

Objectives

We are now in a unique position, with access to an unusually large and heterogeneous dataset of approximately 10,000 affirmation essays from approximately 1,400 middle school and college students, who vary in race and gender, among other characteristics. These data have never been compiled into a single database and systematically analyzed. Funding from ROADS would allow us to compile these affirmation essays and to conduct a systematic textual analysis.

The sheer amount of written data in our cumulative affirmation essays requires us to use automatic methods for analysis. We will leverage recent NLP advances to analyze affirmation essays and determine key ingredients of the intervention’s efficacy. NLP methods offer the ability to leverage large-scale automatic analysis to investigate both coarse holistic patterns (e.g., the topics students write about) and fine-grained details (e.g., the
structure of essays).

This information will be used to identify the precise causal mechanisms that lead affirmed stereotyped students to improved performance.

Consistent with the goals of ROADS, the proposed research combines data science expertise and domain expertise in a novel fashion. Exploring the content and structure of language in the affirmation essays as a mechanism for the success of the intervention offers the potential for a high impact research discovery. In this sense, the proposed work may be considered “high-risk, high-reward.” To mitigate concerns about high-risk, however, it may be helpful to consider preliminary data that supports the ideas proposed here. Preliminary work in our lab which resulted in a published journal article has shown that the content of essays mediates the effectiveness of the affirmation for groups experiencing social-identity threat. Results from this work revealed that (a) students in the values-affirmation condition (regardless of race) spontaneously wrote about social belonging (e.g., “my family gives me love and understanding”) more than those in the control condition, and (b) that writing about social belonging statistically mediated the effect of the treatment on the improved GPA of Black students. In other words, writing about belonging improved performance for Black but not White students.

However, this relatively small study focused on only a handful of essays, and on only one particular type of content (social belonging). Moreover, human-performed coding of affirmation statements is time-consuming, suffers from the potential for human bias, and is limited in the practical range of what can be accomplished. It is therefore clear that the appropriate methodology in this instance requires cross-disciplinary collaboration. Such novel research is beyond the capability of either PI individually. It is this novel, interdisciplinary approach which makes this exciting research high-risk. Traditional funding streams often hesitate to support such risky proposals, necessitating other types of support in launching this collaboration.

Despite this risk, we are confident that the proposed research will lead to transformative insights. Early insights gained from our ROADS proposal can then help us to secure large-scale external funding. In particular, we expect this work will enhance our competitiveness for funding to refine intervention procedures to focus stereotyped students on these mechanisms and customize interventions to maximize efficacy for specific target groups. Furthermore, we anticipate that compiling this data will strengthen our future requests for funding to follow-up with our original middle-school cohorts, many of whom are now working or in college. Although it is possible that treatment effects may decay with time, research suggests that interventions delivered at key developmental transitions, such as early middle school, can set people onto trajectories of lasting change.

Together, the proposed research matches many of the goals of the Institute for Data Sciences and Engineering, including solving one of society’s most challenging problems – the academic achievement gap; improving societal health by enhancing racial and gender equality and improving economic egalitariansim, and protecting the infrastructure of the STEM research pipeline. It will also establish new interdisciplinary collaborations between the Center for Computational Learning Systems and the Department of Psychology at Columbia. The proposal will help fund one Postdoctoral Researcher co-advised by faculty from the two schools, increasing the interdisciplinary nature of a new generation of data science researchers.

Methods

Data Collection and Digitization

Our team has the unique capacity to develop and examine a repository of essays that few other teams have at this time. First, all of our essays have been collected in the context of randomized double-blind experiments, improving the strength of causal inference. Second, stereotyped and non-stereotyped students completing our interventions vary by ethnicity (African American, Latino American, White), age (middle school, high school, college), gender, region of the country, and social class (these data have been previously collected) allowing examination of how student identity and social context moderate the interventions efficacy. Third, for a portion
of these students multiple essays and performance outcomes have been collected for their three-year tenure in middle school, allowing examination of performance trajectories on the actual intervention writing task. Fourth, for middle school student essays, these essays will be linked to comprehensive objective performance outcomes drawn from school and district-wide databases. Outcomes include students’ grade point average (GPA), statewide standardized test scores, and attendance and discipline records. Importantly, a large portion of interventions were conducted in STEM courses and performance outcomes are grades in STEM. Fifth, middle school student essays will also be linked to a dataset that consists of multivariate, longitudinal data obtained from batteries of assessment completed by students at least twice a year through most of their middle school tenure (6th – 8th grade) and subsequently in high school. The inventories measure social-psychological factors thought to affect students’ academic performance, such as self-efficacy, perceptions of belonging in school, and the fear of being stereotyped in school.

Phase 1 of our research will involve an exhaustive compilation and organization of these affirmation essays and the establishment of a database of essays and basic (de-identified) participant information. At present, the individual essays are located primarily in hard copy format in the labs of project PI in different parts of the country (i.e., California, New York, Pennsylvania). Some essays have been entered electronically, but in a variety of different file formats. Thus, the first phase of the proposed project is to gather all the affirmation essays into a single, common electronic format, organize the essays with respect to details of the study and the individual participant, and create a database for use in the proposed project. The database will take the form of a spreadsheet organized by study. For each participant, the spreadsheet will contain an ID number, information about the number and timing of essays, basic demographic information, and outcome summary measures. Each essay will be stored as a separate text file that will be linked to the spreadsheet. These text files will form the raw data to be imported for NLP analyses. All files will be stored on a dedicated server that will be established as part of the proposed project.

**Large Scale, Automatic Content Analysis using Natural Language Processing (NLP)**

The volume and nature of our data motivates us to leverage automatic methods for analyzing the content and structure of the written essays. NLP methods offer the ability to investigate both coarse holistic patterns, such as essay topics, as well as more nuanced details, such as essay structure. Phase 2 of our research will explore the use of topic models for unsupervised analysis of semantic content and essay structure.

**Topic modeling and essay structure.**

One advantage of topic modeling such as Latent Dirichlet Allocation is that it can capture essay structure. As such, we will examine whether a small number of topics account for a large proportion of students’ writing, suggesting a relatively narrow focus, or if writing is associated with a broader number of topics, each accounting for a smaller proportion, suggesting a broader approach to the writing assignment. Anecdotally, for many years our team has speculated that 6th and 7th graders write about a broad number of topics while college students use a more narrow focus (See Table 1). Until now it has not been possible to test whether this pattern is statistically reliable and whether it is predictive of improved performance by age. **Topic entropy** describes the quantification of such an idea\(^4\) that can be used to investigate the relationship between

---

**Table 1: Excerpts from Affirmation Essays**

<table>
<thead>
<tr>
<th>Middle School Participant</th>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dance is important to me, because it is my passion, my life. My second home is the dance studio, my second family is my dance team. My family and friends are so important to me, even more than dance. My family, I can’t live without them. My friends, I am my real self around them (and my sister). I can be silly, goofy, and weird and they don’t care, they accept me for who I am ... And for being creative, I LOVE being creative in dance. When I’m dancing or making a dance it takes me to another place.</td>
<td></td>
</tr>
<tr>
<td>College Participant</td>
<td>Excerpt</td>
</tr>
<tr>
<td>How can one get by without friendship or family? I know I couldn’t. I need that support, at times it can feel like the only thing I have that’s real. At other times I don’t need it, but love and comfort from relationships is something that is always nice ... I was stuck in Keystone this winter and had not [way] of getting back home, I felt helpless ... I didn’t know what to do, so I called a friend and they drove 2 hours out of their way to come help me out, without even thinking twice, without that friend I would have had one bad night. Not the end of the world no, but when in need I fall back on my support, friends and family, without that support I would never stop falling.</td>
<td></td>
</tr>
</tbody>
</table>
breadth and depth of writing and academic performance. Perhaps a broader number of topics may be important for younger students who have yet to crystalize interest. Or, perhaps a broader number of topics may be important for stereotyped students who cannot easily rely on one domain to harness self-integrity given the discrimination targeted at their group. Addressing such questions offers the potential to yield key insights as to why affirmations improve performance.

Topic modeling analysis will unfold in a systematic process starting with empirically derived topics that predict academic performance in STEM courses and statewide standardized exams. We expect additional topics will emerge that have yet to be theorized about as mechanisms that improve the intervention’s efficacy. We can further explore the use of syntactic topic models,\(^8\) which will allow us to take into account both thematic aspects, as well as syntactic and local sentence level aspects of the essays. For example, we can have theme related topics (e.g., *arts, athleticism*) as well as topics related to function words (e.g., *his, her, theirs, I, she*). In the final stage of analysis, we plan to leverage topic modeling to explore the theory-driven topics described above (e.g., social belonging).

**Topic Modeling for Semantic content**

Topic modeling and social belonging. As previously highlighted, our research group has already found evidence that for stereotyped students, social belonging mediates the relationship between affirmation and academic performance.\(^8\) Topic modeling will be used to refine this finding in two ways. First, since topic modeling assigns proportional scores to each essay (i.e., what proportion of each essay includes each topic), we will develop a more sensitive measure of social belonging than the present versus absent categorization previously employed. Second, we will identify which specific words define social belonging and how this varies by intervention type and student characteristics.

<table>
<thead>
<tr>
<th>Table 2: Excerpts from Affirmation Essays</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal Regard – African American College Participant</strong></td>
</tr>
<tr>
<td>Music and art are important to me on a very personal level because they are the things I use to keep my head on straight, vent my emotions and frustrations, and I am rarely without them … I cannot imagine what a different person I would be without a constant soundtrack.</td>
</tr>
<tr>
<td><strong>Group Regard – African American College Participant</strong></td>
</tr>
<tr>
<td>There is … a hierarchy in America, and – like it or not – my group isn’t that high up … You can say this is the most important group to me because I feel a duty &amp; responsibility to it. I feel obligated to do my best to try to push this group up the ladder. There is no way of removing (even if I wanted to) myself from the group, so I feel commitment to it.</td>
</tr>
</tbody>
</table>

Table 2), a phenomenon less likely to appear in the control condition and less likely to occur among other stereotyped groups, such as women and Latinos. Topic modeling could reveal that students who improved performance tend to use affect-laden words in conjunction with their self and with their group (e.g. *African-American* and *commitment*), leading to the detection of a topic that contains a mixture of group-membership and emotion words. If this hypothesis were supported, it could suggest that our affirmation interventions might be tailored in the future to encourage students to reflect on positive self and/or group attributes.

**Data Analysis and Manuscript & External Funding Proposal Creation.**

In Phase 3 we will use the output from the NLP analyses as predictors in a series of multiple regression analyses as proof of concept that will allow us to seek external funding. Primary dependent variables will be short- and long-term STEM grades and performance on standardized tests. Secondary dependent variables will include long-term outcomes collected as part of the follow-up from our earlier middle-school studies (e.g., college status, civic engagement, STEM career) and other variables of interest, such as school disciplinary incidents, health outcomes (as these are related to educational outcomes\(^17\)-\(^19\)), and perseverance in a STEM major. One advantage of the current proposal is our large sample, which allows meaningful subgroup analyses for outcomes available in some but not all studies.
In particular, analyses will focus on the following: (1) examining which NLP variables are associated with the primary and secondary dependent variables; (2) testing which NLP variables (alone or in combination) mediate the relation between affirmation condition (control versus intervention) and academic performance, thus providing an explanatory mechanism for affirmation effectiveness; (3) testing psychological or demographic factors as possible moderators of the proposed mediation pathway, including any difference between our middle school and college samples and any difference by race or gender; and (4) examining how NLP variables unfold longitudinally for students who completed multiple affirmation essays (i.e., the NLP variables themselves become the dependent variable).

Given the effectiveness of the intervention, which is based on the affirmation writing exercise, we are confident that the analysis of essays will yield clues to the affirmation mechanism that will be suitable for publication. In addition, these studies will allow us to extrinsically evaluate the NLP techniques used for the content analysis.

**Conclusion & Suitability for External Funding**

The project, which compiles widely-dispersed already-existing data into a centralized database, allowing for novel analytic approaches meets the targeted goals of (1) combining data science with psychological domain expertise, (2) brings together researchers from computer science and psychology to cross traditional discipline boundaries for a data-centric project, and (3) helps to solve one of society’s most challenging problems – the academic achievement gap. Furthermore, the compilation and initial analysis of this data will strengthen future external large-scale funding applications by delivering concrete evidence that such a unique approach to analyzing these data will work, and allow us to follow-up by not only collecting further longitudinal data on our original cohorts of students, but also test refinements of the intervention in controlled lab settings. Research on affirmation interventions has attracted the attention of funding agencies such as the Spencer Foundation, NSF, the Robert Wood Johnson Foundation, Russel Sage Foundation, and others, and we expect that the work outlined here will serve to increase the attractiveness of funding these ideas. Finally, in addition to the research including a collaboration between two schools (Arts & Sciences and the Fu Foundation School of Engineering and Applied Sciences), the work also compliments the work being done within several centers within the IDSE. For example, although our primary outcomes are academic achievement, we anticipate this grant will enable us to apply for funding which would broaden the scope of our investigations to include health outcomes, making the proposal fit well within the center for health analytics. Additionally, investigating NLP applications to written essays constitutes a form of new media insofar as this form of media has not yet been thoroughly investigated in a field where many tools are trained on other types of media, such as newswire data. This project thus could have connections to the New Media Center. Finally, the project could serve as a motivating factor to establish a center on learning analytics. As this project makes clear, problems in education can clearly benefit from analytic solutions drawing on “big data” approaches and methods in computer science.

The research proposed includes: (1) development of theory-informed and empirically tested educational practices which improve academic outcomes of underrepresented groups in STEM, (2) development and dissemination of scientific knowledge about psychological processes under identity threat, and (3) enhanced understanding of a cost-effective and tested educational practice suitable for large-scale dissemination with potential impact at the national level.
References

1. Archer, L. & Francis, B. Understanding Minority Ethnic Achievement: Race, Gender, Class and 'Success'. (Routledge, 2006).