



# A randomized trial of a lab-embedded discourse intervention to improve research ethics

Dena K. Plemmons<sup>a,1</sup>, Erica N. Baranski<sup>b</sup>, Kyle Harp<sup>c</sup>, David D. Lo<sup>d</sup>, Courtney K. Soderberg<sup>e</sup>, Timothy M. Errington<sup>e</sup>, Brian A. Nosek<sup>e,f</sup>, and Kevin M. Esterling<sup>g,1</sup>

<sup>a</sup>Graduate Division, University of California, Riverside, Riverside, CA 92521; <sup>b</sup>Institute on Place, Wellbeing and Performance, University of Arizona, Tuscon, AZ 85711; <sup>c</sup>Department of Anthropology, University of California, Riverside, CA 92521; <sup>d</sup>School of Medicine Division of Biomedical Sciences, University of California, Riverside, CA 92521; <sup>e</sup>Center for Open Science, Charlottesville, VA 22903 U.S.A.; <sup>f</sup>Department of Psychology, University of Virginia, Charlottesville, VA 22903; and <sup>g</sup>School of Public Policy, University of California, Riverside, CA 92521

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved December 12, 2019 (received for review October 11, 2019)

**We report a randomized trial of a research ethics training intervention designed to enhance ethics communication in university science and engineering laboratories, focusing specifically on authorship and data management. The intervention is a project-based research ethics curriculum that was designed to enhance the ability of science and engineering research laboratory members to engage in reason giving and interpersonal communication necessary for ethical practice. The randomized trial was fielded in active faculty-led laboratories at two US research-intensive institutions. Here, we show that laboratory members perceived improvements in the quality of discourse on research ethics within their laboratories and enhanced awareness of the relevance and reasons for that discourse for their work as measured by a survey administered over 4 mo after the intervention. This training represents a paradigm shift compared with more typical module-based or classroom ethics instruction that is divorced from the everyday workflow and practices within laboratories and is designed to cultivate a campus culture of ethical science and engineering research in the very work settings where laboratory members interact.**

the practice of ethical research (12, 13), a necessity that Lave (14) refers to as “understanding in practice.” Most existing approaches to ethics training are not tailored to enhance the interpersonal skills and capacity of laboratory members to engage in the kind of reasoned communication that is necessary for putting ethical knowledge into practice. As such, we believe that ethical practice is enhanced when training is connected to conversations within the laboratory setting on specific research projects and when ethical discourse governing scientific acts is routinized into the workflow of the laboratory.

The Institutional Re-Engineering of Ethical Discourse in STEM (iREDS) training was designed to be integrated into ongoing projects and the everyday circumstances of the laboratory. The training combines online collaboration technology with project-based and peer-involved conversations that demonstrate the value of effective communication in practice. The training encourages laboratory members to become comfortable with the norms and skills of how best to talk to each other about topics in ethics and to reflect on norms of practice that operate within the

research ethics | randomized trial | authorship | data management

The everyday practice of ethical and responsible research requires members of a laboratory research team to communicate and justify reasons for their scientific decisions and actions, both within the research team and toward external audiences. The general expectations for ethical research in science and engineering are often presented as a set of practices collectively known as the responsible conduct of research (RCR) (1). Some RCR practices necessitate effective within-laboratory communication, such as collaborative research design, authorship attribution and order, and training and mentoring students. Other RCR practices require transparency and effective external communication, such as appropriate conflict of interest reporting, human and animal subjects protections, data integrity, and disclosure (2).

These RCR elements are intended to prevent research misconduct and to promote normative aspirations for good, credible, and responsible science (3, 4). When these aspirations are not met, the consequences to the integrity of science are significant (5–7). However, communication regarding the RCR elements on many university campuses remains anemic at best (8, 9). This disengagement may be because researchers often equate ethics with regulatory compliance and because the delivery of RCR training is often divorced from the everyday reality and social dynamics of the laboratory and collaborations.

Most RCR instruction is delivered to individuals in isolation from the laboratory, usually online, in classrooms, or some combination, and it typically is intended to transfer knowledge of the RCR topics to students (10). The practice of ethics, however, requires reason giving and judgment among all laboratory members to be connected to the actual circumstances in which an action is performed (11). That is, “knowing” and “doing” are inextricable for

## Significance

**The ethical practice of research requires researchers to give reasons and justifications for their actions, both to the other members of their research team as well as to external audiences. We developed a project-based training curriculum intended to make ethics discourse a routine practice in university science and engineering laboratories. Here, we report the results of a randomized control trial implemented among science and engineering laboratories in two research-intensive institutions. We demonstrate that, compared with the control laboratories, treatment laboratory members perceived improvements in the quality of discourse on research ethics within their laboratories as well as enhanced awareness of the relevance and reasons for that discourse for their work as measured in surveys administered 4 mo after the intervention.**

Author contributions: D.K.P., E.N.B., K.H., D.D.L., C.K.S., T.M.E., B.A.N., and K.M.E. designed research; D.K.P., E.N.B., and K.H. performed research; K.M.E. contributed new reagents/analytic tools; E.N.B. and K.M.E. analyzed data; and D.K.P., E.N.B., K.H., D.D.L., C.K.S., T.M.E., B.A.N., and K.M.E. wrote the paper.

Competing interest statement: C.K.S., T.M.E., and B.A.N. are employed by the nonprofit Center for Open Science that has a mission to increase openness, integrity, and reproducibility of research and which maintains the open-source Open Science Framework (OSF) used in the intervention.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The data and code to reproduce all analyses are available at Harvard Dataverse, <https://doi.org/10.7910/DVN/AIPWNU>.

<sup>1</sup>To whom correspondence may be addressed. Email: dena.plemmons@ucr.edu or kevin.esterling@ucr.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1917848117/-DCSupplemental>.

First published January 9, 2020.

laboratory and within the larger discipline. The iREDS training is integrated with a free, open-source, web-based collaboration tool, the Open Science Framework (OSF; <https://osf.io/>), that is maintained by the Center for Open Science. In this project, we focused the training on two topics that align with features of the OSF: authorship attribution and data management (15–17).

We fielded a randomized control trial (RCT) among faculty-led science and engineering research laboratories at two US research-intensive institutions to evaluate the efficacy of the iREDS curriculum. Specifically, we expected that our training will help scientists recognize the relevance of ethics as a part of their work routines and enhance ethical discourse both within teams and toward external audiences, above and beyond the standard training already available. This paper reports the results of our evaluation of this training intervention.

We give a detailed description of the training intervention in *Materials and Methods*, but briefly, two iREDS team members—one expert in RCR and one expert on the OSF—met with each research group in their laboratory and gave two trainings, one for 90 min and one for 1 h, that engaged laboratory members in conversations centered on authorship and data management policies. Each training included a demonstration of the OSF.

The OSF facilitates discussion on the two topics that are the focus of the training: authorship and data management (15–17). A screen capture of an OSF project page is shown in Fig. 1. For authorship, the OSF enables teams to identify laboratory members who contribute to each of a project's components, whether that be the design of the research project, data collection, record keeping, or manuscript preparation. This listing enables teams to be self-conscious of each member's contribution and facilitates laboratory-wide discussions about whether those contributions merit authorship and in what order. For data management, the OSF fosters reproducibility and open science. Researchers use OSF to manage research data, materials, and code for internal use and for documenting the research lifecycle and can make any components of the research available to others either via controlled or open access.

Our intervention is motivated by the core normative principles of deliberation within social communication, which envisions an ideal of reasoned, uncoerced, and transparent argumentation for

and against decisions that are ideally arrived at collectively (18–20). In the laboratory, deliberation prioritizes justifiable and reason-based collaboration for important scientific actions and decisions (21–24). To advance the deliberative ideal, our training did not didactically instruct individuals on what is or is not ethical. Instead, the training proceeded as a decentralized discussion among laboratory participants on how each might apply ethical considerations to the practices that occur within the laboratory.

Research laboratories tend to be highly specialized and hierarchical. Enhancing uncoerced communication and transparency among personnel given this fundamental disparity in authority and expertise is critical to our deliberative approach (25). To address these circumstances, we rely on a peer-involved method of project-based ethics training in which a graduate student volunteer from each laboratory served as a laboratory “ethics peer mentor” who worked with the iREDS trainers to select a project for discussion, to customize the OSF page and script to incorporate that project, and then, to assist in facilitating the discussion during the in-person training. Having the peer mentors assist in the discussion helps to decentralize authority compared with having the principal investigator (PI) leading discussion.

The deliberative approach changes the focus of ethics from educating individuals to changing the culture within the laboratory (12, 26–31). Training methods that deliver skills to enhance communication directly within the context of the research team may help make the ethics training intervention concrete, personal, memorable, and actionable. If the interventions are more actionable, then the implemented behaviors may be more sustainable as part of the laboratory's regular workflow. Moreover, embedding ethical training in teams may foster institutional norms and advance other areas of ethical practice in science (32–34).

## Results

The randomized trial was fielded in active faculty-led laboratories at the University of California, Riverside (UCR) and at Scripps Research (formerly The Scripps Research Institute [TSRI]). Our sample consisted of 184 members of 34 faculty-led research laboratories in science and engineering fields. We provided incentives of a \$1,000 graduate student travel grant per laboratory for completing the study. About 16% of PIs initially indicated an interest in participating in the research, and about 7% of laboratories enrolled and completed the study. *Materials and Methods* discusses recruitment, eligibility, enrollment, incentives, consent procedures, and research design for the RCT. Our sample consisted of faculty PIs, graduate students, postdocs, undergraduate research assistants, research scientists, and support staff from 12 departments in natural, physical, biomedical science, and engineering colleges.

We randomized at the laboratory level, blocking by department. Laboratories randomized to the intervention condition participated in a 1.5-h training followed by a 1-h training scheduled 2 wk apart. Participating laboratories also completed a presurvey, two midpoint surveys, and a postsurvey that was administered 4.5 mo following the training (35). Laboratories randomized to the control condition received the same four surveys over a 6-mo period but did not participate in any intervention. This RCT design identifies the degree to which our training improves on standard practice of ethics training at these two institutions (36). We document the standard practice for our sample using presurvey responses in *Materials and Methods*, noting that half of our sample indicate no prior training.

The pre- and postsurveys consisted of Likert-type questions assessing attitudes and behaviors relating to laboratory climate, communication, data management, and authorship. We reproduce the survey items and response categories in *SI Appendix*. We conducted factor analysis to create meaningful outcome scales that we describe below. Three individual items, however, are of direct interest, and therefore, we analyze them as

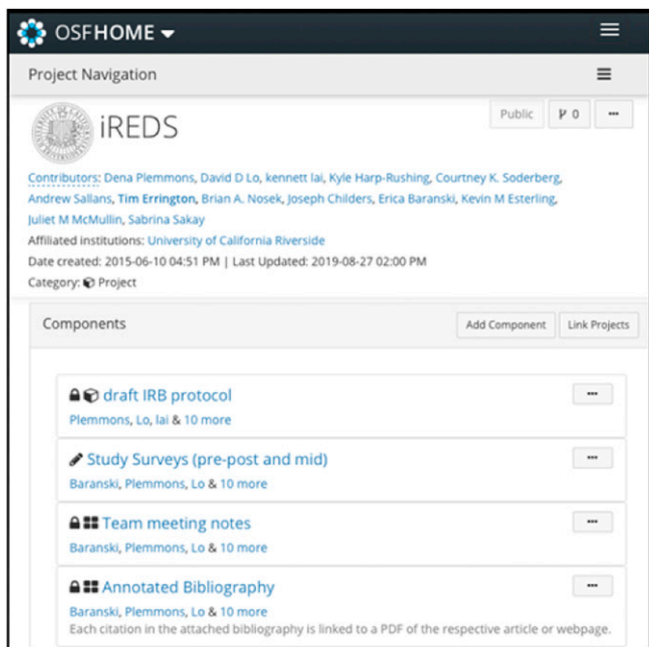
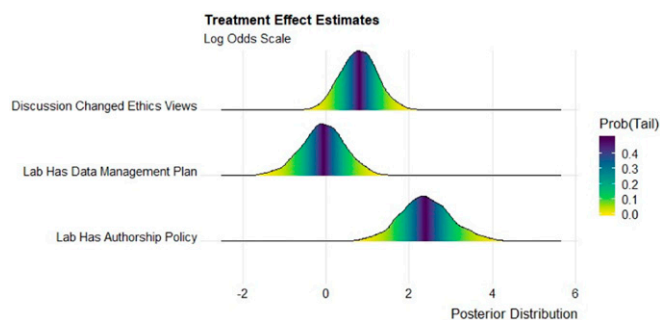


Fig. 1. A screen capture of an OSF project page showing components.



**Fig. 2.** Marginal posterior distributions of causal effect estimates in log odds. Univariate Bayesian conditional autoregressive ordered logit estimates with respondents nested in laboratories.  $n = 184$ . Figure created with ref. 38.

outcomes separately from the scales. Included in both pre- and postsurveys is the question “Have you changed your views about ethical research practices based on discussion within your lab?” This is a summary measure of exposure to the training. Because the trainings centered on awareness of authorship and data management, we included items assessing whether participants understood that their laboratory had an authorship policy and a data management plan.

We used Bayesian nonlinear regression models to estimate the treatment effect for these three individual survey items, nesting respondents within laboratories. Our causal effect identification relies on random assignment nested in a difference-in-differences design. Fig. 2 gives the posterior distributions for the causal effect estimates in log odds for the individual survey items. The figure is a series of marginal density plots of the simulated posterior distribution, and the tail probability is indicated by the color that intersects the vertical line at zero.

We find a significant effect for the item asking if the respondent’s views on ethics changed as a result of discussion in the laboratory, with 97% of the posterior distribution to the right of 0 (by comparison, the maximum likelihood estimate of the log odds = 0.76,  $P = 0.035$ ). We also find significant treatment effect estimates for the item asking the respondent to report the presence of a laboratory authorship policy, with the posterior distribution fully to the right of 0 (by comparison, the maximum likelihood estimate of the log odds = 2.13,  $P < 0.001$ ). This result does not necessarily mean that laboratories created new authorship policies as a result of the training; instead, the training may have raised awareness of what constitutes an authorship policy.

We cannot reject the null hypothesis for the item asking the respondent to report the presence of a laboratory data management policy (maximum likelihood estimate of the log odds = 0.03,  $P = 0.93$ ). This finding suggests that it may be harder to change how individuals and laboratories manage data. This pattern of results could indicate that many scientists already believe that they know best practices for handling data, which is central to their training as scientists, while authorship often is not. To gain inductive insight, we use in *SI Appendix* text analysis of open-ended responses to a follow-up survey question that asked respondents to report their personal data management practices. The text analysis is exploratory and compares the open-ended responses of those who report the presence of a laboratory data management policy with those who report the absence of a policy.

The survey also had batteries of questions that measure six latent dimensions using scaling techniques (*SI Appendix* discusses how we created these scales and convergent validity statistics). Three scales measured the quality of communication within the laboratory. The first scale measures respondents’ beliefs about the relevance of ethics discourse for their research and work, with four

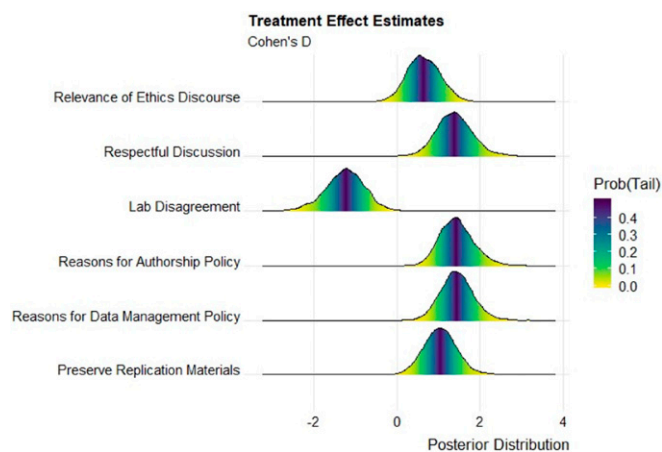
items indicating the importance of ethics to aspects of their work and the frequency of seeking out others in the department or the laboratory to discuss ethics. The second scale measures the extent to which the participants perceived respectful discussion in their laboratory, a critical outcome given the centrality of reason giving, respectful and equal discussion in the concept of deliberation. The five items on this scale were developed in ref. 37 and indicate perceptions that laboratory members make valid arguments, whether everyone has an opportunity to speak, whether laboratory members listen to one another, whether laboratory members understand the respondent’s own views, and confidence in the ethical practice of fellow laboratory members. The third scale measured climate and the amount of laboratory disagreement, which is a measure of the constructiveness of within-laboratory communication, with three items indicating the frequency of disagreements in the laboratory in general and specifically, about authorship and about data management.

The next three scales measure impact of the training content on authorship and data management. The fourth scale measures the respondent’s self-reported understanding of the reasons for an authorship policy using three items asking the respondent to report their understanding of the rationale, importance, and implications for having an authorship plan in the laboratory. The fifth scale had a similar set of questions to measure the respondent’s understanding of the reasons for a data management plan in the laboratory.

Finally, the sixth scale measures the respondent’s perception of the importance of open science practices with respect to the preservation of replication materials, which is a key feature of the OSF and arguably, of good data management practice. The preserve replication materials scale has four items that measure the respondent’s beliefs about the importance of archiving versions of datasets, manuscripts, and laboratory materials and maintaining electronic copies of materials.

We report the causal effect estimates for the scale outcomes in Fig. 3, again with participants nested within laboratories. The figure shows the posterior distributions of the Cohen’s  $D$  marginal effect of the treatment for each of our six scales in a density plot. For each posterior distribution, more than 95% of the mass is different from zero.

The top three lines of Fig. 3 report the treatment effect estimates for the three scales that measure communication outcomes and climate within laboratories. Those individuals who participated in the training were statistically more likely to believe that ethics communication is relevant to their scientific



**Fig. 3.** Marginal posterior distributions of causal effect estimates in Cohen’s  $D$  units. Bayesian conditional autoregressive multilevel regression estimates with respondents nested in laboratories.  $n = 184$ . Figure created with ref. 38.



work, that discourse within their laboratories was respectful and open to diverse views, and that the climate for discussion in the laboratories was more constructive in the sense that disagreements over scientific practices were reduced. These three scales indicate that laboratories that went through the training better approximated deliberative communication ideals in that they perceived discussion surrounding research and ethical practices as more respectful, open, relevant, and constructive.

The bottom three lines of Fig. 3 report the results for three scales that measured the substantive aspects of the training content. Those who participated in the training were more likely to understand the importance of and rationale for having both an authorship policy and a data management policy and report that the laboratory is more likely to archive and preserve replication materials that underlie the research. These items also indicate that laboratories that participated in the training better approximated the deliberative ideals of publicness and transparency and that the laboratory members understood the justification for that openness and transparency.\*

## Discussion

The iREDS ethics training seeks to intervene in how laboratory members communicate with each other about their research ethics beliefs and practices in the context of their day-to-day work. We assert that if laboratory members more routinely talk to each other about the ethical dimensions of their laboratory practices, then they will be more likely to self-consciously select procedures that are not only scientifically credible but also, ethically defensible. The goal of the iREDS training is to make discussions within laboratories about ethical research practices a day-to-day part of research practices in science and engineering.

The present evidence suggests that such an intervention can lead researchers to be more engaged in discussions surrounding the ethical implications of their research, which can foster a culture of more ethically sound science. Overall, the training met its goals of fostering a better climate and deliberative communication within the participating laboratories when compared with the control condition. The results summarized in Figs. 2 and 3 are compelling for two reasons. First, the latent scales that we recover from the survey measure aspects of discourse and culture in laboratories, and this discourse is the central embodiment of the culture of ethics in the laboratory workplace. Second, the goal of the training was to have a durable effect on the culture of ethics and discourse, and we observed detectable effects from the training 4.5 mo after the intervention.

While the statistical estimates demonstrate the impact of the training, the results can only be extended to the kinds of laboratories that voluntarily select into the training and are subsequently randomized into the treatment or control condition. While it would be useful to know the causal effect that is generalizable to all science laboratories, knowing the benefits for those laboratories that voluntarily participate is still meaningful since these are the kinds of laboratories that would be expected to select into a future voluntary training opportunity.

We also note the training was delivered to laboratory participants who were available at the time of the study. We do not know whether and how our results would persist as laboratory members leave and others enter and whether the training would be retained as an institutional memory. Replications of this work will be useful for increasing the precision of the estimated effectiveness of this intervention and for exploring its boundary conditions.

\*We report in *SI Appendix* an exploratory subgroup analysis in which we find no differences in the treatment effect estimates for underrepresented minority, nonmale, and non-PI respondents in comparison with the treatment effects of white male PIs.

This study may help to establish best practices for implementing ethics training in a way that transforms the culture of the university to prioritize discussions of ethical research practices. This training approach might be extendable to other topics in the ethical practice of research, such as collaboration, mentoring, disclosure, and compliance—topics that may present different challenges than authorship and data management. It will also be worth evaluating whether the individualistic, online module training approach has any salutary benefits on ethical practice on its own (10), whether a broader shift toward a conversation-centered, project-based approach to ethics training may be more effective, or whether the best is an integrated approach that combines knowledge transfer with conversational curricula. Rigorous evaluation and deployment of evidence-based ethics training may produce broader impacts for society as scientists become more self-conscious of ethical considerations in the day-to-day practice of their research.

## Materials and Methods

Statistical tests were Bayesian, and we report full posterior distributions. Data collection and analysis were not blind to the conditions of the experiments.

**Experimental Design.** Our sample consisted of 184 members of 34 research laboratories in science and engineering fields at UCR and at TSRI (recently renamed Scripps Research) in La Jolla, California.

Because the laboratory is the unit of analysis, the first step of our recruitment strategy was to reach out to the PI of each laboratory via a direct email combined with presentations during department faculty meetings. Specifically, we emailed invitations individually to all PIs in departments that conduct laboratory-based research in our College of Natural and Agricultural Sciences, the Bourns College of Engineering, and the Biomedical Sciences Department of the Medical School at UCR.<sup>†</sup> If a PI consented to be enrolled in the research, we were then given the opportunity to separately seek consent from each member of the laboratory, including postdocs, graduate students, project scientists, and undergraduate research assistants. If 50% plus 1 of the laboratory members consented to participate, the laboratory met the eligibility criteria and was enrolled. Only those laboratory members who consented participated in the surveys, although nonconsenting members were free to attend the training. Nonconsenting laboratory members are excluded from the surveys.

A total of 30 laboratories from UCR met our eligibility criteria and enrolled in the study. We recruited an additional 4 laboratories from TSRI that also met our eligibility criteria for a total of 34 laboratories enrolled. All laboratories recruited from UCR (regardless of condition assignment) were paid \$1,000 for graduate student travel to conferences for their participation during the 6-mo study period. We sent a total of two follow-up emails after our initial invitation. We emailed invitations to 458 science laboratories at UCR; 75 PIs initially indicated an interest in participating (16%), and 30 laboratories eventually met our eligibility criteria and completed the study (7%). Among the 45 laboratories that began the enrollment process but did not complete the study, 12 failed to meet the eligibility criterion; 31 did not complete the enrollment process for a variety of reasons (lack of interest, no laboratory members, on sabbatical, etc.). Two laboratories enrolled but chose not to complete the study, and their data are not used in the analysis.

In our sample of participants, 53% identified as male, 2% self-identified as African American, 28% self-identified as Asian, 16% self-identified as Hispanic, 44% self-identified as white, and 9% self-identified as other. Our sample consisted of primary investigators (32), graduate students (97), postdoc researchers (12), undergraduate research assistants (23), research scientists (11), and support staff (2); 7 are categorized as other. Thirteen academic disciplines were represented among the respondents. Fifty-four participants took part in the first but not the second assessment (i.e., 6 mo later). Nine participants took part in the second assessment but not the first, and 121 participants completed both assessments. Our procedures for handling the missingness in our sample are discussed below. Of the singles, 7 are PIs, and 56 are laboratory members. The response rate on the pretest is not dependent on random assignment to experimental conditions (94.4% response controls, 95.8% treatment,  $\chi^2 = 0.66$ ), but there is weak evidence of dependence between experimental condition and responding to the

<sup>†</sup>We also sent the invitation email to the Vice Provost of Research at the nearby TSRI, who forwarded it to TSRI faculty.

posttest (76.4% controls, 65.3% treatment,  $\chi^2 = 0.10$ ), although the substantive difference is not large.

### Procedures.

**Randomization procedure and design.** Randomization was at the laboratory level to either the training (treatment) condition or the standard (control) condition. We block randomized within departments to ensure balance. Our randomization procedure required that the first laboratory within a department was randomized to one arm, and the subsequent laboratories within the department were sequentially enrolled in the opposite arm as the previously enrolled laboratory. The one exception was that, through an oversight, we used a simple randomization rather than block randomization among the laboratories within the UCR engineering college.

Laboratories randomized to the intervention condition participated in two trainings, one of which was 90 min and the other of which was 60 min, scheduled 2 wk apart from one another. We did not keep attendance for the training, although the trainings typically occurred at a scheduled laboratory-wide meeting.

**Intervention.** The first of two trainings focused on issues relating to the development and maintenance of a data management policy, and the second training focused on discussing possible frameworks for making manuscript authorship decisions. Both trainings included three elements unique to training in the RCR: 1) in-person delivery, 2) involvement of a peer mentor/training facilitator, and 3) the incorporation of a demonstration of the OSF.

- 1) In-person delivery. Study coordinators met with laboratory members in person in a laboratory-wide meeting to deliver both trainings. The training consisted of a standard training script customized to each laboratory. This script described the ethical dimensions of the specific research practice under consideration and provided opportunities for members of the laboratory to discuss current and possible future relevant behavior. The skeleton script for both data management and authorship trainings can be found at <https://osf.io/h8jcz/>.
- 2) Peer mentor. During the consent process, non-PI laboratory members were asked whether they would like to volunteer to be their laboratory's peer mentor. If more than one laboratory member volunteered, the study coordinator chose one at random. The peer mentor's role was to help with facilitation of the two trainings. They first met with the study coordinator 2 wk prior to the first training and answered a series of questions relating to their laboratory's current data management, authorship, and general laboratory practices (the peer mentor interview prompts can be found at <https://osf.io/5erza/>). During this meeting, the peer mentor and the study coordinator who specialized in OSF together built the initial framework for the OSF demonstration (see below for a detailed description of this demonstration). During each of the trainings, the peer mentor was asked to follow the training script and occasionally read a question to the rest of the laboratory to help facilitate the training discussion.
- 3) Demonstration of the OSF. The training also incorporated a demonstration of the OSF (<https://osf.io>). The OSF is a free online platform that facilitates openness, transparency, and reproducibility in science. It also

serves as a means to share and store data and study materials as well as keep track of individual contributions to a project. Both trainings consisted of the real-time development and demonstration of a project on the OSF (based on the initial project set up during the peer mentor meeting). Visit <https://osf.io/uevcn/> for an example of an OSF project.

**Survey administration.** All surveys were administered online using the Qualtrics survey platform. The presurvey was delivered to participants a month before the first training, the first midpoint survey was delivered 2 wk following the second training, the second midpoint survey was delivered 2 wk plus 2 mo following the second training, and the postsurvey was delivered 2 wk plus 4 mo after the second training. The pre- and postsurveys were identical, and each had a total of 94 items that measure aspects of participants' personal beliefs and opinions about ethical research practices as well as their perceptions of and experiences with various laboratory practices, climate, and communication. Laboratories randomized to the control condition received the same four surveys over the 6-mo study period but did not participate in any intervention. The survey delivery timeline was the same for those randomized to the control condition. The surveys can be found at <https://osf.io/stzgw/>.

**Standard practice for ethics training.** The research design evaluates the difference in response to our survey items for those in the intervention compared with those in the control. Under this design, we identify the changes relative to the existing standard practice. To characterize this standard practice, in the presurvey we ask all respondents to report their exposure to ethics training prior to enrolling in the study. Of the 174 respondents who filled out the presurvey, only 45% report having prior training in RCR or research ethics. Of the 79 respondents who indicated prior training, 49% had online training, and among those, 53% only received training online (22% of the sample). The other modes included a standalone course (14% of the sample), seminar series (9%), lecture (14%), and informal conversations (8%). Those who report having prior training report the total duration up to 2 h (41%), up to 10 h (35%), and more than 10 h (24%). Additionally, among these, 28% reported having training within the previous year, while 28% report having training more than 3 y in the past.

**Human Subjects Institutional Review.** This study was approved by the institutional review board of the University of California, Riverside. We obtained voluntary informed consent from all participants before they took part in the studies.

**ACKNOWLEDGMENTS.** We thank co-PI Juliet McMullin for contributing to the proposal writing, leading the ethnographic component of the project, and helping to develop the survey questionnaire; Joseph Childers for leading the original proposal and planning; Kennett Lai and Andrew Sallans for administrative support; Sabrina Sakay for research assistance; and Esra Kürüm for reviewing the statistical methods. This project was funded by Cultivating Cultures of Ethical STEM Program of the NSF Grant SES-1540440 and by support from the Graduate Division of UCR. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

1. OPRS, PHS policy on instruction in the responsible conduct of research (RCR). [https://oprs.usc.edu/files/2017/04/PHS\\_Policy\\_on\\_RCR1.pdf](https://oprs.usc.edu/files/2017/04/PHS_Policy_on_RCR1.pdf). Accessed 25 December 2019.
2. E. Miguel et al., Social science. Promoting transparency in social science research. *Science* **343**, 30–31 (2014).
3. M. L. Devereaux, Rethinking the meaning of ethics in RCR education. *J. Microbiol. Biol. Educ.* **15**, 165–168 (2014).
4. K. D. Pimple, Six domains of research ethics. A heuristic framework for the responsible conduct of research. *Sci. Eng. Ethics* **8**, 191–205 (2002).
5. B. C. Martinson, M. S. Anderson, R. de Vries, Scientists behaving badly. *Nature* **435**, 737–738 (2005).
6. N. H. Steneck, Fostering integrity in research: Definitions, current knowledge, and future directions. *Sci. Eng. Ethics* **12**, 53–74 (2006).
7. S. L. Titus, J. A. Wells, L. J. Rhoades, Repairing research integrity. *Nature* **453**, 980–982 (2008).
8. D. B. Resnik, G. E. Dinse, Do U.S. Research institutions meet or exceed federal mandates for instruction in responsible conduct of research? A national survey. *Acad. Med.* **87**, 1237–1242 (2012).
9. T. Phillips, F. Nestor, G. Beach, E. Heitman, America COMPETES at 5 years: An analysis of research-intensive universities' RCR training plans. *Sci. Eng. Ethics* **24**, 227–249 (2018).
10. National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Committee on Science, Engineering, Medicine, and Public Policy, Committee on Responsible Science, *Fostering Integrity in Research* (National Academies Press, 2018).
11. M. C. Sloan, Aristotle's nicomachean ethics as the original locus for the septem circumstantiae. *Class. Philol.* **105**, 236–251 (2010).
12. D. K. Plemmons, M. W. Kalichman, Reported goals of instructors of responsible conduct of research for teaching of skills. *J. Empir. Res. Hum. Res. Ethics* **8**, 95–103 (2013).
13. T. Ingold, *Tools, Language and Cognition in Human Evolution*, K. R. Gibson, T. Ingold, Eds. (Cambridge University Press, New York, NY, 1993).
14. J. Lave, "The culture of acquisition and the practice of understanding" in *Situated Cognition: Social, Semiotic, and Psychological Perspectives*, D. Kirshner, J. A. Whitson, Eds. (Lawrence Erlbaum Associates, Mahwah, NJ, 1997), pp. 63–82.
15. M. Kalichman, M. Sweet, D. Plemmons, Standards of scientific conduct: Are there any? *Sci. Eng. Ethics* **20**, 885–896 (2014).
16. M. Kalichman, M. Sweet, D. Plemmons, Standards of scientific conduct: Disciplinary differences. *Sci. Eng. Ethics* **21**, 1085–1093 (2015).
17. S. Vasconcelos, D. Vasgird, I. Ichikawa, D. Plemmons, Authorship guidelines and actual practice: Are they harmonized in different research systems? *J. Microbiol. Biol. Educ.* **15**, 155–158 (2014).
18. A. Gutmann, D. Thompson, *Why Deliberative Democracy* (Princeton University Press, Princeton, NJ, 2004).
19. J. Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (MIT Press, Cambridge, MA, 1996).
20. M. A. Neblo, *Deliberative Democracy Between Theory and Practice* (Cambridge University Press, 2015).
21. J. Cooper, D. Mackie, Cognitive dissonance in an intergroup context. *J. Pers. Soc. Psychol.* **44**, 536–544 (1983).
22. A. J. Elliot, P. G. Devine, On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *J. Pers. Soc. Psychol.* **67**, 382–394 (1994).

23. V. B. Hinsz, R. S. Tindale, D. A. Vollrath, The emerging conceptualization of groups as information processors. *Psychol. Bull.* **121**, 43–64 (1997).
24. D. C. Matz, W. Wood, Cognitive dissonance in groups: The consequences of disagreement. *J. Pers. Soc. Psychol.* **88**, 22–37 (2005).
25. L. M. Sanders, Against deliberation. *Polit. Theory* **25**, 347–376 (1997).
26. B. R. Snyder, *The Hidden Curriculum* (Alfred A. Knopf, New York, NY, 1971).
27. C. Whitbeck, Group mentoring to foster the responsible conduct of research. *Sci. Eng. Ethics* **7**, 541–558 (2001).
28. K. Fryer-Edwards, Addressing the hidden curriculum in scientific research. *Am. J. Bioeth.* **2**, 58–59 (2002).
29. A. M. Peiffer, P. J. Laurienti, C. E. Hugenschmidt, Fostering a culture of responsible lab conduct. *Science* **322**, 1186 (2008).
30. M. Kalichman, A modest proposal to move RCR education out of the classroom and into research. *J. Microbiol. Biol. Educ.* **15**, 93–95 (2014).
31. D. K. Plemmons, M. W. Kalichman, Mentoring for responsible research: The creation of a curriculum for faculty to teach RCR in the research environment. *Sci. Eng. Ethics* **24**, 207–226 (2018).
32. M. D. Mumford et al., Environmental influences on ethical decision making: Climate and environmental predictors of research integrity. *Ethics Behav.* **17**, 337–366 (2007).
33. K. Okkenhaug, Integrity in scientific research: Creating an environment that promotes responsible conduct. *BMJ* **327**, 935 (2003).
34. B. K. Sovacool, Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? *J. Bioeth. Inq.* **5**, 271–282 (2008).
35. K. Esterling, Replication data for: Enhancing ethical discourse in STEM labs: A randomized trial of project-based ethics training. Harvard Dataverse. <https://doi.org/10.7910/DVN/AIPWNU>. Deposited 9 July 2019.
36. M. Castro, Placebo versus best-available-therapy control group in clinical trials for pharmacologic therapies: Which is better? *Proc. Am. Thorac. Soc.* **4**, 570–573 (2007).
37. K. M. Esterling, A. Fung, T. Lee, How much disagreement is good for democratic deliberation? *Pol. Comm.* **34**, 529–551 (2015).
38. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (2016). <https://ggplot2.tidyverse.org>. Accessed 25 December 2019.