

The Moderating Effect of Faculty Mentorship on Undergraduate Students' Summer Research Outcomes

Angelica Monarrez,[†] Danielle Morales,[‡] Lourdes E. Echegoyen,[§] Diego Seira,[†] and Amy E. Wagler^{†*}

[†]Department of Mathematical Sciences, [‡]Department of Sociology and Anthropology, and

[§]Department of Chemistry, University of Texas at El Paso, El Paso, TX 79968

ABSTRACT

Summer undergraduate research experience (SURE) programs are proven interventions that provide undergraduate students with opportunities to develop research skills under the mentorship of a faculty member. These are essential programs, particularly for members of underrepresented minorities, because SUREs are known to broaden their participation and increase retention. We present the results of a study investigating the influence of faculty mentorship quality on the quality of research presentations for undergraduate students attending a 10-week, distributed, multi-institutional SURE program focused on biomedical research training. Upon returning to the home institution, students presented research posters at a local symposium. Poster presentations were judged using a scale validated as part of this project. Combining collected information on student demographics and their self-reported assessments of research gains and belonging to the scientific community, we made use of data analytics methodologies to appropriately merge and analyze the data to address the overarching research question: What are the independent and combined effects of the quality of faculty mentorship and student characteristics on the quality of SURE student poster presentations? Results show that faculty mentor quality functions as a moderating influence for student characteristics on research presentation quality. Implications and recommendations for SURE program implementation are discussed.

INTRODUCTION

Due to the current shortage of biomedical researchers and severe lack of diversity in the field (Oh *et al.*, 2015), there is a call to broaden participation in science, technology, engineering, and mathematics (STEM)-related research, especially in the biomedical sciences (Valantine and Collins, 2015). Federal agencies such as the National Science Foundation and the National Institutes of Health (NIH) have funded various undergraduate research programs at universities to recruit and retain more students, especially members of minoritized groups, in biomedical sciences. While there have been improvements (Chesler *et al.*, 2010), members of minoritized groups continue to be highly underrepresented in biomedical sciences (Chesler *et al.*, 2010; McGee *et al.*, 2012; Valantine *et al.*, 2016). To reverse course, directors of biomedical research programs have been developing and implementing high-quality interventions to broaden the participation of underrepresented minorities (URMs). Among them, faculty-mentored summer undergraduate research experience (SURE) programs are proven interventions that provide undergraduate students with opportunities to engage in transformative research experiences and enhance their knowledge and professional credentials (Adedokun *et al.*, 2012). While SURE programs are known to be effective, the vast majority of studies have focused on mid- or long-term outcomes for students, such as improved research skills (Mabrouk, 2009; Thiry and Laursen, 2011; Daniels *et al.*, 2016), increased science self-efficacy (e.g., Russell *et al.*, 2007; Junge *et al.*, 2010;

Graham F. Hatfull, *Monitoring Editor*

Submitted May 4, 2020; Revised Aug 27, 2020; Accepted Sep 1, 2020

CBE Life Sci Educ December 1, 2020 19:ar56

DOI:10.1187/cbe.20-04-0081

*Address correspondence to: Amy E. Wagler (awagler2@utep.edu).

© 2020 A. Monarrez *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Lopatto, 2010; Fakayode *et al.*, 2014; Ghee *et al.*, 2016), or likelihood to enroll in graduate school (e.g., Gonzalez-Espada and LaDue, 2006; Russell *et al.*, 2007; Pender *et al.*, 2010; Eagan *et al.*, 2013; Carpi *et al.*, 2017). Less research has focused on short-term outcomes of SURE programs, such as students' presentations (e.g., Falconer and Holcomb, 2008; Burge and Hill, 2014; Corwin, Graham, and Dolan, 2015). Yet those short-term outcomes are known to increase research self-efficacy of undergraduate researchers, and, in the long-term, improve their science identity (Robnett *et al.*, 2015; Estrada *et al.*, 2018). In fact, little is known about the effects of specific program components, such as faculty mentorship, on specific student outcomes, such as student poster presentations. To fill these gaps, we present in this article the results of our investigation on how faculty mentorship quality influences the quality of poster presentations for students attending a multi-institutional SURE program.

It is well documented that students need to develop multiple skills to rise to the level of experts as researchers, and communicating research effectively is one of those skills. For undergraduate researchers, conducting successful presentations at a research symposium or conference increases their communication skills, which in turn leads to an increase in their science identity, research self-efficacy, and sense of belonging to the scientific community (Fakayode *et al.*, 2014; Corwin *et al.*, 2015; Robnett *et al.*, 2015). In fact, many SURE programs offer opportunities for participants to present their summer research projects at internal research symposia or national conferences (Gonzalez-Espada and LaDue, 2006; Junge *et al.*, 2010; Fakayode *et al.*, 2014). However, very little is known about the quality of those presentations. Because presenting at scientific conferences is a significant part of academic life, it is essential to examine not only the quantity of presentations SURE students deliver, but also the quality of the presentations.

Although there are complicating factors involved in assessing students' research quality, as the measurement of research quality is quite complex, measuring the quality of a presentation should be straightforward. However, SURE programs use varying judging rubrics that are usually not validated, so the conceptual integrity and internal consistency become an issue. To our knowledge, in peer-reviewed publications, there is currently no information about the construct validity and reliability of poster-judging scales. Without this information, SURE programs should not be using these scales for assessing the quality of student research or, even worse, ranking student research quality.

Finally, one crucial aspect of SURE programs is mentoring relationships between faculty mentors and undergraduate students. Previous research has suggested that, in the context of undergraduate research experiences, faculty mentorship has the potential to increase retention of undergraduate students, especially URMs in STEM degrees (Daniels *et al.* 2016, 2019; Estrada *et al.*, 2018; Morales *et al.*, 2019). Particularly, when matched with a faculty mentor, an undergraduate student is more likely to continue in the STEM field and graduate (Haeger and Fresquez, 2016) as well as enroll in graduate school (Morley *et al.*, 1998; Gonzalez-Espada and LaDue, 2006; Morales *et al.*, 2019). Some researchers have measured the quality of faculty mentorship by examining interview data from students about their relationships with their faculty mentors (Adedokun *et al.*, 2012; Daniels *et al.*, 2019). Others have collected survey

data about students' satisfaction with the mentoring relationship, time spent with faculty mentors, and freedom during the research experience (Cox and Andriot, 2009; Daniels *et al.*, 2016; Ghee *et al.*, 2016). While the effect of faculty mentorship on SURE students is well understood, little is known about the moderating effect of faculty mentoring on undergraduate students' research outcomes. In contrast, at the graduate level, there are studies suggesting that mentoring relationships mediate student characteristics on graduate research outcomes (Mansson and Myers, 2012; Brill *et al.*, 2014). In this study, we used the Mentoring Competency Assessment (MCA) to measure the quality of faculty mentorship (Fleming *et al.*, 2013) and investigate how the quality of faculty mentoring mediates the relationships between student characteristics and student research products, specifically the quality of their poster presentations, in the context of undergraduate research programs.

Taken together, this study contributes to the literature by linking the quality of faculty mentorship to the quality of students' research presentations in the context of a distributed, multi-institutional SURE program. With this goal in mind, we address the research question: What are the independent and combined effects of the quality of faculty mentorship and student characteristics on the quality of SURE student poster presentations? To address this research question, we developed a standard poster-judging rating (PJR) scale to measure the quality of SURE students' posters at a research symposium. The PJR scale is a slightly modified version of the scale originally developed by Rachel Hayes-Harb at the University of Utah (Hayes-Harb, 2018). Then, we analyzed scores from the PJR scale to assess construct validity of the item set. Finally, the PJR scale scores were used to identify moderating effects of faculty mentorship (based on the MCA survey) on student characteristics of participants in this study. These results were then used to generate pragmatic program recommendations for institutions wishing to improve the quality of research outcomes, such as presentations.

METHODOLOGY

Study Context

We focused on an NIH-funded multi-institutional (distributed) SURE program for students from the University of Texas at El Paso (UTEP). UTEP is unique among research institutions as a Hispanic-serving institution (HSI) that enrolls more than 25,000 students, 21,000 of whom are undergraduates. Around 85% of the undergraduate student population is Hispanic, and approximately 55% of those students are the first in their families to attend college. The central mission of UTEP is to ensure access to high-quality education, research, and training programs to prepare students to meet their professional goals. The goal of the SURE program is to broaden participation in biomedical research careers by involving students in transformative undergraduate research experiences. For this SURE program, undergraduate students participated as apprentices in 10 weeks of traditional summer research experiences with faculty mentors (Lopatto, 2003) in biomedical sciences at 14 different institutions. Participating students received various levels of research training, depending on their classifications during the academic year, to put their preparation for the summer experience into perspective. Specifically, freshmen participated in a freshman-year research-intensive sequence consisting of a research foundations course and one or two course-based undergraduate research

experiences (CUREs; Auchincloss *et al.*, 2014); sophomores participated in mentee training (Morales *et al.*, 2020), participated in CUREs if they had not done so as freshmen, and were encouraged to participate in research as apprentices with established faculty researchers at UTEP, which some elected to do; the latter was not an option for junior students, as they were required to participate in research projects with faculty. All students, regardless of classification, participated in approximately 30 hours of professional development through a series of workshops that took place in early January. For their summer experiences, the students were matched with faculty mentors at one of the 14 different institutions. Matching was done based on both student and faculty research interests. Of the 14 locations where students conducted summer research, although four are designated as HSIs, 13 are predominantly white; nine institutions are doctoral universities with very high research activity (Carnegie classification); four are special focus medical schools (Carnegie classification); and one is a large pharmaceutical company. Although nine of the research locations had well-established summer research programs, all 14 had faculty/research scientists experienced in mentoring undergraduate students in research. Moreover, through the SURE program, all mentors had access to mentoring resources, and students were also provided training about what to expect in a research mentoring experience and were informed about the expectations of the research mentors. This kind of structure provided a context to allow for individual heterogeneity with regard to mentoring style and approach, but also supported mentors so they could be effective and provide the structure needed for a quality research experience. Students received a stipend and were required to attend weekly mentee training, seminars, and research/professional development workshops. Toward the end of the summer, students worked closely with their (external) summer faculty mentors to submit abstract about their research projects for a presentation at their home institution undergraduate research symposium. Preparing the posters and their presentations in most cases took place 1.5 to 2 months after returning to the home institution. Therefore, rehearsing the presentation with the mentor could only take place virtually, if at all. Visiting program directors from partner institutions, and home institution faculty members, postdoctoral fellows, and graduate students voluntarily served as judges for the symposium. Before the symposium, all judges received both the PJR and detailed instructions on how to use it. Students presented their posters to at least two judges, and using the PJR scale, the judges provided scores for students. In summer 2018, 100 students participated in the SURE program, and 84 of them presented at the symposium and received scores.

Data Collection

An evaluation survey was administered to students at the end of the SURE program but before the symposium. It included questions on students' research experiences and mentoring relationships. The survey was paper based or online for those who could not attend the paper-based session. Students were allowed all the time they needed to complete the survey. Of the 100 students who participated in the SURE program, 67 took the evaluation survey, and 84 presented their research findings at a symposium. Taken together, 40 students took the survey and were judged at the symposium. We included these 40 students in our sample. Student demographics, including gender,

race/ethnicity, first-generation status, classification, major, and prior research experience are presented in Table 1 ($n = 40$).

This research was approved by the University of Texas at El Paso's Institutional Review Board under protocol no. 746424.

Instruments

Three categories of information about SURE student experiences were collected: 1) PJRs; 2) information about mentor competency using the MCA; and 3) individual characteristics about the students from the Undergraduate Research Student Self-Assessment (URSSA) scale and from Estrada *et al.*'s (2011) scale of science self-efficacy and science identity. Details about these are outlined in the following sections.

PJR

To assess the ability of students to present their research projects, we used the judges' scores from the poster presentation at the symposium (Hayes-Harb, 2018). During the symposium, judges answered 10 questions on the PJR scale about the students' presentations. The 10 questions were created based on the following learning outcomes: 1) the student identifies and uses relevant previous work that supports the research, scholarly, or creative work; 2) the student articulates a timely or important research question or creative objective; 3) the student identifies and uses appropriate methods to address the research question or creative objective; and 4) the student presents the research effectively in a conference setting. For judges to score each question, we used four Likert-scale items from 1 = needs improvement ("the student is in the very early stages of development with respect to this learning outcome") through 4 = outstanding ("the student has mastered this learning outcome without any flaws"). The sum scores for identified subscales for the PJR are used as described in the next section.

An important point to keep in mind is that students may have received help from mentors in preparing their presentations and may have rehearsed and memorized them. However, answering direct questions from judges cannot be rehearsed. The ability to answer research questions is considered a medium- to high-level skill strength according to the Researcher Development Framework (Vitae, 2014; Kneale *et al.*, 2016). Within the fourth learning outcome of the PJR, we had a specific item measuring a student's ability to answer judges' questions during the presentation.

MCA

To account for the effect of mentoring competency on student gains, we used questions from the MCA instrument (Fleming *et al.*, 2013). The MCA is a validated skills inventory that consists of 26 items; students can rate their faculty mentors using a 7-point Likert-type scale in which 1 = not at all skilled, 4 = moderately skilled, and 7 = extremely skilled. The MCA covers the following competencies: maintaining effective communication, aligning expectations, assessing understanding, addressing diversity, promoting professional development, and fostering independence. Sum scores of the MCA are used for each student-faculty pair in the analysis described in *Statistical Analysis*.

URSSA

We used the four constructs from the URSSA instrument (Weston and Laursen, 2015): 1) "thinking and working like a

TABLE 1. Student demographic information

Variable	Number	%	Mean score
Classification			
Freshmen	0	0.00	N/A
Sophomores	6	15.00	31.23
Juniors	15	37.50	31.23
Seniors	19	47.50	31.19
Did not specify	0	0.00	N/A
Major			
Biochemistry	3	7.50	29.43
Psychology and sociology	1	2.50	N/A
Biological sciences	12	30.00	31.5
Mechanical engineering	4	10.00	33.00
Microbiology	1	2.50	28.00
Sociology	2	5.00	32.17
Cellular and molecular biochemistry	4	10.00	32.09
Physics	1	2.50	35.67
Chemistry	1	2.50	30.33
Kinesiology	2	5.00	29.50
Health promotion	0	0.00	N/A
Electrical engineering	2	5.00	27.57
Psychology	5	12.50	29.94
Social work	1	2.50	35.33
Computer science	1	2.50	36.00
Pre-engineering	0	0.00	N/A
Engineering leadership	0	0.00	N/A
Previous experience: total research weeks			
0	2	5.00	28.83
1–20	6	15.00	30.23
21–40	13	32.50	31.78
41–60	8	20.00	28.44
61–80	10	25.00	30.71
Did not specify	1	2.50	N/A
Gender			
Male	9	22.50	31.60
Female	29	72.50	30.46
Did not specify	2	5.00	29.67
Race/ethnicity			
American Indian	1	2.50	36.50
White-Hispanic	31	77.50	31.40
White-non -Hispanic	2	5.00	32.00
Other choices	4	10.00	29.55
Did not specify	2	5.00	29.67
First-generation college			
Yes	19	47.50	
No	18	45.00	
Did not specify	3	7.50	
Total	40		

scientist,” 2) “personal gains related to research work,” 3) “skills,” and 4) “attitudes and behaviors,” because they are related to undergraduate students’ research experiences. 1) “Thinking and working like a scientist” refers specifically to student reports of growth in applying scientific knowledge and skills, understanding the scientific research process, and improving their intellectual understanding of the field. 2) “Personal gains” relates to student reports of improvement in comfort and

ability working within the scientific field. 3) “Gains in knowledge and skills” measures student reports of acquisition of new skills and knowledge within the field and expansion of their existing knowledge outside the field. 4) “Attitudes and behaviors” refers to students’ reports on attitudes and behaviors about working in a scientific community, feelings of creativity, independence, and responsibility around working on scientific projects. As per URSSA, the gains scales items on the survey were

rated on a five-point Likert scale ranging from 1 = no gain to 5 = a great gain. For each of the scales used, we made use of the preprogram scores and the difference (post–pre) scores. The pre scores helped us capture the baseline status of the students, while the difference determines the change in these measures during the SURE. Various measures are used that demonstrate strong predictive power to predict PJR scores in the analysis.

Science Self-Efficacy, Science Identity, and Research Self-Efficacy

To measure science self-efficacy and science identity, the survey contained items from Estrada *et al.* (2011). Science self-efficacy assesses students' ability to function as a scientist in a variety of tasks. Items include "use technical science skills (use of tools, instruments, and/or techniques)," "generate a research question to answer," "figure out what data/observations to collect and how to collect them," "create explanations for the results of the study," "use scientific literature and/or reports to guide research," and "develop theories (integrate and coordinate results from multiple studies)." Each item is rated on a scale of 1 (not at all confident) to 5 (absolutely confident). The science identity items included "I have a strong sense of belonging to the community of scientists," "I derive great personal satisfaction from working on a team that is doing important research," "I have come to think of myself as a 'scientist,'" "I feel like I belong in the field of science," and "the daily work of a scientist is appealing to me." Each item was rated on scale of 1 (strongly disagree) to 5 (strongly agree) to the extent each statement was true for the student. To measure research self-efficacy, some items were taken from the research self-efficacy scale (Bieschke *et al.*, 1996), which assesses students' perceptions of performance capability based on the following factors: 1) find and research an idea; 2) present and write the idea; 3) finalize the research idea and method; 4) conduct the research; 5) analyze data; and 6) write and present results.

Statistical Analysis

After data cleaning, validation, and merging of the surveys and poster-judging scores, the data analysis proceeded through three stages: in stage 1, a modified parallel analysis was performed on the poster-judging scores to assess the dimensionality of the poster-judging scale; in stage 2, the emerging factors from the poster-judging scores were predicted using recursive partitioning models to identify a subset of strong predictors and identify reasonable data-driven interactions from the full set of URSSA variables; and in stage 3, censored regression (Tobit) models were used to describe the relationships between the identified factors from stage 2 with the poster-judging scores. Each of these steps was necessary to fully address the nature of the scale data and sort through the 289 potential predictor variables available for modeling PJR scores.

Modified Parallel Analysis of the PJR

Each scale involved in this study was assessed for dimensionality. This included well-understood scales, such as those for science identity and research self-efficacy, but also the judging score scale, which is not a previously validated scale. Hence, we conducted modified parallel analysis to assess the dimensionality and structure of the PJR scale. All modified parallel analyses were performed in *R* using the *psych* package (Revelle, 2018).

Recursive Partitioning Models

Recursive partitioning models are predictive models that make use of tree structures to understand relationships between variables. These models are particularly well suited for scenarios in which there are a preponderance of explanatory variables that may help predict a response variable. The procedure requires minimal model assumptions and is highly effective for settings where the predicted variable is ordinal. Moreover, this procedure is particularly adept at identifying potential interaction (moderations) in the data for a particular predictor variable. The suitable predictors and interactions identified using the partitioning models were included in the censored regression models described below. All recursive partitioning models reported in the *Statistical Analysis* section were estimated in *R* using the *rpart* package (Therneau and Atkinson, 2018).

Censored Regression (Tobit) Models

Following the recursive partitioning analysis, the identified predictor variables and potential interactions were included in censored regression models. Tobit models, a censored regression model, were used for predicting the judging subscales. This is an appropriate mode of analysis, as the judging subscales are left- and right-censored variables (see Table 4 later in the article for the potential range of scores for each factor). Models specific to ordinal-level measurements, such as proportional odds or cumulative logit models, would not be parsimonious and tend toward overfitting. The Tobit model does not censor observed scores but adjusts the model to take into account the censoring imposed by the potential range of values for sum score measures. It allows for modeling of data when there are potential ceiling or floor effects in the data and is a powerful tool for detecting differences when censoring is imposed by the data-collection process. The model-fitting procedure was carried out using the following algorithm:

- Step 1. Null and full models were fit using the Tobit model. In particular, the null model includes all of the predictor variables identified in the recursive partitioning models (null), and the full model used the same set of variables but allowed for second-order interactions between all terms.
- Step 2. A stepwise variable selection algorithm was applied to the null and full model sets so that a most-parsimonious model using a subset of the linear and interaction terms was identified. Both Akaike's information criterion and the Bayesian information criterion were used to simultaneously assess model parsimony.
- Step 3. Once a best-fit model was identified, the model was studied to answer the research question. In addition, appropriate mean comparisons and explanatory plots were used to assist in the model explanation. All Tobit analyses were performed in *R* using the *censReg* package (Henningsen, 2017). Pseudo- R^2 will indicate a model fit for each model, and Cohen's f^2 effect-size measure indicates the practical significance of individual factor effects. Based upon Cohen's (1988) guidelines, $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ represent small, medium, and large effect sizes, respectively.

RESULTS

Using the described analyses, the following results emerged to answer the central research question: What are the independent and combined effects of the quality of faculty mentorship and student characteristics on the quality of SURE student poster presentations?

Modified Parallel Analysis Results

Three dominant factors emerged conforming relatively well to the hypothesized structure. The first factor (communicates previous research) includes the items that were designed to assess the students' ability to communicate the prior literature pertaining to their research project. The second factor (communicates research objective and methods) identifies the three indicators asking about how students frame the research question, research objective, and research methods. Finally, the last and third factor (communicating research results) consisted of all indicators specifying how well the student presents the research results in both oral and written/visual form (i.e., the poster). While these empirical factors do not exactly conform to the hypothesized structure, it is very reasonable and close to the reasoning used for the hypothesized structure. These three factors—factor 1 (communicates previous research), factor 2 (communicates research objective and methods), and factor 3 (communicates research results)—are outlined in Table 2. This table also includes the factor loadings and uniqueness statistics for each identified factor. Reliability of the three factors is high, ranging from 0.80 to 0.91 ($\alpha_1 = 0.85$, $\alpha_2 = 0.80$, $\alpha_3 = 0.91$). The three-factor model fit well to the observed data (Tucker Lewis index [TLI] = 0.968, root mean square error of approximation [RMSEA] = 0.078), indicating a relatively good fit and stable loading structure. We note that a one-factor loading structure

was also tested and indicated a reasonable fit (TLI = 0.946, RMSEA = 0.102). However, due to theoretical considerations, we chose the three-factor model in order to separate what we regard to be tangibly different components of poster research quality (e.g., prior literature, research objective and methods, research results).

Recursive Partitioning Results

Using the three identified factor scores from the PJR scale, a multivariate recursive partitioning model was run on the full predictor set. Table 3 provides a list of variables identified using all three factor scores, in order of importance. This set of variables was used for the remaining analysis, in which each individual factor score for research quality was modeled separately. The recursive model also indicated probable interactions between the sum score for MCA and the following set of variables: post–pre difference in likelihood of pursuing a biomedical research career, a sum score for diversity, post–pre difference in research self-efficacy, post–pre difference in science identity, sum score for personal gains, sum score for satisfaction with research mentorship, and sum score for thinking and working like a scientist. These interactions will be considered for inclusion in the censored regression (Tobit) models.

Tobit Modeling Results

Due to the censored nature of scale data, we modeled the three identified factors for research quality using Tobit models. Proportional odds models were also fit but resulted in an unstable fit due to the semicontinuous nature of the sum scale scores. A linear model that does not correct for censoring in the data would not be well suited and was not fit to the data. The model results are summarized in Table 4. In the table, three models are

TABLE 2. PJR scale proposed structure, indicators, and empirical factors

Proposed outcome	Indicators (loading; uniqueness)	Empirical factors (predictor variables)
Identify and use relevant previous work that supports the research, scholarly, or creative work	1. Presenter provides sufficient background information to place the project in an appropriate scholarly context. (0.72; 0.22)	Factor 1: Communicates previous research
	2. Presenter effectively communicates the significance of the project and contribution to the field or society. (0.70; 0.23)	
Articulate a timely or important research question or creative objective	3. Presenter clearly articulates the research question or creative objective. (0.56; 0.31)	Factor 2: Communicates research question and methodology
	4. The research question or creative objective follows logically from the previous work cited. (0.61; 0.38)	
Identify and use appropriate methods to address the research question or creative objective	5. Presenter clearly explains the methods and links methods to the project objective. (0.50; 0.47)	Factor 3: Communicates research results
	6. Presenter effectively communicates the project progress and results, and interprets results with respect to the research question or creative objective. (0.60; 0.29)	
Present the research effectively in a conference setting	7. Presentation materials, performance, or visuals are relevant and of professional quality. (0.52; 0.34)	
	8. Presentation is structured, organized, and flows logically. (0.55; 0.39)	
	9. Presenter has command of the topic and can easily answer questions. (0.64; 0.26)	
	10. Presenter is clear, enthusiastic, and effectively engages the audience. (0.73; 0.26)	

TABLE 3. Important predictor variables for PJR factor scores

Variable name	Variable description
Diff in SE	Post-pre difference in research self-efficacy
Diff in BM	Post-pre difference in likelihood of pursuing a biomedical research career
SATIS	Sum score for satisfaction with mentoring relationship
ATT	Sum score for attitudes toward science
Diff in SI	Post-pre difference in science identity
DIV	Sum score for diversity
preSI	Pre sum score for science identity
PG	Sum score for personal gains
RS	Sum score for research skills
TWS	Sum score for thinking and working like a scientist
MCA	Sum score for mentor competency scale
DIV	Sum score for mentors' skills in discussing diversity with the mentees and valuing and respecting cultural differences
TR	Total number of months of prior research experience
Born in US	Indicator variable, with "1" indicating born in the United States

presented that provide reasonable fit and explanatory power for each of the three latent factors for the PJR (e.g., prior literature, research objective and methods, and research results). As indicated in the table, all models show adequate fit, given the nature of the data (pseudo $R^2 = 0.45$ [factor 1], 0.46 [factor 2], and 0.40 [factor 3]). For all three models, the data are right-censored, indicating that many ratings are at the top level of the scale. However, the log regression standard error is highly significant for all three models, indicating that the standard error of the Tobit regression differs from zero ($p_{\log \sigma} < 0.001$ for all three models) and is commensurate with the ordinary least-squares mean square error. Taken together, these results indicate the Tobit model is an appropriate choice. The Tobit models include both main effects and interaction terms. We note that, in general, the interaction terms can be interpreted using the following guidance: 1) When the interaction effect for MCA and another factor is positive, it indicates that high levels of research mentor quality augments the effect of the other variable, 2) When the interaction is negative, it implies that when MCA is high, it decreases research quality for correspondingly high levels of the other variable.

Beginning with the interaction terms identified in the recursive partitioning models involving MCA, we note that the estimated slope for thinking and working like a scientist (TWS) and MCA is positive for all three factors ($p < 0.01$ for all three factors). This positive slope implies that the impact of TWS is magnified by correspondingly high levels of mentor competency. A similar positive slope is located for the post-pre difference in pursuing a biomedical research career (diffBM) and MCA as well ($p < 0.01$ for factor 1 and $p < 0.001$ for factors 2 and 3). These two results demonstrate the added benefit of quality mentoring on research quality, even for those with high levels of difference scores for pursuing a biomedical research career and TWS.

In contrast, the satisfaction with research experience (SATIS) and MCA interaction has a negative slope ($p < 0.01$ for factors 1 and 3 and $p < 0.05$ for factor 2). This indicates that the effect of quality mentoring on research qualities diminishes at higher

TABLE 4. Tobit analysis results for communicating research literature (factor 1), research questions (factor 2), and research results (factor 3)

	Factor 1 model	Factor 2 model	Factor 3 model
(Intercept)	-9.18* (4.58)	-7.65 (5.85)	12.88 (20.51)
MCA	3.08*** (0.78)	2.77** (0.98)	0.30 (3.64)
Diff in BM	-4.69** (1.81)	-8.15*** (2.30)	-16.40*** (3.96)
SATIS	10.28** (3.63)	8.26* (4.15)	36.25** (11.92)
Diff in SI	-3.68* (1.52)		-18.88 (10.50)
DIV	0.45 (0.24)		0.32* (0.14)
preSI	-1.34* (0.56)	-3.13 (1.93)	-21.22 (11.78)
TR	0.03** (0.01)	0.13*** (0.04)	0.36*** (0.11)
SE	0.81 (0.43)	1.27** (0.48)	
TWS	-5.71** (2.20)	-11.21** (3.61)	-19.98** (6.41)
ATT			7.22 (4.32)
Diff in SE		1.49** (0.51)	2.49* (0.98)
Born in US		0.26* (0.11)	0.38 (0.21)
PG		7.60** (2.75)	
MCA:Diff in BM	0.82** (0.28)	1.25*** (0.36)	2.58*** (0.60)
MCA:SATIS	-1.79** (0.58)	-1.38* (0.68)	-5.64** (1.80)
MCA:Diff in SI	0.59* (0.26)		2.85 (1.61)
MCA:DIV	-0.11* (0.05)		
MCA:TWS	1.06** (0.40)	1.73** (0.57)	3.01** (1.08)
MCA:preSI		0.57 (0.31)	3.41 (1.78)
MCA: PG		-1.41** (0.47)	
MCA: TR		-0.02*** (0.01)	-0.06** (0.02)
MCA: ATT			-1.31 (0.74)
logSigma	0.42*** (0.08)	0.55*** (0.07)	1.12*** (0.07)
Number observed	115	115	115
Left-censored	0	0	0
Uncensored	86	102	98
Right-censored	29	13	17

* $p < 0.05$.** $p < 0.01$.*** $p < 0.001$.

satisfaction levels. In addition to these interactions, personal gains due to research (PG) negatively interacted with MCA for factor 2 (communicating research question and methods, $p < 0.01$) and can be interpreted similarly. Additionally, MCA and total research experience (TR) had a negative interaction for factor 2 as well as factor 3 (communicating research results, $p < 0.001$ for factor 2 and $p < 0.01$ for factor 3). These negative interactions also indicate a diminished return on mentor competency at high levels of total research experience or perceived personal gains.

Main effects that warrant interpretation include the following. First, the difference in likelihood of pursuing a biomedical research career (diffBM) has a large negative effect for all three factors ($p < 0.01$ for factor 1 and $p < 0.001$ for factors 2 and 3) and small practical significance levels using Cohen's f^2 (all smaller than 0.07). Recall too that MCA and the biomedical research career variable have a positive interaction effect; however, the magnitude of the negative main effect is noteworthy. Similarly, thinking and working like a scientist (TWS) also has a significant negative main effect slope ($p < 0.01$ for all factors) and corresponding positive interaction with MCA. This indicates that those entering the research program with high levels of TWS and diffBM are less likely to have high levels of research quality by the end. However, the magnitude of the negative effect is diminished by high levels of mentor competency. Satisfaction with the research experience program also results in high research-quality scores. This effect is statistically significant ($p < 0.01$ for factors 1 and 3 and $p < 0.05$ for factor 2) and also demonstrates smaller levels of practical significance (Cohen's $f^2 = 0.050, 0.031, \text{ and } 0.028$ for factors 1, 2, and 3, respectively). Similarly, the total amount of research experience also has a statistically significant positive effect ($p < 0.01$ for factor 1 and $p < 0.001$ for factors 2 and 3) and has practical significance (Cohen's $f^2 = 0.042, 0.043, \text{ and } 0.030$ for factors 1, 2, and 3, respectively). Other variables that indicate some effect on any of the three factors include a moderate effect on factor 1 (communicating research literature) from difference in science identity (Diff in SI, $p < 0.05$) and prescore for science identity (preSI, $p < 0.05$), on factor 2 (communicating research question and methods) from the sum score of research self-efficacy (RSE, $p < 0.01$) and personal gains (PG, $p < 0.01$), and on factor 3 (communicating research results) from diversity sum score (DIV, $p < 0.05$). All of the practical significance values for these terms exceed 0.02 but are also smaller than 0.07, indicating a small practical effect. Due to the preponderance of variables included in each model, the lack of strong practical significance is not a surprise. Instead, these results demonstrate that all models show a strong cumulative effect due to the combined effects of the included variables.

DISCUSSION

The results of the analysis illuminate the role of research mentoring quality in SURE program outcomes, with a focus on the quality of research presentations. They also illustrate how mentoring can have differential effects on research product (presentation) outcomes based on student characteristics. Though there is substantial evidence in the literature regarding the efficacy of research mentoring on student research products, this study provides a more detailed analysis about how mentoring can bring added value and when it has a less comparative

impact. The results of this study can help program directors and faculty research mentors to better allocate time to maximize impact on student research (presentation) outcomes. The presentation outcomes explored in this study include the student's ability to communicate the previous research results, the research methodology, and the research results. Note that communicating the student's research results includes answering extemporaneous questions. This provides evidence about the ability of the scale to measure the quality of the research presentation.

A host of student characteristic variables, such as science identity and self-efficacy, among others, are moderated by MCA. Those with positive interaction slopes can be regarded as evidence of quality mentoring being a "value-added" component contributing to quality research (presentation) products of SURE students. The variables with a positive, and hence value-added interpretation, include: difference score for intent to pursue biomedical research career (diffBM), thinking and working like a scientist (TWS), difference in science self-efficacy (Diff in SE), diversity sum score (DIV), and difference in science identity (Diff in SI). From the data, there is evidence that high levels of these student characteristics are magnified by strong, quality mentoring. For the cases with a negative slope, one can reasonably assume that the effect of quality mentoring has a negative effect on undergraduate researchers who have elevated self-reported levels of such measures. These include the students' personal gains (PG), total research experience (TR), and satisfaction with research experience (SATIS). These results indicate that students who begin a SURE with relatively low levels of self-reported satisfaction, total research experience, and personal gains benefit more from quality mentoring than students with higher self-reported measures of these values. We note that this holds true even when the model takes into account the potential ceiling effect of the data reported for these measures. We speculate that the negative slope is probably due to a subset of students with extremely high levels of these variables and should not be regarded as evidence of a true negative effect of quality mentoring.

The negative main effect for the post-pre difference in the intent to pursue a biomedical degree is somewhat puzzling, but it may be interpreted as an indicator of student overconfidence. We define overconfidence as an attribute of novices who operate in a "beginner's bubble," but tend to experience difficulties that do not conform to their prior self-reported judgments when confronted with the "real world" (Pallier et al., 2002). Novice students often experience these difficulties, which may include a temporary decrease in performance, motivation, efficacy, and even identity (Sanchez and Dunning, 2018). Therefore, the negative main effect should not be interpreted as if intention to pursue a biomedical degree is associated with poorer research quality. Instead, the data are revealing a psychological artifact common among novices, which is apparently ameliorated by high-quality mentoring, as evidenced by the positive interaction terms for MCA and difference score for intent to pursue a biomedical career.

The most salient positive main effects on research quality are MCA, thinking and working like a scientist (TWS), and satisfaction with research (SATIS). This indicates that focusing on these attributes could have large payoffs and positive effects for

programs wishing to improve SURE. Other researchers have found that SURE programs help improve students' gains about thinking and working like a scientist (e.g., Mabrouk, 2009; Daniels *et al.*, 2016). Our results also echoed other studies about the importance of quality mentorship (Adedokun *et al.*, 2012; Linn *et al.*, 2015). Moreover, URM students, who historically have less access to SURE programs, should benefit even more from program planners paying attention to these factors. Regarding mentorship quality, Haeger and Fresquez (2016) found that quality mentoring practices increase levels of inclusion in undergraduate research programs. While this does not tie directly into the quality of the research presentations, the research results indicate that increasing mentoring time increases research skills and independence in research. To better understand the relationship between mentors and mentees, a future study will focus on collecting qualitative data of students talking about their experiences with their mentors and how those experiences might have impacted their poster presentations.

Recommendations

This research study provides empirical evidence that exposing students to high-quality mentoring has a positive moderating effect on research products (poster presentations) for students involved in SUREs. Though the effect of high levels of mentor competency is not always positively interacting with student characteristics, the overall model suggests that all three dimensions of research quality are overwhelmingly positively affected by high levels of mentor competency. Additionally, there are student-level characteristics, such as a positive change in the intent to pursue a biomedical research career and thinking and working like a scientist, that set the stage for students having major gains in research quality when provided access to a quality mentor. In contrast, students with already high levels of research experience or particularly high levels of satisfaction or personal gains associated with the research experience, are not as assisted by having access to a high-quality mentor, though they still benefit overall. Practically speaking, these results provide guidance to SURE programs to allocate top mentors to students who lack prior research experience and, at the same time, demonstrate a strong desire to pursue biomedical research careers. Similarly, for students who lack practical experience, but also strongly identify as a biomedical researcher or show high self-efficacy as a researcher, the effect of quality mentoring is magnified. Conversely, it appears that more-experienced students who already are satisfied with their current research experiences will fare relatively well even with a less competent mentor. This guidance can help programs improve the matching of mentors and students based on shared research interests as well as student attributes (e.g., science identity, self-efficacy) and prior experience.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences of the NIH under linked award numbers RL5GM118969, TL4GM118971, and UL1GM118970. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Dr. Guadalupe Corral and Joseph Ramos for assistance with data collection in this project.

REFERENCES

- Adedokun, O. A., Zhang, D., Parker, L. C., Bessenbacher, A., Childress, A., & Burgess, W. D. (2012). Understanding how undergraduate research experiences influence student aspirations for research careers and graduate education. *Journal of College Science Teaching*, 42(1), 82–90.
- Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., & ... Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, 13, 29–40.
- Bieschke, K. J., Bishop, R. M., & Garcia, V. L. (1996). The utility of the Research Self-Efficacy Scale. *Journal of Career Assessment*, 4(1), 59–75. <https://doi.org/10.1177/106907279600400104>
- Brill, J. L., Balcanoff, K. K., Land, D., Gogarty, M., & Turner, F. (2014). Best practices in doctoral retention: Mentoring. *Higher Learning Research Communications*, 4(2), 26. <https://doi.org/10.18870/hlrc.v4i2.186>
- Burge, S. K., & Hill, J. H. (2014). The medical student summer research program in family medicine. *Family Medicine*, 46(1), 45–48.
- Carpi, A., Ronan, D. M., Falconer, H. M., & Lents, N. H. (2017). Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. *Journal of Research in Science Teaching*, 54(2), 169–194.
- Chesler, N. C., Barabino, G., Bhatia, S. N., & Richards-Kortum, R. (2010). The pipeline still leaks and more than you think: A status report on gender diversity in biomedical engineering. *Annals of Biomedical Engineering*, 38(5), 1928–1935.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, 14(1), 1–13. <https://doi.org/10.1187/cbe.14-10-0167>
- Cox, M. F., & Andriot, A. (2009). Mentor and undergraduate student comparisons of students' research skills. *Journal of STEM Education*, 10(1–2), 31–39.
- Daniels, H., Grineski, S. E., Collins, T. W., Morales, D. X., Morera, O., & Echegoyen, L. (2016). Factors influencing student gains from undergraduate research experiences at a Hispanic-serving institution. *CBE—Life Sciences Education*, 15(3), 1–12.
- Daniels, H. A., Grineski, S. E., Collins, T. W., & Frederick, A. H. (2019). Navigating social relationships with mentors and peers: Comfort and belonging among men and women in STEM summer research programs. *CBE—Life Sciences Education*, 18(2), ar17.
- Eagan, M. K., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal*, 50(4), 683–713.
- Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, 17(1), ar9.
- Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, 103(1), 206–222. <https://doi.org/10.1037/a0020743>
- Fakayode, S. O., Yakubu, M., Adeyeye, O. M., Pollard, D. A., & Mohammed, A. K. (2014). Promoting undergraduate STEM education at a historically Black college and university through research experience. *Journal of Chemical Education*, 91(5), 662–665.
- Falconer, J., & Holcomb, D. (2008). Understanding undergraduate research experiences from the student perspective: A phenomenological study of a summer student research program. *College Student Journal*, 42(3)
- Fleming, M., House, S., Shewakramani, V., Yu, L., Garbutt, J., McGee, R., ... Rubio, D. M. (2013). The mentoring competency assessment: Validation of a new instrument to evaluate skills of research mentors. *Academic Medicine*, 88(7), 1002–1008.
- Ghee, M., Keels, M., Collins, D., Neal-Spence, C., & Baker, E. (2016). Fine-tuning summer research programs to promote underrepresented students' persistence in the STEM pathway. *CBE—Life Sciences Education*, 15(3), 1–11.

- Gonzalez-Espada, W. J., & LaDue, D. S. (2006). Evaluation of the impact of the NWC REU program compared with other undergraduate research experiences. *Journal of Geoscience Education*, 54(5), 541–549.
- Haeger, H., & Fresquez, C. (2016). Mentoring for inclusion: The impact of mentoring on undergraduate researchers in the sciences. *CBE—Life Sciences Education*, 15(3), 1–9.
- Hayes-Harb, R. (2018). *Assessment of undergraduate research learning outcomes: Poster presentations as artifacts*. Council on Undergraduate Research Conference 2018. Oral presentation. Washington DC, July. Conference Paper, Refereed, Presented, 07/2018.
- Henningsen, A. (2017). *Package "censReg."* Retrieved September 2019, from <https://cran.rproject.org/web/packages/censReg/censReg.pdf>
- Junge, B., Quiñones, C., Kakietek, J., Teodorescu, D., & Marsteller, P. (2010). Promoting undergraduate interest, preparedness, and professional pursuit in the sciences: An outcomes evaluation of the SURE program at Emory University. *CBE—Life Sciences Education*, 9(2), 119–132.
- Kneale, P., Edwards-Jones, A., Walkington, H., & Hill, J. (2016). Evaluating undergraduate research conferences as vehicles for novice researcher development. *International Journal for Researcher Development*, 7(2), 159–177. <https://doi.org/10.1108/IJRD-10-2015-0026>
- Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science*, 347(6222), 627–633.
- Lopatto, D. (2003). The essential features of undergraduate research. *Council on Undergraduate Research Quarterly*, 24, 139–142.
- Lopatto, D. (2010). Undergraduate research as a high-impact student experience. *Peer Review*, 12(2), 27–30.
- Mabrouk, P. A. (2009). Survey study investigating the significance of conference participation to undergraduate research students. *Journal of Chemical Education*, 86(11), 1335–1340.
- Mansson, D. H., & Myers, S. A. (2012). Using mentoring enactment theory to explore the doctoral student–advisor mentoring relationship. *Communication Education*, 61(4), 309–334. <https://doi.org/10.1080/03634523.2012.708424>
- McGee, R., Jr., Saran, S., & Krulwich, T. A. (2012). Diversity in the biomedical research workforce: Developing talent. *Mount Sinai Journal of Medicine*, 79(3), 397–411.
- Morales, D. X., Grineski, S. E., & Collins, T. W. (2019). Effects of mentoring relationship heterogeneity on student outcomes in summer undergraduate research *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2019.1639041>
- Morales, D. X., Wagler, A. E., & Monarrez, A. (2020). BUILD peer mentor training model: developing a structured peer-to-peer mentoring training for biomedical undergraduate researchers. *Understanding Interventions*, 11(1), 1–18.
- Morley, R. L., Havick, J. J., & May, G. S. (1998). An evaluation of the Georgia Tech summer undergraduate program of research in electrical engineering for minorities. *Journal of Engineering Education*, 87(3), 321–325.
- Oh, S. S., Galanter, J., Thakur, N., Pino-Yanes, M., Barcelo, N. E., White, M. J., ... & Burchard, E. G. (2015). Diversity in clinical and biomedical research: A promise yet to be fulfilled. *PLoS Medicine*, 12(12), e1001918. <https://doi.org/10.1371/journal.pmed.1001918>
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, 129(3), 257–299.
- Pender, M., Marcotte, D. E., Sto. Domingo, M. R., & Maton, K. I. (2010). The STEM pipeline: The role of summer research experience in minority students' Ph.D. aspirations. *Education Policy Analysis Archives*, 18(30), 1–36.
- Revelle, W. (2018). *Package "psych."* Retrieved September 2019, from <https://cran.r-project.org/web/packages/psych/psych.pdf>
- Robnett, R. D., Chemers, M. M., & Zurbruggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, 52(6), 847–867.
- Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science*, 316(5824), 548–549.
- Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, 114(1), 10–28.
- Therneau, T., & Atkinson, B. (2018). *Package "rpart."* Retrieved September 2019, from <https://cran.r-project.org/web/packages/rpart/rpart.pdf>
- Thiry, H., & Laursen, S. L. (2011). The role of student–advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. *Journal of Science Education and Technology*, 20, 771–784.
- Valantine, H. A., & Collins, F. S. (2015). National Institutes of Health addresses the science of diversity. *Proceedings of the National Academy of Sciences USA*, 112(40), 12240–12242.
- Valantine, H. A., Lund, P. K., & Gammie, A. E. (2016). From the NIH: A systems approach to increasing the diversity of the biomedical research workforce. *CBE—Life Sciences Education*, 15(3), fe4.
- Vitae. (2014). About the Vitae Researcher Development Framework. Retrieved August 2020, from www.vitae.ac.uk/researchers-professional-development/about-the-vitae-researcher-development-framework
- Weston, T. J., & Laursen, S. L. (2015). The Undergraduate Research Student Self-Assessment (URSSA): Validation for use in program evaluation. *CBE—Life Sciences Education*, 14(3), ar33. <https://doi.org/10.1187/cbe.14-11-0206>