Original Manuscript

Assessing the Efficacy of a Training Intervention to Reduce Acceptance of Questionable Research Practices in Psychology Graduate Students

Journal of Empirical Research on Human Research Ethics
1–10

DOI: 10.1177/1556264619840525
journals.sagepub.com/home/jre
SAGE

Donald F. Sacco[1] and Mitch Brown[1]

Abstract

We designed and tested the efficacy of a 1-hr training session to mitigate endorsement of questionable research practices (QRPs), research practices that raise ethical concerns and are detrimental to reproducible science, in psychology graduate students. We assessed attitudes toward QRPs 1 week prior to the training, 1 week following the training, and at 2-month follow-up. Participants reported QRPs as less ethically defensible 1 week following the intervention compared with 1 week prior, with attitudes at 2-month follow-up falling in between these time points. Results were maintained even when controlling for socially desirable responding. Participants who rated the training more favorably demonstrated greater attitude change toward detrimental research practices. These results provide evidence that an intervention to educate graduate students about QRPs and their negative impact on science can mitigate consideration of such practices as ethically defensible, although such benefits may not hold over time without additional training sessions.

Keywords

questionable research practices, ethics training, responsible conduct of research, intervention, detrimental research practices

Given the immense societal value of scientific research, scientists and policy makers have increasingly expressed interest in strategies to improve the ethical conduct of science. Attempts are mirrored by expanding awareness of the frequency of ethically questionable behavior that permeates research and the related concerns with reproducibility and nonvalidity of scientific research, particularly in psychology, with scientists becoming especially cognizant of these behaviors' detrimental effects (Bouter, Tijdink, Axelsen, Martinson, & ter Riet, 2016; Godecharle, Fieuws, Nemery, & Dierickx, 2018; Open Science Collaboration, 2015; Simmons, Nelson, & Simonsohn, 2011). Traditional attempts to facilitate ethical research conduct have demonstrated efficacy in fostering an understanding of ethical research, such as mandatory Responsible Conduct of Research (RCR) training for researchers at universities who receive federal funding (e.g., Schmaling & Blume, 2009; for a meta-analysis, see Watts et al., 2017). However, such efforts pose various limitations (Marusic, Wager, Utrobic, Rothstein, & Sambunjak, 2016), with an increasing quantity of scientific retractions due to misconduct and high rates of self-reported ethically questionable research practices (QRPs), indicating existing attempts to facilitate ethical research conduct are insufficient (Fang, Steen, & Casadevall, 2012; John, Loewenstein, & Prelec, 2012).

One potential shortcoming of these efforts may be related to limited education on what constitutes QRPs, specifically research practices with the potential to prove detrimental, during scientists' graduate training.[1] This limitation would necessitate the implementation of training modules that could facilitate a more comprehensive understanding of research ethics and the consequences of QRPs for researchers-in-training. Given the efficacy of such training modules in fostering awareness of the reproducibility crisis in undergraduates (Chopik, Bremner, Defever, & Keller, 2018), it would seem sensible to predict that similar training modules would be similarly effective for graduate students whose work is more directly affected by this knowledge. The current research tested the efficacy of one such training module that focused on an understanding of what constitutes QRPs for psychology graduate students. Specifically, we sought to determine whether this training module reduced students' QRP endorsement as ethically defensible and the extent to

[1]The University of Southern Mississippi, Hattiesburg, USA

Corresponding Author:
Donald F. Sacco, School of Psychology, The University of Southern Mississippi, 118 College Drive #5025, Hattiesburg, MS 39406, USA.
Email: Donald.Sacco@usm.edu

which any positive gains from the training were maintained 2 months following the intervention.

Pervasiveness of QRPs

Concern has grown over the past decade regarding a trend in human subjects research, whereby research findings appear to be less reproducible than previously thought, suggesting a reproducibility "crisis" in science. Reproducibility projects in psychology suggest that in only approximately 60% of research replicates and of studies that do, effect sizes are about half the magnitude of the original demonstration (Open Science Collaboration, 2015). Such issues become further problematic by independent researchers failing to reproduce several high-profile findings (e.g., Lynott et al., 2014; Ritchie, Wiseman, & French, 2012). Corroborating these findings, more than 70% of scientists indicate being unable to replicate or reproduce another's results, whereas 50% have failed to replicate their own experiments (Baker, 2016). Given that statistical models usually afford 5% probability that significant results are not actual effects, the noticeably higher rates of failed replication and reproducibility than would be predicted by chance alone suggest additional factors accounting for such low reproducibility.

Factors Influencing the Reproducibility of Research

A confluence of factors has likely contributed to low reproducibility rates in scientific research, including publication biases, a lack of direct replications, or a focus on statistical significance without considering sampling error. However, some evidence suggests certain factors significantly influence engagement in research behaviors that may undermine ethical scientific conduct and contribute to lower research reproducibility. For example, competitive academic environments are associated with not only increased productivity but also increased publication bias, or the increasing number of publications featuring results that confirm hypotheses (Fanelli, 2010). Furthermore, perceptions of publication pressure, described by respondents as impossibly high, are associated with increased admission of scientific misconduct (fabrication, falsification, and plagiarism). In addition, such behaviors were reported with greater frequency by early-career scientists, suggesting field-specific pressures may disproportionately compromise the research behavior of scientists trying to establish themselves in their field (Tijdink, Verbeke, & Smulders, 2014).

Furthermore, this competitive culture has additional negative downstream consequences, as other research finds that graduate students are more willing to engage in unethical behavior based on its institutional normativity (Langlais & Bent, 2014). Compounding these findings is another work demonstrating that various QRPs, which may be ambiguous in their ethicality (e.g., using covariates without theoretical justification), can increase rates of false-positive results by up to 50% beyond the $p < .05$ cutoff for significance (Simmons et al., 2011). In fact, more than 90% of survey participants reported having personally engaged in at least one of these detrimental research practices (John et al., 2012). Taken together, additional training that focuses on these more nefarious detrimental research practices appears warranted and may be especially critical for graduate students and early-career research scientists.

Impact of Detrimental Research Practices

While it has been traditionally argued that science is a self-correcting system capable of counteracting detrimental research practices through its self-policing and self-correcting nature, such mechanisms may be insufficient to mitigate unethical practices. For example, the majority of retractions in highly cited journals (67.4%) are due to misconduct compared with only 21.3% of retractions being attributed to honest error, suggesting that self-policing and self-correction alone are not an adequate deterrent (Fang et al., 2012). Tangible costs of misconduct and detrimental research practices are notable as well, with direct costs of dealing with misconduct cases often exceeding US$500,000, with an estimated US$110 million in annual expense at institutions to handle misconduct cases (Michalek, Hutson, Wicher, & Trump, 2010). Of even greater concern is the fact that misconduct and detrimental research practices can undermine public trust in the scientific enterprise, which is the very population of which scientists are to be stewards. Indeed, recent polling of the general public has revealed that only 35% of sampled individuals reported having "a lot" of trust in scientists, and the number of people who reported "not at all" trusting scientists increased by more than 50% from a similar 2013 poll (Tsipursky, 2018). This increasing level of mistrust toward scientific enterprises may subsequently foster reluctance toward initiatives to maintain or increase public spending on science. Such information could prove advantageous in encouraging scientists-in-training to consider the consequences of QRPs, as it could remind them of the potential caustic ramification of publishing nonreproducible science (e.g., the dubious link between vaccines and autism; Zeidler, 2016).

The Current Study

Given the high rates of reported detrimental research practices in science, their negative impact on scientific findings, and the greater propensity for their occurrence by early-career scientists, we designed a training module that focused specifically on educating graduate students on

QRPs. Specifically, this training focused on the frequency of these behaviors in science, their negative impact on scientific findings, the impact they may have on the public's view of science, the potential impact they may have on scientists' professional reputation, as well as appeals to strategies to avoid such practices in one's own research. Following presentation of such information, we further fostered a sense of consistency between one's values as a scientist to find the truth and future intentions to act ethically in research, given a notion that research scientists enter their field in pursuit of the truth (e.g., Parzuchowski & Wojciszke, 2014; Zeidler, 2016). We assessed graduate students' perceptions of various QRPs' ethical defensibility before and after the implementation of this module as well as assessed such effects' longevity 2 months following the intervention. Importantly, we further assessed participants' proclivity toward socially desirable responding (Paulhus & Reid, 1991) to covary out the influence of potential responses biases, response style that could undermine the efficacy of the intervention, thus affording the opportunity to determine the fullest extent of our intervention's efficacy.

We predicted a reduction in participants' endorsement of QRPs as ethically defensible following the postintervention (Time 2) compared with preintervention (Time 1), even after statistically controlling for socially desirable responding. Furthermore, we predicted this reduction in endorsement would show some degree of maintenance at a 2-month follow-up assessment (Time 3). More specifically, such reductions in endorsement should be especially apparent for QRPs with ambiguous ethicality (Simmons et al., 2011) relative to QRPs that are considered unambiguously unethical (e.g., refusing to share data and materials with other researchers). Finally, we also predicted that positive gains from the intervention would be associated with favorable attitudes toward it.

Method

Participants

We recruited a total of 49 psychology graduate students, from four different graduate programs (experimental, clinical, counseling, and school), from a midsized public university in Southeastern United States to participate in an ethics training session that spanned over the course of 2 weeks. A total of 41 participants completed the entirety of the study, that is, who provided baseline measures a week prior, participated in the session, and provided postintervention responses a week later (M_{Age} = 24.93, SD = 3.03; 32 women, nine men; 92.7% White). A medium effect sized power analysis (Cohen's f = 0.25, β = 0.80) indicated 34 participants would suffice to detect effects for a pre-/post-design. Participants completed this study in exchange for up to US$30.00 in Amazon gift cards, with their compensation contingent upon the amount of baselines they provided. Procedures for this study were approved by The University of Southern Mississippi Institutional Review Board. Informed consent was obtained from participants before each baseline measure.

Materials and Procedure

Socially desirable responding. Participants initially completed an index for individual differences in a motivation to appear socially desirable at Time 1, the Balanced Inventory for Desirable Responding–16 (BIDR-16; Hart, Ritchie, Hepper, & Gebauer, 2015). This ensured we could consider participants' proclivity toward presenting a positive self-image via self-reports, possibly at the expense of their actual behavior (Leary & Kowalski, 1990). This 16-item measure operated along 7-point scales (1 = *strongly disagree*, 7 = *strongly agree*; eight items reverse-scored), with higher scores indicating more interest in appearing socially desirable. Items exhibited acceptable reliability (α = .84).

QRP endorsement. At Times 1 and 2, participants indicated the extent to which they perceived 31 different QRPs as defensible along 7-point scales (1 = *completely indefensible*, 7 = *completely defensible*). QRPs were further divided into two separate factors based on the level of ambiguity in defensibility using a measure previously validated with a sample of federally funded research scientists: unambiguously unethical (UU, 16 items) and ambiguously unethical (AU, 15 items; Sacco, Bruton, & Brown, 2018). Examples include "Refusing to share data or materials with other researchers to prevent questions about the quality of your work from being raised," for UU, and "Adding additional research participants because the results collected thus far are not yet statistically significant," for AU. Across both times, the measures were reliable ($UU_{Time\ 1}$, α = .95; $AU_{Time\ 1}$, α = .91; $UU_{Time\ 2}$, α = .97; $AU_{Time\ 2}$, α = .92).

For Time 1, a week prior to the session, we emailed participants a link to the study, whereby they could complete procedures confidentially. Consenting participants initially created a unique personal identification code comprising their mother's maiden name, hometown, and birth year; before each measure, participants were required to provide the same identification to link their data for each baseline, thus ensuring we could accurately track individuals' changes. Then, they completed the BIDR-16. Participants were then initially placed into one of two counterbalances to introduce one of two controversial effects in psychology. In random order, participants then indicated their attitudes toward psychological research as a filler task to encourage thinking about research (Chopik et al., 2018) and their endorsement of QRPs (item orders counterbalanced within the randomization). This was followed by demographics.

Finally, participants were directed to a separate survey link to provide their email address to receive compensation that was not connected to the information they provided (Sacco et al., 2018).

One week later, participants completed the 1-hr training session directed by the first author, which addressed various aspects of QRPs (e.g., small sample sizes, replication crisis) and how to mitigate one's endorsement in their own research. Part of this session included the broad social implications of detrimental research practices (e.g., Diederick Stapel's rampant fraud) and the ambiguous nature of some QRPs and their consequences if not theoretically justified (e.g., use of certain covariates). One part of this training involved an exercise in which participants were encouraged to foster consistency between their identities as research scientists and ethical research practices (Sacco, Bruton, & Brown, 2019). Specifically, this exercise emphasized participants to consider their previously espoused ideals for why they entered psychology (e.g., a desire for knowledge or to help people) and how that ideal for the truth should exhibit congruity with one's research practices and their outlook on how to do research ethically. In other words, participants were instructed to consider how their previously espoused ideals should inform their future ethicality (Parzuchowski & Wojciszke, 2014).[2] Following the session, participants provided basic feedback about their experience for six items along 7-point scales assessing positive attitudes (e.g., How much more knowledgeable are you of the information presented in this session compared with prior to this session? 1 = *not at all*, 7 = *very much*; α = .90) and one assessing satisfaction (1 = *not at all satisfied*, 7 = *very satisfied*). Participants received compensation via email following participation.

Another week following the session, participants received the other reading about controversial effects in psychology before responding to the attitude primer and QRP questions. This was followed by demographics, debriefing, and compensation information. Two months following this postintervention measure, we distributed another postintervention assessment that simply included only the QRP measure and the opportunity to provide demographic information before a debriefing.

Results

Initial Analysis

For our initial pre-/post-analysis, we conducted a 2 (intervention: pre vs. post) × 2 (QRP: UU vs. AU) custom repeated ANCOVA with BIDR-16 scores as a covariate to test for interactivity with two categorical predictors to determine how socially desirable responding influences QRP endorsement. Analyses were conducted using IBM SPSS, Version 23. Because our analyses involved nonindependent observations (i.e., within-subjects design), we report partial eta squared as the effect size throughout (for discussion, see Brown, 2008).

A main effect of QRP emerged, such that participants endorsed AU-QRPs (M = 2.42, SD = 1.04) more than UU-QRPs (M = 1.59, SD = 0.95), $F(1, 39) = 21.49$, $p < .001$, $\eta_p^2 = .175$. Another main effect of intervention emerged, such that QRP endorsement significantly decreased from pre- (M = 2.08, SD = 0.98) to postintervention (M = 1.93, SD = 1.01), $F(1, 39) = 9.37$, $p = .004$, $\eta_p^2 = .194$. QRP endorsement still decreased from pre- (estimated marginal means [EMM] = 2.09, SE = 0.14) to postintervention (EMM = 1.93, SE = 1.93) when covarying out BIDR-16 scores, $M_{\text{Difference}} = 0.162$, $p = .018$, $\eta_p^2 = .175$.[3]

We further found it prudent to consider how socially desirable responding influences QRP endorsement directly. We conducted a supplemental analysis considering the extent to which individuals learned about QRPs from Time 1 to Time 2 addressed the differences in the endorsement of both AU- and UU-QRPs by individually correlating difference scores between Time 1 and Time 2 for both QRP types. BIDR-16 scores with UU-QRP difference scores indicated a marginal negative correlation between socially desirable responding and UU-QRP endorsement, $r = -.289$, $p = .067$, and a significant negative correlation between socially desirable responding and AU-QRP endorsement, $r = -.353$, $p = .024$.

We also considered how positive reactions to the intervention were associated with learning from before and after the intervention. Thus, we correlated the extent to which individuals learned in the training (i.e., difference between QRP endorsement from pre- to postintervention) with their positive reactions. Specifically, we found that attitudes toward the training positively correlated with the amount learned, such that more positive attitudes correlated with more learning, $r = .47$, $p < .01$. We also considered participants' satisfaction with the intervention using a single item and correlated it with the amount learned; a similar positive association emerged, such that satisfaction with the training was associated with more learning, $r = .42$, $p < .01$.

Two-Month Follow-Up

Along with our two initial time points, we collected an additional time point 2 months following the intervention to determine the long-lasting efficacy of these effects. Of those 41 participants who completed both original time points, only 33 provided data for the UU- (α = .80) and AU-QRP scales (α = .89) 2 months later. Despite this reduction in statistical power, we found it prudent to conduct an exploratory analysis of these findings as a means of providing tentative evidence for the efficacy of the training

over time. Thus, we submitted our data to an additional 3 (intervention: pre vs. post vs. follow-up) × 2 (QRP: UU vs. AU) custom repeated ANCOVA with BIDR-16 as the covariate.

A main effect of QRP emerged, such that participants endorsed AU-QRPs ($M = 2.37$, $SD = 0.81$) more than UU-QRPs ($M = 1.48$, $SD = 0.42$), $F(1, 62) = 6.57$, $p = .015$, $\eta_p^2 = .175$. Another main effect of intervention emerged, $F(1.36, 42.19) = 5.78$, $p = .013$, $\eta_p^2 = .355$. Post hoc least significant difference (LSD) tests indicated that endorsement of QRPs significantly reduced from pre- ($M = 2.03$, $SD = 0.59$) to postintervention ($M = 1.84$, $SD = 0.61$), $p = .025$, $d = 0.31$. No difference emerged between postintervention and the 2-month follow-up ($M = 1.90$, $SD = 0.65$), and there was no difference between preintervention and the 2-month follow-up, ps $> .20$, ds < 21. These data suggest that participants' endorsement of QRPs did not return to baseline. When covarying out socially desirable responding, a main effect of intervention nonetheless persisted with Time 1 having the largest endorsement ($EMM = 2.03$, $SE = 0.08$), followed by Time 3 ($EMM = 1.91$, $SE = 0.11$) and then Time 2 ($EMM = 1.85$, $SE = 0.10$), $|M_{\text{Differences}}| > .120$, $p = .031$, $\eta_p^2 = .206$. No other effects emerged in this omnibus analysis, Fs < 2.62, ps $> .08$.

Effects were further qualified by an Intervention × BIDR-16 interaction, $F(1.36, 42.19) = 4.39$, $p = .031$, $\eta_p^2 = .124$. We decomposed this interaction by individually correlating the differences in QRP endorsement for Times 1 and 2, 2 and 3, and 1 and 3 for the 33 participants who responded to the 2-month follow-up. Socially desirable responding was negatively associated with QRP endorsement for both of Time 1's difference scores (with Time 2, $r = -.415$, $p = .016$; with Time 3, $r = -.343$, $p = .051$). In other words, participants' motivation to appear socially desirable reduced endorsement of QRPs following introduction of knowledge identifying their problematic nature, thus motivating them to espouse additional socially desirable behavior. No association emerged for BIDR-16 with the difference for Times 2 and 3, $r = .006$, $p = .972$. This suggests that socially desirable responding was associated with the lack of positive change from Time 2 to Time 3 in the intervention.

We conducted subsequent analyses to determine the associations between learning about QRPs with both attitudes toward the intervention and participants' satisfaction with the training by individually correlating learning differences with both attitudinal measures. The difference in QRP endorsement at Times 1 and 3 was positively correlated with both positive attitudes toward and satisfaction with the intervention, rs $> .420$, ps $< .015$. However, the difference between Times 2 and 3 elicited correlations with neither attitudes nor satisfaction, rs $< .045$, ps $> .820$.

Discussion

Our research sought to determine the efficacy of a QRP training session for psychology graduate students. We assessed attitudes toward various research practices that varied in their ethicality from clearly unethical (e.g., withholding relevant methodological details from publication) to those ambiguous in their ethicality (e.g., including covariates in a statistic model). Importantly, we assessed these attitudes 1 week prior to training, 1 week following the training, and 2 months later. We found mixed, but promising, evidence for the efficacy of our training. First, results of our statistical analyses indicated participants reported less favorable attitudes toward the QRPs we assessed at Time 2 compared with Time 1. This suggests broadly that our training was capable of promoting ethical attitudes when assessed in relatively close time proximity to the training. Such findings have consonance with other recent findings demonstrating efficacy in how education about QRPs temporally shifts attitudes toward certain practices (e.g., Chopik et al., 2018). Importantly, whereas previous findings focused primarily on undergraduate populations, the current results directly addressed a population more involved in the research process with graduate students.

Interestingly, the reduction of perceived defensibility in QRPs was general and not specifically for either AU-QRPs or UU-QRPs. That is, QRP category did not moderate training efficacy, suggesting the training demonstrated statistically similar effects on both clearly unethical research practices and ethically ambiguous practices. Several factors could explain this null finding. It is possible that because the training made mention of more unethical practices and communicated the severity of such behaviors in a similar capacity to ambiguous ethical practices, participants may have ultimately equivocated UU- and AU-QRPs. In addition, these results may suggest that even for clearly unethical behaviors for which training already exists (e.g., Langlais & Bent, 2014), supplemental training may be of additional value.

Unsurprisingly, socially desirable responding was associated with espoused aversion to detrimental research practices. Those motivated to respond in socially desirable ways as a means of facilitating others' positive impressions of them are aware of how to respond to facilitate such a goal (Leary & Kowalski, 1990), often by downplaying one's own actual involvement in controversial behavior (e.g., Meston, Heiman, Trapnell, & Paulhus, 1998); given the detrimental nature of QRPs, it would be advantageous for those seeking to appear socially desirable to report an abstention from these behaviors. This necessitated our inclusion of a measure to assess socially desirable responding to covary its influence out of our findings. Nonetheless, even when considering the influence of socially desirable

responding as a covariate, the training reduced endorsement of QRPs as ethically defensible. As such, the intervention's efficacy cannot be explained by participants just becoming more motivated to respond in socially desirable ways as the intervention itself had an independent impact beyond this response bias.

Using aggregate measures of positivity toward the training session, we found preliminary evidence suggesting that favorability toward the intervention correlated with reductions in QRP endorsement in some capacities. Those reporting greater favorability toward the training session immediately following it also demonstrated the largest change in detrimental research attitudes from Time 1 to Time 2. However, favorable attitudes toward the intervention did not correlate with differences in detrimental research practice attitudes from Time 2 to Time 3, suggesting a level of maintenance following the intervention. Nonetheless, the association between learning from Time 1 to Time 3 and favorability toward the intervention was also significant, with positivity toward the intervention serving as a buffer from reconsidering QRPs as defensible, as evidenced by Time 3's endorsement of QRPs not returning to baseline levels. Such findings are sensible, given previous research indicating the benefits of positivity in attentional adhesion to important details (Fredrickson & Branigan, 2005) and better understanding of "big picture" concepts, which could include the impact of QRPs (Gasper & Clore, 2002).

It should also be noted that the positive change was rather small following the intervention after covarying out socially desirable responding ($M_{\text{Difference}} = 0.162$), and how that resulted in a small effect size, along with larger standard deviations, is a point that necessitates additional research to demonstrate how robust these effects are. For example, socially desirable responding toward a major issue in psychological research could pose considerable limitations in gauging an accurate statistical representation of our findings, given the social consequences of endorsing QRPs. Nonetheless, we find such small findings could be considered impressive, given how difficult certain attitudinal changes may be to attain (e.g., Petty & Krosnick, 2014; Prentice & Miller, 1992). Not only was the baseline endorsement of QRPs already low, given the controversial nature of such practices, but also participants may have felt compelled to respond in socially desirable capacities regardless of an intervention. The fact that our intervention was capable of reducing endorsement of behaviors with an inherently low base rate suggests a level of efficacy in the training to warrant further refinement of similar training modules. Furthermore, the mean change from covarying out socially desirable responding ultimately elicited decrease in QRP endorsement from Time 1 to Time 2, a decrease that occurred because of a lack of floor effect at Time 1. Taken together, this small effect could be best described as preliminary in terms of such a manipulation being effective in eliciting positive changes, especially considering the effects at Time. This necessitates further research to ensure its robustness.

Maintenance of Effects

Unexpectedly, participants' reduced endorsement of QRPs was not significantly maintained at Time 3. Although we did not obtain clear empirical support for long-term reduction in QRP endorsement, some evidence emerged to suggest at least partial descriptive support. That is, despite a difference between Times 1 and 2, attitudes toward QRPs neither differed between Time 1 and Time 3, nor did attitudes differ from Time 2 to Time 3. Thus, attitudes toward QRPs fell roughly between those expressed at Time 1 and Time 2 at Time 3. Although participants' belief that QRPs are unethical did not perfectly maintain from 1-week postintervention to 2-month follow-up, they also did not return to the level of Time 1 either. Had the intervention provided no long-term positive effects, Times 2 and 3 would have been significantly different from each other with Time 3 returning the Time 1 baseline. This may provide indirect evidence that the long-term attitudinal gains from the training are present, even if not statistically significant. Nonetheless, future research would benefit from considering how to maintain the positive change more significantly.

Several factors may explain why positive gains were not maintained. It may be the case that without more consistent training and QRP education, the benefits of single training sessions are relatively short-lived. It may be beneficial to consider brief booster trainings to help maintain positive attitude gains from the intervention over time. Alternatively, participants may have also received additional, unassessed training in between the QRP training and the 2-month follow-up. Indeed, 60.6% of our participants in our sample were in their first 2 years of doctoral training and thus actively enrolled in statistics and research methods courses. It is possible that while enrolled in these courses, or through mentor interactions, they learned of the nuances of some of the practices originally assessed pre- and post-training. Ambiguous QRPs' ethical defensibility is context-specific. That is, the motive underlying decisions to engage in a practice determines its ethical defensibility. For example, excluding a participant from a statistical analysis simply because it improves statistical significance would be less ethically defensible than excluding a participant because an outlier analysis suggests their exclusion is warranted. Past research shows that more experienced researchers who are aware of the contextual factors making these ambiguous practices more or less ethically defensible may ultimately impart this knowledge to their mentees (Sacco,

Brown, & Bruton, in press). It is possible that participants' ratings of detrimental research practices, as descriptively more ethically defensible at 2-month follow-up compared with directly following training, are due to increased sensitivity to the nuances and context specificity of some of these behaviors.

Best Practices

These results provide initial evidence on the efficacy of a QRP training session in educating graduate students on the nuances of research ethics that also highlight the potential detriments of such practices. Research ethics training has traditionally focused on practices that meet the federal definition of research misconduct in falsification, fabrication, and plagiarism and the consequences of these behaviors, with less focus on ambiguous practices that could be equally detrimental. Such initiatives themselves may be limited in their scope of fostering research integrity, given the considerably larger number of researchers self-reporting engagement in ambiguous QRPs compared with those admitting to federally defined misconduct (John et al., 2012). Specifically, this training gave more focus on the so-called researcher degrees of freedom, which could inflate Type I error rates and lead to the publication of nonreproducible science (e.g., Open Science Collaboration, 2015; Simmons et al., 2011). In other words, this training may more directly address actual ethical dilemmas faced by research scientists on a daily basis, as practices ambiguous in their ethicality are ultimately perceived as having a degree of ethical defensibility (Sacco et al., in press; Sacco et al., 2018). Such training could itself potentially foster an environment, wherein graduate students can develop an understanding of best research practices and therefore continue these practices as a core part of their research identity.

Limitations

While promising, several limitations are worth mentioning, which themselves could provide direction for future research. One potential limitation includes the lack of an actual control group with positive change being the product of a time effect or the effects occurring through the passage of time irrespective of any treatment condition. Future studies would benefit from conducting a study that utilizes a control condition in which some participants either do not participate in training or engage in traditional ethics training (i.e., simply addressing the federal definition of research misconduct; falsification, fabrication, and plagiarism). Given challenges in fostering ethical decision making (e.g., Schmaling & Blume, 2009), it would seem sensible to predict that the current intervention could be a complementary, yet necessary, module alongside others that demonstrate efficacy in other domains (e.g., Chopik et al., 2018). The current module would benefit from integrating the positive aspects of other training to develop optimal education programs. For example, training modules that implement team-based learning instruction produce positive changes in ethical decision making (McCormack & Garvan, 2014). Future versions of this intervention could potentially utilize sessions in which participants consider how groups consider QRPs as a collective to reduce their pervasiveness. Such an integration of training modules would be advantageous, given the importance of collaboration in ensuring ethical research conduct (Lungeanu, Huang, & Contractor, 2014).

Our sample, although adequately powered, was relatively small and homogeneous as the psychology graduate students comprised largely of women. It would be beneficial to test this intervention, or variation of this intervention, with graduate students from a variety of science, technology, engineering, and mathematics (STEM) disciplines to further elucidate its potential efficacy. Research indicates that men are overrepresented in research misconduct cases (Fang, Bennett, & Casadevall, 2013), despite not being more likely to commit misconduct (Fanelli, Costas, & Larivière, 2015). However, one possible outlet to reduce this overrepresentation could be to develop an intervention with greater consideration of how men respond. Future research would benefit from increasing gender symmetry in sampling. In addition, we only assessed participants' attitudes toward QRPs; however, the current study was unable to determine whether the intervention promotes more actual ethical research behavior. Future research should include assessments that more closely approximate research behavior, such as vignettes in which participants determine whether a colleague should address a potential detrimental research practice prior to submitting a manuscript or not.

Furthermore, the intervention itself was relatively brief and broad in its communication of QRPs and how to avoid them. Future research could create several trainings that focus on each piece contained in the current training to see which one has the most positive impact (e.g., education about what QRPs are, education about their negative impact on science, and strategies to avoid them). This intervention may also be able to address the nuances in QRPs by considering the specific contextual factors that reduce their defensibility and provide the opportunity to understand when graduate students reduce their endorsement of QRPs (Sacco et al., in press). Such an empirical test would also be better able to determine whether multiple training sessions over time allow for the maintenance of training gains over time.

Research Agenda

One concern for future research is identifying the mechanisms through which the training was able to foster a reduction in QRP endorsement. Indeed, education of what constitutes QRPs necessarily heightens awareness of ethical research practices, but previous research indicates such measures may not actually foster ethical research behavior, particularly if ethical research is not normative within a department (Langlais & Bent, 2014; Schmaling & Blume, 2009). Along with explicitly indicating what constitutes QRPs, the training provided an exercise through which individuals fostered a sense of consistency between their researcher identity and ethicality, a similar exercise previously demonstrated as effective in reducing QRP endorsement among early-career researchers (Sacco et al., 2019). Future research would benefit from explicitly measuring the extent to which this consistency between different identities fosters endorsement of ethical research practices. From a self-perception theory framework (Bem, 1972; Chaiken & Baldwin, 1981), previous engagement in prosocial behaviors should elicit an understanding of oneself as chronically prosocial and therefore inform subsequent behaviors that seek to perpetuate that identity. This would ultimately form the basis of an ethical identity in researchers who were reminded of their intentions to seek the truth in science. Such processes could potentially be assessed by considering inclusion of ethicality within one's self-concept, which serves as a proxy for congruity between one's research identity and an ideal (Aron & Aron, 1997; for self-ideal congruity measure, see Derrick, Gabriel, & Tippin, 2008).

Educational Implications

Taken together, the results of the current study suggest that a specific training to educate graduate students with respect to detrimental research practices can be effective in promoting ethical research attitudes similar to previous efforts employed with undergraduates (Chopik et al., 2018). Our training led participants to view detrimental research practices as less ethically defensible at 1 week after the intervention compared with 1 week prior, with attitudes at 2-month follow-up following in between. The results were maintained even when controlling for socially desirable responding and were of equivalent magnitude for both clearly unambiguously unethical practices and more ethically ambiguous practices. This suggests this intervention was effective in reducing endorsement of QRPs and would benefit from further refinement to ensure longer lasting positive gains, thereby providing an environment to ensure more ethical STEM practices, a factor that undermines the efficacy of previously implemented ethics training (Langlais & Bent, 2014; Marusic et al., 2016).

Conclusion

With the increasing importance of research integrity in STEM fields, it becomes necessary to develop pedagogical approaches that foster integrity for scientists-in-training. Such training would serve to facilitate an understanding of ethical science and ultimately reduce the amount of nonreproducible research in the literature by mitigating perceptions of QRPs as ethically defensible. The current study provided preliminary evidence demonstrating the efficacy of such training initiatives in facilitating an initial sense of integrity among early-career scientists.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a 2018 Lucas Award funded by The University of Southern Mississippi.

Notes

1. As has been convention, our training referred to these research behaviors as questionable research practices. However, we use this term and detrimental research practices interchangeably throughout this article, given the recent recommendation to refer to practices using the latter term due to their potential negative impact on science (The National Academy of Sciences, Engineering, and Medicine, 2017).
2. Presentation, materials, and data available: https://osf.io/p4v7n
3. Effects were further qualified by a Time × Balanced Inventory for Desirable Responding–16 (BIDR-16) interaction, $F(1, 39) = 6.66$, $p = .014$, $\eta_p^2 = .146$. Individually correlating the effects of BIDR-16 scores at Times 1 and 2 separately indicated high social desirability scores negatively correlated, albeit nonsignificantly, with questionable research practice (QRP) endorsement at Time 1, $r = -.107$, $p = .506$, whereas such responding positively and nonsignificantly correlated with QRP endorsement, $r = .074$, $p = .648$. This suggests the basis of the interaction was merely a difference in directionality of nonsignificant associations, or that one correlation was positive and the other was negative. Another QRP × BIDR-16 interaction emerged, $F(1, 39) = 6.38$, $p = .016$, $\eta_p^2 = .141$. Individually correlating BIDR-16 scores with endorsement of both types of QRP separately indicated socially desirable responding negatively correlated with ambiguously unethical (AU)-QRP endorsement, $r = -.123$, $p = .445$, and positively with unambiguously unethical (UU)-QRP endorsement, $r = .099$, $p = .538$; similarly, neither of these effects are significant, again suggesting the basis of the interaction being a difference in directionality.

These interactions suggest that socially desirable responding influenced participants' responses, thus necessitating consideration of the effects when covarying out this response bias. No other main effects or interactions emerged in this model, Fs < 2.80, ps $> .100$.

ORCID iD

Donald F. Sacco https://orcid.org/0000-0001-6017-5070

References

Aron, A., & Aron, E. N. (1997). Self-expansion motivation and including other in the self. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research and interventions* (pp. 251-270). Hoboken, NJ: John Wiley.

Baker, M. (2016). Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help. *Nature*, *533*, 452-455.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*, 1-62.

Bouter, L. M., Tijdink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review*, *1*, Article 17.

Brown, J. D. (2008). Effect size and eta squared. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *12*, 38-43.

Chaiken, S., & Baldwin, M. W. (1981). Affective-cognitive consistency and the effect of salient behavioral information on the self-perception of attitudes. *Journal of Personality and Social Psychology*, *41*, 1-12.

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, *45*, 158-163.

Derrick, J. L., Gabriel, S., & Tippin, B. (2008). Parasocial relationships and self-discrepancies: Faux relationships have benefits for low self-esteem individuals. *Personal Relationships*, *15*, 261-280.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from U.S. States Data. *PLoS ONE*, *5*(4), e10271.

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS ONE*, *10*(6), e0127556.

Fang, F. C., Bennett, J. W., & Casadevall, A. (2013). Males are overrepresented among life science researchers committing scientific misconduct. *MBio*, *4*(1), e00640-12.

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 17028-17033.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, *19*, 313-332.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *13*, 34-40.

Godecharle, S., Fieuws, S., Nemery, B., & Dierickx, K. (2018). Scientists still behaving badly? A survey within industry and universities. *Science and Engineering Ethics*, *24*, 1697-1717.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16), *5*, 1-9. doi:10.1177/2158244015621113

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524-532.

Langlais, P. J., & Bent, B. J. (2014). Individual and organizational predictors of the ethicality of graduate students' responses to research integrity issues. *Science and Engineering Ethics*, *20*, 897-921.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*, 34-47.

Lungeanu, A., Huang, Y., & Contractor, N. S. (2014). Understanding the assembly of interdisciplinary teams and its impact on performance. *Journal of Informetrics*, *8*, 59-70.

Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology*, *45*(2), 216.

Marusic, A., Wager, E., Utrobic, A., Rothstein, H. R., & Sambunjak, D. (2016). Interventions to prevent misconduct and promote integrity in research and publication. *Cochrane Database of Systematic Reviews*, *4*, Article MR000038.

McCormack, W. T., & Garvan, C. W. (2014). Team-based learning instruction for responsible conduct of research positively impacts ethical decision-making. *Accountability in Research*, *21*, 34-49.

Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1998). Socially desirable responding and sexuality self-reports. *Journal of Sex Research*, *35*, 148-157.

Michalek, A. M., Hutson, A. D., Wicher, C. P., & Trump, D. L. (2010). The costs and underappreciated consequences of research misconduct: A case study. *PLoS Medicine*, *7*(8), e1000318.

The National Academy of Sciences, Engineering, and Medicine. (2017). *Fostering integrity in research*. Washington, DC: The National Academies Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 1-8.

Parzuchowski, M., & Wojciszke, B. (2014). Hand over heart primes moral judgments and behavior. *Journal of Nonverbal Behavior*, *38*, 145-165.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*, 307-317.

Petty, R. E., & Krosnick, J. A. (2014). *Attitude strength: Antecedents and consequences*. New York, NY: Psychology Press.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160-164.

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's "retroactive facilitation of recall" effect. *PLoS ONE*, *7*(3), e33423.

Sacco, D. F., Brown, M., & Bruton, S. V. (in press). Grounds for ambiguity: Justifiable bases for engaging in questionable research practices. *Science and Engineering Ethics*.

Sacco, D. F., Bruton, S. V., & Brown, M. (2018). In defense of the questionable: Defining the basis of research scientists' engagement in questionable research practices. *Journal of Empirical Research on Human Research Ethics*, *13*, 101-110.

Sacco, D. F., Bruton, S. V., & Brown, M. (2019, February). *Contextual and psychological factors influencing perceptions of questionable research practice acceptability*. Poster presented at 2019 Society for Personality and Social Psychology Conference, Portland, OR.

Schmaling, K. B., & Blume, A. W. (2009). Ethics instruction increases graduate students' responsible conduct of research knowledge but not moral reasoning. *Accountability in Research*, *16*, 268-283.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366.

Tijdink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication pressure and scientific misconduct in medical scientists. *Journal of Empirical Research on Human Research Ethics*, *9*(5), 64-71.

Tsipursky, G. (2018, July 5). (Dis)trust in science. *Psychology Today*. Retrieved from https://www.psychologytoday.com/us/blog/intentional-insights/201807/distrust-in-science

Watts, L. L., Medeiros, K. E., Mulhearn, T. J., Steele, L. M., Connelly, S., & Mumford, M. D. (2017). Are ethics training programs improving? A meta-analytic review of past and present ethics instruction in the sciences. *Ethics & Behavior*, *27*, 351-384.

Zeidler, D. L. (2016). STEM education: A deficit framework for the twenty first century? A sociocultural socioscientific response. *Cultural Studies of Science Education*, *11*, 11-26.

Author Biographies

Donald F. Sacco is an associate professor of social psychology at the University of Southern Mississippi, USA. His research interests include the impact of financial conflict of interest disclosure on perceptions of researcher ethicality, as well as how educational experience and academic background influence these perceptions. He was involved in the conceptualization of the research project, contributed to the creation of the study materials, served a primary role in data analysis, and assisted in the writing of the article.

Mitch Brown is a graduate student in social psychology at The University of Southern Mississippi, USA. His research primarily concerns the various bases of social perception, including those pertaining to questionable research practices. For the current article, he recruited participants via email, contributed to data analysis, and assisted in writing the article.